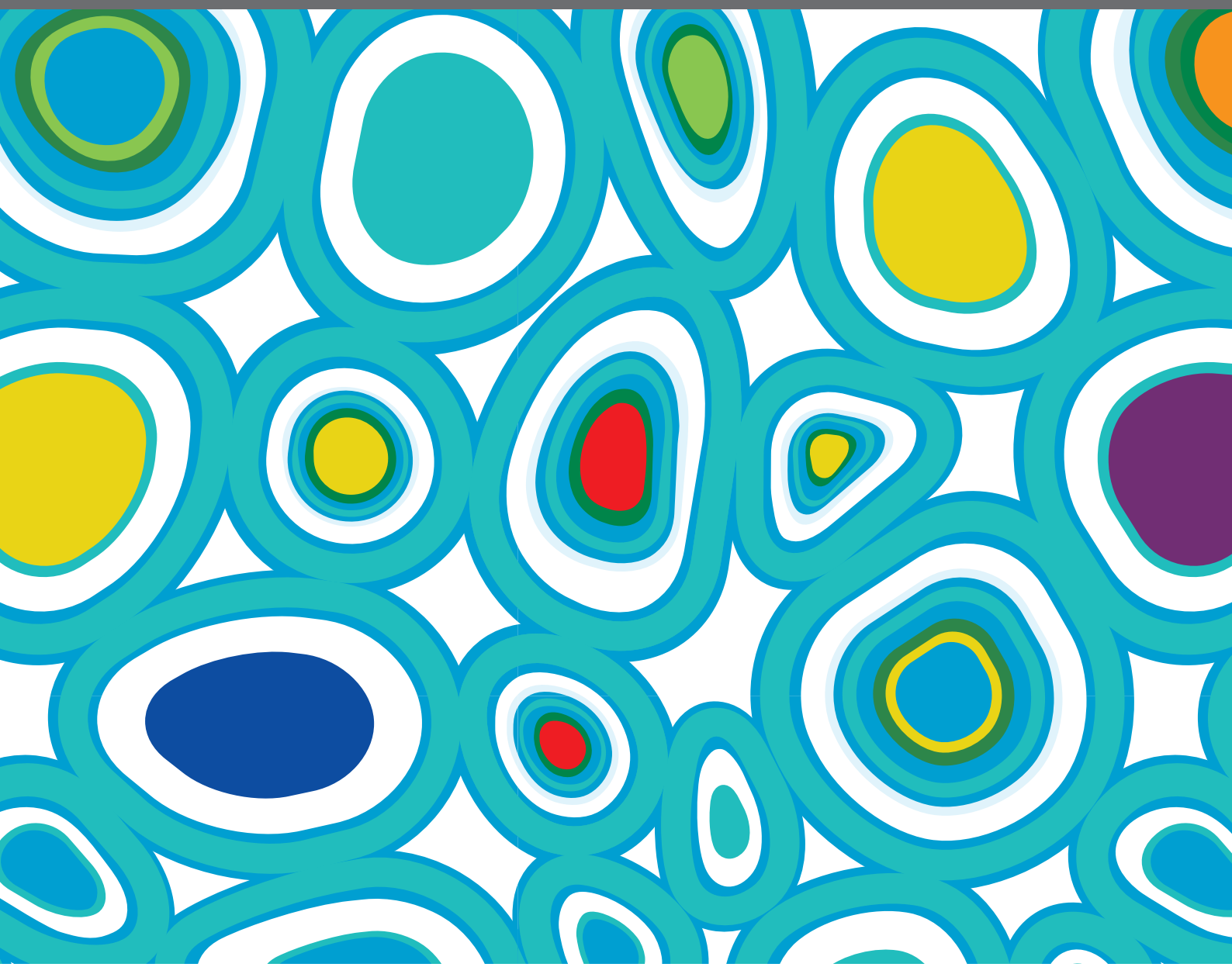
CHROMATIN REGULATION IN CELL FATE DECISIONS

EDITED BY: José Luis Sardina, Justin Brumbaugh and Bruno Di Stefano
PUBLISHED IN: Frontiers in Cell and Developmental Biology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-510-7
DOI 10.3389/978-2-88971-510-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CHROMATIN REGULATION IN CELL FATE DECISIONS

Topic Editors:
José Luis Sardina, Josep Carreras Leukaemia Research Institute (IJC), Spain
Justin Brumbaugh, University of Colorado Boulder, United States
Bruno Di Stefano, Baylor College of Medicine, United States

Citation: Sardina, J. L., Brumbaugh, J., Di Stefano, B., eds. (2021). Chromatin Regulation in Cell Fate Decisions. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-510-7

Table of Contents

frontiers in Cell and Developmental Biology

EDITORIAL
published: 01 September 2021
doi: 10.3389/fcell.2021.734020

Editorial: Chromatin Regulation in Cell Fate Decisions

Justin Brumbaugh[1,2,3]*, *Bruno Di Stefano*[4,5,6]* and *José Luis Sardina*[7]*

[1] Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO, United States, [2] Molecular, Cellular and Developmental Biology, University of Colorado Comprehensive Cancer Center, Aurora, CO, United States, [3] Charles C. Gates Center for Regenerative Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States, [4] Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, United States, [5] Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX, United States, [6] Stem Cells and Regenerative Medicine Center, Baylor College of Medicine, Houston, TX, United States, [7] Epigenetic Control of Haematopoiesis Group, Josep Carreras Leukaemia Research Institute, Barcelona, Spain

Keywords: chromatin, gene regulation, epigenetics, pluripotency, differentiation, cell fate, development, stem cells

Edited by:
Atsushi Asakura,
University of Minnesota Twin Cities, United States

Reviewed by:
Qiang Wu,
Macau University of Science and Technology, Macao, SAR China
Eran Meshorer,
Hebrew University of Jerusalem, Israel

****Correspondence:***
Justin Brumbaugh
justin.brumbaugh@Colorado.EDU
Bruno Di Stefano
bruno.distefano@bcm.edu
José Luis Sardina
jsardina@carrerasresearch.org

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *30 June 2021*
Accepted: *12 August 2021*
Published: *01 September 2021*

Citation:
Brumbaugh J, Di Stefano B and Sardina JL (2021) Editorial: Chromatin Regulation in Cell Fate Decisions. Front. Cell Dev. Biol. 9:734020. doi: 10.3389/fcell.2021.734020

Editorial on the Research Topic

Chromatin Regulation in Cell Fate Decisions

Cell fate decisions, including events that occur naturally (e.g., development/embryogenesis, differentiation, regeneration, homeostasis) and experimentally (e.g., reprogramming to induced pluripotent stem (iPS) cells, directed differentiation, transdifferentiation), are typically mediated by transcription factors in concert with epigenetic modifications (Apostolou and Hochedlinger, 2013). These crucial regulators direct gene expression changes to establish cell-type specific transcriptional profiles. Mechanistically, a wide range of chromatin-related processes are involved (DNA and histone modifications, histone variants, genome topology, RNA processing and many more). The interplay between all of these layers of epigenetic regulation finally controls the transcriptional output of the conversion process, thus determining the final cell fate. This special edition of Frontiers in Cell and Developmental Biology examines the role of chromatin-based regulation in controlling cell fate decisions.

Many epigenetic marks dramatically change during embryogenesis, suggesting that they have fundamental roles in developmental decisions. For example, the early embryo experiences an initial, widespread loss of DNA methylation that is reacquired around the time of implantation. In this edition, Greenberg reviews the mechanisms responsible for these dynamics, with a focus on key regions that escape DNA methylation during its reestablishment. It remains unresolved whether this early round of DNA demethylation is passive (i.e., DNA methylation is diluted over multiple rounds of cell division) or active (i.e., resulting from enzymatic action). In any case, the Ten-eleven translocation (TET) class of DNA demethylases, and specifically TET2, are crucial for exiting pluripotency and initiating differentiation (Dai et al., 2016). Garcia-Outeiral et al. provide an updated perspective on TET2's role in these processes as well as reprogramming, and discuss the possibility that TET2 is also responsible for oxidizing RNA methylation. Together, these publications highlight the importance of DNA/RNA (de)methylation in the early embryo.

Pluripotent stem cells (PSCs) provide an experimentally tractable system to study the association between chromatin and cell fate decisions. In recent years, multiple cell culture conditions were established to maintain self-renewing PSCs (both mouse and human) but with distinct molecular and phenotypic properties. Although all of the culture conditions support pluripotency, each is believed to represent a different developmental stage and collectively, these tools have broadened our ability to model early embryonic development. Correspondingly, cells cultured in each

condition have their own epigenetic signature and transitioning between states leads to large rearrangements in the chromatin landscape. Sun et al. summarize the recent characterization of these states and also compare culture conditions and epigenetic regulation in mouse and human. Work in PSCs has also shed light on the proteins responsible for modifying chromatin during developmental transitions. Franklin et al. provide a comprehensive review of histone chaperones and their role in cell fate specification, with a focus on proteins responsible for depositing histone H3. Once deposited, histones are further regulated by nucleosome remodelers, many of which have been linked to development and differentiation. Separate articles by the Ye et al. and Pagliaroli and Trizzino labs lay out the role of mammalian SWI/SNF complexes in pluripotency, differentiation and developmental disorders. Cis-regulatory elements are likewise important in guiding cell fate decisions. Agrawal and Rao discuss the contribution of super-enhancers and CTCF-mediated three-dimensional genome organization in regulating cell-type specific expression programs. These articles provide important insight into the role of chromatin in directing embryonic cell fate decisions.

During the formation and maintenance of adult tissues, several molecular mechanisms coordinate gene expression programs to ensure proper lineage specification. For example, chromatin organization is tightly controlled by different epigenetic regulators during brain development. In this issue, Mastrototaro et al. show that the nuclear receptor TBL1XR1 modulates the stability of the epigenetic repressive complex NCOR, thus impacting neural progenitor self-renewal and differentiation. DNA methylation at CpG residues is another important regulator of neuronal development. Recently, non-CpG DNA methylation (mCH) was detected in the vertebrate nervous system (de Mendoza et al., 2021); yet, its biological relevance remains unclear. Ross et al. found that mCH accumulates at gene bodies and transposable elements during zebrafish brain development and potentially plays a role in regulating the transcription at these genomic regions. Similar to its role in controlling pluripotency, TET-mediated DNA demethylation is a prominent mechanism of chromatin-based regulation in brain development and function. MacArthur and Dawlaty discuss the role of TET enzymes in neural differentiation and review how TET dysregulation contributes to neurological disorders.

As in the brain, differentiation of hematopoietic stem cells to specialized cells is accompanied by extensive chromatin remodeling. Interestingly, genes coding for regulators of DNA methylation (i.e., *DNMT3A and TET2*) are frequently mutated in a wide range of hematological malignancies, suggesting that DNA methylation is key for establishment and maintenance of hematopoietic cell identity. In this edition, Tsiouplis et al. provide a comprehensive review on TET enzymes as key regulators of the immune system in homeostasis and in pathological conditions. Chromatin modifying complexes including Polycomb group (PcG) of proteins have also been involved in hematopoietic cell differentiation. In their primary research article, García-Montolio et al. show that the epigenetic regulator PHF19 (a PRC2-associated factor) controls the balance between proliferation and differentiation of erythroid progenitors. These articles expand on a growing research interest in correlation between hematopoiesis and chromatin regulation.

Unlike the brain or blood, current evidence suggests that the liver does not have a dedicated stem cell compartment to mediate homeostasis and regeneration. However, following injury or resection, hepatic cells undergo substantial changes in proliferative capacity and are capable of transdifferentiation or de-differentiation. Mediating these changes are a variety of epigenetic regulators. In this issue, Aloia discusses the role of epigenetic mechanisms in liver regeneration and expands on how alterations in these mechanisms lead to complex disorders, including cancer.

Finally, DNA and histone methylation is a key epigenetic mechanism that potently influences transcriptional outputs. DNA, RNA, and histone methyltransferases require S-Adenosyl methionine (SAM) as a donor for the methyl group. Therefore, SAM plays an essential role in regulating the chromatin status of cells, consequently linking methionine metabolism with cell fate. SAM generation is catalyzed by the Adenosylhomocysteinase (AHCY) that reportedly binds to chromatin in pluripotent stem cells (Aranda et al., 2019). Vizán et al. review the evolution and biochemical properties of AHCY across all living organisms and highlight its functions in homeostasis and disease.

Together, the manuscripts in this Research Topic provide an overview of the latest research on the impact of chromatin regulation in instructing a full spectrum of cell fate decisions, ranging from early embryogenesis to regeneration.

AUTHOR CONTRIBUTIONS

JB, BD, and JS contributed to the editing of this Research Topic. All authors contributed to the article and approved the submitted version.

FUNDING

JB was funded by the National Institutes of Health (R35GM142884), Boettcher Foundation's Webb-Waring Biomedical Research Awards program, the CU Cancer Center Joint Multi-PI Pilot Project Program and the University of Colorado Research and Innovation Seed Grant Program. BD was supported by funds from CPRIT-RR200079. JS was funded by Instituto de Salud Carlos III through the project CP19/00176 (co-funded by European Social Fund, Investing in your future) and the Spanish Ministry of Science, Innovation and Universities (PID2019-111243RA-I00/AEI/10.13039/501100011033).

ACKNOWLEDGMENTS

We thank CERCA Programme/Generalitat de Catalunya for institutional support.

REFERENCES

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. *Nature* 502, 462–471. doi: 10.1038/nature12749

Aranda, S., Alcaine-Colet, A., Blanco, E., Borràs, E., Caillot, C., Sabidó, E., et al. (2019). Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci. Adv.* 5:eaav2448. doi: 10.1126/sciadv.aav2448

Dai, H.-Q., Wang, B.-A., Yang, L., Chen, J.-J., Zhu, G.-C., Sun, M.-L., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. *Nature* 538, 528–532. doi: 10.1038/nature20095

de Mendoza, A., Poppe, D., Buckberry, S., Pflueger, J., Albertin, C. B., Daish, T., et al. (2021). The emergence of the brain non-CpG methylation system in vertebrates. *Nat. Ecol. Evol.* 5, 369–378. doi: 10.1038/s41559-020-01371-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 07 January 2021
doi: 10.3389/fcell.2020.629068

Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals

*Maxim V. C. Greenberg**

Centre National de la Recherche Scientifique, Institut Jacques Monod, Université de Paris, Paris, France

Vertebrate genomes are marked by notably high levels of 5-cytosine DNA methylation (5meC). The clearest function of DNA methylation among members of the subphylum is repression of potentially deleterious transposable elements (TEs). However, enrichment in the bodies of protein coding genes and pericentromeric heterochromatin indicate an important role for 5meC in those genomic compartments as well. Moreover, DNA methylation plays an important role in silencing of germline-specific genes. Impaired function of major components of DNA methylation machinery results in lethality in fish, amphibians and mammals. Despite such apparent importance, mammals exhibit a dramatic loss and regain of DNA methylation in early embryogenesis prior to implantation, and then again in the cells specified for the germline. In this minireview we will highlight recent studies that shine light on two major aspects of embryonic DNA methylation reprogramming: (1) The mechanism of DNA methylation loss after fertilization and (2) the protection of discrete loci from ectopic DNA methylation deposition during reestablishment. Finally, we will conclude with some extrapolations for the evolutionary underpinnings of such extraordinary events that seemingly put the genome under unnecessary risk during a particularly vulnerable window of development.

Keywords: DNA methylation, epigenetics, reprogramming, mammalian development, embryonic stem cells (ESC)

OPEN ACCESS

Edited by:
Bruno Di Stefano, Baylor College of Medicine, United States

Reviewed by:
Zachary Smith, Yale University, United States
Ozren Bogdanovic, University of Western Australia, Australia
Tuncay Baubec, University of Zurich, Switzerland

***Correspondence:**
Maxim V. C. Greenberg
maxim.greenberg@ijm.fr

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *13 November 2020*
Accepted: *09 December 2020*
Published: *07 January 2021*

Citation:
Greenberg MVC (2021) Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals. Front. Cell Dev. Biol. 8:629068. doi: 10.3389/fcell.2020.629068

INTRODUCTION

5-cytosine DNA methylation (5meC) is a modification conserved across all kingdoms of eukaryotes. Most generally it is found in the CpG dinucleotide context, and there are factors that ensure faithful daughter strand methylation after each round of DNA synthesis (Bostick et al., 2007; Sharif et al., 2007; Woo et al., 2008). Given the tight link with DNA replication, 5meC has high potential to exhibit epigenetic stability. In flowering plants, so-called "epialleles"—alleles that differ in 5meC content, not their DNA sequence—can persist for an indefinite number of generations (Weigel and Colot, 2012). Recently it was demonstrated in the yeast *Cryptococcus neoformans* that DNA methylation patterns have endured for millions of years exclusively through a maintenance mechanism, as there is an absence of *de novo* DNA methylation enzymes in the genome (Catania et al., 2020).

Standing apart from other eukaryotic lineages, it has been known for decades that mammals exhibit two rounds of dramatic DNA methylation reprogramming during embryonic development: first immediately after fertilizations, and a second time in the germline (Monk et al., 1987). The

most facile explanation for the initial 5meC erasure event is that the embryo must level the high DNA methylation asymmetry exhibited by the paternal and maternal gametic genomes that arrive in the zygote (Wang et al., 2014), thus mitigating dosage discrepancies between alleles. By the blastocyst stage, which coincides with naive pluripotency, residual DNA methylation (∼20% of CpGs) is largely restricted to genomic imprints and TEs (Wang et al., 2014, 2018; Zhu et al., 2018). As the embryo implants in the uterus, the *de novo* DNA methyltransferases, DNMT3A and DNMT3B, rapidly remethylate the genome to 70–80% CpG methylation, establishing a pattern that is globally maintained in somatic tissue-types (Borgel et al., 2010; Seisenberger et al., 2012; Smith et al., 2017; Zhang et al., 2018). This whole process repeats itself in primordial germ cells (PGCs), with a key difference being that in the germline, genomic imprints are erased and then reset in a sex-specific manner (Hajkova et al., 2002; Lee et al., 2002).

The last decade has seen advances in sequencing protocols and technology that have allowed for stunning temporal resolution of allele-specific DNA methylation patterns in early mammalian development (Wang et al., 2014, 2018; Gkountela et al., 2015; Zhu et al., 2018; Grosswendt et al., 2020). Nevertheless, much of the mechanistic underpinning of these processes remains lacking. To this end, embryonic stem cells (ESCs), which are derived from the inner cell mass (ICM) of the blastocyst, remain a powerful model for exploring the bases of the phenomenology of DNA methylation reprogramming. In this mini-review, we will highlight recent findings made in mouse ESCs (mESCs) that may help explain (1) how rapid global demethylation occurs, and (2) how promoters remain protected from the onslaught of DNA methylation establishment during implantation. Finally, given that DNA methylation reprogramming is a peculiarity—one that does not even appear to occur in our non-mammalian vertebrate cousins—we will discuss a possible clue that might explain the evolutionary rise of counterintuitive events.

PASSIVE AGGRESSIVE: DNA DEMETHYLATION AFTER FERTILIZATION

DNA demethylation can occur via either passive or active means. Passive demethylation simply requires the impairment of maintenance DNA methylation machinery, which results in 2-fold dilution of methyl-CpGs during each round of DNA synthesis. In mammals, active DNA demethylation occurs through action of Ten-eleven translocase (TET) family of dioxygenases, although the mechanism is far from intuitive: iterative oxidation of 5meC to hydroxymethylcytosine (5hmC), formylcytosine (5fC), and finally carboxylcytosine (5caC) will trigger the thymine DNA glycosylase-dependent base excision repair (TDG BER) pathway to replace modified cytosines with unmodified versions (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009; He et al., 2011; Maiti and Drohat, 2011; Weber et al., 2016). Oxidized forms of 5meC also impede DNA methylation maintenance (Hashimoto et al., 2012; Otani et al., 2013; Ji et al., 2014), thus in that sense, TET proteins contribute to passive DNA demethylation, as well.

The extent to which the TET enzymes contribute to the global demethylation exhibited during embryonic 5meC reprogramming remains an active, and somewhat controversial, area of research. Soon after fertilization, the DNA within the paternal pronucleus becomes strongly enriched for TET3-dependent 5hmC, relative to its maternal counterpart (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). Moreover, paternal DNA methylation is rapidly erased in the zygote, before passive dilution can even occur (Mayer et al., 2000; Oswald et al., 2000). Therefore, it seems quite logical to posit that TET3 is responsible for demethylating the paternal genome. However, careful genetic dissection indicates that this is not entirely the case (Amouroux et al., 2016). Instead, TET3 activity may help protect the paternal genome from DNMT3A-dependent *de novo* DNA methylation (Amouroux et al., 2016; Albert et al., 2020). Of course, this would then imply there is an undiscovered mechanism of active DNA demethylation that must be occurring in the zygote. It should also be noted that in the early embryo, TET3 activity can lead to a passive loss of DNA methylation, independently of base-excision repair (Guo et al., 2014; Shen et al., 2014).

In fact, impaired DNA methylation maintenance plays an undeniably important role in reprogramming during the first cell divisions of preimplantation development. The maintenance pathway primarily consists of two proteins: the methyltransferase DNMT1 and its co-factor Ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1, also known as NP95). While there are many intricate layers to UHRF1/DNMT1 regulation, at its most fundamental level the mechanism is elegantly simple: UHRF1 recognizes hemimethylated CpG sites after DNA replication, and recruits DNMT1 to methylated the cytosine on the daughter strand (Bostick et al., 2007; Sharif et al., 2007) (**Figure 1**).

In the past few years, studies have emerged that show UHRF1 disruption may be an important means of DNA demethylation. On the spectrum of differentiated cell types, oocytes are quite DNA hypomethylated: roughly 50% of CpGs are methylated in mouse and human oocytes, compared to 70–80% in somatic tissues (Kobayashi et al., 2012; Shirane et al., 2013; Okae et al., 2014). Recently it was demonstrated that the protein Developmental Pluripotency-Associated 3 (DPPA3, also known as STELLA and PGC7) shuttles UHRF1 to the cytoplasm of mouse oocytes (Li et al., 2018a). In the *Dppa3* KO oocytes, UHRF1 returns to the nucleus, and oocyte methylation is increased (Li et al., 2018a; Han et al., 2019). It has been known for decades in mouse preimplantation embryos, an oocyte-specific isoform form of DNMT1 (oDNMT1) is enriched in the cytoplasm, hence its access to DNA is restricted (Carlson et al., 1992; Howell et al., 2001). While immunofluorescence indicates that oDNMT1 still remains largely cytoplasmic in *Dppa3* mutants, a proportion does enter the nucleus as well—that is, DPPA3 activity is the best explanation for the lowly methylated state of the oocyte genome.

Our understanding of the DPPA3 regulation of UHRF1 has been greatly elucidated in a mESC system. When cultured in serum-free conditions and in the presence of MAPK and GSK3β inhibitors (2i), mESCs exhibit globally low levels of DNA methylation—more or less on par with the ICM cells from which

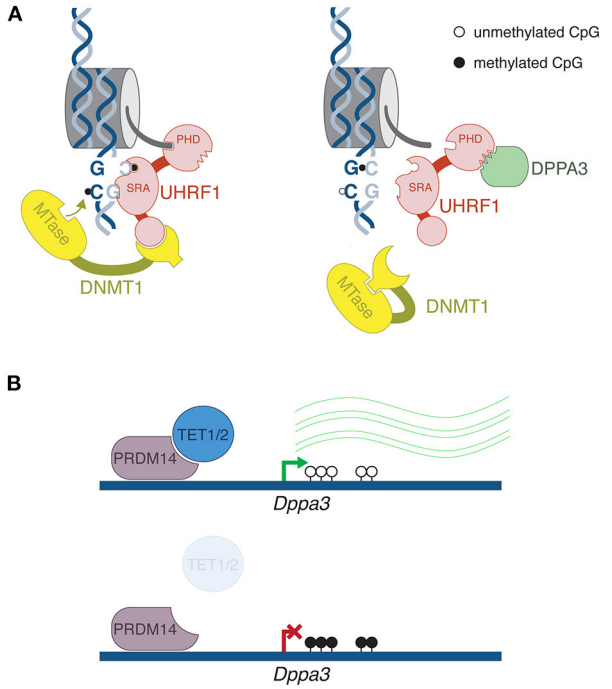


FIGURE 1 | Mechanism of DPPA3-mediated DNA demethylation. **(A)** After DNA replication, the SRA domain of UHRF1 binds to hemimethylated CpG, flipping out the methylated cytosine. Recruitment of DNMT1 directs methylation of the symmetrical unmethylated cytosine on the daughter strand. Crucially, UHRF1 must bind to histone H3 through its PHD domain. DPPA3 binds to the PHD domain, disrupting UHRF1's interaction with chromatin leading to impaired DNA maintenance methylation. **(B)** Top panel: In mouse ESCs, PRDM14 recruits TET1 and TET2 roughly 2 kb upstream of the *Dppa3* promoter, leading to its demethylation and activation. Bottom panel: in *Tet1/2* double mutants, increased DNA methylation at the *Dppa3* gene corresponds with repression. Thus, active demethylation of this one gene is required for global passive demethylation. This mechanism likely does not occur during embryonic DNA methylation reprogramming, but may help explain the wave of DNA demethylation that occurs in the male and female germlines.

ESCs are derived (Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013). It was demonstrated that the underlying reason for such low methylation was impaired UHRF1 activity (von Meyenn et al., 2016). While initially UHRF1 was shown to be unstable at the post-translational level in 2i ESCs (von Meyenn et al., 2016), subsequently it was demonstrated that DPPA3 is the factor that is impeding UHRF1 function (Mulholland et al., 2020). Through a rigorous set of biochemical and microscopy-based experiments, Mulholland et al. showed that DPPA3 binds to UHRF1, thus impairing the latter's ability to bind to chromatin (**Figure 1**). In the mESC system, *Dppa3* mutants are marked by reduction of UHRF1 localization to the cytoplasm, reminiscent of the phenotype in oocytes. However, this aspect of DPPA3-mediated UHRF1 regulation is much less drastic in the cell culture system, thus it is not completely clear if the mechanism of action is exactly the same between oocytes and mESCs. Moreover, DPPA3 can lead to demethylation without its nuclear export function (Du et al., 2019). Regardless, like in oocytes, absence of DPPA3 leads to an increased nuclear fraction of UHRF1 and a gain of DNA methylation. Consistently, when DPPA3 is overexpressed in other cell culture systems, there is a decrease in DNA methylation (Funaki et al., 2014). It should be noted that in addition to DPPA3 action, a number of features of 2i-cultured mESCs likely contribute to the globally DNA hypomethylated state, such as TET protein activity, *de novo* DNA methyltransferase repression (Leitch et al., 2013), and a chromatin state refractory to DNA methylation (van Mierlo et al., 2019), to name three pertinent examples. However, DPPA3 has clearly emerged as a dominant player for this particular feature.

Does the role of DPPA3 in mESCs reflect a role in the DNA demethylation occurring in preimplantation development? While formal demonstration awaits, there are some intriguing indications that indeed DPPA3 performs an analogous function *in vivo*: during normal development, expression of certain classes of TEs are important for activation of the proper embryonic transcription program (Macfarlan et al., 2012; Ishiuchi et al., 2015; Jachowicz et al., 2017); *Dppa3* mutant mice exhibit repression of at least a proportion of these elements at the 2-cell stage (Huang et al., 2017). Given the important role of 5meC in transposon silencing, it is not outlandish to suggest that maintenance of DNA methylation on TEs at the heart of the *Dppa3* transposon expression phenotype, with the caveat that it is difficult to tease apart the maternal from the embryonic effect, given these embryos were generated from *Dppa3* mutant oocytes. Moreover, it should be noted that DPPA3 has been reported to *protect* the maternal genome from TET3-mediated demethylation in zygotes, i.e., the inverse phenomenon (Nakamura et al., 2007, 2012; Han et al., 2019). While not trivial given the severe phenotype of *Dppa3* maternal/zygotic mutants (Payer et al., 2003), hopefully future work will help resolve these apparently contradictory functions.

Finally, Mulholland et al. showed that TET1 and TET2 are required for demethylation of *Dppa3* regulatory regions, thus its activation (**Figure 1**) (Mulholland et al., 2020). In other words, targeted demethylation of one gene supports global passive demethylation. While this is a compelling finding, it is likely not the mechanism occurring after fertilization. Firstly, DPPA3 is already present in the maternal store of protein inherited from the oocyte (Li et al., 2018a); secondly, TET3 is the active TET enzyme after fertilization, not TET1 or 2; and thirdly, the *Dppa3* gene arrives unmethylated from the oocyte, thus does not require further demethylation (Wang et al., 2014). However, this indeed might be the mechanism to activate *Dppa3* prior to the germline DNA demethylation program. *Dppa3* is expressed in the early stages of germline specification, and importantly, *in vivo* genetic analyses have revealed a role for DPPA3-mediated demethylation in PGCs, although the link with UHRF1 was not made (Nakashima et al., 2013). Incidentally, it has been reported that UHRF1 is downregulated during PGC progression at both the RNA and protein level (Kagiwada et al., 2013; Ohno et al., 2013). It will be necessary to eventually demonstrate if the DPPA3 phenomenology observed in mESCs occur at the relevant stages of *in vivo* development.

ALWAYS WEAR PROTECTION: KEEPING DEVELOPMENTAL GENES DNA METHYLATION FREE

Shortly after reaching its lowest global levels of DNA methylation, the embryo implants into the uterine wall, which coincides with upregulation of the *de novo* DNA methylation program. In a few short days, the genome becomes highly DNA methylated. DNMT3A and DNMT3B show preference for histone 3 lysine 36 di- and trimethylation (H3K36me2/3)-marked regions, respectively, via binding by their PWWP domains (Dhayalan et al., 2010). H3K36me2/3 are broadly deposited in the genome, and may serve to enhance DNMT3A/B activity (Baubec et al., 2015; Weinberg et al., 2019). In general though, the *de novo* methyltransferases exhibit very little discrimination for genomic compartments, with one key exception: CpG island (CGI) promoters, which are distinguished by their markedly high CpG content. Roughly two-thirds of promoters are CGIs, and comprise most housekeeping and developmental genes (Gardiner-Garden and Frommer, 1987; Larsen et al., 1992; Ku et al., 2008). There is nothing about the sequence, *per se*, that should repel *de novo* DNA methylation; in fact, DNMT3A/B show a preference for CpG-rich sequences (Baubec et al., 2015). However, keeping promoters free of methylation is absolutely crucial for proper cellular function; DNA methylation is a very stable epigenetic mark, and ectopic promoter methylation can lead to long-term silencing of important genes. The ADD domain, also harbored by both enzymes, is repelled by H3K4 methylation (Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010; Guo et al., 2015). Given that H3K4me3 is strongly linked with active promoters, therein lies a simple mechanism to protect promoter sequences from DNA methylation deposition.

During the dramatic flux of DNA methylation in embryogenesis, there must be a means by which genes that are not expressed during reprogramming do not become unwilling targets of DNMT3A/B. Several years ago the discovery was made in mESCs that a large proportion of developmental gene promoters are "bivalently" marked—that is marked by both H3K4me3 and the polycomb repressive complex 2 (PRC2)-deposited H3K27me3 (Azuara et al., 2006; Bernstein et al., 2006). Also known as poised genes, these genes are silent, thanks to PRC2-mediated repression, but at the same time protected from DNA methylation and ready to activate upon the proper developmental cue. Moreover, there is some evidence that H3K27me3 marked regions may be refractory to *de novo* methylation independently of the H3K4 mark (Greenberg et al., 2017; Li et al., 2018b). Finally, TET1 is enriched at bivalent gene promoters (Manzo et al., 2017; Gu et al., 2018). Incidentally, TET1 contains a CxxC domain, which binds specifically to unmethylated CpG-rich sequences, however this domain does not appear to determine TET1 CGI localization (Zhang et al., 2016). Both TET proteins and KDM2B—a CxxC-domain containing complex associated with PRC1—have been demonstrated to protect CGIs from *de novo* DNA methylation (Boulard et al., 2015; Verma et al., 2018). Thus, there are several layers of regulation to prevent DNA methylation-based silencing.

In addition to TET and polycomb action, there must be a sequence-based recruitment of H3K4 methyltransferase complexes, i.e., through transcription factors. CGIs serve as important platforms of transcription factor binding, which is associated with alterations in DNA methylation level (Lienert et al., 2011; Krebs et al., 2014). Indeed, integration of a number of genome-wide datasets indicated the main determinant of sequences that do not exhibit DNA methylation is transcription factor binding (Kremsky and Corces, 2020). Recently, two studies converged on two factors that play a role in protecting bivalent genes: DPPA2 and DPPA4 (Eckersley-Maslin et al., 2020; Gretarsson and Hackett, 2020). While both studies utilized an mESC system, the discoveries were through different means. DPPA2/4 are heterodimeric transcription factors that are known to play a role in zygotic genome activation (De Iaco et al., 2019; Eckersley-Maslin et al., 2019; Yan et al., 2019), but they also are bound to bivalent genes (Engelen et al., 2015; Hernandez et al., 2018; Klein et al., 2018). Intriguingly, *Dppa2/4* mutant mice exhibit developmental defects far after the embryonic stages where developmental genes are enriched for bivalent marks (Madan et al., 2009; Nakamura et al., 2011). Following up on this curiosity, Eckersley-Maslin et al. showed that in *Dppa2/4* mutant mESCs, a subset of DPPA2/4 targets lose both H3K4me3 and H3K27me3 enrichment, indicating a role for these transcription factors in recruiting both silencing and activating complexes (Eckersley-Maslin et al., 2020) (**Figure 2**). Moreover, during differentiation to embryoid bodies, this subset of genes

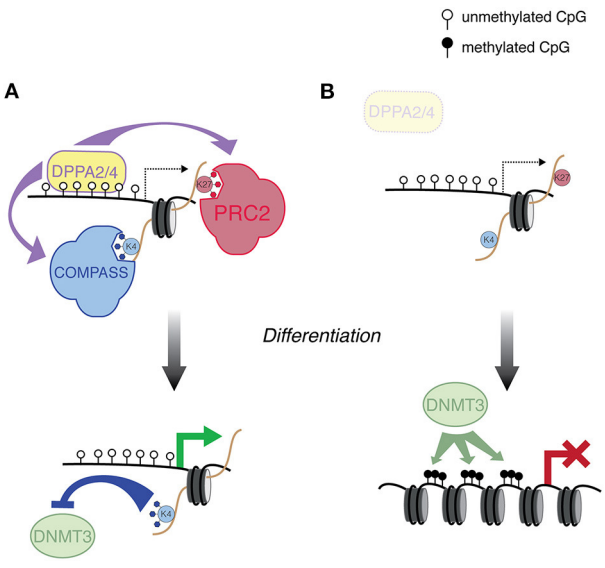


FIGURE 2 | DPPA2/4-mediated regulation of developmental genes. **(A)** DPPA2/4 bind a subset of developmental genes, and recruit the COMPASS and PRC2 complexes, which deposit H3K4me3 and H3K27me3, respectively. These genes are then poised to activate in the proper developmental context. The H3K4me3 protects this class of promoter from *de novo* DNA methylation. **(B)** In *Dppa2/4* mutant ESCs, genes that are normally targeted by DPPA2/4 are susceptible to aberrant *de novo* DNA methyltransferase activity, preventing the ability of this class of genes to activate during differentiation.

acquire DNA methylation and fail to activate, likely because they no longer are protected from *de novo* DNA methyltransferases (**Figure 2**).

Gretarsson and Hackett ultimately arrived at similar conclusions, although they discovered a role for DPPA2/4 through different means. Using a clever mESC-based CRISPR screen, they searched for mutants that failed to activate a methylated reporter during an *in vitro* model of global DNA demethylation (Gretarsson and Hackett, 2020). *Dppa2/4* were two of the top hits from the screen, but mutations did not lead to global DNA methylation perturbations. Consistent with the study from Eckersley-Maslin, DNA methylation abnormalities were restricted to a proportion of bivalently marked gene promoters. Curiously, DPPA2/4 also target LINE1 retrotransposons, although their role in transposon regulation is less clear. It should be noted that DNA methylation defects do not occur at all DPPA2/4 targets in the double mutant background; moreover, there are bivalent promoters that are not DPPA2/4 targets. That is all to say, DPPA2/4 are only two of the transcription factors important for DNA methylation protection during epigenetic reprogramming and there is a high probability others exist. Nevertheless, these are two important studies that undergird a model wherein sequence-specific transcription factors not only prime genes for proper expression during development, but also play a protective role against ectopic silencing.

CONCLUDING REMARKS: DPPA3, NOT JUST A BASIC PROTEIN

DNA methylation, if misregulated, can have dire consequences. As discussed above, the improper *de novo* DNA methylation of promoters can lead to stable silencing of key developmental regulators. Perhaps more worrisome, the absence of DNA methylation can result not only in the ectopic expression of protein coding genes, but massive transposon derepression, which can have far ranging and deleterious effects (Walsh et al., 1998). Mouse embryos lacking either *de novo* or maintenance DNA methylation machinery die shortly after gastrulation (Li et al., 1992; Okano et al., 1999). Therefore, mammals must have evolved compensatory mechanisms to control the genome during not one, but two waves of DNA methylation reprogramming (Walter et al., 2016; Hill et al., 2018). What is more, the lowest levels of DNA methylation occur in pluripotent stem cells and primordial germ cells—the cell types that give rise to all somatic tissues and the germline, respectively. It is difficult to imagine cells that are more important for proper organismal development and transmission of genetic material to the next generation. Why then do mammals take such extraordinary apparent risks with their genome? Even among vertebrates this phenomenon is odd: zebrafish, by comparison, undergo nothing so drastic with their methylome during embryonic development (Skvortsova et al., 2019).

In investigating the evolutionary conservation of the *Dppa3* gene, Mulholland et al. revealed it is not found outside mammals. Amazingly, when the mouse DPPA3 protein was incubated with the egg extracts of the amphibian *Xenopus laevis*, the mouse protein inhibited the frog UHRF1. Furthermore, when the fertilized eggs from the model fish medaka were injected with *Dppa3* mRNA, the embryos exhibited dramatic hypomethylation (Mulholland et al., 2020). Therefore, DPPA3 has evolved as a potent DNA demethylation factor that can disrupt UHRF1 function in distant species. The authors suggest that perhaps *Dppa3* arose in the mammalian lineage in concert with the role of transposon expression regulating the early transcription program. This certainly is possible, however it does not explain why global demethylation occurs exclusively in mammals, and also why it happens twice in development.

It is notable that *Dppa3* is only present in placental mammals; marsupials and egg-laying monotremes lack the gene. Classic experiments in the 1980s demonstrated that mouse paternal and maternal genomes are not equivalent—androgenetic diploids exhibit robust placenta, whereas conversely gynogenetic diploids harbor severely undersized placenta (Surani and Barton, 1983; Barton et al., 1984; McGrath and Solter, 1984; Surani et al., 1984). This result is consistent with genetic conflict theory, which states that there is a conflict between the parental genes with regards to offspring development (Moore and Haig, 1991). In this case, the paternally expressed genes promote larger placenta leading to greater resource allocation to the developing embryo and fetus at the mother's expense, and for the maternally expressed genes it is the inverse. This theory was developed to explain the existence of genomic imprinting, which is parent-specific gene expression controlled by DNA methylation patterns inherited from the gametes. Pertinently, neither androgenetic nor gynogenetic embryos are viable. Given the stark DNA methylation asymmetry between gametes, perhaps *Dppa3* evolved in order to prevent either the paternal or maternal genome exerting too much control with regards to inherent conflict. Imprint control regions, in turn, evolved mechanisms to escape DNA methylation erasure during embryogenesis. In the case of marsupials, there is evidence of germline reprogramming and genomic imprinting (Renfree et al., 2008; Suzuki et al., 2013; Ishihara et al., 2019), however evidence for embryonic reprogramming in both marsupials and monotremes is limited. Hopefully future studies will interrogate dynamic DNA methylation after conception in non-placental mammals, which will not only provide insights into DPPA3 function, but into evolutionary theory.

AUTHOR CONTRIBUTIONS

MVCG wrote the manuscript and prepared the figures.

FUNDING

MVCG was supported by a Laboratoire d'excellence Who Am I? (Labex 11-LABX-0071) Emerging Teams Grant and the European Research Council (ERC-StG-2019 DyNAmecs).

REFERENCES

Albert, J. R., Au Yeung, W. K., Toriyama, K., Kobayashi, H., Hirasawa, R., Brind'Amour, J., et al. (2020). Maternal DNMT3A-dependent *de novo* methylation of the paternal genome inhibits gene expression in the early embryo. *Nat. Commun.* 11:5417. doi: 10.1038/s41467-020-19279-7

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W., D'Souza, Z., et al. (2016). *De novo* DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* 18, 225–233. doi: 10.1038/ncb3296

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* 8, 532–538. doi: 10.1038/ncb1403

Barton, S. C., Surani, M. A., and Norris, M. L. (1984). Role of paternal and maternal genomes in mouse development. *Nature* 311, 374–376. doi: 10.1038/311374a0

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., et al. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520, 243–247. doi: 10.1038/nature14176

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326. doi: 10.1016/j.cell.2006.02.041

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., et al. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* 42, 1093–1100. doi: 10.1038/ng.708

Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764. doi: 10.1126/science.1147939

Boulard, M., Edwards, J. R., and Bestor, T. H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* 47, 479–485. doi: 10.1038/ng.3272

Carlson, L. L., Page, A. W., and Bestor, T. H. (1992). Subcellular localization and properties of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev.* 6, 2536–2541. doi: 10.1101/gad.6.12b.2536

Catania, S., Dumesic, P. A., Pimentel, H., Nasif, A., Stoddard, C. I., Burke, J. E., et al. (2020). Evolutionary persistence of DNA methylation for millions of years after ancient loss of a *de novo* methyltransferase. *Cell* 180, 263–277.e20. doi: 10.1016/j.cell.2020.02.006

De Iaco, A., Coudray, A., Duc, J., and Trono, D. (2019). DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep.* 20:e47382. doi: 10.15252/embr.201847382

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., et al. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* 285, 26114–26120. doi: 10.1074/jbc.M109.089433

Du, W., Dong, Q., Zhang, Z., Liu, B., Zhou, T., Xu, R.-M., et al. (2019). Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger–type E3 ubiquitin ligase UHRF1. *J. Biol. Chem.* 294, 8907–8917. doi: 10.1074/jbc.RA119.008008

Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., and Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* 33, 194–208. doi: 10.1101/gad.321174.118

Eckersley-Maslin, M. A., Parry, A., Blotenburg, M., Krueger, C., Ito, Y., Franklin, V. N. R., et al. (2020). Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. *Nat. Struct. Mol. Biol.* 27, 696–705. doi: 10.1038/s41594-020-0443-3

Engelen, E., Brandsma, J. H., Moen, M. J., Signorile, L., Dekkers, D. H. W., Demmers, J., et al. (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* 6:7155. doi: 10.1038/ncomms8155

Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* 13, 351–359. doi: 10.1016/j.stem.2013.06.004

Funaki, S., Nakamura, T., Nakatani, T., Umehara, H., Nakashima, H., and Nakano, T. (2014). Inhibition of maintenance DNA methylation by stella. *Biochem. Biophys. Res. Commun.* 453, 455–460. doi: 10.1016/j.bbrc.2014.09.101

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* 196, 261–282. doi: 10.1016/0022-2836(87)90689-9

Gkountela, S., Zhang, K. X., Shafiq, T. A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., et al. (2015). DNA demethylation dynamics in the human prenatal germline. *Cell* 161, 1425–1436. doi: 10.1016/j.cell.2015.05.012

Greenberg, M. V. C., Glaser, J., Borsos, M., Marjou, F. E., Walter, M., Teissandier, A., et al. (2017). Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat. Genet.* 49, 110–118. doi: 10.1038/ng.3718

Gretarsson, K. H., and Hackett, J. A. (2020). Dppa2 and Dppa4 counteract *de novo* methylation to establish a permissive epigenome for development. *Nat. Struct. Mol. Biol.* 27, 706–716. doi: 10.1038/s41594-020-0445-1

Grosswendt, S., Kretzmer, H., Smith, Z. D., Kumar, A. S., Hetzel, S., Wittler, L., et al. (2020). Epigenetic regulator function through mouse gastrulation. *Nature* 584, 102–108. doi: 10.1038/s41586-020-2552-x

Gu, T., Lin, X., Cullen, S. M., Luo, M., Jeong, M., Estecio, M., et al. (2018). DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol.* 19:88. doi: 10.1186/s13059-018-1464-7

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610. doi: 10.1038/nature10443

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* 15, 447–459. doi: 10.1016/j.stem.2014.08.003

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640–644. doi: 10.1038/nature13899

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* 13, 360–369. doi: 10.1016/j.stem.2013.06.002

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., et al. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15–23. doi: 10.1016/S0925-4773(02)00181-8

Han, L., Ren, C., Zhang, J., Shu, W., and Wang, Q. (2019). Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. *Cell Discov* 5:9. doi: 10.1038/s41421-019-0081-2

Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., et al. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucl. Acids Res.* 40, 4841–4849. doi: 10.1093/nar/gks155

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi: 10.1126/science.1210944

Hernandez, C., Wang, Z., Ramazanov, B., Tang, Y., Mehta, S., Dambrot, C., et al. (2018). Dppa2/4 facilitate epigenetic remodeling during reprogramming to pluripotency. *Cell Stem Cell* 23, 396–411.e8. doi: 10.1016/j.stem.2018.08.001

Hill, S., Leitch, H. G., Requena, C. E., Sun, Z., Amouroux, R., Roman-trufero, M., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* 555, 392–396. doi: 10.1038/nature25964

Howell, C. Y., Bestor, T. H., Ding, F., Latham, K. E., Mertineit, C., Trasler, J. M., et al. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* 104, 829–838. doi: 10.1016/S0092-8674(01)00280-X

Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylicz, J. J., et al. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *Elife* 6:e22345. doi: 10.7554/eLife.22345

Iqbal, K., Jin, S.-G., Pfeifer, G. P., and Szabó, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3642–3647. doi: 10.1073/pnas.1014033108

Ishihara, T., Hickford, D., Shaw, G., Pask, A. J., and Renfree, M. B. (2019). DNA methylation dynamics in the germline of the marsupial tammar wallaby, *Macropus eugenii*. *DNA Res.* 26, 85–94. doi: 10.1093/dnares/dsy040

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., et al. (2015). Early embryonic-like cells are induced by

downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671. doi: 10.1038/nsmb.3066

Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., and Torres-Padilla, M. E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* 49, 1502–1510. doi: 10.1038/ng.3945

Ji, D., Lin, K., Song, J., and Wang, Y. (2014). Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. BioSyst.* 10, 1749–1750. doi: 10.1039/c4mb00150h

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* 32, 340–353. doi: 10.1038/emboj.2012.331

Klein, R. H., Tung, P.-Y., Somanath, P., Fehling, H. J., and Knoepfler, P. S. (2018). Genomic functions of developmental pluripotency associated factor 4 (Dppa4) in pluripotent stem cells and cancer. *Stem Cell Res.* 31, 83–94. doi: 10.1016/j.scr.2018.07.009

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., et al. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish Oocyte-specific heritable marks. *PLoS Genet.* 8:e1002440. doi: 10.1371/journal.pgen.1002440

Krebs, A. R., Dessus-Babus, S., Burger, L., and Schübeler, D. (2014). High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *Elife* 3:e04094. doi: 10.7554/eLife.04094.019

Kremsky, I., and Corces, V. G. (2020). Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. *Genome Biol.* 21:118. doi: 10.1186/s13059-020-02036-w

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4: e1000242. doi: 10.1371/journal.pgen.1000242

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics* 13, 1095–1107. doi: 10.1016/0888-7543(92)90024-M

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., et al. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* 129, 1807–1817.Available online at: https://dev.biologists.org/content/129/8/1807.

Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* 20, 311–316. doi: 10.1038/nsmb.2510

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926. doi: 10.1016/0092-8674(92)90611-F

Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., et al. (2018a). Stella safeguards the oocyte methylome by preventing *de novo* methylation mediated by DNMT1. *Nature* 564, 136–140. doi: 10.1038/s41586-018-0751-5

Li, Y., Zheng, H., Wang, Q., Zhou, C., Wei, L., Liu, X., et al. (2018b). Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol.* 19:18. doi: 10.1186/s13059-018-1390-8

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* 43, 1091–1097. doi: 10.1038/ng.946

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi: 10.1038/nature11244

Madan, B., Madan, V., Weber, O., Tropel, P., Blum, C., Kieffer, E., et al. (2009). The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. *Mol. Cell Biol.* 29, 3186–3203. doi: 10.1128/MCB.01970-08

Maiti, A., and Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338. doi: 10.1074/jbc.C111.284620

Manzo, M., Wirz, J., Ambrosi, C., Villaseñor, R., Roschitzki, B., and Baubec, T. (2017). Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J.* 36, 3421–3434. doi: 10.15252/embj.201797038

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* 403, 501–502. doi: 10.1038/35000656

McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37, 179–183. doi: 10.1016/0092-8674(84)90313-1

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99, 371–382.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* 7, 45–49. doi: 10.1016/0168-9525(91)90040-W

Mulholland, C. B., Nishiyama, A., Ryan, J., Nakamura, R., Yigit, M., Glück, I. M., et al. (2020). Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nat. Commun.* 11:5972. doi: 10.1038/s41467-020-19603-1

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* 9, 64–71. doi: 10.1038/ncb1519

Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., et al. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415–419. doi: 10.1038/nature11093

Nakamura, T., Nakagawa, M., Ichisaka, T., Shiota, A., and Yamanaka, S. (2011). Essential roles of ECAT15-2/Dppa2 in functional lung development. *Mol. Cell Biol.* 31, 4366–4378. doi: 10.1128/MCB.05701-11

Nakashima, H., Kimura, T., Kaga, Y., Nakatani, T., Seki, Y., Nakamura, T., et al. (2013). Effects of dppa3 on DNA methylation dynamics during primordial germ cell development in mice. *Biol. Reprod.* 88:125. doi: 10.1095/biolreprod.112.105932

Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., et al. (2013). A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* 140, 2892–2903. doi: 10.1242/dev.093229

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., et al. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* 10:e1004868. doi: 10.1371/journal.pgen.1004868

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 99, 247–257. doi: 10.1016/S0092-8674(00)81656-6

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* 448, 714–717. doi: 10.1038/nature05987

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., et al. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* 10, 475–478. doi: 10.1016/S0960-9822(00)00448-6

Otani, J., Kimura, H., Sharif, J., Endo, T. A., Mishima, Y., Kawakami, T., et al. (2013). Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PLoS ONE* 8:e82961. doi: 10.1371/journal.pone.0082961

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep.* 10, 1235–1241. doi: 10.1038/embor.2009.218

Payer, B., Saitou, M., Barton, S. C., Thresher, R., Dixon, J. P. C., Zahn, D., et al. (2003). Stella is a maternal effect gene required for normal early development in mice. *Curr. Biol.* 13, 2110–2117. doi: 10.1016/j.cub.2003.11.026

Renfree, M. B., Ager, E. I., Shaw, G., and Pask, A. J. (2008). Genomic imprinting in marsupial placentation. *Reproduction* 136, 523–531. doi: 10.1530/REP-08-0264

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., et al. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* 48, 849–862. doi: 10.1016/j.molcel.2012.11.001

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912. doi: 10.1038/nature06397

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* 15, 459–471. doi: 10.1016/j.stem.2014.09.002

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., et al. (2013). Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet*. 9:e1003439. doi: 10.1371/journal.pgen.1003439

Skvortsova, K., Tarbashevich, K., Stehling, M., Lister, R., Irimia, M., Raz, E., et al. (2019). Retention of paternal DNA methylome in the developing zebrafish germline. *Nat. Commun*. 10:3054. doi: 10.1038/s41467-019-10895-6

Smith, Z. D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., et al. (2017). Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* 549, 543–547. doi: 10.1038/nature23891

Surani, M. A., and Barton, S. C. (1983). Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. *Science* 222, 1034–1036. doi: 10.1126/science.6648518

Surani, M. A., Barton, S. C., and Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308, 548–550. doi: 10.1038/308548a0

Suzuki, S., Shaw, G., and Renfree, M. B. (2013). Postnatal epigenetic reprogramming in the germline of a marsupial, the tammar wallaby. *Epigenetics Chromatin* 6:14. doi: 10.1186/1756-8935-6-14

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116

van Mierlo, G., Dirks, R. A. M., De Clerck, L., Brinkman, A. B., Huth, M., Kloet, S. L., et al. (2019). Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-state pluripotency. *Cell Stem Cell* 24, 123–137.e8. doi: 10.1016/j.stem.2018.10.017

Verma, N., Pan, H., Doré, L. C., Shukla, A., Li, Q. V., Pelham-Webb, B., et al. (2018). TET proteins safeguard bivalent promoters from *de novo* methylation in human embryonic stem cells. *Nat. Genet*. 50, 83–95. doi: 10.1038/s41588-017-0002-y

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., et al. (2016). Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* 62, 848–861. doi: 10.1016/j.molcel.2016.04.025

Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet*. 20, 116–117. doi: 10.1038/2413

Walter, M., Teissandier, A., Pérez-Palacios, R., Bourc'his, D., Akalin, A., Kormaksson, M., et al. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* 5:e11418. doi: 10.7554/eLife.11418

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., et al. (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol*. 20, 620–631. doi: 10.1038/s41556-018-0093-4

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell* 157, 979–991. doi: 10.1016/j.cell.2014.04.017

Weber, A. R., Krawczyk, C., Robertson, A. B., Kuśnierczyk, A., Vågbø, C. B., Schuermann, D., et al. (2016). Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* 7:10806. doi: 10.1038/ncomms10806

Weigel, D., and Colot, V. (2012). Epialleles in plant evolution. *Genome Biol*. 13:249. doi: 10.1186/gb-2012-13-10-249

Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573, 281–286. doi: 10.1038/s41586-019-1534-3

Woo, H. R., Dittmer, T. A., and Richards, E. J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet*. 4:e1000156. doi: 10.1371/journal.pgen.1000156

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., et al. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2:1240. doi: 10.1038/ncomms1240

Yan, Y.-L., Zhang, C., Hao, J., Wang, X.-L., Ming, J., Mi, L., et al. (2019). DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program. *PLoS Biol*. 17:e3000324. doi: 10.1371/journal.pbio.3000324

Zhang, W., Xia, W., Wang, Q., Towers, A. J., Chen, J., Gao, R., et al. (2016). Isoform switch of TET1 regulates DNA demethylation and mouse development. *Mol. Cell* 64, 1062–1073. doi: 10.1016/j.molcel.2016.10.030

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucl. Acids Res*. 38, 4246–4253. doi: 10.1093/nar/gkq147

Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., et al. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat. Genet*. 50, 96–105. doi: 10.1038/s41588-017-0003-x

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., et al. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet*. 50, 12–19. doi: 10.1038/s41588-017-0007-6

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

REVIEW
published: 15 January 2021
doi: 10.3389/fcell.2020.623948

TET-Mediated Epigenetic Regulation in Immune Cell Development and Disease

*Nikolas James Tsiouplis[1], David Wesley Bailey[1,2,3], Lilly Felicia Chiou[4], Fiona Jane Wissink[1] and Ageliki Tsagaratou[1,2,3,4,5,6]**

[1] University of North Carolina Lineberger Comprehensive Cancer Center, Chapel Hill, NC, United States, [2] University of North Carolina Center of Translational Immunology, Chapel Hill, NC, United States, [3] University of North Carolina Institute of Inflammatory Disease, Chapel Hill, NC, United States, [4] University of North Carolina Curriculum in Genetics and Molecular Biology, Chapel Hill, NC, United States, [5] University of North Carolina Department of Genetics, Chapel Hill, NC, United States, [6] University of North Carolina Department of Microbiology and Immunology, Chapel Hill, NC, United States

TET proteins oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and further oxidation products in DNA. The oxidized methylcytosines (oxi-mCs) facilitate DNA demethylation and are also novel epigenetic marks. TET loss-of-function is strongly associated with cancer; *TET2* loss-of-function mutations are frequently observed in hematological malignancies that are resistant to conventional therapies. Importantly, TET proteins govern cell fate decisions during development of various cell types by activating a cell-specific gene expression program. In this review, we seek to provide a conceptual framework of the mechanisms that fine tune TET activity. Then, we specifically focus on the multifaceted roles of TET proteins in regulating gene expression in immune cell development, function, and disease.

Keywords: TET proteins, epigenetics, 5hmC, immune cell development, cancer

OPEN ACCESS

Edited by:
José Luis Sardina, Josep Carreras Leukaemia Research Institute (IJC), Spain

Reviewed by:
Luisa Cimmino, University of Miami, United States
Jose Javier Fuster, Spanish National Centre for Cardiovascular Research, Spain
Andrew Das, University of Otago, New Zealand

***Correspondence:**
Ageliki Tsagaratou
ageliki_tsagaratou@med.unc.edu

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *30 October 2020*
Accepted: *21 December 2020*
Published: *15 January 2021*

Citation:
Tsiouplis NJ, Bailey DW, Chiou LF, Wissink FJ and Tsagaratou A (2021) TET-Mediated Epigenetic Regulation in Immune Cell Development and Disease. Front. Cell Dev. Biol. 8:623948. doi: 10.3389/fcell.2020.623948

INTRODUCTION

Gene expression in mammalian cells is a highly regulated process whereby transcription factors (TFs) bind to specific DNA-binding motifs within promoters and enhancers in distinct cell types, causing them to differentiate and acquire new cell fates (Roadmap Epigenomics Consortium et al., 2015). These processes are both spatially and temporally specific, resulting in the varied assortment of cell types observed in mammals. While some of the TFs are ubiquitously expressed, others exhibit a cell-specific expression pattern. In some cases, the same TF can regulate different genes in different cells, highlighting the dynamic nature of regulatory networks across the organism (Lambert et al., 2018). Epigenetic markers provide an additional component of regulation to this process by modifying the accessibility of the histones surrounding DNA, or even the DNA itself (Bernstein et al., 2007). Two of the primary epigenetic modifications are histone post-translational modifications (Zhou et al., 2011) and DNA methylation (Smith and Meissner, 2013). In mammalian cells, transcription of the vast majority of protein-coding genes starts at promoters, which are rich in CG sequences (Bogdanovic and Lister, 2017). DNA methylation occurs on cytosine bases within CpG islands (Lister et al., 2009). DNA methylation of cytosine involves the covalent addition of a methyl group at position 5 of the pyrimidine ring of cytosine and is achieved by the catalytic activity of the family of DNA methyltransferases (DNMTs) (Goll and Bestor, 2005), which consists of DNMT1, DNMT2, DNMT3a, and DNMT3b. In the human genome, 60–80% of 28 million CpG dinucleotides are methylated (Lister et al., 2009; Ziller et al., 2013).

Genome-wide studies using bisulfite sequencing to assess cytosine methylation have established that highly transcribed genes have sparsely methylated CpG promoters, whereas silenced, non-transcribed genes show high levels of cytosine methylation in the CpG context of their promoters (Lister et al., 2009; Laurent et al., 2010). Methylation of repetitive DNA sequences, found close to centromeres, is instrumental in the maintenance of genomic integrity. In mice, repetitive DNA can be distinguished in major satellites (243 bp repeat sequences) found in the pericentromeric region as well as in minor satellites (120 bp repeat sequences) found in the centromeric region (Guenatri et al., 2004). Aberrations in DNA demethylation are a hallmark of cancer and can result in silencing of tumor suppressor genes by increasing the methylated cytosines at their promoters. Conversely, global hypomethylation leads to genomic instability (Baylin and Jones, 2011).

Previously, it was believed that DNA methylation was an irreversible event that could only be removed passively *via* dilution during DNA replication. However, the Ten Eleven Translocation (TET) family of proteins has been shown to catalyze the subsequent oxidations of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Tahiliani et al., 2009; He et al., 2011; Ito et al., 2011) (**Figure 1**). TET proteins therefore provide an active pathway for DNA demethylation and consequently have relevance for regulation of gene expression. TET proteins mediate "active" (replication-independent) DNA demethylation *via* excision of 5fC and 5caC by thymine DNA glycosylase (TDG). Afterwards, base excision repair machinery substitutes the excised base with an unmethylated cytosine (Branco et al., 2012; Pastor et al., 2013). Notably, the majority of 5hmC is passively diluted *via* replication (Tsagaratou et al., 2014; Nestor et al., 2015) (**Figure 1**).

Indicating a conserved role in controlling DNA demethylation, representatives of the TET/JBP superfamily have been reported in every metazoan organism (Iyer et al., 2009; Pastor et al., 2013). In mammalian cells specifically, there are three TET proteins: TET1, TET2, and TET3. TET1 was identified as a fusion partner of the mixed-lineage leukemia (MLL) gene from the breakpoint of chromosomal translocation t(10;11)(q22;q23) in acute myeloid leukemia (AML) (Lorsbach et al., 2003). Studying mouse models over the life course has shown that TET1 and TET2 are most highly expressed in the inner cell mass and embryonic stem (ES) cells (Tahiliani et al., 2009; Koh et al., 2011). TET2 is expressed at lower levels than TET1 in ES cells, and its expression first drops and then increases upon differentiation; it is expressed in numerous differentiated organs and cell types in the adult (Pastor et al., 2013; Tsagaratou and Rao, 2014). TET1 is also highly expressed in primordial germ cells (PGCs) (Hackett et al., 2013; Vincent et al., 2013), while TET2 and TET3 are highly expressed throughout the remainder of development. TET3 exhibits high expression in oocytes and zygotes (Gu et al., 2011), and loss of TET3 in mice results in perinatal lethality (Pastor et al., 2013). Both TET1 and TET2 are implicated in cancer. TET1 is an MLL partner in cases of acute myeloid (AML) and lymphoid (ALL) leukemias, while loss of function of TET2 is strongly associated with myelodysplastic syndromes, myeloproliferative neoplasms, and myeloid leukemias (Ko et al., 2010).

TET proteins arose from a common ancestral gene that underwent triplication in jawed vertebrates. TET1 and TET3 have greater structural similarities, as they share an N-terminal CXXC DNA binding domain. However, TET2 lacks a CXXC domain and thus cannot directly bind to DNA. During evolution, the ancestral *Tet2* gene underwent a chromosomal inversion that resulted in separation of the TET2 CXXC DNA binding domain from the rest of the protein. The CXXC DNA binding domain of TET2 became a separate gene known as IDAX (Iyer et al., 2009; Ko et al., 2013) (**Figure 2**). The core catalytic domain on each TET protein is comprised of a cysteine-rich domain, a conserved double-stranded β-helix (DSBH) domain, and binding sites for the cofactors α-ketoglutarate (α-KG) and Fe (II) (Pastor et al., 2013). Studies have indicated that these catalytic domains preferentially bind to cytosines on CpG islands without interacting with adjacent bases (Pastor et al., 2013) (**Figure 2**).

5hmC is found at different levels in mammalian cells. It is most abundant in Purkinje neurons, where it comprises ∼40% of 5mC levels (Kriaucionis and Heintz, 2009). In ES cells, the levels of 5hmC vary between 5 and 10% of the levels of 5mC, whereas it is present at only 1% of the total level of 5mC in some immune populations (Ko et al., 2010). 5hmC in gene bodies and enhancers has been positively correlated with increased gene expression in various cell types such as neural cells (Mellen et al., 2012; Lister et al., 2013), T cells (Tsagaratou et al., 2014; Ichiyama et al., 2015), B cells (Lio et al., 2016; Orlanski et al., 2016), and spermatogenic cells (Gan et al., 2013). TET-mediated DNA demethylation in distal enhancers occurs at a higher rate than passive demethylation (Ginno et al., 2020). 5fC and 5caC are even less abundant compared to 5hmC (He et al., 2011). In addition to their role in mediating DNA demethylation, the oxidative derivatives of TET function−5hmC, 5fC, and 5caC—are also stable epigenetic marks (Bachman et al., 2014, 2015) that can be specifically recognized and preferentially bound by readers—mainly transcriptional regulators—to impact transcriptional elongation, genomic integrity, and DNA repair (Yildirim et al., 2011; Mellen et al., 2012; Iurlaro et al., 2013; Spruijt et al., 2013; Hashimoto et al., 2014; Xiong et al., 2016; Chen et al., 2017; Cimmino and Aifantis, 2017; Tsagaratou et al., 2017a; Wu and Zhang, 2017; Parry et al., 2020; Shukla et al., 2020) (**Figure 1**). For instance, Ubiquitin-like protein containing PHD and RING finger domains 1 (UHRF1) as well as methyl-CpG-binding protein 2 (MeCP2) can bind 5mC and 5hmC (Frauer et al., 2011; Mellen et al., 2012). It was suggested that 5hmC binding by MeCP2 in neural cells alters chromatin structure and facilitates gene expression (Mellen et al., 2012). Experiments in mESCs revealed that 5fC and 5caC are involved in specific binding interactions with a greater number of proteins in comparison to 5hmC (Spruijt et al., 2013). During their cell cycle-independent removal by the base excision repair pathway, 5fC and 5caC recruit an increased number of DNA repair proteins compared to 5hmC (Spruijt et al., 2013). Moreover, 5caC can be recognized by basic helix-loop-helix proteins such as MAX and TCF4 (Wang et al., 2017; Yang et al., 2019) as well as Wilms tumor protein (Hashimoto et al., 2014).

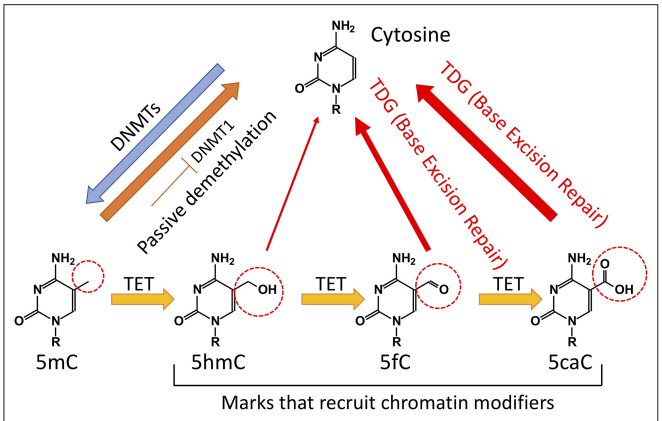


FIGURE 1 | Regulation of DNA methylation in mammalian cells. Cytosine (C) is methylated by DNA methyltransferases (DNMTs) to 5-methylcytosine (5mC). Cytosine demethylation can occur in the absence of enzymatic activity during cell division. In addition, Ten Eleven Translocation (TET) proteins can oxidize 5mC to 5-hydroxymethylcytosine (5hmC). A significant portion of 5hmC will be diluted during cell division. TET proteins can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The TDG through the Base Excision Repair (BER) can convert 5fC and 5caC to unmodified C.

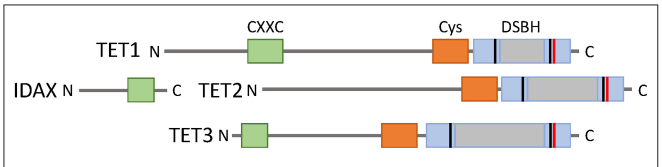


FIGURE 2 | The TET family of proteins. TET1, TET2, and TET3 share a C-terminal catalytic domain consisting of cysteine-rich (orange) and double stranded β-helix (gray) domains, and binding sites for cofactors Fe(II) (black) and 2-oxoglutarate (red). TET1 and TET3 have an N-terminal CXXC DNA binding domain, but this was lost in TET2 from a chromosomal inversion and became a separate protein IDAX.

In addition to promoting binding of some transcription factors, modified cytosines can inhibit binding of transcription factors and transcriptional activators to suppress gene expression. For instance, the presence of 5hmC within the sequence of a cAMP response element (CRE) at an artificial promoter can decrease the binding of the transcriptional activator c-AMP Response Element Binding (CREB) protein, resulting in decreased expression of the target genes (Kitsera et al., 2017). On the other hand, oxi-mCs can also prevent binding of transcriptional repressors and thus promote gene expression. MDB1 can specifically bind to 5mC but not to oxi-mCs and recruit the histone methyltransferase SETDB1 to promote H3K9 methylation and repress the expression of proopiomelanocortin (*Pomc*) (Marco et al., 2016). Thus, the presence of oxi-mCs prevents binding of SETDB1 and promotes the *Pomc* expression.

In the next sections we will first summarize the known mechanisms that regulate TET function. Although the majority of the described mechanistic studies have been performed in embryonic stem cells or cell lines, we should note that these mechanisms might be applicable to an array of other cell types, including immune cells, the major focus of this review. Then, we will discuss in detail how TET proteins shape immune cell development and function.

MECHANISMS OF TET FUNCTION

Competition Among TET Proteins and DNMTs

Mammalian genomes maintain high levels of CpG methylation (Lister et al., 2009; Ziller et al., 2013) even though the enzymes that regulate DNA methylation, DNMTs, and TET proteins are concomitantly expressed. Various studies have suggested a dynamic regulation of DNA methylation that is achieved through focal competition between TET proteins and DNMTs in pluripotent cells. Bivalent promoters are marked simultaneously by H3K4me3 and H3K27Ac but exhibit low levels of DNA methylation (Mikkelsen et al., 2007). However, loss of TET1, TET2, and TET3 resulted in aberrant hypermethylation of bivalent promoters given that DNMT3B could act without any competition from TET proteins on these genomic loci (Verma et al., 2018). As a result, the expression of lineage-specifying transcription factors was prohibited, and proper cellular differentiation was hindered (Verma et al., 2018). TET1-specific chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments in mESCs revealed that TET1 binds to bivalent promoters (Xu et al., 2011). TET1 seems to exclude DNMT3A1, the longer isoform of DNMT3A, from proximal promoters and canyons where TET1 seems to preferentially bind in embryonic stem cells (Gu et al., 2018). Moreover, it was shown that TET proteins compete with DNMT proteins to regulate the methylation status of enhancers (Verma et al., 2018; Charlton et al., 2020).

TET Proteins and Interacting Partners

TET proteins mediate a cell-specific, focal DNA demethylation. This is broadly achieved by interaction with transcription factors that mediate the recruitment of TET proteins onto the DNA. For instance, SALL4 in enhancers of mouse embryonic stem cells (ESCs) interacts with TET1 and binds to 5hmC (Xiong et al., 2016). Subsequently, SALL4 mediates the recruitment of TET2 that further oxidizes 5hmC (Xiong et al., 2016). This stepwise oxidation of 5mC to other oxi-mCs tightly regulates gene expression of developmental genes in mouse ESCs (Xiong et al., 2016). In addition, TET1 and TET2 can interact with Nanog to enhance reprogramming efficiency in a catalytic-dependent manner (Costa et al., 2013). RINF (also known as CXXC5) can form a complex with NANOG, OCT4, TET1, and TET2, facilitating the recruitment of the complex to the DNA; RINF also regulates the expression of TET proteins (Ravichandran et al., 2019).

TET proteins interact with various heterochromatin-associated proteins such as SIN3A, HDAC1, and HDAC2 (Ficz et al., 2011; Williams et al., 2011). This can affect chromatin modifications and ultimately impact gene expression. For example, TET proteins interact with OGlcN-Acetyl Transferase (OGT), subsequently impacting histone modifications and gene expression (Chen et al., 2013; Deplus et al., 2013). TET

interaction with OGT can also impact TET protein stability (Shi et al., 2013) and activity (Vella et al., 2013).

Moreover, TET proteins interact with components of the Base Excision Repair Complex (BER), such as PARP1, LIG3, and XRCC1 (Muller et al., 2014), as well as DNA glycosylases, such as Thymine DNA glycosylase, NEIL, and MBD4 (Muller et al., 2014), therefore suggesting a role in DNA repair. 5hmC has been found to be increased in cells upon treatment with DNA-damaging agents in cell lines (Kafer et al., 2016). Deletion of TET1 results in increased accumulation of DNA breaks as evaluated by increased staining for γH2Ax (Kafer et al., 2016).

Post-translational Modifications and DNA Binding

How the tri-dimensional structure of TET proteins is controlled remains less understood. Recent studies in the past years revealed that TET proteins are post-translationally modified. The interaction with OGT results in O-GlcNAcylation of TET1 and TET2 in ESCs (Vella et al., 2013). In addition, all three TET proteins can be phosphorylated (Bauer et al., 2015) at their N-terminus as well as at the low-complexity insert region between the two parts responsible for dioxygenase activity. Interestingly, there seems to be competition between O-GlcNAcylation and phosphorylation. Indeed, some peptides have both post-translational modifications (PTMs) (Bauer et al., 2015). These protein sequences within TET proteins could act as PTM switches that influence the PTM pattern on the neighboring amino acid (Bauer et al., 2015). For TET2 and TET3, a variety of PTMs are observed in highly modified regions. In the case of TET2, phosphorylation and O-GlcNacylation do not co-occur at the same amino acid. For TET3, however, the same amino acid could have both PTMs in some cases. PTMS in TET1 were more isolated. This dynamic interplay of phosphorylation and O-GlcNAcylation could facilitate dynamic changes in TET protein localization, activity, or targeting to genomic loci in response to external stimuli or environmental cues (Bauer et al., 2015).

AMP-activated protein kinase (AMPK) can phosphorylate murine TET2 at the serine residue 97 (Wu et al., 2018; Zhang et al., 2019). This phosphorylation event stabilizes TET2 (Wu et al., 2018; Zhang et al., 2019) which can then demethylate enhancers as C2C12 cells differentiate to myotubes (Zhang et al., 2019).

Moreover, it has been reported that TET conformation and DNA-binding ability can be affected by ubiquitination (Nakagawa et al., 2015). Specifically, VprBP interacts with TET2 by binding to the C-terminal dioxygenase catalytic domain of TET2. VprBP can also bind to the catalytic domain of TET1 and TET3 (Nakagawa et al., 2015). Notably, deletion of VprBP in mouse embryonic fibroblasts (MEFs) results in reduction of 5hmC, suggesting that VprBP is essential for TET protein function (Nakagawa et al., 2015).

TET2 can be acetylated by p300 at lysine K110 and deacetylated by HDAC1/2 (Zhang et al., 2017). Acetylation increases TET2 activity and stability as well as the interaction of TET2 with DNMT1, which targets TET2 to chromatin (Zhang et al., 2017). Importantly, oxidative stress can target the TET2/DNMT1 complex to chromatin, resulting in elevated DNA methylation and hydroxymethylation (Zhang et al., 2017). Loss of TET2 and subsequent induction of oxidative stress results in aberrant gain of methylation at CGI promoters and enhancers. Acetylation of TET2 can also increase interaction with DNMT3b; however, DNMT3b cannot target TET2 to chromatin as DNMT1 (Zhang et al., 2017). In addition, p300 can acetylate TET1 and TET3, but this most likely occurs in different lysine residues since K110 is not conserved among TET proteins (Zhang et al., 2017). Deacetylation of TET2 results in disassembly from DNMT3, polyubiquitination, and proteasome degradation (Zhang et al., 2017).

TET Proteins and RNA Modification

5hmC has also been detected in RNA (Delatte et al., 2016; Zhang et al., 2016). It has been reported to preferentially mark polyadenylated RNAs in Drosophila (Delatte et al., 2016). Studies suggest that 5hmC in the RNA can facilitate mRNA translation (Delatte et al., 2016). In addition, TET2 has been shown bind to RNA in mESCs; this is mediated *via* its interaction with the RNA-binding protein Paraspeckle component 1 (PSPC1) (Guallar et al., 2018). TET2 and PSPC1 mediate the silencing of endogenous retroviruses (ERVs). MERVL was among the ERVs that were repressed in the aforementioned studies. It was shown that catalytic activity of TET2 was required for repression (Guallar et al., 2018). The PSPC1 and TET2 complex could bind to both 5mC and 5hmC RNAs but had higher affinity for 5mC-containing RNAs (Guallar et al., 2018). The PSPC1-TET2 mediated 5hmC deposition on MERVL transcripts resulted in their destabilization and subsequent degradation (Guallar et al., 2018). In addition, both TET1 and TET2 deposit 5hmC in mRNAs in genes that are fundamental for pluripotency, such as *Jarid2* and *Eed*, and can result in reduced mRNA stability (Lan et al., 2020). As a consequence of the transcript destabilization, pluripotency genes that would be expressed too highly acquire appropriate expression levels and adequately repress the expression of lineage-specifying factors (Lan et al., 2020). As the ES cells receive differentiation cues, the pluripotency factors are rapidly turned off and the lineage-specifying factors are upregulated to drive the differentiation process with efficiency (Lan et al., 2020).

TET Proteins and Catalytic-Independent Roles

TET proteins can also regulate gene expression in a catalytic-independent manner *via* interactions with other proteins that affect chromatin architecture and transcription. An example is the formation of the TET1-SIN3A complex (Williams et al., 2011). The SIN3A complex—together with its components, histone deacetylases 1 and 2 (HDAC1/2)—can repress transcription by mediating histone deacetylation. In addition, TET3 can also interact with SIN3A *via* a TET SIN3A interaction domain (SID) that interfaces directly with the paired amphipathic helix (PAH) domain of SIN3A (Deplus et al., 2013; Chandru et al., 2018). The SID domain is necessary for TET1 to suppress gene expression (Chandru et al., 2018)

and is not part of its catalytic domain. Interestingly, this domain is present in TET1 and TET3 but not TET2 (Chandru et al., 2018). Furthermore, TET1 was shown to interact with Hypoxia Factor (HIF)−1a and HIF-2a to act as co-activator and promote gene expression in a catalytic-independent manner (Tsai et al., 2014). The CXXC DNA binding domain of TET1 is required for this interaction (Tsai et al., 2014). In addition, TET3 fine-tunes adult neurogenesis in a catalytic-independent manner (Montalban-Loro et al., 2019). TET3 prevents premature differentiation of neural stem cells (NSCs) into astrocytes in the adult subventricular zone by inhibiting the expression of Small nuclear ribonucleoprotein-associated polypeptide (*Snrpn*) (Montalban-Loro et al., 2019). This is achieved by direct binding of TET3 to the promoter of *Snrpn* and subsequent suppression of gene expression without any alterations in 5hmC distribution (Montalban-Loro et al., 2019).

TET PROTEINS IN IMMUNE CELL DEVELOPMENT AND DISEASE

Consistent with their multifaceted regulatory roles, TET proteins have been implicated in various developmental procedures in immune cells (Tsagaratou et al., 2017a; Lio and Rao, 2019) (**Figure 3**). However, while the implication of TET proteins in DNA demethylation is well-established, the full spectrum of mechanisms that regulate TET proteins in immune cells is yet to be revealed. Immune cell development is a well-characterized process during which progenitor cells, committed in a given pathway of differentiation, give rise to progeny cells (Cumano et al., 2019). This process of differentiation and lineage commitment is irreversible under physiological conditions (Cumano et al., 2019). However, during tumorigenesis, cells de-differentiate or transdifferentiate, frequently resulting in novel cell types that represent a mix of multiple lineages (Le Magnen et al., 2018). Strikingly, TET loss-of-function is strongly associated with hematological malignancies (Cimmino et al., 2011; Shih et al., 2012; Ficz and Gribben, 2014; Huang and Rao, 2014; Ko et al., 2015): *TET2* loss-of-function mutations are frequently observed in myelodysplastic syndromes and myeloid malignancies (Ko et al., 2010; Cimmino et al., 2011; Shih et al., 2012) as well as in certain peripheral T-cell lymphomas (PTCL) (Couronne et al., 2012; Palomero et al., 2014; Sakata-Yanagimoto et al., 2014), which are a heterogeneous and poorly understood group of aggressive non-Hodgkin lymphomas that are resistant to conventional therapies (Armitage, 2012). Loss-of-function studies in mice have been instrumental in unraveling the biological roles of TET proteins in immune cell development, function, and malignant transformation.

TET Proteins and T-Cell Lineage Fate

In T-cells, loss of TET proteins results in a variety of developmental phenotypes that can compromise immune function or trigger malignant transformation. During the process of T-cell development and lineage specification, 5hmC exhibits dynamic enrichment as precursor cells differentiate to progeny (Tsagaratou et al., 2014). In the thymus, 5hmC is increased in the gene body of lineage-specifying transcription factors such as ThPOK, the factor that seals the fate of CD4 lineage, and RUNX3, the factor that determines the CD8 cell lineage, specifically at the cell stage at which these factors are expressed (Tsagaratou et al., 2014). It has been reported that murine T-cells that lack TET2 exhibit compromised differentiation toward helper lineages such as Th1 and Th17 (Ichiyama et al., 2015) in addition to reduced expression of cytokines such as IFNγ and IL-10 (Ichiyama et al., 2015). Loss of TET2 results in increased representation of CD8 memory T-cells (Carty et al., 2018). *In vitro* polarization of human, naïve CD4 T-cells toward helper lineages demonstrates that DNA demethylation and 5hmC remodeling across the genome occur early after activation and before any differentiation (Nestor et al., 2016; Monticelli, 2019; Vincenzetti et al., 2019). Studies using T-cell polarization cultures suggest that 5hmC plays an important role in directing the specification toward helper lineages but is not necessary for expansion (Vincenzetti et al., 2019).

The most profound phenotypes in T-cells have been observed upon concomitant deletion of at least two TET members, suggesting functional redundancy (Tsagaratou et al., 2017a; Lio and Rao, 2019). Deletion of TET2 and TET3 at the DP cell stage using CD4-cre mice results in a striking increase of the iNKT cell lineage (Tsagaratou et al., 2017b; Tsagaratou, 2018). Furthermore, the *Tet2/3* DKO iNKT cells show lineage skewing in addition to an increase in the NKT17 cell lineage (Tsagaratou et al., 2017b; Tsagaratou, 2019). Moreover, the NKT1 cells are functionally impaired and maintain high expression of stemness genes, such as *Lef1*, *Lmo1,* and *Myc* (Tsagaratou et al., 2017b), that are usually expressed in earlier stages of iNKT cell development. In this setting, TET proteins regulate the deposition of 5hmC across the gene body of *Tbx21* and *Zbtb7b*, which encode for the lineage specifying factors T-bet and ThPOK, respectively. Upregulation of T-bet is critical for establishing the NKT1 cell fate. At a genome-wide level, loss of TET proteins in iNKT cells does not result in massive DNA demethylation, but rather exerts a focal impact on differential DNA methylation (Tsagaratou et al., 2017b).

These *Tet2/3* DKO iNKT cells can mediate a TCR driven, transmissible T-cell lymphoma upon transfer to fully immunocompetent congenic recipients (Tsagaratou et al., 2017b). However, transplantation of the *Tet2/3* DKO iNKT cells to Cd1dKO mice—incapable of expressing CD1d and presenting antigens to iNKT cells—fails to recapitulate the expansion, indicating an instrumental role of TCR activation in the expansion process. Further analysis revealed that the *Tet2/3* DKO iNKTs that have been transplanted and expanded in congenic recipients show accumulation of DNA breaks and R-loops, signifying that they are undergoing replication stress (Lopez-Moyado et al., 2019).

TET Proteins and Stability of Regulatory T-Cells

Tet2-deficient mice show reduced generation of regulatory T-cells (Tregs) (Nair et al., 2016). Indeed, loss of TET1 and TET2 significantly impairs Tregs by compromising the demethylation of the CNS2 locus of FOXP3: both TET1 and TET2 can bind to this locus (Yang et al., 2015). Concomitant loss of TET2 and TET3 at the DP cell stage using CD4-cre mice exerts more

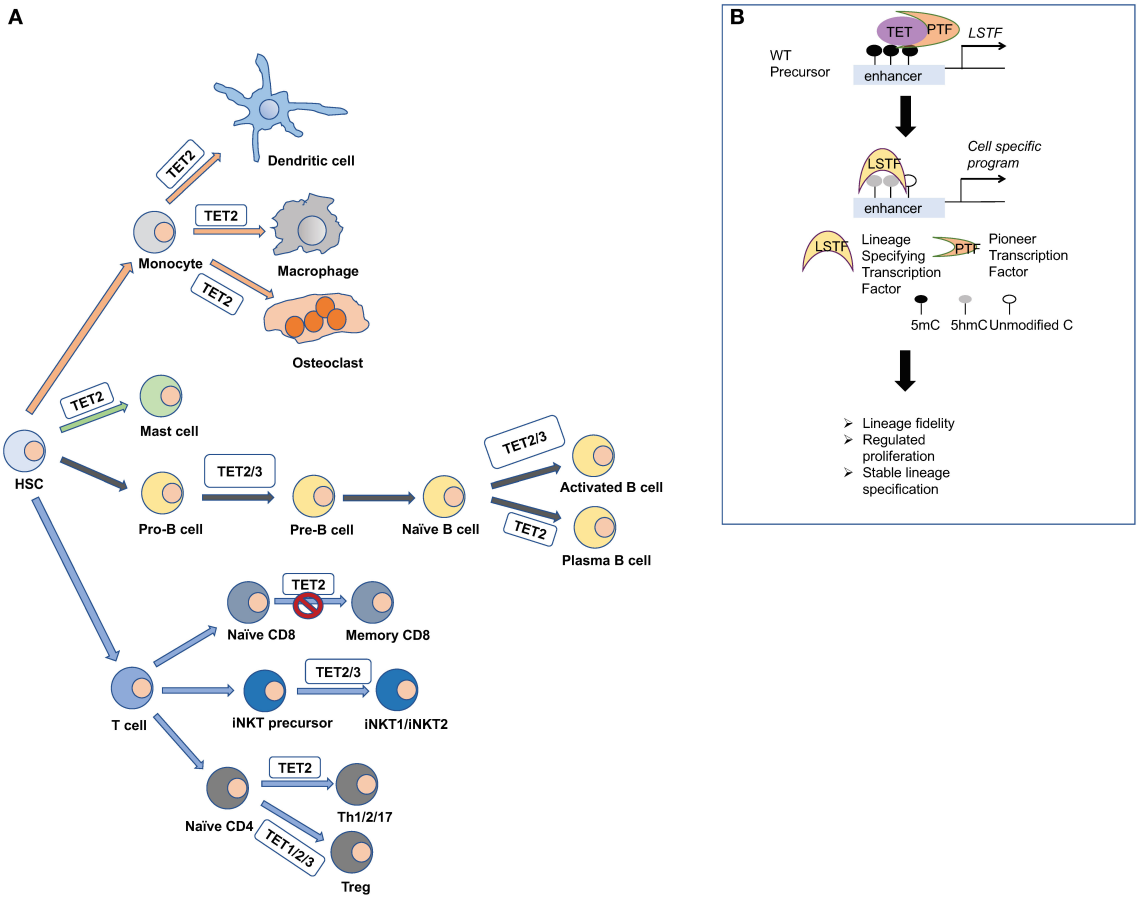


FIGURE 3 | TET proteins orchestrate the differentiation of immune cells. **(A)** Hematopoietic stem cells give rise to the various lineages of our immune system. Mice that are deficient for TET proteins have been used to explore their impact in immune cell development. TET1, TET2, and TET3 have been shown to regulate the methylation status of FOXP3, and they impact the stability of the regulatory T cells (Treg). TET2 and TET3 regulate the iNKT cell lineage specification and are critical for NKT1 and NKT2 cell differentiation. TET2 regulates the formation of memory and effector CD8 cells upon viral infection. During B cell differentiation, TET2 and TET3 orchestrate B cell maturation and function. TET2 also regulates mast cell differentiation and function in both a catalytic-dependent and -independent manner. Moreover, TET2 regulates the function of monocytic populations such as dendritic cells, macrophages, and osteoclasts. **(B)** Mechanistically, TET proteins are recruited by pioneer transcription factors (PTF) at cell-specific enhancers to oxidize 5mC and induce the expression of lineage specifying transcription factors (LSTF). Then, the LSTF execute their cell-specific gene expression program and shape cell identity.

severe impact on the stability of the Foxp3 expression due to aberrant methylation of the CNS2 locus (Yue et al., 2016). Enhancing catalytic activity of TET proteins with vitamin C can promote *in vitro* demethylation of the CNS2 locus, resulting in the generation of induced Tregs (iTregs) with superior stability compared to iTregs generated in vitamin C-absent culture (Xue et al., 2016). This observation is valid for both murine and human iTregs (Yue et al., 2016). Deleting TET2 and TET3 specifically at Tregs using Foxp3-cre mice not only compromises the stability of the Foxp3 lineage, but it also results in gain of effector function and aberrant hyperactivation of the *Tet2/3* deficient Tregs; this leads to increased inflammation and ultimately death of the mice (Nakatsukasa et al., 2019; Yue et al., 2019). In addition, hydrogen sulfide (H2S) was found to induce expression of TET1 and TET2 by regulating binding of nuclear transcription factor y subunit B (NFYB) in the promoters of *Tet1* and *Tet2* (Yang et al., 2015). H2S deficiency results in reduced expression of TET1 and TET2 in T-cells and impaired Treg differentiation due to hypermethylation of the CNS2 locus (Yang et al., 2015). In Tregs, TET recruitment to the CNS2 locus is mediated by transcription factor STAT4 (Yang et al., 2015).

Interestingly, altered metabolism in T-cells can impact the methylation status of regulatory loci in genes that encode for lineage-specifying transcription factors, ultimately affecting the lineage choice of T-cells. Indeed, it has been reported that the glutamate oxalo-acetate transaminase 1 (GOT1) is preferentially expressed in differentiating Th17 cells and catalyzes the conversion of glutamate into a-ketoglutaric acid, resulting in increased levels of 2-hydroxyglutarate (2-HG) (Xu et al., 2017). 2-HG inhibits TET catalytic activity, resulting in increased methylation and reduced expression of FOXP3, the key transcription factor that shapes the Treg lineage (Xu et al., 2017).

TET Proteins in B-Cell Development and Disease

TET-dependent 5hmC deposition and DNA demethylation are important sources of epigenetic regulation in B-cell development. TET protein expression is regulated dynamically throughout B- lymphopoiesis in humans and mice. Expression of TET1 is drastically reduced in pro-B-cells (Cimmino et al., 2015), while expression of TET2 and TET3 increases progressively over B-cell maturation and during activation (Schoeler et al., 2019). Tet-mediated 5hmC accumulation in B-cells was shown to occur within gene bodies (Cimmino et al., 2015; Schoeler et al., 2019) and at enhancer regions (Lio et al., 2016; Orlanski et al., 2016), additionally correlating with H3K4me1 histone modifications and increased transcriptional activity (Lio et al., 2016; Orlanski et al., 2016). Loss of TET1 in hematopoietic stem cells promotes differentiation with a lymphoid bias (Cimmino et al., 2015). *In vitro* analysis of *Tet1*$^{-/-}$ cells resulted in more self-renewing pro-B-cell colonies compared to pre-B-cell colonies (Cimmino et al., 2015). These proliferating TET1-deficient pro-B-cells show increased accumulation of DNA breaks as attested by increased staining for γH2Ax (Cimmino et al., 2015). In the long-term, germline deletion of TET1 results in lymphocytosis in TET1 deficient mice by 18–24 months of age (Cimmino et al., 2015). Interestingly, transplantation of TET1-deficient cells isolated from the spleen or the lymph nodes of the TET1KO mice to congenic recipients could fully recapitulate the disease within 12 weeks, thereby establishing TET1 as a tumor suppressor of B-cell malignancy (Cimmino et al., 2015).

TET2 is frequently mutated in diffuse large B-cell lymphoma (DLBCL) (Reddy et al., 2017). Deletion of *Tet2* using either Vav-cre or CD19-cre resulted in germinal center hyperplasia (Dominguez et al., 2018). Loss of TET2-mediated 5hmC deposition in enhancer regions of genes involved in exiting the germinal center (GC) reaction also correlated with reduced transcriptional activity (Dominguez et al., 2018). TET2 was instrumental for class switch recombination (CSR) and affinity maturation. TET2 deficient GC B-cells showed a defect in plasma cell differentiation (Dominguez et al., 2018). Moreover, loss of TET2 resulted in downregulation of *Prdm1,* which encodes for BLIMP1. Interestingly, reconstitution of the expression of BLIMP1 in *Tet2*-deficient naïve B-cells by retroviral transduction could rescue the differentiation defects of *Tet2* KO cells (Dominguez et al., 2018). Collectively, these data establish TET2 as a tumor suppressor of B-cell lymphomas (Dominguez et al., 2018).

Interestingly, TET2 mutations in human DLBCLs result in altered gene expression, reminiscent of the *Tet2*-deficient GC gene signature (Dominguez et al., 2018). Comparative analysis of gene expression profiles revealed strong similarities with cases that had mutations in the histone acetyltransferase CREBBP (Dominguez et al., 2018). Thus, TET2 and CREBBP could potentially collaborate to regulate enhancer activation by generating 5hmC and H3K27Ac (Dominguez et al., 2018).

Consistent with observations in T-cells, simultaneous deletion of TET2 and TET3 resulted in more severe B-cell phenotypes. During bone marrow development, ablation of TET2 and TET3 in B-cells using the Mb1-cre mice inhibited B-cell maturation; *Tet2/3* DKO mice exhibited an accumulation of pro- and pre-B-cells, while the mature B-cells were significantly decreased (Lio et al., 2016; Orlanski et al., 2016). TET2 and TET3 were shown to play a critical role in DNA demethylation of the enhancers of Igk light chains (Lio et al., 2016; Orlanski et al., 2016). The recruitment of TET proteins at the enhancer was mediated by the pioneer transcription factor PU.1 (Lio et al., 2016). In addition, TET2 and TET3 regulate the expression of IRF4 and IRF8 that are involved in Igk rearrangement (Lio et al., 2016). Addition of ascorbic acid promotes the differentiation of germinal center B cells to plasma cells both *in vitro* and *in vivo* (Qi et al., 2020). This is achieved by enhancing TET2 and TET3 catalytic activity to demethylate enhancers that control the expression of *Prdm1* (Qi et al., 2020).

TET2 and TET3 proteins regulate somatic hypermutation (SHM) and CSR through Tet-dependent upregulation of Activation Induced Deaminase (AID) in activated B-cells (Lio et al., 2019a; Schoeler et al., 2019). These studies showed that TET proteins were recruited to two sites within the AID super-enhancer, *TetE1* and *TetE2* (Lio et al., 2019a), by the basic leucine zipper transcription factor, ATF-like (BATF) (Lio et al., 2019a; Schoeler et al., 2019).

TET Proteins in Innate Cell Development and Disease

Loss of TET2 in hematopoietic stem cells (HSCs) results in increased stem cell self-renewal, increased number of progenitor cells, and skewed development toward the monocyte/macrophage lineage (Ko et al., 2011; Moran-Crusio et al., 2011) (**Figure 4**). Additionally, TET2 loss impacts mast cell differentiation and cytokine production as well as proliferation (Montagner et al., 2016). Interestingly, other TET proteins can compensate for altered cell differentiation, suggesting functional redundancy. However, proliferation is exclusively TET2-dependent and independent of its catalytic activity (Montagner et al., 2016). The precise mechanism by which TET2 non-catalytic function is achieved remains unknown, but a plausible scenario is that TET2 could form a complex with other proteins involved in regulating gene expression.

Differentiation of monocytes to osteoclasts is characterized by dynamic changes in DNA methylation (de la Rica et al., 2013). Genomic regions that exhibit changes in DNA methylation during osteoclastogenesis were enriched for PU.1, NF-κB, and AP-1 DNA binding motifs (de la Rica et al., 2013). PU.1 motifs were highly enriched in both hypo- and hyper-methylated regions; it was shown that PU.1 could interact with both TET2 and DNMT3b, thus playing a critical role in recruiting these proteins across the genome to regulate osteoclastogenesis (de la Rica et al., 2013). In addition, differentiation of peripheral blood mononuclear cells (PBMCs) to macrophages and osteoclasts revealed that both cell types exhibit similar dynamic changes of DNA methylation and hydroxymethylation (Garcia-Gomez et al., 2017). However, TET2 and TDG exert a dual function to establish the distinct phenotypes of macrophages and osteoclasts. TET2 further oxidizes 5hmC to oxi-mCs, followed by TDG

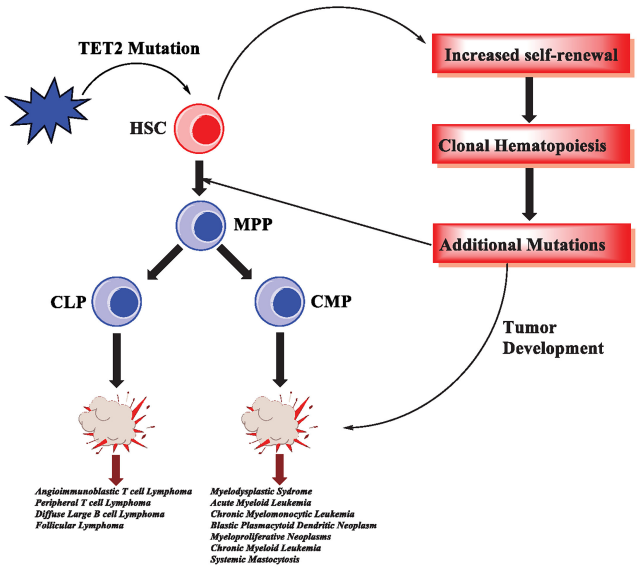


FIGURE 4 | TET2 mutations in hematological malignancies. In hematopoietic stem cells, TET2 mutations are an early event that results in increased self-renewal and subsequently clonal hematopoiesis. As the cells acquire additional mutations, malignant transformation and tumorigenesis occur. TET2 mutations have been reported in various hematological malignancies, affecting myeloid cells as well as T-cell and B-cell lymphomas.

mediating the generation of unmodified C (Garcia-Gomez et al., 2017). In addition, TET2 can mediate the recruitment of the H3K4 histone methyltransferase SETD1A to promote histone methylation (H3K4me3) at genes that are specifically expressed in osteoclasts (Garcia-Gomez et al., 2017). Similarly, *in vitro* differentiation of human monocytes to dendritic cells requires TET2-dependent DNA demethylation (Klug et al., 2013).

Interestingly, *Tet2* haploinsufficiency contributes to transformation *in vivo*, consistent with the fact that *Tet2* monoallelic loss is an important pathogenic event in myeloid malignancies (Moran-Crusio et al., 2011). Although TET2 mutations can lead to malignancies, they are often an early event in a series of mutations (Huang and Rao, 2014; Rasmussen and Helin, 2016) (**Figure 4**). TET2 mutations can lead to clonal hematopoiesis (CH), a physiological state in which a specific lineage, or clone, of cells expands at a greater rate than other lineages (Challen and Goodell, 2020). Importantly, acute loss of TET proteins using a system of inducible deletion in mice resulted in the rapid emergence of aggressive myeloid leukemia (An et al., 2015).

Enhancing the activity of TET proteins with vitamin C (**Figure 5**) can protect hematopoietic stem cells that have *Tet2* mutations from aberrant proliferation *in vitro* and leukemia progression *in vivo* (Agathocleous et al., 2017; Cimmino et al., 2017; Das et al., 2019) (reviewed in Ang et al., 2018; Cimmino et al., 2018; Yue and Rao, 2020). It has been shown that when $Tet2^{+/-}$ or $Tet2^{-/-}$ HSCs are cultured *in vitro* in the presence of vitamin C, there is an increase of 5hmC compared to *Tet2*-deficient HSCs that are cultured in the absence of vitamin C (Cimmino et al., 2017). The increased 5hmC levels in $Tet2^{+/-}$ HSCs are due to residual TET2 activity and enhanced catalytic activity of TET3, whereas in $Tet2^{-/-}$ HSCs the catalytic activity of TET3 is required to oxidize 5mC to 5hmC. The restoration of 5hmC controls cell proliferation (Cimmino et al., 2017). Similarly, *in vivo* administration of vitamin C in xenograft experiments in mice diminished the proliferation rate of *Tet2*-deficient HSCs and reduced tumorigenesis (Cimmino et al., 2017). These findings have significant clinical implications (Ang et al., 2018; Cimmino et al., 2018; Yue and Rao, 2020). Patients with hematological malignancies are often vitamin C-deficient. Oral administration of vitamin C in patients with myeloid malignancies who were on treatment with the DNMT inhibitor azacytidine significantly increased the ratio of 5hmC/5mC in their plasma and restored vitamin C concentration to normal levels (Gillberg et al., 2019).

Mutations in TET2 are not the only pathway that leads to dysfunction. As mentioned earlier, its catalytic activity is Fe (II), alpha-ketoglutarate (α-KG), and oxygen dependent (Tahiliani et al., 2009). Vitamin C (ascorbate) enhances TET activity *in vitro* and *in vivo* (Das et al., 2020). Recent studies have shed light on the connections between metabolism and epigenetic modifiers in physiological and pathological conditions (Chisolm and Weinmann, 2018; Lio et al., 2019b). Disruptions in important metabolic pathways also result in disease states. Mutations in isocitrate dehydrogenase 1 and 2 (IDH), an enzyme in the TCA cycle that converts isocitrate to α-KG, often are gain-of-function (GOF), allowing the enzyme to produce the oncometabolite 2-hydroxyglutarate (2-HG). 2-HG competitively inhibits binding of α-KG to TET2, severely impairing its function (**Figure 5**). It has been shown that IDH1/2 GOF and TET2 LOF mutations show similar phenotypes in mouse models, with reduced genome-wide 5hmC levels and dysregulated HSC differentiation (Figueroa et al., 2010; Moran-Crusio et al., 2011; Lio et al., 2019b). In myeloid neoplasms, TET2 and IDH1/2 mutations are usually mutually exclusive (Figueroa et al., 2010; Shih et al., 2015; Inoue et al., 2016), but they are often paired together in Angioimmunoblastic T-cell Lymphoma (AITL) (Wang et al., 2015; Cortes and Palomero, 2020). This indicates that there are some differing oncogenic mechanisms at play even given the phenotypic similarities of the two mutations.

Importantly, loss of TET2 can affect the inflammatory response. In fact, TET2 has been implicated in repression of the proinflammatory cytokine interleukin 6 (IL-6), both in macrophages and dendritic cells (Zhang et al., 2015). This is achieved by TET2 interaction with Iκbζ, which permits binding to the *IL6* promoter. Subsequently, TET2, independently of its DNA demethylation activity, mediates the recruitment of the histone deacetylase 2 (HDAC2) to repress *IL6* expression (Zhang et al., 2015). As a result, $Tet2^{-}$ deficient mice are more susceptible to endotoxin-induced septic shock, induced by administration of lipopolysaccharide (LPS), and colitis compared to control mice, due to exacerbated inflammation (Zhang et al., 2015). Notably, *Tet2*-deficient tumor infiltrating macrophages exhibit defective immunosuppressive capacity in a mouse melanoma model as a result of altered cytokine expression profile (Pan et al., 2017).

It has been reported that CH can result in a 30–40% increased mortality risk unrelated to blood cancers but instead attributed

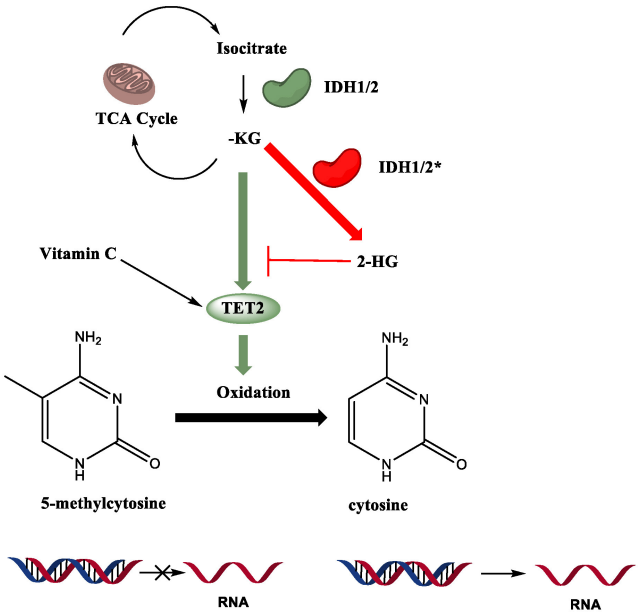


FIGURE 5 | Regulatory mechanisms of TET enzymatic activity. TET2 catalytic activity is Fe (II), alpha-ketoglutarate (α-KG), and oxygen dependent. Mutations in isocitrate dehydrogenase 1 and 2 (IDH), an enzyme in the TCA cycle that converts isocitrate to α-KG, often are gain-of-function (GOF), allowing the enzyme to produce the oncometabolite 2-hydroxyglutarate (2-HG). 2-HG competitively inhibits binding of α-KG to TET2, compromising its function. Vitamin C can enhance the catalytic activity of TET proteins.

to higher cardiovascular mortality from coronary heart disease and ischemic stroke (Jaiswal et al., 2014; Fuster and Walsh, 2018). Further studies revealed a causal link between TET2 mutations in hematopoietic stem cells, increased inflammation, and atherosclerosis. Competitive transfer of *Tet2*-deficient bone marrow cells resulted in enlarged atherosclerotic lesions in irradiated, atherosclerosis-prone mice that are deficient for low-density lipoprotein receptor ($Ldrl^{-/-}$) (Fuster et al., 2017; Jaiswal et al., 2017). *Tet2*-deficient macrophages secreted increased amounts of the cytokine IL-1β in a NLRP3 inflammasome-dependent manner (Fuster et al., 2017). Inhibition of NLRP3 provided enhanced protection from atherosclerosis preferentially to the $Ldrl^{-/-}$ mice that had received *Tet2*-deficient bone marrow cells (Fuster et al., 2017).

Similarly, in experimental heart failure mouse models, hematopoietic *Tet2* deficiency followed by competitive transfer or myeloid-specific *Tet2* deficiency resulted in severely impaired cardiac remodeling, accompanied by an NLRP3 inflammasome-dependent increase in IL-1β (Sano et al., 2018). Adoptive transfer of unfractionated *Tet2*-deficient bone marrow cells in non-irradiated recipients revealed that *Tet2* deficiency alters the phenotype of macrophages present in the heart and promotes cardiomyopathy in steady state conditions in aged mice without pre-existing cardiovascular injury (Wang et al., 2020). Gene expression analysis of *Tet2*-deficient derived macrophages 8 months after transfer revealed that IL1-β was upregulated (Wang et al., 2020). The aged mice showed signs of cardiac dysfunction and increased inflammation (Wang et al., 2020). In addition, competitive transfer of TET2KO bone marrow cells exacerbates insulin resistance in aging and obesity in an IL-1β NLRP3 inflammasome-dependent manner (Fuster et al., 2020). Increased inflammation due to *Tet2* loss has recently been associated with pulmonary arterial hypertension in humans as well as in *Tet2*-deficient mice (Potus et al., 2020).

PERSPECTIVES

TET Proteins Regulate Focal DNA Demethylation

Various studies using TET-deficient mice demonstrated that loss of TET proteins has only a mild impact on global DNA methylation (An et al., 2015; Cimmino et al., 2015; Tsagaratou et al., 2017b). However, when focusing on regions that are differentially methylated across development, a robust increase in DNA methylation was observed upon TET loss (Tsagaratou et al., 2017b). This observation is consistent with the report that only 21.8% of autosomal CpGs exhibit dynamic changes in their methylation status across development (Ziller et al., 2013). This primarily occurs in loci genomically distant from the TSS (Ziller et al., 2013). Loss of TET proteins contributed to increased DNA methylation even in regulatory areas with high methylation levels in T-cell subsets, suggesting that TET proteins compete with DNMTs to avoid aberrant hypermethylation (Tsagaratou et al., 2017b). Maintaining a certain threshold of DNA methylation and/or generating 5hmC could stabilize the enhancers in a poised state, allowing the rapid initiation of gene expression at subsequent developmental stages or following certain environmental cues. This concomitant existence of two opposing epigenetic marks is reminiscent of the poised bivalent promoters that have been extensively described mainly in embryonic stem cells and are characterized by coexistence of the H3K4me3 histone mark, which positively correlates with gene expression, and the suppressing mark H3K27me3 (Bernstein et al., 2006).

The focal activity of TET proteins in DNA demethylation strongly suggests that TET proteins are recruited and targeted to the DNA *via* transcription factors to regulate the DNA demethylation of regulatory elements that control the expression of key genes involved in the cell-specific program of a given immune cell. Indeed, in regulatory T-cells, members of the STAT family act as pioneer transcription factors that exert TET recruitment at enhancers (Yang et al., 2015). In B-cells, PU.1, EBF1, and BATF can mediate TET recruitment to regulatory elements (Lio et al., 2016, 2019a). Open chromatin conformation as well as chromatin accessibility correlates with increased 5hmC levels across a variety of leukocytes (Lio et al., 2016; Tsagaratou et al., 2017b). TET proteins affect TF binding at regulatory elements, including enhancers, by virtue of their cell type-specific binding motifs and role in modifying chromatin accessibility (Rasmussen et al., 2019).

TET Proteins and Lineage Specification

TET proteins play a critical role in regulating lineage specification of various cell types (Tsagaratou et al., 2017a; Wu and Zhang,

2017). For instance, TET proteins deposit intragenic 5hmC in *Zbtb7b* and *Tbx21*, genes that produce the crucial lineage-specifying factors of T-cell differentiation: ThPOK and RUNX3, respectively. 5hmC enrichment slowly decreases over time upon commitment to a given cell fate (Tsagaratou et al., 2014).

Importantly, TET protein loss results in abnormal development, failure to progress beyond precursor cell stages, and unregulated cell division (Cimmino et al., 2015; Lio et al., 2016; Orlanski et al., 2016; Tsagaratou et al., 2017b). Investigation of TET loss in mutant mice indicated that malignant transformation occurs due to maintenance of a stemness gene expression program instead of commitment to a lineage-specific program.

In addition, TET proteins are instrumental in safeguarding stability of gene expression, preventing de-differentiation of cells. For instance, TET proteins prevent aberrant methylation of regulatory elements to stabilize the expression of the Treg lineage-specifying factor FOXP3. During thymic development, TET1 and TET3 can regulate the cytosine methylation status of enhancers that permit *Cd4* gene expression at later stages in the periphery (Issuree et al., 2018). Presumably, deposition of 5hmC in enhancers can prime these regulatory elements to become fully activated and promote gene expression at subsequent developmental stages.

TET Proteins and Functional Redundancy

Analysis of various mouse models strongly suggests that TET proteins exhibit redundancy. For example, development proceeds normally in many cases upon deletion of a single TET protein. It seems that TET proteins act in complement to regulate enhancers and lineage-specifying TFs, leading to activation of a cell-specific gene expression program. In addition, mice that lack a single TET protein develop cancer slowly over the course of several years (Cimmino et al., 2015; Lio et al., 2019b). However, simultaneous deletion of two or more TET proteins results in rapid, malignant transformation of a gamut of immune cell lineages (An et al., 2015; Zhao et al., 2015; Lio et al., 2016; Tsagaratou et al., 2017b).

FUTURE DIRECTIONS

Since TET proteins are recruited at specific genomic loci by interacting partners, it is critical to unveil the cell-specific interactome of TET proteins that will allow us to gain appreciation of the full spectrum of TET-regulated cell properties. We do anticipate that these interactions will reveal novel, unexpected roles of TET proteins in immune cell development that extend beyond the regulation of DNA demethylation. For example, until recently, it was thought that only mutations in the catalytic region of TET2 could induce oncogenic transformations. However, recent studies have shown that TET2 knockout mice and TET2 mutant mice (with a mutation rendering the catalytic domain non-functional) produce different disease states (Ito et al., 2019). The former resulted in both myeloid and lymphoid malignancies, while the latter produced primarily myeloid malignancies (Ito et al., 2019). This suggests that TET2 has roles as a tumor suppressor independent of its catalytic function. Further investigation is needed to identify its other physiological functions that safeguard the proper differentiation and proliferation of cells. Along these lines, we anticipate that TET1 and TET3 might also exert catalytic-independent functions in the context of immune cell differentiation and function.

An additional future direction of research that will allow us to fully understand the regulatory impact of TET proteins on gene expression is the fact that multiple loci affected by TET proteins are enhancers; thus, it is challenging to qualify the genes that are directly affected by TET loss (Tsagaratou et al., 2017b). It is now largely accepted that enhancers can regulate genes that are located far across the genome. Methods such as Hi-C (Lieberman-Aiden et al., 2009; Rao et al., 2014), ChiA-PET (Fullwood et al., 2009), and HiChIP (Mumbach et al., 2016) not only are expensive ways to assess genome-wide topological associations but also, in many cases, are not easily adjustable to the small numbers of primary cells that can be purified. Future work that can precisely determine the genes comprising the TET regulome will elucidate the causal mechanisms underlying the abnormal immune cell phenotypes present in biological systems lacking TET proteins.

After identifying potential regulatory elements and assigning them to genes that they might regulate, it is critical to confirm experimentally if these enhancers are instrumental for gene expression. Novel genome editing technologies can be employed to test enhancer activity, such as clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) (Catarino and Stark, 2018). Briefly, CRISPR/Cas9 creates double-stranded breaks in target DNA sequences that are specified by sequence complementary guide RNAs (Jinek et al., 2012).

CONCLUSIONS

To conclude, in the last decade we have witnessed significant progress in our understanding of the biology of TET proteins. Besides the well-established enzymatic function of TET proteins that contributes to DNA demethylation, we have started to appreciate additional roles that these proteins assume to regulate gene expression and establish cell identity. As we move forward, it is critical to dissect the unique versus the shared functions of TET proteins, unravel the cell-specific interactome of each TET protein, and decipher the regulatory elements that they control. Ultimately, by harnessing TET enzymatic and non-enzymatic activity, we hope to be able to epigenetically reprogram cells, preventing their hyperproliferation and malignant transformation that ultimately results in tumorigenesis. Moreover, deciphering the mechanisms by which TET2 loss and clonal hematopoiesis result in increased inflammation and age-related cardiovascular diseases can pave the way for therapeutic intervention.

AUTHOR CONTRIBUTIONS

NJT and AT wrote the manuscript with contributions from DB, LFC, and FJW. NJT and LFC prepared the

figures with input from AT. All authors approved the submitted manuscript.

FUNDING

AT is supported by a University of North Carolina (UNC) Lineberger Comprehensive Cancer Center (LCCC) startup fund, National Institute of Health grant R35 GM138289 (National Institute of General Medical Sciences), and Junior Faculty Development Award. NJT was the recipient of a UNC Summer Undergraduate Research Fellowship (SURF), LFC is supported in part by a grant from the National Institute of General Medical Sciences under award 1T32GM135128, and FJW was the recipient of a UNC SURF, Program for Precision Medicine in Health Care (PPMH).

REFERENCES

Agathocleous, M., Meacham, C. E., Burgess, R. J., Piskounova, E., Zhao, Z., Crane, G. M., et al. (2017). Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature* 549, 476–481. doi: 10.1038/nature23876

An, J., Gonzalez-Avalos, E., Chawla, A., Jeong, M., Lopez-Moyado, I. F., Li, W., et al. (2015). Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat. Commun.* 6:10071. doi: 10.1038/ncomms10071

Ang, A., Pullar, J. M., Currie, M. J., and Vissers, M. C. M. (2018). Vitamin C and immune cell function in inflammation and cancer. *Biochem. Soc. Trans.* 46, 1147–1159. doi: 10.1042/BST20180169

Armitage, J. O. (2012). The aggressive peripheral T-cell lymphomas: 2012 update on diagnosis, risk stratification, and management. *Am. J. Hematol.* 87, 511–519. doi: 10.1002/ajh.23144

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H. E., Iurlaro, M., Reik, W., et al. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* 11, 555–557. doi: 10.1038/nchembio.1848

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* 6, 1049–1055. doi: 10.1038/nchem.2064

Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., et al. (2015). Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J. Biol. Chem.* 290, 4801–4812. doi: 10.1074/jbc.M114.605881

Baylin, S. B., and Jones, P. A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* 11, 726–734. doi: 10.1038/nrc3130

Bernstein, B. E., Meissner, A., and Lander, E. S. (2007). The mammalian epigenome. *Cell* 128, 669–681. doi: 10.1016/j.cell.2007.01.033

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326. doi: 10.1016/j.cell.2006.02.041

Bogdanovic, O., and Lister, R. (2017). DNA methylation and the preservation of cell identity. *Curr. Opin. Genet. Dev.* 46, 9–14. doi: 10.1016/j.gde.2017.06.007

Branco, M. R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* 13, 7–13. doi: 10.1038/nrg3080

Carty, S. A., Gohil, M., Banks, L. B., Cotton, R. M., Johnson, M. E., Stelekati, E., et al. (2018). The loss of TET2 promotes CD8(+) T cell memory differentiation. *J. Immunol.* 200, 82–91. doi: 10.4049/jimmunol.1700559

Catarino, R. R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* 32, 202–223. doi: 10.1101/gad.310367.117

Challen, G., and Goodell, M. A. (2020). Clonal hematopoiesis: mechanisms driving dominance of stem cell clones. *Blood* 136, 1590–1598. doi: 10.1182/blood.2020006510

Chandru, A., Bate, N., Vuister, G. W., and Cowley, S. M. (2018). Sin3A recruits Tet1 to the PAH1 domain via a highly conserved Sin3-interaction domain. *Sci. Rep.* 8:14689. doi: 10.1038/s41598-018-32942-w

Charlton, J., Jung, E. J., Mattei, A. L., Bailly, N., Liao, J., Martin, E. J., et al. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* 52, 819–827. doi: 10.1038/s41588-020-0639-9

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561–564. doi: 10.1038/nature11742

Chen, R., Zhang, Q., Duan, X., York, P., Chen, G. D., Yin, P., et al. (2017). The 5-Hydroxymethylcytosine (5hmC) reader UHRF2 is required for normal levels of 5hmC in mouse adult brain and spatial learning and memory. *J. Biol. Chem.* 292, 4533–4543. doi: 10.1074/jbc.M116.754580

Chisolm, D. A., and Weinmann, A. S. (2018). Connections between metabolism and epigenetics in programming cellular differentiation. *Annu. Rev. Immunol.* 36, 221–246. doi: 10.1146/annurev-immunol-042617-053127

Cimmino, L., Abdel-Wahab, O., Levine, R. L., and Aifantis, I. (2011). TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell* 9, 193–204. doi: 10.1016/j.stem.2011.08.007

Cimmino, L., and Aifantis, I. (2017). Alternative roles for oxidized mCs and TETs. *Curr. Opin. Genet. Dev.* 42, 1–7. doi: 10.1016/j.gde.2016.11.003

Cimmino, L., Dawlaty, M. M., Ndiaye-Lobry, D., Yap, Y. S., Bakogianni, S., Yu, Y., et al. (2015). TET1 is a tumor suppressor of hematopoietic malignancy. *Nat. Immunol.* 16, 653–662. doi: 10.1038/ni.3148

Cimmino, L., Dolgalev, I., Wang, Y., Yoshimi, A., Martin, G. H., Wang, J., et al. (2017). Restoration of TET2 function blocks aberrant self-renewal and leukemia progression. *Cell.* 170, 1079–1095.e20. doi: 10.1016/j.cell.2017.07.032

Cimmino, L., Neel, B. G., and Aifantis, I. (2018). Vitamin C in stem cell reprogramming and cancer. *Trends Cell Biol.* 28, 698–708. doi: 10.1016/j.tcb.2018.04.001

Cortes, J. R., and Palomero, T. (2020). Biology and molecular pathogenesis of mature T-cell lymphomas. *Cold Spring Harb. Perspect. Med.* a035402. doi: 10.1101/cshperspect.a035402. [Epub ahead of print].

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495, 370–374. doi: 10.1038/nature11925

Couronne, L., Bastard, C., and Bernard, O. A. (2012). TET2 and DNMT3A mutations in human T-cell lymphoma. *N. Engl. J. Med.* 366, 95–96. doi: 10.1056/NEJMc1111708

Cumano, A., Berthault, C., Ramond, C., Petit, M., Golub, R., Bandeira, A., et al. (2019). New molecular insights into immune cell development. *Annu. Rev. Immunol.* 37, 497–519. doi: 10.1146/annurev-immunol-042718-041319

Das, A. B., Kakadia, P. M., Wojcik, D., Pemberton, L., Browett, P. J., Bohlander, S. K., et al. (2019). Clinical remission following ascorbate treatment in a case of acute myeloid leukemia with mutations in TET2 and WT1. *Blood Cancer J.* 9:82. doi: 10.1038/s41408-019-0242-4

Das, A. B., Smith-Diaz, C. C., and Vissers, M. C. M. (2020). Emerging epigenetic therapeutics for myeloid leukemia: modulating demethylase activity with ascorbate. *Haematologica.* 106:14–25 doi: 10.3324/haematol.2020.259283

de la Rica, L., Rodriguez-Ubreva, J., Garcia, M., Islam, A. B., Urquiza, J. M., Hernando, H., et al. (2013). PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol.* 14:R99. doi: 10.1186/gb-2013-14-9-r99

Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., et al. (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* 351, 282–285. doi: 10.1126/science.aac5253

Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Mendez, J., Murphy, N., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* 32, 645–655. doi: 10.1038/emboj.2012.357

Dominguez, P. M., Ghamlouch, H., Rosikiewicz, W., Kumar, P., Beguelin, W., Fontan, L., et al. (2018). TET2 deficiency causes germinal center hyperplasia, impairs plasma cell differentiation, and promotes B-cell lymphomagenesis. *Cancer Discov*. 8, 1632–1653. doi: 10.1158/2159-8290.CD-18-0657

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., et al. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398–402. doi: 10.1038/nature10008

Ficz, G., and Gribben, J. G. (2014). Loss of 5-hydroxymethylcytosine in cancer: cause or consequence? *Genomics* 104, 352–357. doi: 10.1016/j.ygeno.2014.08.017

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567. doi: 10.1016/j.ccr.2010.11.015

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., et al. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* 6:e21306. doi: 10.1371/journal.pone.0021306

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 58–64. doi: 10.1038/nature08497

Fuster, J. J., MacLauchlan, S., Zuriaga, M. A., Polackal, M. N., Ostriker, A. C., Chakraborty, R., et al. (2017). Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* 355, 842–847. doi: 10.1126/science.aag1381

Fuster, J. J., and Walsh, K. (2018). Somatic mutations and clonal hematopoiesis: unexpected potential new drivers of age-related cardiovascular disease. *Circ. Res.* 122, 523–532. doi: 10.1161/CIRCRESAHA.117.312115

Fuster, J. J., Zuriaga, M. A., Zorita, V., MacLauchlan, S., Polackal, M. N., Viana-Huete, V., et al. (2020). TET2-loss-of-function-driven clonal hematopoiesis exacerbates experimental insulin resistance in aging and obesity. *Cell Rep*. 33:108326. doi: 10.1016/j.celrep.2020.108326

Gan, H., Wen, L., Liao, S., Lin, X., Ma, T., Liu, J., et al. (2013). Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat. Commun*. 4:1995. doi: 10.1038/ncomms2995

Garcia-Gomez, A., Li, T., Kerick, M., Catala-Moll, F., Comet, N. R., Rodriguez-Ubreva, J., et al. (2017). TET2- and TDG-mediated changes are required for the acquisition of distinct histone modifications in divergent terminal differentiation of myeloid cells. *Nucleic Acids Res.* 45, 10002–10017. doi: 10.1093/nar/gkx666

Gillberg, L., Orskov, A. D., Nasif, A., Ohtani, H., Madaj, Z., Hansen, J. W., et al. (2019). Oral vitamin C supplementation to patients with myeloid cancer on azacitidine treatment: normalization of plasma vitamin C induces epigenetic changes. *Clin. Epigenetics* 11:143. doi: 10.1186/s13148-019-0739-5

Ginno, P. A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., et al. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun*. 11:2680. doi: 10.1038/s41467-020-16354-x

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem*. 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Gu, T., Lin, X., Cullen, S. M., Luo, M., Jeong, M., Estecio, M., et al. (2018). DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol*. 19:88. doi: 10.1186/s13059-018-1464-7

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610. doi: 10.1038/nature10443

Guallar, D., Bi, X., Pardavila, J. A., Huang, X., Saenz, C., Shi, X., et al. (2018). RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat. Genet*. 50, 443–451. doi: 10.1038/s41588-018-0060-9

Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol*. 166, 493–505. doi: 10.1083/jcb.200403109

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., et al. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339, 448–452. doi: 10.1126/science.1229277

Hashimoto, H., Olanrewaju, Y. O., Zheng, Y., Wilson, G. G., Zhang, X., and Cheng, X. (2014). Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev*. 28, 2304–2313. doi: 10.1101/gad.250746.114

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi: 10.1126/science.1210944

Huang, Y., and Rao, A. (2014). Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet.* 30, 464–474. doi: 10.1016/j.tig.2014.07.005

Ichiyama, K., Chen, T., Wang, X., Yan, X., Kim, B. S., Tanaka, S., et al. (2015). The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. *Immunity* 42, 613–626. doi: 10.1016/j.immuni.2015.03.005

Inoue, S., Lemonnier, F., and Mak, T. W. (2016). Roles of IDH1/2 and TET2 mutations in myeloid disorders. *Int. J. Hematol*. 103, 627–633. doi: 10.1007/s12185-016-1973-7

Issuree, P. D., Day, K., Au, C., Raviram, R., Zappile, P., Skok, J. A., et al. (2018). Stage-specific epigenetic regulation of CD4 expression by coordinated enhancer elements during T cell development. *Nat. Commun*. 9:3594. doi: 10.1038/s41467-018-05834-w

Ito, K., Lee, J., Chrysanthou, S., Zhao, Y., Josephs, K., Sato, H., et al. (2019). Non-catalytic roles of Tet2 are essential to regulate hematopoietic stem and progenitor cell homeostasis. *Cell Rep*. 28, 2480–2490.e4. doi: 10.1016/j.celrep.2019.07.094

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303. doi: 10.1126/science.1210597

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., et al. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol*. 14:R119. doi: 10.1186/gb-2013-14-10-r119

Iyer, L. M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8, 1698–1710. doi: 10.4161/cc.8.11.8580

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med*. 371, 2488–2498. doi: 10.1056/NEJMoa1408617

Jaiswal, S., Natarajan, P., Silver, A. J., Gibson, C. J., Bick, A. G., Shvartz, E., et al. (2017). Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med*. 377, 111–121. doi: 10.1056/NEJMoa1701719

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Kafer, G. R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I., et al. (2016). 5-Hydroxymethylcytosine marks sites of DNA damage and promotes genome stability. *Cell Rep*. 14, 1283–1292. doi: 10.1016/j.celrep.2016.01.035

Kitsera, N., Allgayer, J., Parsa, E., Geier, N., Rossa, M., Carell, T., et al. (2017). Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. *Nucleic Acids Res*. 45, 11033–11042. doi: 10.1093/nar/gkx718

Klug, M., Schmidhofer, S., Gebhard, C., Andreesen, R., and Rehli, M. (2013). 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. *Genome Biol*. 14:R46. doi: 10.1186/gb-2013-14-5-r46

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Aijo, T., Pastor, W. A., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126. doi: 10.1038/nature12052

Ko, M., An, J., Pastor, W. A., Koralov, S. B., Rajewsky, K., and Rao, A. (2015). TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol. Rev*. 263, 6–21. doi: 10.1111/imr.12239

Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., et al. (2011). Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. U.S.A*. 108, 14566–14571. doi: 10.1073/pnas.1112317108

Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839–843. doi: 10.1038/nature09586

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 8, 200–213. doi: 10.1016/j.stem.2011.01.008

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172, 650–665. doi: 10.1016/j.cell.2018.01.029

Lan, J., Rajan, N., Bizet, M., Penning, A., Singh, N. K., Guallar, D., et al. (2020). Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nat. Commun*. 11:4956. doi: 10.1038/s41467-020-18729-6

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res*. 20, 320–331. doi: 10.1101/gr.101907.109

Le Magnen, C., Shen, M. M., and Abate-Shen, C. (2018). Lineage plasticity in cancer progression and treatment. *Annu. Rev. Cancer Biol*. 2, 271–289. doi: 10.1146/annurev-cancerbio-030617-050224

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lio, C. J., and Rao, A. (2019). TET enzymes and 5hmC in adaptive and innate immune systems. *Front. Immunol*. 10:210. doi: 10.3389/fimmu.2019.00210

Lio, C. J., Shukla, V., Samaniego-Castruita, D., Gonzalez-Avalos, E., Chakraborty, A., Yue, X., et al. (2019a). TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. *Sci Immunol*. 4:eaau7523. doi: 10.1126/sciimmunol.aau7523

Lio, C. J., Yuita, H., and Rao, A. (2019b). Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood* 134, 1487–1497. doi: 10.1182/blood.2019791475

Lio, C. W., Zhang, J., Gonzalez-Avalos, E., Hogan, P. G., Chang, X., and Rao, A. (2016). Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. *Elife* 5:e18290. doi: 10.7554/eLife.18290.030

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* 341:1237905. doi: 10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 462, 315–322. doi: 10.1038/nature08514

Lopez-Moyado, I. F., Tsagaratou, A., Yuita, H., Seo, H., Delatte, B., Heinz, S., et al. (2019). Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proc. Natl. Acad. Sci. U.S.A*. 116, 16933–16942. doi: 10.1073/pnas.1903059116

Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., and Downing, J. R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17, 637–641. doi: 10.1038/sj.leu.2402834

Marco, A., Kisliouk, T., Tabachnik, T., Weller, A., and Meiri, N. (2016). DNA CpG methylation (5-methylcytosine) and its derivative (5-hydroxymethylcytosine) alter histone posttranslational modifications at the pomc promoter, affecting the impact of perinatal diet on leanness and obesity of the offspring. *Diabetes* 65, 2258–2267. doi: 10.2337/db15-1608

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430. doi: 10.1016/j.cell.2012.11.022

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560. doi: 10.1038/nature06008

Montagner, S., Leoni, C., Emming, S., Della Chiara, G., Balestrieri, C., Barozzi, I., et al. (2016). TET2 regulates mast cell differentiation and proliferation through catalytic and non-catalytic activities. *Cell Rep*. 15, 1566–1579. doi: 10.1016/j.celrep.2016.04.044

Montalban-Loro, R., Lozano-Urena, A., Ito, M., Krueger, C., Reik, W., Ferguson-Smith, A. C., et al. (2019). TET3 prevents terminal differentiation of adult NSCs by a non-catalytic action at Snrpn. *Nat. Commun*. 10:1726. doi: 10.1038/s41467-019-09665-1

Monticelli, S. (2019). DNA (hydroxy)methylation in T helper lymphocytes. *Trends Biochem. Sci*. 44, 589–598. doi: 10.1016/j.tibs.2019.01.009

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 20, 11–24. doi: 10.1016/j.ccr.2011.06.001

Muller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res*. 42, 8592–8604. doi: 10.1093/nar/gku552

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., et al. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods*13, 919–922. doi: 10.1038/nmeth.3999

Nair, V. S., Song, M. H., Ko, M., and Oh, K. I. (2016). DNA demethylation of the Foxp3 enhancer is maintained through modulation of ten-eleven-translocation and DNA methyltransferases. *Mol. Cells* 39, 888–897. doi: 10.14348/molcells.2016.0276

Nakagawa, T., Lv, L., Nakagawa, M., Yu, Y., Yu, C., D'Alessio, A. C., et al. (2015). CRL4(VprBP) E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. *Mol. Cell* 57, 247–260. doi: 10.1016/j.molcel.2014.12.002

Nakatsukasa, H., Oda, M., Yin, J., Chikuma, S., Ito, M., Koga-Iizuka, M., et al. (2019). Loss of TET proteins in regulatory T cells promotes abnormal proliferation, Foxp3 destabilization and IL-17 expression. *Int. Immunol*. 31, 335–347. doi: 10.1093/intimm/dxz008

Nestor, C. E., Lentini, A., Hagg Nilsson, C., Gawel, D. R., Gustafsson, M., Mattson, L., et al. (2016). 5-hydroxymethylcytosine remodeling precedes lineage specification during differentiation of human CD4(+) T cells. *Cell Rep*. 16, 559–570. doi: 10.1016/j.celrep.2016.05.091

Nestor, C. E., Ottaviano, R., Reinhardt, D., Cruickshanks, H. A., Mjoseng, H. K., McPherson, R. C., et al. (2015). Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. *Genome Biol*. 16:11. doi: 10.1186/s13059-014-0576-y

Orlanski, S., Labi, V., Reizel, Y., Spiro, A., Lichtenstein, M., Levin-Klein, R., et al. (2016). Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. *Proc. Natl. Acad. Sci. U.S.A*. 113, 5018–5023. doi: 10.1073/pnas.1604365113

Palomero, T., Couronne, L., Khiabanian, H., Kim, M. Y., Ambesi-Impiombato, A., Perez-Garcia, A., et al. (2014). Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat. Genet*. 46, 166–170. doi: 10.1038/ng.2873

Pan, W., Zhu, S., Qu, K., Meeth, K., Cheng, J., He, K., et al. (2017). The DNA methylcytosine dioxygenase Tet2 sustains immunosuppressive function of tumor-infiltrating myeloid cells to promote melanoma progression. *Immunity* 47, 284–297.e5. doi: 10.1016/j.immuni.2017.07.020

Parry, A., Rulands, S., and Reik, W. (2020). Active turnover of DNA methylation during cell fate decisions. *Nat. Rev. Genet*. 22, 59–66. doi: 10.1038/s41576-020-00287-8

Pastor, W. A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol*. 14, 341–356. doi: 10.1038/nrm3589

Potus, F., Pauciulo, M. W., Cook, E. K., Zhu, N., Hsieh, A., Welch, C. L., et al. (2020). Novel mutations and decreased expression of the epigenetic regulator TET2 in pulmonary arterial hypertension. *Circulation* 141, 1986–2000. doi: 10.1161/CIRCULATIONAHA.119.044320

Qi, T., Sun, M., Zhang, C., Chen, P., Xiao, C., and Chang, X. (2020). Ascorbic acid promotes plasma cell differentiation through enhancing TET2/3-mediated DNA demethylation. *Cell Rep*. 33:108452. doi: 10.1016/j.celrep.2020.108452

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Rasmussen, K. D., Berest, I., Kessler, S., Nishimura, K., Simon-Carrasco, L., Vassiliou, G. S., et al. (2019). TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. *Genome Res*. 29, 564–575. doi: 10.1101/gr.239277.118

Rasmussen, K. D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev*. 30, 733–750. doi: 10.1101/gad.276568.115

Ravichandran, M., Lei, R., Tang, Q., Zhao, Y., Lee, J., Ma, L., et al. (2019). Rinf regulates pluripotency network genes and tet enzymes in embryonic stem cells. *Cell Rep*. 28, 1993–2003. e5. doi: 10.1016/j.celrep.2019.07.080

Reddy, A., Zhang, J., Davis, N. S., Moffitt, A. B., Love, C. L., Waldrop, A., et al. (2017). Genetic and functional drivers of diffuse large B cell lymphoma. *Cell* 171, 481–494.e15. doi: 10.1016/j.cell.2017.09.027

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. doi: 10.1038/nature14248

Sakata-Yanagimoto, M., Enami, T., Yoshida, K., Shiraishi, Y., Ishii, R., Miyake, Y., et al. (2014). Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat. Genet*. 46, 171–175. doi: 10.1038/ng.2872

Sano, S., Oshima, K., Wang, Y., MacLauchlan, S., Katanasaka, Y., Sano, M., et al. (2018). Tet2-mediated clonal hematopoiesis accelerates heart failure through a mechanism involving the IL-1beta/NLRP3 inflammasome. *J. Am. Coll. Cardiol*. 71, 875–886. doi: 10.1016/j.jacc.2017.12.037

Schoeler, K., Aufschnaiter, A., Messner, S., Derudder, E., Herzog, S., Villunger, A., et al. (2019). TET enzymes control antibody production and shape the mutational landscape in germinal centre B cells. *FEBS J*. 286, 3566–3581. doi: 10.1111/febs.14934

Shi, F. T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M. A., et al. (2013). Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J. Biol. Chem*. 288, 20776–20784. doi: 10.1074/jbc.M113.460386

Shih, A. H., Abdel-Wahab, O., Patel, J. P., and Levine, R. L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer*. 12, 599–612. doi: 10.1038/nrc3343

Shih, A. H., Jiang, Y., Meydan, C., Shank, K., Pandey, S., Barreyro, L., et al. (2015). Mutational cooperativity linked to combinatorial epigenetic gain of function in acute myeloid leukemia. *Cancer Cell* 27, 502–515. doi: 10.1016/j.ccell.2015.03.009

Shukla, V., Halabelian, L., Balagere, S., Samaniego-Castruita, D., Feldman, D. E., Arrowsmith, C. H., et al. (2020). HMCES functions in the alternative end-joining pathway of the DNA DSB repair during class switch recombination in B cells. *Mol. Cell* 77, 384–394.e4. doi: 10.1016/j.molcel.2019.10.031

Smith, Z. D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet*. 14, 204–220. doi: 10.1038/nrg3354

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159. doi: 10.1016/j.cell.2013.02.004

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116

Tsagaratou, A. (2018). TET mediated epigenetic regulation of iNKT cell lineage fate choice and function. *Mol. Immunol*. 101, 564–573. doi: 10.1016/j.molimm.2018.08.020

Tsagaratou, A. (2019). Unveiling the regulation of NKT17 cell differentiation and function. *Mol. Immunol*. 105, 55–61. doi: 10.1016/j.molimm.2018.11.013

Tsagaratou, A., Aijo, T., Lio, C. W., Yue, X., Huang, Y., Jacobsen, S. E., et al. (2014). Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc. Natl. Acad. Sci. U.S.A*. 111, E3306–E3315. doi: 10.1073/pnas.1412327111

Tsagaratou, A., Gonzalez-Avalos, E., Rautio, S., Scott-Browne, J. P., Togher, S., Pastor, W. A., et al. (2017b). TET proteins regulate the lineage specification and TCR-mediated expansion of iNKT cells. *Nat. Immunol*. 18, 45–53. doi: 10.1038/ni.3630

Tsagaratou, A., Lio, C. J., Yue, X., and Rao, A. (2017a). TET methylcytosine oxidases in T cell and B cell development and function. *Front. Immunol*. 8:220. doi: 10.3389/fimmu.2017.00220

Tsagaratou, A., and Rao, A. (2014). TET proteins and 5-methylcytosine oxidation in the immune system. *Cold Spring Harb. Symp. Quant. Biol*. 78, 1–10. doi: 10.1101/sqb.2013.78.020248

Tsai, Y. P., Chen, H. F., Chen, S. Y., Cheng, W. C., Wang, H. W., Shen, Z. J., et al. (2014). TET1 regulates hypoxia-induced epithelial-mesenchymal transition by acting as a co-activator. *Genome Biol*. 15:513. doi: 10.1186/s13059-014-0513-0

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., et al. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* 49, 645–656. doi: 10.1016/j.molcel.2012.12.019

Verma, N., Pan, H., Dore, L. C., Shukla, A., Li, Q. V., Pelham-Webb, B., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet*. 50, 83–95. doi: 10.1038/s41588-017-0002-y

Vincent, J. J., Huang, Y., Chen, P. Y., Feng, S., Calvopina, J. H., Nee, K., et al. (2013). Stage-specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell*. 12, 470–478. doi: 10.1016/j.stem.2013.01.016

Vincenzetti, L., Leoni, C., Chirichella, M., Kwee, I., and Monticelli, S. (2019). The contribution of active and passive mechanisms of 5mC and 5hmC removal in human T lymphocytes is differentiation- and activation-dependent. *Eur. J. Immunol*. 49, 611–625. doi: 10.1002/eji.201847967

Wang, C., McKeithan, T. W., Gong, Q., Zhang, W., Bouska, A., Rosenwald, A., et al. (2015). IDH2R172 mutations define a unique subgroup of patients with angioimmunoblastic T-cell lymphoma. *Blood* 126, 1741–1752. doi: 10.1182/blood-2015-05-644591

Wang, D., Hashimoto, H., Zhang, X., Barwick, B. G., Lonial, S., Boise, L. H., et al. (2017). MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res*. 45, 2396–2407. doi: 10.1093/nar/gkw1184

Wang, Y., Sano, S., Yura, Y., Ke, Z., Sano, M., Oshima, K., et al. (2020). Tet2-mediated clonal hematopoiesis in nonconditioned mice accelerates age-associated cardiac dysfunction. *JCI Insight* 5:e135204. doi: 10.1172/jci.insight.135204

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappsilber, J., et al. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348. doi: 10.1038/nature10066

Wu, D., Hu, D., Chen, H., Shi, G., Fetahu, I. S., Wu, F., et al. (2018). Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer. *Nature* 559, 637–641. doi: 10.1038/s41586-018-0350-5

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet*. 18, 517–534. doi: 10.1038/nrg.2017.33

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., et al. (2016). Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. *Mol. Cell* 64, 913–925. doi: 10.1016/j.molcel.2016.10.013

Xu, T., Stewart, K. M., Wang, X., Liu, K., Xie, M., Ryu, J. K., et al. (2017). Metabolic control of TH17 and induced Treg cell balance by an epigenetic mechanism. *Nature* 548, 228–233. doi: 10.1038/nature23475

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42, 451–464. doi: 10.1016/j.molcel.2011.04.005

Xue, S., Liu, C., Sun, X., Li, W., Zhang, C., Zhou, X., et al. (2016). TET3 inhibits type I IFN Production Independent of DNA demethylation. *Cell Rep*. 16, 1096–1105. doi: 10.1016/j.celrep.2016.06.068

Yang, J., Horton, J. R., Li, J., Huang, Y., Zhang, X., Blumenthal, R. M., et al. (2019). Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. *Nucleic Acids Res*. 47, 8375–8387. doi: 10.1093/nar/gkz381

Yang, R., Qu, C., Zhou, Y., Konkel, J. E., Shi, S., Liu, Y., et al. (2015). Hydrogen sulfide promotes Tet1- and Tet2-mediated Foxp3 demethylation to drive regulatory T cell differentiation and maintain immune homeostasis. *Immunity* 43, 251–263. doi: 10.1016/j.immuni.2015.07.017

Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Ee, L. S., et al. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510. doi: 10.1016/j.cell.2011.11.054

Yue, X., Lio, C. J., Samaniego-Castruita, D., Li, X., and Rao, A. (2019). Loss of TET2 and TET3 in regulatory T cells unleashes effector function. *Nat. Commun*. 10:2011. doi: 10.1038/s41467-019-09541-y

Yue, X., and Rao, A. (2020). TET family dioxygenases and the TET activator vitamin C in immune responses and cancer. *Blood* 136, 1394–1401. doi: 10.1182/blood.2019004158

Yue, X., Trifari, S., Aijo, T., Tsagaratou, A., Pastor, W. A., Zepeda-Martinez, J. A., et al. (2016). Control of Foxp3 stability through modulation of TET activity. *J. Exp. Med*. 213, 377–397. doi: 10.1084/jem.20151438

Zhang, H. Y., Xiong, J., Qi, B. L., Feng, Y. Q., and Yuan, B. F. (2016). The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. *Chem Commun.* 52, 737–740. doi: 10.1039/C5CC07354E

Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., et al. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* 525, 389–393. doi: 10.1038/nature15252

Zhang, T., Guan, X., Choi, U. L., Dong, Q., Lam, M. M. T., Zeng, J., et al. (2019). Phosphorylation of TET2 by AMPK is indispensable in myogenic differentiation. *Epigenetics Chromatin* 12:32. doi: 10.1186/s13072-019-0281-x

Zhang, Y. W., Wang, Z., Xie, W., Cai, Y., Xia, L., Easwaran, H., et al. (2017). Acetylation Enhances TET2 function in protecting against abnormal DNA methylation during oxidative stress. *Mol. Cell* 65, 323–335. doi: 10.1016/j.molcel.2016.12.013

Zhao, Z., Chen, L., Dawlaty, M. M., Pan, F., Weeks, O., Zhou, Y., et al. (2015). Combined loss of Tet1 and Tet2 promotes B cell, but not myeloid malignancies, in mice. *Cell Rep*. 13, 1692–1704. doi: 10.1016/j.celrep.2015.10.037

Zhou, V. W., Goren, A., and Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet*. 12, 7–18. doi: 10.1038/nrg2905

Ziller, M. J., Gu, H., Muller, F., Donaghey, J., Tsai, L. T., Kohlbacher, O., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481. doi: 10.1038/nature12433

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 18 January 2021
doi: 10.3389/fcell.2020.630754

The Complexity of TET2 Functions in Pluripotency and Development

*Vera Garcia-Outeiral[1], Cristina de la Parte[2], Miguel Fidalgo[1] and Diana Guallar[2]**

[1] Stem Cells and Human Diseases Group, Department of Physiology, Center for Research in Molecular Medicine and Chronic Diseases, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, [2] Epitranscriptomics and Ageing Group, Department of Biochemistry and Molecular Biology, Center for Research in Molecular Medicine and Chronic Diseases, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Ten-eleven translocation-2 (TET2) is a crucial driver of cell fate outcomes in a myriad of biological processes, including embryonic development and tissue homeostasis. TET2 catalyzes the demethylation of 5-methylcytosine on DNA, affecting transcriptional regulation. New exciting research has provided evidence for TET2 catalytic activity in post-transcriptional regulation through RNA hydroxymethylation. Here we review the current understanding of TET2 functions on both DNA and RNA, and the influence of these chemical modifications in normal development and pluripotency contexts, highlighting TET2 versatility in influencing genome regulation and cellular phenotypes.

Keywords: TET2, epigenetic, epitranscriptomic, pluripotency, reprogramming, development, 5hmC

OPEN ACCESS

Edited by:
Justin Brumbaugh, University of Colorado Boulder, United States

Reviewed by:
Masaki Yagi, Massachusetts General Hospital and Harvard Medical School, United States
Simão Teixeira da Rocha, University of Lisbon, Portugal

***Correspondence:**
Diana Guallar
diana.guallar@usc.es

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *18 November 2020*
Accepted: *22 December 2020*
Published: *18 January 2021*

Citation:
Garcia-Outeiral V, de la Parte C, Fidalgo M and Guallar D (2021) The Complexity of TET2 Functions in Pluripotency and Development. Front. Cell Dev. Biol. 8:630754. doi: 10.3389/fcell.2020.630754

INTRODUCTION

TET2 belongs to the Ten-eleven-translocation (TET) family of proteins, which also includes TET1 and TET3. The expression of the three TET enzymes differs during early embryonic development, with TET3 mostly restricted to oocytes and zygotes, while TET1 and TET2 are highly expressed later in preimplantation embryos (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). TET proteins have been extensively characterized as α-ketoglutarate and Fe(II)-dependent dioxygenases capable of catalyzing the iterative oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009; Ito et al., 2010), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011) on DNA. Importantly, 5hmC is now not only considered a demethylation intermediate but also an epigenetic mark by itself. TET2 plays thus key roles in shaping the methylome of a cell and establishing novel 5hmC-enriched genomic regions, including enhancers, for chromatin and transcriptional regulation. Moreover, recently we and others have shown that TET2 can also oxidize m^5C (to distinguish it from 5mC in DNA) in RNAs of ESCs and during myelopoiesis for the regulation of RNA stability (Guallar et al., 2018; Shen et al., 2018; He et al., 2020; Lan et al., 2020) Here we review the current understanding of TET2 functions, providing an overview of its roles in development and pluripotency contexts. Although multiple studies have approached 5hmC function through double and triple *Tet1/2/3* depletion (Dawlaty et al., 2013, 2014; Lu et al., 2014; Kang et al., 2015; Li et al., 2016; Verma et al., 2018; Charlton et al., 2020) we focus on *Tet2* single mutants to try to better dissect TET2 specific functions on both DNA and RNA by integrating structure and function, providing a comprehensive overview of this versatile epigenetic and epitranscriptomic regulator.

TET2 PROTEIN STRUCTURE AND FUNCTION

TET2 protein contains an amino-terminal domain and a C-terminal catalytic domain which consists of a Cys-rich region and a double-stranded β helix (DSBH) with a large low-complexity insert (**Figure 1A**). Contrary to TET1 and TET3, during evolution TET2 lost its CXXC zinc finger domain, involved in binding of unmethylated CpG sequences, through a chromosomal inversion that resulted in the appearance of a new gene: IDAX (also known as CXXC4) (Iyer et al., 2009, 2011) (**Figure 1A**). Possibly due to the absence of a CXXC domain, TET2 is more associated to gene bodies and enhancers than to CpG-rich promoters (Hon et al., 2014; Huang et al., 2014). The DSBH domain contains key residues for the interaction of TET2 with its cofactors Fe(II) and 2-oxoglutarate (An et al., 2017) that are required for its catalytic function (**Figure 1A**). The interaction of TET2 with DNA has been investigated through the crystallization of the catalytic domain of human TET2 (TET2 CD) with 5mC-, 5hmC-, and 5fC-modified DNA (Hu et al., 2013, 2015). Importantly, TET2 catalytic cavity does not discriminate between 5mC and its oxidative derivatives, thus allowing iterative oxidation steps (Hu et al., 2013, 2015). Mutations introduced in TET2 DNA-interacting key residues or cofactor binding sites [i.e., Fe(II) and α-KG] have been shown to abolish its catalytic activity *in vitro* and *in vivo* (Ito et al., 2010; Ko et al., 2010; Hu et al., 2013; Shen et al., 2018) (**Figure 1A**). Although the minimal catalytically active fragment of TET2 is located in its C-terminal domain, the full-length protein shows a higher activity on DNA than an N-terminal truncation, suggesting important functions of the N-terminus of TET2 for its catalytic function (Hu et al., 2013; He et al., 2016). Indeed, not only the C-terminus, but also the N-terminal domain of TET2 was shown to be heavily post-translationally modified (Bauer et al., 2015; An et al., 2017), pointing to important regulatory functions of this region in TET2 regulation and function on DNA.

Besides its binding to DNA, more recently TET2 has also been shown to interact with RNA *in vitro* and *in vivo* (Fu et al., 2014; Guallar et al., 2018; Shen et al., 2018; He et al., 2020; Lan et al., 2020). Indeed, and in line with the observation that TET2 is less strongly associated to chromatin than TET1 (Vella et al., 2013), its recruitment to chromatin was recently shown to be facilitated by RNA and decreased upon global transcriptional inhibition (Guallar et al., 2018). While TET2 interaction with RNA was reported to be dependent on an RNA-binding region (RBR) (He et al., 2016) (**Figure 1A**), the absence of the RBR in a C-terminal truncation of TET2 shows reduced, but not fully abrogated TET2-RNA interaction (He et al., 2016). In line with this, TET2 interaction with RNA is greatly reduced in ESCs in absence of one of its most confident interactors: the RNA binding protein PSPC1 (Guallar et al., 2018). It is plausible then that TET2 possesses the intrinsic ability to bind RNA but requires intermediate RNA-binding interactors for stabilizing this interaction and allowing m^5C oxidation. It remains to be determined whether the catalytic activity of TET2 on RNA is only dependent on its RBR or on other protein domains that have not been identified yet. Of note, TET2 presents a greater affinity *in vitro* for m^5C than hm^5C (Guallar et al., 2018), similar to what happens with DNA (Hu et al., 2015). Although TET2 DNA-binding domain (DBD) and RBR domain have been mapped to different regions of the protein (**Figure 1A**) both were found to fold in proximity in human TET2 (Hu et al., 2013). Thus, it will be important to address how TET2 affinity and preference for RNA or DNA is regulated in the context of chromatin, where both nascent RNAs and genomic DNA can be oxidized (Chen et al., 2013; Hon et al., 2014; Guallar et al., 2018; Lan et al., 2020). Importantly, overexpression of full-length TET2 in HEK293T cells showed no catalytic activity in m^5C in RNA, in contrast to TET2 catalytic domain, suggesting negative regulatory functions of the N-terminal domain on TET2 function on RNA oxidation (Fu et al., 2014). On the other hand, Hu and colleagues showed that the GS-linker which replaced the low complexity insert of the TET2-DNA in the crystal structure, was located distant from the core catalytic structure, suggesting that the low complexity insert is positioned on the exterior surface of the catalytic domain (Hu et al., 2013). Low complexity domains have been previously found in RNA-binding proteins and are involved in the formation of liquid-liquid phase-separation (LLPS) droplets by allowing multivalent and weak interactions (Franzmann and Alberti, 2019; Roden and Gladfelter, 2020). Given TET2's ability to interact with RNA, it is tempting to speculate that its low complexity insert could introduce a new layer of regulation through the formation of LLPS condensates. Future work will address this intriguing possibility and will determine the precise contribution and regulation of the different protein domains of TET2 to its roles as an epigenetic and epitranscriptomic regulator.

TET2 FUNCTION IN EARLY DEVELOPMENT: DO WE KNOW EVERYTHING?

Epigenetic modifications, such as DNA methylation and hydroxylation, are critical for normal early embryonic development (Monk et al., 1987; Howlett and Reik, 1991; Kafri et al., 1992; Hajkova et al., 2002; Guibert et al., 2012). During early embryogenesis, although readily detectable at the zygote, TET2 expression peaks in the inner cell mass (ICM) of the preimplantation blastocyst (Tang et al., 2010; Gu et al., 2011; Wossidlo et al., 2011). After implantation, TET2 expression is downregulated and is later only detected in primordial germ cells (PGCs) and in several adult tissues (Ito et al., 2010; Hackett et al., 2013; Guallar et al., 2018). Different groups have generated *Tet2* mutant mice, all consistently reporting that mice lacking TET2 are viable and fertile and can give rise to litters with normal size compared to their wild type counterparts (Ko et al., 2011; Li et al., 2011; Quivoron et al., 2011) (**Table 1**). Several studies have pointed to potential compensatory effects of different TET members during early embryogenesis, based on the observation that single *Tet1* or *Tet2* knock-outs are viable while combined loss of both enzymes in embryos leads to mid-gestation abnormalities and perinatal lethality (Dawlaty

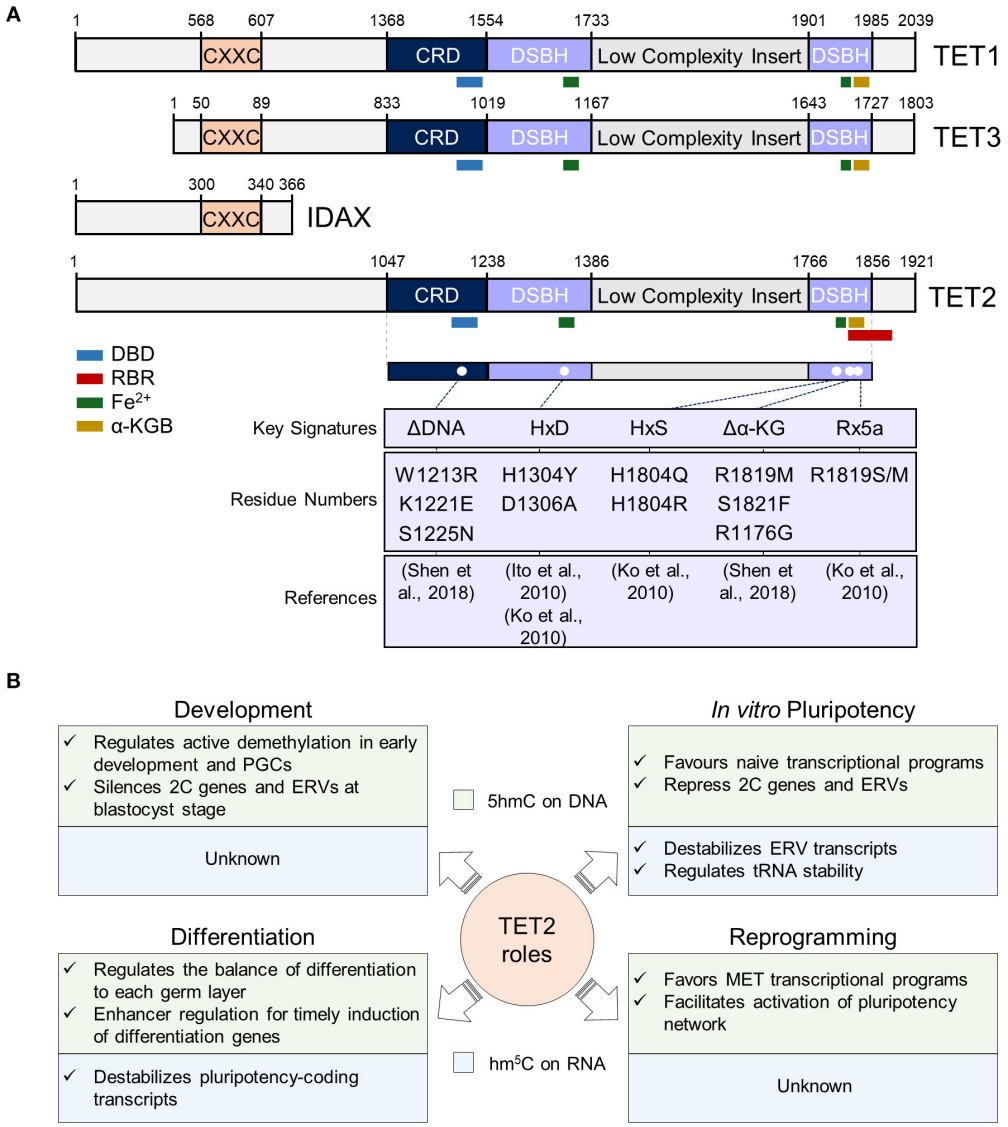


FIGURE 1 | Overview of TET2 structure and functions in different cellular contexts. **(A)** TET2 structure in the context of TET proteins. In the top panel, sequences of mouse TET1 (NP_001240786.1), TET2 (ACY38292.1), TET3 (NP_001334242.1), and IDAX (NP_001004367.2) proteins are represented. The numbers represent the amino acid numbers. The two conserved domains cysteine-rich (CRD) and double-stranded beta-helix (DSBH) domains are indicated [based in Hu et al. (2013) and Iyer et al. (2009) alignments]. The conserved CXXC domain is also represented (coordinates were obtained from NCBI database). The TET2 RNA Binding Region (RBR) sequence described by He et al. (2016), was mapped first in the isoform 1 of TET2 (NP_001035490.2), and then in ACY38292.1 TET2 sequence. The interaction sites with DNA (DBD), RNA, Fe^{2+} and α-Ketoglutarate (α-KG) cofactors are also represented, and their sequences, also shown by Ito et al. (2010) and Hu et al. (2013), were obtained from the NCBI database. In the lower panel, described mouse TET2 mutations are shown. Since the indicated mutations have been mapped in the ACY38292.1 sequence, possible discrepancies may appear in the mutated positions according to the reference article. **(B)** Summary of TET2 functions in pluripotency and during development, differentiation, and somatic cell reprogramming. TET2 functions in these processes at the DNA and RNA level is indicated in the light green and blue squares, respectively. PGCs, primordial germ cells; 2C, 2-cell embryo; ERVs, endogenous retroviruses; MET, mesenchymal-to-epithelial transition.

et al., 2013; Dai et al., 2016), though a substantial fraction of them being viable and fertile (Dawlaty et al., 2013). Given that no upregulation of *Tet1* or *Tet3* was observed upon *Tet2* depletion in mice (Li et al., 2011; Quivoron et al., 2011), it remains to be determined how TET1/3 are relocated to cover for TET2 loss for epigenetic and/or epitranscriptomic regulation *in vivo*. Despite these observations, no detailed analysis has been performed on the effects of *Tet2* depletion at the earliest cell fate transitions *in vivo*. Indeed, although *Tet2* silencing with siRNAs in the zygote was compatible with normal development *in vitro*, it leads to an abnormal upregulation of 2-cell (2C) embryo-specific genes and endogenous retroviral elements up to the preimplantation blastocyst stage in mouse (Guallar et al., 2018) (**Figure 1B**). Aberrant control of endogenous retrovirus at critical stages during development can lead to mobilization of these genomic parasites with implications on genomic stability and regulation

(Rebollo et al., 2012; Faulkner and Garcia-Perez, 2017). These data thus support that a detailed analysis of TET2 targets and function during early preimplantation development is needed to study previously unappreciated fine-tuning functions of this protein which could be relevant for later generations. Later on in development, TET2 participates, together with TET1, in the active DNA demethylation at PGCs, and their deletion leads to defects in 5mC erasure at imprinting sites, as well as LINE1 and IAP endogenous retroviruses *loci* (Hackett et al., 2013) (**Figure 1B**). Nevertheless, the specific contribution of TET2 to this process is waiting to be addressed.

Most importantly, loss of function of *Tet2* in mice results in abnormal DNA hydroxymethylation patterns in bone marrow cells and leads to the emergence of myeloid malignancies between 2 and 4 months of age (Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011; Mulholland et al., 2020a). In depth discussion of TET2 role in normal and malignant hematopoiesis can be found in several excellent reviews (An et al., 2017; Bowman and Levine, 2017; Lio et al., 2019). Moreover, TET2 deficient mice show a mortality of around one-third in homozygosity and an 8% in heterozygosity within the first year of life (Li et al., 2011), implying other yet uncharacterized key roles of TET2 during aging. Additionally, TET2 has been shown to play a role in adult stem cell maintenance and differentiation (Ko et al., 2010, 2011; Li et al., 2011, 2017; Moran-Crusio et al., 2011; Quivoron et al., 2011; Zhang et al., 2016; Gontier et al., 2018; Yang et al., 2018). Future work with single-cell 5hmC/hm^{5}C mapping at the DNA and RNA level, together with the identification of TET2 DNA and RNA targets during development, tissue homeostasis and aging will shed light into unknown functions of TET2.

AN EXPANDED VIEW OF TET2 IN PLURIPOTENCY *IN VITRO*

Pluripotent embryonic stem cells (ESCs) of mouse origin are derived from the ICM of the preimplantation blastocyst (Evans and Kaufman, 1981; Martin, 1981) and can be maintained indefinitely *in vitro* without affecting their properties of self-renewal and differentiation potential. When maintained in conventional culture conditions (i.e., serum/LIF), ESCs are a heterogeneous population of cells with different developmental potential, including cells with totipotency-like features and the duet of *naïve* and primed pluripotent subtypes (Weinberger et al., 2016; Li and Izpisua Belmonte, 2018). Thus, ESCs undoubtfully represent a useful *in vitro* model to study molecular events taking place during early development, and a great amount of the data accumulated during these years regarding how pluripotency is governed have been derived from investigating ESCs. TET2 expression is high in pluripotent ESCs and decreases during differentiation (Ficz et al., 2011; Koh et al., 2011; Hon et al., 2014). Indeed, *Tet2* gene in ESCs is bound and activated by the master pluripotency factor OCT4 (i.e., *Pou5f1*), with *Oct4* silencing causing a reduction of *Tet2* expression (Koh et al., 2011; Wu et al., 2013). Nevertheless, *Tet2* depletion does not affect ESC colony morphology or pluripotent marker gene expression and only moderately induces some differentiation markers such as *Pax6*, *Neurod1* or *Lefty1/2* which is compatible with pluripotency maintenance (Ito et al., 2010; Koh et al., 2011).

As mentioned above, ESCs cultured in presence of serum/LIF are heterogenous, and include a small population (<5%) of cells with an expanded potency (Macfarlan et al., 2012), known as 2C-like cells (2CLCs). These 2CLCs, in contrast to pluripotent stem cells, have the ability to contribute to the trophectoderm. 2CLCs present a similar transcriptome and proteome to that of cells of the 2C-embryo, including high levels of endogenous retroviruses, particularly MERVL elements (Macfarlan et al., 2012; Eckersley-Maslin et al., 2016). TET2 is required for proper silencing of 2C-genes and MERVL elements in ESCs, and its depletion leads to an increase of 2CLCs population *in vitro* (Guallar et al., 2018). Importantly, TET2 contributes to both transcriptional and epitranscriptomic regulation of MERVL elements by recruiting HDACs to chromatin and oxidizing m^{5}C in MERVL transcripts, respectively (**Figure 1B**). Additionally, a very recent study by He et al. (2020) has shown that TET2 can also modify tRNAs to modulate tRNA fragmentation. Given that tRNA fragments have been previously involved in ERV control (Schorn et al., 2017), it is tempting to speculate that TET2-mediated hm^{5}C deposition on tRNA contributes to ERV regulation in pluripotent cells. This exciting possibility remains to be addressed. On the other hand, we and others have shown that TET1 and TET2 not only display specific patterns of expression in defined pluripotent subtypes (Fidalgo et al., 2016; Pantier et al., 2019; Mulholland et al., 2020b), but most importantly, play opposite roles in regulating *naïve* and primed pluripotency (Fidalgo et al., 2016).While TET1 is important for primed pluripotency, TET2 is a key determinant of *naïve* one. In fact, *Tet2* overexpression in primed post-implantation epiblast stem cells (EpiSCs) is sufficient to restore *naïve* pluripotency (Fidalgo et al., 2016). Thus, exquisite control of TET proteins during early development has been revealed through the study of their function in ESCs, exposing a critical role for TET2 in limiting previous developmental stages in ESCs (i.e., 2CLCs) and defining the unique pluripotency features of *naïve* pluripotency. Future studies are needed to address how TET2 catalytic-dependent and independent activities orchestrate pluripotency heterogeneity on DNA and RNA regulatory layers.

TET2 ROLE IN DIFFERENTIATION: THE DNA AND RNA SIDES OF THE STORY

In contrast to TET2 dispensability for ESC maintenance, it is at the exit of pluripotency that a negative effect of loss of *Tet2* has been documented in a variety of systems. On the one hand, *Tet2* depleted ESCs give rise to large hemorrhagic teratomas that grow more aggressively than controls and are accompanied by a greater neuroectoderm contribution, thus reflecting a skewed differentiation potential (Koh et al., 2011). On the other hand, absence of *Tet2* during differentiation of ESCs to neural progenitors (NPCs) leads to a delay in the induction of differentiation genes, molecularly explained by a gain in 5mC at enhancers in absence of

TABLE 1 | Overview of TET2 studies deciphering its role and functions in mouse, murine embryonic stem cells (mESCs), human embryonic stem cells (hESCs) and during somatic cell reprogramming.

Model	Genotype	Phenotype/most relevant findings	References
Mouse	*Tet2* KO	- Reduction in 5hmC levels in the genomic DNA of BM cells - 1/3 of *Tet2*$^{-/-}$ *and 8% of Tet2*$^{+/-}$died within 1 year of age - Development of myeloid malignancies	Li et al., 2011
	Tet2 KO	- Amplification and competitive advantage of hematopoietic stem and progenitor cells - Pleiotropic alterations of the hematopoietic compartments including both lymphoid and myeloid lineages - Myeloid malignancies with differentiation abnormalities	Quivoron et al., 2011
	Tet2 KO	- No evidence for embryonic lethality - Decreased genomic levels of 5hmC - Increase in the size of the HSPC pool in a cell-autonomous manner. - Competitive advantage of HSCs, leading to an enhanced hematopoiesis into both lymphoid and myeloid lineages.	Ko et al., 2011
	Tet2 KD (Early embryo)	- No observable delay in embryonic development - Derepression of MERVL elements and MERVL-associated genes	Guallar et al., 2018
mESCs	*Tet2* KD	- ∼40% of decrease in genomic 5hmC levels - No changes in ESC maintenance - Formation of large hemorrhagic and aggressive teratomas, with greater contribution from neuroectoderm.	Koh et al., 2011
	Tet2 KD/KO	- ∼50% of decrease in genomic 5hmC levels in *Tet2*-deficient mESCs - Loss of 5hmC in gene bodies and increased 5hmC levels at promoter/TSS regions in *Tet2*-deficient mESCs - No detectable 5hmC levels in *Tet2*$^{-/-}$ mESCs	Huang et al., 2014
	Tet2 KO	- >90% loss of global 5hmC levels - Enhancers hypermethylation - Delay in early differentiation markers expression	Hon et al., 2014
	Tet2 KD	- Telomere shortening - Chromosomal instability	Yang et al., 2016
	Tet2 KO	- Significant decrease in global 5hmC levels in both *naïve* ESCs and primed EpiLCs	Mulholland et al., 2020b
	Tet2 KO	- Significant decrease in global 5hmC levels in *naïve* ESCs - Significant increase in DNA methylation at LINE1/L1 elements - Severe hypermethylation at promoters, gene bodies, and repetitive elements - Premature repression of *Dppa3*	Mulholland et al., 2020a
hESCs	*Tet2* KD	- Increased expression of neuroectoderm markers and decreased expression of mesoderm and endoderm makers during EBs differentiation - No detectable alterations in pluripotency	Langlois et al., 2014
Somatic cell reprogramming	*Tet2* KD *MEF+OSKM*	- Abolishment of iPSC generation - Decrease of H3K4me2 at pluripotency loci - TET2 binds to *Nanog* and *Esrrb loci* during SCR and catalyze 5hmC	Doege et al., 2012
	Tet2 KD *mEGC+hB cell fusion*	- Reduction of ∼ 50% in reprogramming capacity - Decrease in 5hmC levels at the somatic *Oct4* promoter	Piccolo et al., 2013
	preiPSC +NANOG	- TET2 overexpression enhances NANOG-mediated reprogramming - TET2 KD abolishes the reprogramming synergy of NANOG with a catalytically deficient mutant of TET1	Costa et al., 2013
	Tet2 KO *MEF+OSKM*	- Reduction of ∼70% in reprogramming capacity	Hu et al., 2014
	B cell^{+} CEBPα^{+} OSKM	- TET2 overexpression enhances SCR - CEBP*α* activates *Tet2* expression - TET2 binds and demethylates pluripotency loci (e.g. *Oct4*)	Di Stefano et al., 2014
	Tet2 KO B cell+ CEBPα+ iOSKM	- TET2 is required for B cell activation by CEBPα and for reprogramming by OSKM - TET2 interactors (i.e. CEBPα, KLF4, TFCP2l1) contribute with TET2 to reprogramming	Sardina et al., 2018
	2nd MEF+ iOSKM	- ZSCAN4F binds to TET2 and recruits it to 5hmC-modify genes involved in glycolysis and proteasome activity	Cheng et al., 2020

Model of study, genotype, phenotype, and mayor findings are included along with the reference article. HSCs, hematopoietic stem cells; HSPCs, hematopoietic stem progenitor cells; KD, knockdown; KO, knockout; iPSCs, induced pluripotent stem cells. 2nd MEF, secondary system expressing inducible OSKM.

TET2 activity (Hon et al., 2014) (**Figure 1B**). This lack of phenotype of *Tet2* absence in pluripotency maintenance but its requirement for proper differentiation has also been shown to be conserved in human ESCs (Langlois et al., 2014), supporting an evolutionary conserved requirement of TET2 for cell commitment.

Remarkably, although both *Tet1* and *Tet2* are highly expressed in ESCs, many studies show a higher contribution of TET2 to 5hmC levels of genomic DNA (Hon et al., 2014; Huang et al., 2014; Mulholland et al., 2020b). At the molecular level, TET1 and TET2 have been shown to target different genomic regions, and thus be responsible for the regulation of 5mC at distinct targets (Chen et al., 2013; Hon et al., 2014; Huang et al., 2014). In particular, TET2 contributes to enhancer hypomethylation by 5mC oxidation, which is important for timely induction of lineage genes upon differentiation (Hon et al., 2014) (**Figure 1B**). Moreover, TET2-mediated deposition of 5hmC in gene bodies in ESCs has been related to exon inclusion/exclusion regulation (Huang et al., 2014) (**Table 1**). These observations, together with the absence of *Tet1* upregulation upon *Tet2* depletion in ESCs, and *vice versa*, (Koh et al., 2011; Huang et al., 2014), point toward non-redundant functions of TET1/2 in pluripotency maintenance and exit (Huang et al., 2014; Mulholland et al., 2020b). This is in sharp contrast to the absence of strong phenotypes in embryonic development in single *Tet1* or *Tet2* knock-out animals, which only becomes obvious when both TET1/2 proteins are absent and a partially penetrant perinatal lethality appears (Dawlaty et al., 2013). Intrinsic differences of *in vitro* and *in vivo* studies may be behind these apparently contradictory observations. It may be as well possible that the effects of TET1 or TET2 individual loss-of-function may only be visible *in vivo* after a number of generations. Indeed, this could be expected according to several studies which reported defects in telomere maintenance in ESCs in absence of one or several of the TET members (Lu et al., 2014; Yang et al., 2016), which would have minor effects on F0-F1 offsprings, but could lead to the appearance of genomic instability and propensity to cancer development in later generations.

Besides its well-known presence on DNA as an epigenetic mark, hm^5C has also been detected in RNA of ESCs (Fu et al., 2014; Guallar et al., 2018; Lan et al., 2020). Besides 2C-specific transcripts and MERVL RNAs (Guallar et al., 2018), around 800 mRNAs, including more than 100 pluripotency-related transcripts, are hm^5C-modified in ESCs (Lan et al., 2020) (**Figure 1B**). Importantly, characterization of flag-tagged endogenous TET1 and TET2 RNA interactome revealed overlapping and distinct targets, suggesting common and specific roles of TET proteins on RNA regulation (Lan et al., 2020). Deletion of TET2 RNA-binding region (RBR) (He et al., 2016) resulted in a 47% decrease in TET2 interaction with ESC transcripts, supporting that additional TET2 domains and/or interactors, such as PSPC1, could be required for its stable interaction with RNA (Guallar et al., 2018; Lan et al., 2020). In line with the observation that transcriptional inhibition greatly reduces TET2 chromatin occupancy (Guallar et al., 2018), hm^5C modified-RNAs were found to be enriched in nuclear nascent-chromatin associated RNAs, pointing to a co-transcriptional mechanism of RNA modification by TETs (Lan et al., 2020). In sharp contrast, Legrand et al. (2017) did not detect hm^5C modification in mRNAs when analyzing total RNA in ESCs. A potential explanation could be that hm^5C-modified RNAs are likely less abundant compared to their non-modified counterparts, given that this modification has been linked to RNA destabilization (Guallar et al., 2018; Shen et al., 2018; Lan et al., 2020), and thus more difficult to detect. Future studies will be needed to further clarify these apparent contradictions. Importantly, RNA destabilization of pluripotency-coding transcripts, through m^5C oxidation to hm^5C including some encoding for Polycomb proteins, is important for contributing to transcriptome flexibility necessary for differentiation (Lan et al., 2020). This is in agreement with previous observations in both pathological and physiological contexts which have reported a stabilizing role of m^5C on mRNA (Chen et al., 2019; Yang et al., 2019). Future studies are required to interrogate hm^5C patterns and dynamics during early development, coinciding with TET expression, and its potential role in fine-tuning cell fate transitions through mRNA modifications.

TET2 FUNCTIONS IN THE ESTABLISHMENT OF PLURIPOTENCY

The process of somatic cell reprogramming (SCR) to regain pluripotency from somatic cells was first achieved by overexpressing four transcription factors (*Oct4, Sox2, Klf4,* and c-*Myc*) called OSKM or Yamanaka factors (Takahashi and Yamanaka, 2006). This complex process of cell fate transitions leads to the generation of induced pluripotent stem cells (iPSCs) and involves wide remodeling cell fate steps, including the shutting down of somatic transcriptional programs and reactivation of pluripotency genes (Apostolou and Stadtfeld, 2018), and a metabolic rewiring toward a more glycolytic and less oxidative cellular metabolism (Yoshida et al., 2009; Zhu et al., 2010; Folmes et al., 2011; Mathieu et al., 2014). *Tet2* expression is induced as early as day 2 of OSKM reprogramming in mouse embryonic fibroblasts (MEFs) (Doege et al., 2012). TET2 has been reported to play key roles in a myriad of reprogramming systems, which include reprogramming of adult B cells through cell fusion (Piccolo et al., 2013), pre-iPSC reprogramming by NANOG (Costa et al., 2013), B cell reprogramming through CEBP and OSKM overexpression (Di Stefano et al., 2014; Sardina et al., 2018) and the classical Yamanaka reprogramming (Doege et al., 2012; Hu et al., 2014; Cheng et al., 2020) (**Table 1**). All these studies reflect a critical role of TET2 as an epigenetic regulator for pluripotency reacquisition in somatic cells. At the molecular level, TET2 has been shown to be important for 5hmC deposition on DNA and transcriptional regulation of pluripotency genes, mesenchymal-to-epithelial facilitators (i.e., miR200 cluster), proteasomal subunits and glycolytic regulators (Doege et al., 2012; Piccolo et al., 2013; Di Stefano et al., 2014; Hu et al., 2014; Sardina et al., 2018; Cheng et al., 2020) (**Figure 1B** and **Table 1**). Importantly, TET2 is required both at the earliest phases of reprogramming, and during later stages

of pluripotency reacquisition, further underlining the relevance of TET2 as a central reprogramming player (Doege et al., 2012; Costa et al., 2013; Di Stefano et al., 2014; Hu et al., 2014; Sardina et al., 2018; Cheng et al., 2020). Importantly, TET2 has been shown to mediate its roles in reprogramming in coordination with several transcription factors including NANOG, CEBPα, KLF4, TFCP2l1, and ZSCAN4F (Costa et al., 2013; Sardina et al., 2018; Cheng et al., 2020), though for some of them definitive proof of direct recruitment has not been provided. Although at this stage there is no doubt of TET2 relevance for iPSC generation as a transcriptional regulator through DNA hydroxymethylation, its contribution to pluripotency reacquisition through RNA modification is currently unexplored and will need further studies.

SUMMARY AND IMPLICATIONS

TET2 function as an epigenetic modifier through 5mC to 5hmC, 5fC and 5caC oxidation is currently fully accepted. The relevance of TET2-mediated catalytic-dependent functions, has been implicated *in vivo* in the correct function of hematopoietic compartment (Lio and Rao, 2019) and *in vitro* in the appropriate establishment of the epigenetic landscape that endows a correct timing of differentiation program expression as well as facilitates pluripotency reacquisition through SCR (Doege et al., 2012; Piccolo et al., 2013; Di Stefano et al., 2014; Hon et al., 2014; Hu et al., 2014; Langlois et al., 2014; Sardina et al., 2018). In contrast, only a handful of studies have shown the implication of TET2 in these physiological and pathological contexts in terms of RNA m^5C oxidation (**Figure 1B**). The determination of the TET2-RNA complex structure, combined with the known TET2-DNA complex (Hu et al., 2013, 2015) may provide additional mechanistic insights into differences and similarities between TET-mediated oxidation of m^5C/5mC in RNA and DNA. Moreover, neither the enzyme(s) responsible for depositing m^5C on RNA for TET2 oxidation, nor the hm^5C readers in charge of destabilizing TET2-targeted RNAs are currently known. Furthermore, the observation that a fraction of its interactome is comprised by unmodified RNAs raises the possibility that TET2 could also have catalytic-independent functions on RNA, similar to what has been previously described on DNA (Chen et al., 2013; Zhang et al., 2015; Guallar et al., 2018; Ito et al., 2019). The development of technology that allows mapping of chromatin regulators using low input samples, coupled with the appearance of sequencing techniques which can faithfully distinguish 5mC/m^5C from 5hmC/hm^5C at a single base resolution on DNA and RNA, respectively, opens up exciting possibilities to explore TET2 function as an epigenetic and epitranscriptomic regulator during early development, in ESC subpopulations and in tissue heterogeneity in the near future.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was funded by grants from the Spanish Agencia Estatal de Investigación, co-funded by the FEDER Program of the EU (RTI2018-096708-J-I00 to DG, and BFU2016-80899-P and PID2019-105739GB-I00 to MF) (AEI/FEDER, UE), the 2020 Leonardo Grant for Researchers and Cultural Creators BBVA Foundation (2020-PO041 to DG) and the Xunta de Galicia-Consellería de Cultura, Educación e Ordenación Universitaria (ED431F 2016/016 to MF). DG and MF are recipients of a Marie Skłodowska-Curie (MSCA-IF-EF-RI-895984) from the European Commission and Ramón y Cajal (RYC-2014-16779) awards, respectively. VG-O (Ministerio de Ciencia, Innovación y Universidades, FPU17/01131) and CP (Xunta de Galicia, ED481A-2020/187) are recipients of fellowships.

ACKNOWLEDGMENTS

We thank members of the Guallar and Fidalgo laboratories for critical review of this manuscript.

REFERENCES

An, J., Rao, A., and Ko, M. (2017). TET family dioxygenases and DNA demethylation in stem cells and cancers. *Exp. Mol. Med.* 49:e323. doi: 10.1038/emm.2017.5

Apostolou, E., and Stadtfeld, M. (2018). Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Curr. Opin. Genet. Dev.* 52, 77–85. doi: 10.1016/j.gde.2018.06.002

Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., et al. (2015). Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J. Biol. Chem.* 290, 4801–4812. doi: 10.1074/jbc.M114.605881

Bowman, R. L., and Levine, R. L. (2017). TET2 in normal and malignant hematopoiesis. *Cold Spring Harb. Perspect. Med.* 7:a026518. doi: 10.1101/cshperspect.a026518

Charlton, J., Jung, E. J., Mattei, A. L., Bailly, N., Liao, J., Martin, E. J., et al. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* 52, 819–827. doi: 10.1038/s41588-020-0639-9

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561–564. doi: 10.1038/nature11742

Chen, X., Li, A., Sun, B. F., Yang, Y., Han, Y. N., Yuan, X., et al. (2019). 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat. Cell Biol.* 21, 978–990. doi: 10.1038/s41556-019-0361-y

Cheng, Z. L., Zhang, M. L., Lin, H. P., Gao, C., Song, J. B., Zheng, Z., et al. (2020). The Zscan4-Tet2 transcription nexus regulates metabolic rewiring and enhances proteostasis to promote reprogramming. *Cell Rep.* 32:107877. doi: 10.1016/j.celrep.2020.107877

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495, 370–374. doi: 10.1038/nature11925

Dai, H. Q., Wang, B. A., Yang, L., Chen, J. J., Zhu, G. C., Sun, M. L., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. *Nature* 538, 528–532. doi: 10.1038/nature20095

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., et al. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* 29, 102–111. doi: 10.1016/j.devcel.2014.03.003

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* 24, 310–323. doi: 10.1016/j.devcel.2012.12.015

Di Stefano, B., Sardina, J. L., van Oevelen, C., Collombet, S., Kallin, E. M., Vicent, G. P., et al. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* 506, 235–239. doi: 10.1038/nature12885

Doege, C. A., Inoue, K., Yamashita, T., Rhee, D. B., Travis, S., Fujita, R., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* 488, 652–655. doi: 10.1038/nature11333

Eckersley-Maslin, M. A., Svensson, V., Krueger, C., Stubbs, T. M., Giehr, P., Krueger, F., et al. (2016). MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* 17, 179–192. doi: 10.1016/j.celrep.2016.08.087

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156. doi: 10.1038/292154a0

Faulkner, G. J., and Garcia-Perez, J. L. (2017). L1 Mosaicism in mammals: extent, effects, and evolution. *Trends Genet.* 33, 802–816. doi: 10.1016/j.tig.2017.07.004

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., et al. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398–402. doi: 10.1038/nature10008

Fidalgo, M., Huang, X., Guallar, D., Sanchez-Priego, C., Valdes, V. J., Saunders, A., et al. (2016). Zfp281 Coordinates opposing functions of Tet1 and Tet2 in pluripotent states. *Cell Stem Cell* 19, 355–369. doi: 10.1016/j.stem.2016.05.025

Folmes, C. D., Nelson, T. J., Martinez-Fernandez, A., Arrell, D. K., Lindor, J. Z., Dzeja, P. P., et al. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab.* 14, 264–271. doi: 10.1016/j.cmet.2011.06.011

Franzmann, T. M., and Alberti, S. (2019). Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* 294, 7128–7136. doi: 10.1074/jbc.TM118.001190

Fu, L., Guerrero, C. R., Zhong, N., Amato, N. J., Liu, Y., Liu, S., et al. (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* 136, 11582–11585. doi: 10.1021/ja505305z

Gontier, G., Iyer, M., Shea, J. M., Bieri, G., Wheatley, E. G., Ramalho-Santos, M., et al. (2018). Tet2 rescues age-related regenerative decline and enhances cognitive function in the adult mouse brain. *Cell Rep.* 22, 1974–1981. doi: 10.1016/j.celrep.2018.02.001

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610. doi: 10.1038/nature10443

Guallar, D., Bi, X., Pardavila, J. A., Huang, X., Saenz, C., Shi, X., et al. (2018). RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat. Genet.* 50, 443–451. doi: 10.1038/s41588-018-0060-9

Guibert, S., Forne, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* 22, 633–641. doi: 10.1101/gr.130997.111

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., et al. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339, 448–452. doi: 10.1126/science.1229277

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., et al. (2002). Epigenetic reprogramming in mouse primordial germ cell. *Mech. Dev.* 117, 15–23. doi: 10.1016/S0925-4773(02)00181-8

He, C., Bozler, J., Janssen, K. A., Wilusz, J. E., Garcia, B. A., Schorn, A. J., et al. (2020). TET2 chemically modifies tRNAs and regulates tRNA fragment levels. *Nat. Struct. Mol. Biol.* doi: 10.1038/s41594-020-00526-w. [Epub ahead of print].

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A., et al. (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* 64, 416–430. doi: 10.1016/j.molcel.2016.09.034

Hon, G. C., Song, C. X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., et al. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* 56, 286–297. doi: 10.1016/j.molcel.2014.08.026

Howlett, S. K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. *Development* 113, 119–127.

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., et al. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 155, 1545–1555. doi: 10.1016/j.cell.2013.11.020

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., et al. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature* 527, 118–122. doi: 10.1038/nature15713

Hu, X., Zhang, L., Mao, S. Q., Li, Z., Chen, J., Zhang, R. R., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* 14, 512–522. doi: 10.1016/j.stem.2014.01.001

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., et al. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1361–1366. doi: 10.1073/pnas.1322921111

Iqbal, K., Jin, S. G., Pfeifer, G. P., and Szabo, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3642–3647. doi: 10.1073/pnas.1014033108

Ito, K., Lee, J., Chrysanthou, S., Zhao, Y., Josephs, K., Sato, H., et al. (2019). Non-catalytic roles of Tet2 are essential to regulate hematopoietic stem and progenitor cell homeostasis. *Cell Rep.* 28, 2480–2490.e4. doi: 10.1016/j.celrep.2019.07.094

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133. doi: 10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303 doi: 10.1126/science.1210597

Iyer, L. M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 101, 25–104. doi: 10.1016/B978-0-12-387685-0.00002-0

Iyer, L. M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8, 1698–1710. doi: 10.4161/cc.8.11.8580

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., et al. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.* 6, 705–714. doi: 10.1101/gad.6.5.705

Kang, J., Lienhard, M., Pastor, W. A., Chawla, A., Novotny, M., Tsagaratou, A., et al. (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 112, E4236–E4245. doi: 10.1073/pnas.1510510112

Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., et al. (2011). Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. U. S. A.* 108, 14566–14571. doi: 10.1073/pnas.1112317108

Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839–843. doi: 10.1038/nature09586

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 8, 200–213. doi: 10.1016/j.stem.2011.01.008

Lan, J., Rajan, N., Bizet, M., Penning, A., Singh, N. K., Guallar, D., et al. (2020). Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nat. Commun.* 11:4956. doi: 10.1038/s41467-020-18729-6

Langlois, T., da Costa Reis Monte-Mor, B., Lenglet, G., Droin, N., Marty, C., Le Couedic, J. P., et al. (2014). TET2 deficiency inhibits mesoderm and hematopoietic differentiation in human embryonic stem cells. *Stem Cells* 32, 2084–2097. doi: 10.1002/stem.1718

Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M., et al. (2017). Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* 27, 1589–1596. doi: 10.1101/gr.210666.116

Li, M., and Izpisua Belmonte, J. C. (2018). Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.* 20, 382–392. doi: 10.1038/s41556-018-0067-6

Li, X., Yao, B., Chen, L., Kang, Y., Li, Y., Cheng, Y., et al. (2017). Ten-eleven translocation 2 interacts with forkhead box O3 and regulates adult neurogenesis. *Nat. Commun.* 8:15903. doi: 10.1038/ncomms15903

Li, X., Yue, X., Pastor, W. A., Lin, L., Georges, R., Chavez, L., et al. (2016). Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc. Natl. Acad. Sci. U. S. A.* 113, E8267–E76. doi: 10.1073/pnas.1617802113

Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., et al. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* 118, 4509–4518. doi: 10.1182/blood-2010-12-325241

Lio, C. J., and Rao, A. (2019). TET enzymes and 5hmC in adaptive and innate immune systems. *Front. Immunol.* 10:210. doi: 10.3389/fimmu.2019.00210

Lio, C. J., Yuita, H., and Rao, A. (2019). Dysregulation of the TET family ofepigenetic regulators in lymphoid and myeloid malignancies. *Blood* 134, 1487–1497. doi: 10.1182/blood.2019791475

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* 28, 2103–2119. doi: 10.1101/gad.248005.114

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi: 10.1038/nature11244

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 78, 7634–7638. doi: 10.1073/pnas.78.12.7634

Mathieu, J., Zhou, W., Xing, Y., Sperber, H., Ferreccio, A., Agoston, Z., et al. (2014). Hypoxia-inducible factors have distinct and stage-specific roles during reprogramming of human cells to pluripotency. *Cell Stem Cell* 14, 592–605. doi: 10.1016/j.stem.2014.02.012

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99, 371–382.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 20, 11–24. doi: 10.1016/j.ccr.2011.06.001

Mulholland, C. B., Nishiyama, A., Ryan, J., Nakamura, R., Yigit, M., Glück, I. M., et al. (2020a). Recent evolution of a TET-controled and DPPA3/STELLA-driven pathway of passive demethylation in mammals. *bioRxiv*. doi: 10.1038/s41467-020-20453-0

Mulholland, C. B., Traube, F. R., Parsa, E., Eckl, E.-M., Schönung, M., Modic, M., et al. (2020b). Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. *Sci. Rep.* 10:12066. doi: 10.1038/s41598-020-68600-3

Pantier, R., Tatar, T., Colby, D., and Chambers, I. (2019). Endogenous epitope-tagging of Tet1, Tet2 and Tet3 identifies TET2 as a naive pluripotency marker. *Life Sci Alliance* 2:e201900516. doi: 10.26508/lsa.201900516

Piccolo, F. M., Bagci, H., Brown, K. E., Landeira, D., Soza-Ried, J., Feytout, A., et al. (2013). Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. *Mol. Cell* 49, 1023–1033. doi: 10.1016/j.molcel.2013.01.032

Quivoron, C., Couronne, L., Della Valle, V., Lopez, C. K., Plo, I., Wagner-Ballon, O., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* 20, 25–38. doi: 10.1016/j.ccr.2011.06.003

Rebollo, R., Romanish, M. T., and Mager, D. L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* 46, 21–42. doi: 10.1146/annurev-genet-110711-155621

Roden, C., and Gladfelter, A. S. (2020). RNA contributions to the form and function of biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* doi: 10.1038/s41580-020-0264-6

Sardina, J. L., Collombet, S., Tian, T. V., Gomez, A., Di Stefano, B., Berenguer, C., et al. (2018). Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell* 23, 727–741.e9. doi: 10.1016/j.stem.2018.08.016

Schorn, A. J., Gutbrod, M. J., LeBlanc, C., and Martienssen, R. (2017). LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* 170, 61–71 doi: 10.1016/j.cell.2017.06.013

Shen, Q., Zhang, Q., Shi, Y., Shi, Q., Jiang, Y., Gu, Y., et al. (2018). Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature* 554, 123–127. doi: 10.1038/nature25434

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676. doi: 10.1016/j.cell.2006.07.024

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., et al. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6, 468–478. doi: 10.1016/j.stem.2010.03.015

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., et al. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* 49, 645–656. doi: 10.1016/j.molcel.2012.12.019

Verma, N., Pan, H., Dore, L. C., Shukla, A., Li, Q. V., Pelham-Webb, B., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* 50, 83–95. doi: 10.1038/s41588-017-0002-y

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J. H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* 17, 155–169. doi: 10.1038/nrm.2015.28

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., et al. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2:241. doi: 10.1038/ncomms1240

Wu, Y., Guo, Z., Liu, Y., Tang, B., Wang, Y., Yang, L., et al. (2013). Oct4 and the small molecule inhibitor, SC1, regulates Tet2 expression in mouse embryonic stem cells. *Mol. Biol. Rep.* 40, 2897–2906. doi: 10.1007/s11033-012-2305-5

Yang, J., Guo, R., Wang, H., Ye, X., Zhou, Z., Dan, J., et al. (2016). Tet enzymes regulate telomere maintenance and chromosomal stability of mouse ESCs. *Cell Rep.* 15, 1809–1821. doi: 10.1016/j.celrep.2016.04.058

Yang, R., Yu, T., Kou, X., Gao, X., Chen, C., Liu, D., et al. (2018). Tet1 and Tet2 maintain mesenchymal stem cell homeostasis via demethylation of the P2rX7 promoter. *Nat. Commun.* 9:2143. doi: 10.1038/s41467-018-04464-6

Yang, Y., Wang, L., Han, X., Yang, W. L., Zhang, M., Ma, H. L., et al. (2019). RNA 5-Methylcytosine facilitates the maternal-to-zygotic transition by preventing maternal mRNA decay. *Mol. Cell* 75, 1188–1202.e11. doi: 10.1016/j.molcel.2019.06.033

Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T., and Yamanaka, S. (2009). Hypoxia enhances the generation of induced pluripotent stem cells. *Cell Stem Cell* 5, 237–241. doi: 10.1016/j.stem.2009.08.001

Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., et al. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* 525, 389–393. doi: 10.1038/nature15252

Zhang, X., Su, J., Jeong, M., Ko, M., Huang, Y., Park, H. J., et al. (2016). DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat. Genet.* 48, 1014–1023. doi: 10.1038/ng.3610

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., et al. (2010). Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell Stem Cell* 7, 651–655. doi: 10.1016/j.stem.2010.11.015

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 18 January 2021
doi: 10.3389/fcell.2020.626383

Mammalian SWI/SNF Chromatin Remodeling Complexes in Embryonic Stem Cells: Regulating the Balance Between Pluripotency and Differentiation

Ying Ye[1], *Xi Chen*[2]* *and Wensheng Zhang*[1]*

[1] Cam-Su Genomic Resource Center, Medical College of Soochow University, Suzhou, China, [2] Department of Biology, Southern University of Science and Technology, Shenzhen, China

OPEN ACCESS

Edited by:
Bruno Di Stefano,
Baylor College of Medicine,
United States

Reviewed by:
Davide Seruggia,
Boston Children's Hospital,
United States
Octavio A. Romero-Ferraro,
Josep Carreras Leukaemia Research Institute (IJC), Spain

***Correspondence:**
Xi Chen
chenx9@sustech.edu.cn
Wensheng Zhang
zhangwensheng@suda.edu.cn

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: 05 November 2020
Accepted: 30 November 2020
Published: 18 January 2021

Citation:
Ye Y, Chen X and Zhang W (2021) Mammalian SWI/SNF Chromatin Remodeling Complexes in Embryonic Stem Cells: Regulating the Balance Between Pluripotency and Differentiation. Front. Cell Dev. Biol. 8:626383. doi: 10.3389/fcell.2020.626383

The unique capability of embryonic stem cells (ESCs) to maintain and adjust the equilibrium between self-renewal and multi-lineage cellular differentiation contributes indispensably to the integrity of all developmental processes, leading to the advent of an organism in its adult form. The ESC fate decision to favor self-renewal or differentiation into specific cellular lineages largely depends on transcriptome modulations through gene expression regulations. Chromatin remodeling complexes play instrumental roles to promote chromatin structural changes resulting in gene expression changes that are key to the ESC fate choices governing the equilibrium between pluripotency and differentiation. BAF (Brg/Brahma-associated factors) or mammalian SWI/SNF complexes employ energy generated by ATP hydrolysis to change chromatin states, thereby governing the accessibility of transcriptional regulators that ultimately affect transcriptome and cell fate. Interestingly, the requirement of BAF complex in self-renewal and differentiation of ESCs has been recently shown by genetic studies through gene expression modulations of various BAF components in ESCs, although the precise molecular mechanisms by which BAF complex influences ESC fate choice remain largely underexplored. This review surveys these recent progresses of BAF complex on ESC functions, with a focus on its role of conditioning the pluripotency and differentiation balance of ESCs. A discussion of the mechanistic bases underlying the genetic requirements for BAF in ESC biology as well as the outcomes of its interplays with key transcription factors or other chromatin remodelers in ESCs will be highlighted.

Keywords: SWI/SNF (BAF) complex, embryonic stem cells, pluripotency, differentiation, chromatin remodeling complex

FUNDAMENTALS OF EMBRYONIC STEM CELLS

Embryonic stem (ES) cells are pluripotent cells derived from the inner cell mass of blastocyst-stage embryos (Evans and Kaufman, 1981; Martin, 1981; Thomson et al., 1998). Their importance to basic biology and translational medicine derives from two unique characteristics that distinguish them from all other cell types. First, they can be maintained as a self-renewing stem cell population *in vitro*. Second, they have the capacity to differentiate into every cell type of the body. For decades, the

mechanism underlying the self-renewal and pluripotency of ESCs has been the focus of intensive research in the field of stem cell biology.

Mouse ESCs were initially established and maintained by co-culture with mouse embryonic fibroblasts (Evans and Kaufman, 1981; Martin, 1981). Subsequent studies identified leukemia inhibitory factor (LIF) as one of the feeder-cell-derived molecules that support the growth of undifferentiated ESCs through gp130-mediated activation of STAT3 (Smith et al., 1988; Williams et al., 1988; Stewart et al., 1992; Niwa et al., 1998; Matsuda et al., 1999). In contrast to mouse ESCs, LIF and STAT3 appear to be dispensable for the self-renewal of primed human ESCs (Thomson et al., 1998; Reubinoff et al., 2000; Dahéron et al., 2004). Furthermore, serum could be replaced by BMP4, which activates Smad and subsequently induces the expressions of helix–loop–helix ID factors (Ying et al., 2003). ESCs cultivated in a serum-free medium with MAPK/ERK pathway inhibitor PD0325901 and glycogen synthase kinase 3β (GSK3β) pathway inhibitor CHIR99021 (called 2i), and LIF represent naïve state and exhibit greater and homogenous pluripotent gene expression than those cultivated in serum with LIF (Ying et al., 2008). With these developments, it is now possible to grow ESCs with defined factors in the absence of serum or feeder cells.

Numerous studies demonstrate the importance of transcription factors (TFs) on the maintenance of ESCs and their pluripotency, among which OCT4, SOX2, and NANOG form a core transcriptional regulatory circuit (Martello and Smith, 2014). Ablation of their expression disrupts the pluripotency network, leading to the exit from pluripotency and initiation of differentiation of ESCs (Okamoto et al., 1990; Schöler et al., 1990; Nichols et al., 1998; Chambers et al., 2003; Mitsui et al., 2003; Masui et al., 2007; Silva et al., 2009). In addition, downregulation of epiblast-specific TFs such as TBX3, KLF2/4/5, TFCP2L1, and ESRRB disturb the self-renewal of ESCs, demonstrating their supporting roles in the maintenance of ESC identity (Ivanova et al., 2006; Jiang et al., 2008; Festuccia et al., 2012, 2018; Martello et al., 2013; Yeo et al., 2014).

CHROMATIN REMODELING COMPLEX

Besides signaling and TFs, chromatin remodeling complexes play instrumental roles on maintaining the identity of ESCs (Papatsenko et al., 2018). At least three epigenetic mechanisms allow regulation of DNA expression and chromatin accessibility, which include DNA methylation (Winata et al., 2018), histone modifications (Lawrence et al., 2016), and ATP-dependent chromatin remodeling (Clapier et al., 2017). This mini-review will focus on the SWI/SNF family of ATP-dependent chromatin remodeling complexes and its role in the maintenance of ESCs and their differentiation.

The ATP-dependent SWI/SNF complexes were first discovered in yeast in genetic screens aimed at uncovering factors responsible for the regulation of mating type switching (Stern et al., 1984) and those being able to allow changing of nutrient sources used for energy supply (Carlson et al., 1981; Neigeborn and Carlson, 1984, 1987), therefore termed SWI/SNF complex (short for SWItch/sucrose non-fermentable) (Alfert et al., 2019). In *Drosophila melanogaster*, this complex was first discovered in screens to uncover genes that are able to suppress phenotypes caused by mutations in Polycomb genes (PcGs) (Tamkun et al., 1992; Elfring et al., 1994).

The BAF (BRG1/BRM-associated factor) complex, the mammalian homolog of the SWI/SNF complex, is one of four ATP-dependent chromatin remodeling complex families known in mammals (the other three are INO80/SWR1, ISWI, and CHD complexes) (Clapier and Cairns, 2009). Three mammalian BAF complexes have been identified based on their different subunit compositions. The subunits are encoded by 29 genes (Centore et al., 2020). The PBAF (Polybromo-associated BAF complex) is distinguished from the cBAF (canonical BAF complex) by the incorporation of BAF200 instead of BAF250A/B and of BAF180 (Yan et al., 2005). Furthermore, PBAF lacks SS18 but includes the PBAF-specific subunits BAF45A and BRD7 (Kaeser et al., 2008; Middeljans et al., 2012). Recently, a third class, called ncBAF (for non-canonical BAF complex) or GBAF (after its distinctive subunits GLTSCR1/1L), has been identified, which is characterized by the incorporation of BRD9 and GLTSCR1/1L (Alpsoy and Dykhuizen, 2018), but lacks the cBAF subunits such as BAF47, BAF57, and BAF250 and the PBAF-specific subunits BAF180 and BRD7 (Clapier et al., 2017; Alpsoy and Dykhuizen, 2018; Mashtalir et al., 2018).

FUNCTION OF BAF COMPONENTS IN EMBRYONIC STEM CELLS

BAF complexes are made up of multiple subunits that are assembled in a combinatorial manner to tailor their functions, regulating specific developmental events (Ho and Crabtree, 2010). The BAF complexes in different tissues are distinctive for their specific subunit compositions (Lickert et al., 2004; Lessard et al., 2007; Vogel-Ciernia et al., 2013; Harada et al., 2017; Sokpor et al., 2018; Akerberg and Pu, 2020). Hence, it is not only the BAF complex itself that controls biological processes, but the expressions of distinct BAF complexes with unique subunit compositions are also a major part of the regulatory process.

The assembly of an ESC-specific BAF (esBAF) complex is required for the regulation of the ESC transcriptome, therefore controlling the self-renewal and differentiation of ESCs (Ho et al., 2009a). The esBAF complex depends on BRG1 as the ATPase subunit, as BRM does not express in ESCs (Ho et al., 2009b). Moreover, esBAF can be distinguished by the incorporation of Baf250a not 250b, Baf60a/b not 60c, and a Baf155::155 homodimer instead of a Baf155::170 heterodimer (Kaeser et al., 2008; Ho et al., 2009b). In human ESCs, BAF170, and not BAF155, seems to play an important role in the maintenance of pluripotency (Zhang et al., 2014).

A possible way of elucidating the role and importance of individual subunits of multiprotein complexes is the deletion or downregulation of genes encoding their subunits. Genetic

TABLE 1 | BAF subunits and their role in embryonic stem cells (ESCs).

Subunit	Type of mutant	Phenotype	References
BAF250a (SMARCF1)	*Baf250a*$^{fl/-}$ mES cells	Inhibit self-renewal, promote differentiation into primitive endoderm-like cells, are defective in differentiating into fully functional mesoderm-derived cardiomyocytes and adipocytes, but are capable of differentiating into ectoderm-derived neurons.	Gao et al., 2008
BAF250b	*Baf250b*$^{-/-}$ mES cells	Reduced proliferation rate and an abnormal cell cycle. Deficient in the self-renewal capacity of undifferentiated ES cells and exhibit certain phenotypes of differentiated cells.	Yan et al., 2008
BRG1 (SMARCA4)	*Brg1* shRNA; *Brg1*$^{fl/fl}$ mES cells	Essential for ES cell self-renewal and pluripotency genes, and upregulation of differentiation genes.	Ho et al., 2009b; Kidder et al., 2009
DPF2 (BAF45d)	*Dpf2*$^{fl/fl}$ mES cells	Impaired meso-endoderm differentiation but promoted neuro-ectoderm differentiation.	Zhang et al., 2019
Srg3(BAF155) (SMARCC1)	*Srg3*$^{-/-}$; *Baf155* shRNA mES cells	Mutant blastocysts hatch, adhere, and form a layer of trophoblast giant cells, degenerated inner cell mass after prolonged culture, facilitate ESC differentiation; decrease proliferation; and increase apoptosis of ES cells.	Kim et al., 2001; Ho et al., 2009b; Schaniel et al., 2009
BAF47 (SNF5) (SMARCB1)	*SNF5/INI1* null mouse embryos; *Baf47* shRNA and ectopic expression	Die between 3.5 and 5.5 days postcoitum; and Ini1-null blastocysts fail to hatch, form the trophectoderm, or expand the inner cell mass when cultured *in vitro*; knockdown Baf47 block differentiation; overexpression of Baf47 enhances differentiation of mES cells.	Guidi et al., 2001; You et al., 2013; Sakakura et al., 2019
BAF53a	*Baf53a* knockdown; *Baf53a* cKO mES cells	Cell growth repressed, induced cell death and reduction of mouse ES cell viability; Baf53b rescued cell death of Baf53a-deficient mES cells.	Zhu et al., 2017
BRD9	*Brd9* shRNA, BRD9 inhibitor	Preserving the naïve pluripotency of mouse ESCs and preventing transition to the primed state.	Gatchalian et al., 2018
BAF170 (SMARCC2)	BAF170 ectopic expression	Defects in pluripotency of mouse ES cells.	Ho et al., 2009b
BAF60c (SMARCD3)	*Baf60c* knockdown	Impaired anterior/secondary heart field, and abnormal cardiac and skeletal muscle differentiation.	Lickert et al., 2004
hBAF250a	*hBaf250a*$^{-/-}$	Disrupted cardiomyocyte differentiation.	Lei et al., 2020
hBRG1, hBAF170	*hBrg1*, *Baf170* knockdown human ESCs	Defects in self-renewal of human ES cells.	Zhang et al., 2014

inactivation of specific subunit of BAF complex leads to diverse aberrant phenotypes in ESCs (**Table 1**).

Both *Brg1* and *Baf155* knockout mice are lethal at the pre-implantation stage (Bultman et al., 2000; Kim et al., 2001), suggesting that they play a key role in the maintenance of pluripotency. Consistently, depletion of either *Brg1* or *Baf155* expression in ESCs leads to the downregulation of the key pluripotent TFs *Oct4*, *Sox2*, and *Nanog*, indicating that BAF155 and BRG1 cooperate to maintain ESC identity (Fazzio et al., 2008; Ho et al., 2009b; Kidder et al., 2009). Corresponding to the unique subunit composition of esBAF, neither *Brm* nor *Baf170* overexpression can rescue *Brg1* or *Baf155* knockout, respectively (Ho et al., 2009b). Different from mouse esBAF complex, human *Baf170* deficiency led to the differentiation of human ESCs, demonstrating that the BAF170-containing BAF complex was required for the self-renewal of human ESCs (Zhang et al., 2014). Ho et al. reported that neuro-ectodermal differentiation was impaired and mesodermal differentiation was delayed in *Brg1* knockout embryoid bodies (Ho et al., 2009a). In contrast, knockdown of *Brg1* in ESCs promoted the expression of differentiation marker genes (Kim et al., 2001). These results might indicate the distinct role of BRG1 in ESCs and differentiating cells. *Baf47* knockout mice are also lethal at the pre-implantation stage (Klochendler-Yeivin et al., 2000; Guidi et al., 2001). The negative regulation of *Oct4* by *Baf47* may control the balance between pluripotency and differentiation of ESCs (You et al., 2013). A recent contradicting report indicates the upregulation of *Cdx2* expression in *Baf47* KO ESCs (Sakakura et al., 2019). BAF250a and BAF250b are two mutually exclusive esBAF subunits. Inactivation of either of them decreases expression of *Oct4* and *Sox2* or *Nanog*, thereby inhibiting the self-renewal of ESCs (Gao et al., 2008; Yan et al., 2008). Knockout of *Baf250a* upregulates primitive endoderm maker genes, such as *Gata4*, *Gata6*, and *Sox17* in mouse ESCs but impairs mesodermal lineage differentiation of both mouse and human ESCs (Gao et al., 2008; Lei et al., 2020). In contrast, knockout of *Baf250b* increased the expression of mesoderm marker genes in mouse ESCs, *Gata2* and *Esx1* (Yan et al., 2008). This may indicate the balance role of BAF250a- and BAF250b-containing BAF complexes on mesoderm differentiation of ESCs. The deletion

of esBAF subunit *Baf45d* only perturbs the self-renewal of ESCs, whereas its knockout impairs the differentiation of ESCs to all three germ lineages (Zhang et al., 2019).

In addition to the long-known esBAF, the newly discovered ncBAF complex also plays an important role in the regulation of the ESC transcriptome. Inhibition of *Brd9*, the specific ncBAF subunit, changed the morphology of ESCs to that resembles primed or epiblast ESCs (EpiESCs), reduced colony-forming capability, and downregulated expressions of *Nanog* and *Klf4*, indicating that BRD9 has an important role in maintaining the naïve pluripotent state of ESCs (Gatchalian et al., 2018).

Consistent to its functions on the maintenance and differentiation of ESCs, BAF complexes also play important roles in the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs). Depletion of *Brg1* was associated with failures in reprogramming (Hansis et al., 2004; Egli and Eggan, 2010). Overexpression of *Brg1* and *Baf155* achieves euchromatin, enhances binding of OCT4, and increases the reprogramming efficiency of MEFs to iPSCs (Singhal et al., 2010). In contrast, downregulation of *Brm* and *Baf170* improves reprogramming efficiency and promotes complete reprogramming of immature iPSCs (Jiang et al., 2015). Therefore, similar to the distinct roles of different BAF subunits for the maintenance and differentiation of ESCs, different BAF components may play different roles in the reprogramming.

In summary, esBAF complex is crucial for the maintenance of ESCs, with distinct effects from the deletion of different subunits. The knockout of different subunits of esBAF leads to defects of ES differentiations to different lineages, though the precise molecular mechanisms underlying the different phenotypes upon the deletion of different subunits need further investigations.

MECHANISTIC INSIGHTS INTO BAF COMPLEXES IN EMBRYONIC STEM CELLS

Inactivation of individual esBAF subunits downregulates expression of pluripotent TFs (Gao et al., 2008; Yan et al., 2008; Ho et al., 2009b; You et al., 2013; Zhang et al., 2019), indicating that esBAF controls the self-renewal of ESCs via regulating pluripotent factors. esBAF subunits BRG1, BAF155, BAF250a, and BAF45d are bound at sites engaged by the core pluripotency TFs OCT4, SOX2, and NANOG (Ho et al., 2009a; Gatchalian et al., 2018; Zhang et al., 2019). The expression of the core TFs *Nanog*, *Oct4*, and *Sox2* as well as a variety of other factors governs the maintenance of pluripotency in ESCs (Martello and Smith, 2014). Specifically, NANOG, OCT4, and SOX2 have been shown to repress the expression of developmental genes while modulating their own expression levels by binding to each other's promoter regions (Saunders et al., 2013). Both BRG1 and BAF155 are located near the transcriptional starting site (TSS) of core pluripotency factors *Oct4*, *Nanog*, and *Sox2* (Ho et al., 2009b). The binding of OCT4, SOX2, and NANOG is impaired in *Dpf2* KO ESCs (Zhang et al., 2019). Therefore, esBAF complex may collaborate with the core TFs to regulate the expression of pluripotency TFs, thereby controlling the maintenance of ES self-renewal.

LIF/STAT3 pathway is essential for the maintenance of mouse ESCs (Niwa et al., 1998; Matsuda et al., 1999). It also plays a role in naïve or murine-like human ESC pluripotency (Hanna et al., 2010; Buecker et al., 2014). BRG1, DPF2, and STAT3 binding sites display a substantial genome-wide overlap in mouse ESCs. STAT3 binding is considerably impaired in *Brg1*- or *Dpf2* (*Baf45d*)-depleted ESCs, leading to the downregulation of *Stat3* target genes (Ho et al., 2011; Zhang et al., 2019). esBAF stabilizes the binding of STAT3 and thereby helps the maintenance of ES self-renewal (Ho et al., 2011).

esBAF also regulates gene expression in ESCs (Ho et al., 2011; Gatchalian et al., 2018; Zhang et al., 2019). esBAF preferably binds to enhancers and regulates their H3K27ac deposition. Loss of esBAF subunit Dpf2 changes the activity of enhancers and the target gene expression (Ho et al., 2011; Zhang et al., 2019). On the other hand, ncBAF predominantly binds to H3K4-trimethylated promoter regions and is associated with the TFs *Klf4* (Kruppel-like factor 4) (Gatchalian et al., 2018), indicating a distinct mechanism to regulate gene expression. One of the most striking differences is, however, that ncBAF binds to TAD (topologically associating domain) boundaries and CTCF sites, potentially contributing to the regulation of genome topology (Gatchalian et al., 2018). Thus, esBAF and ncBAF complexes might regulate ESC identity coordinately via distinct mechanisms that future studies need to elucidate.

COLLABORATION OF BAF COMPLEX WITH PRC2 COMPLEX IN EMBRYONIC STEM CELLS

The PcG family has first been discovered in *Drosophila* followed by the observation of male flies with ectopic sex combs (Margueron and Reinberg, 2011). In mammals, the multiprotein-containing Polycomb repressive complex 2 (PRC2) has repressive influence on the genome (Margueron and Reinberg, 2011). PRC2 is dispensable for the maintenance of undifferentiated mouse ESCs, as the deletion of PRC2 components has little effect on their morphology and self-renewal, although a subset of PRC2 target genes are derepressed (Boyer et al., 2006; Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008; Leeb et al., 2010). Similarly, deletion of EZH2, the catalytic subunit of PRC2 complex, in human ESCs also causes misexpression of developmental genes but severely affects the self-renewal of human ESCs (Collinson et al., 2016).

The cooperative function of BAF complex with PRC2 in ESCs has been revealed (Ho et al., 2011; Zhang et al., 2019). Ho et al. report that the core subunit of BAF complex, BRG1, in ESCs potentiates LIF signaling by opposing PRC2 complex (Ho et al., 2011). The opposing regulation of BAF and PRC2 subunits DPF2 and EED on *Tbx3* expression is critical to the proper differentiation of ESCs to mesoendoderm. The other PRC2 subunit EZH2 also opposes DPF2-dependent differentiation through a distinct mechanism involving *Nanog* repression (Zhang et al., 2019). Contrary to the opposing function of BAF and PRC2 complexes, BRG1 facilitates PRC2 to reinforce the repression on its target genes in ESCs. Therefore, esBAF not

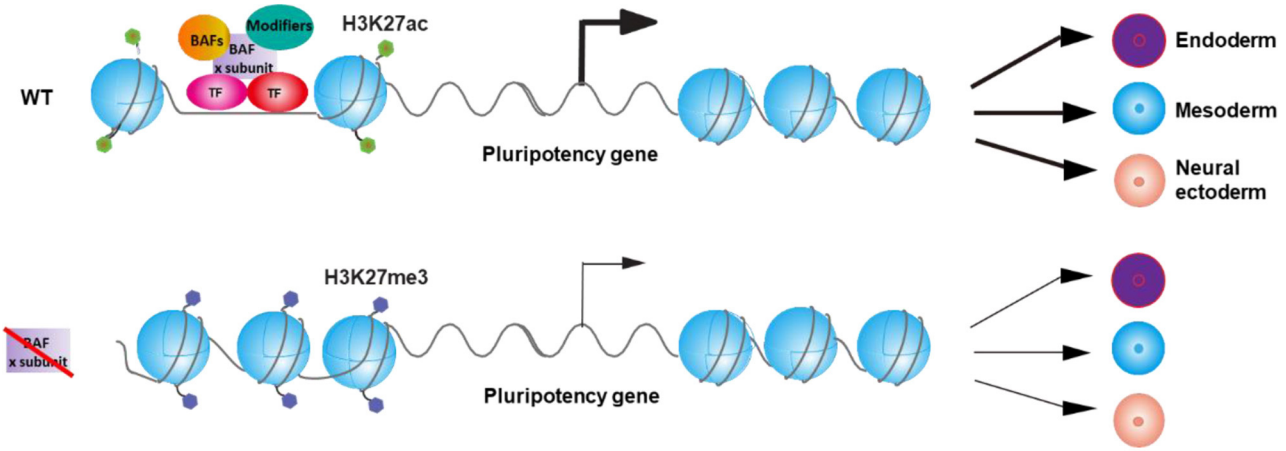


FIGURE 1 | Model for the regulation of the balance between pluripotency and differentiation of embryonic stem cells (ESCs) by BAF complexes *via* the control of pluripotency gene expression. BAF complex, transcription factors, and other chromatin modifiers regulate the expression of specific pluripotency gene(s) and thereby control the balance between pluripotency and differentiation of ESCs. Inactivation of specific BAF subunit leads to the deregulation of the expression of specific pluripotency gene(s) and therefore results in the differential differentiation defects of ESC.

only simply antagonizes PcG but also acts synergistically with the common goal of supporting pluripotency (Ho et al., 2011).

Inactivation of different subunit of BAF complex differentially affects the expression of pluripotency TFs (**Table 1**). Furthermore, DPF2 opposingly regulates differentiation of ESCs via controlling different pluripotency TFs with distinct components of PRC2 complex. As a result, distinct BAF subunits may regulate the expression of different pluripotency TFs collaboratively with other TFs and chromatin modifiers and therefore leads to different differentiation defects of ESCs upon the deletion of different BAF components (**Figure 1**). Interestingly, knockout of *Dpf2* only affects about 8% of BRG1 binding sites on the genome (Zhang et al., 2019), indicating that the loss of a specific BAF subunit only affects the binding of a specific portion of BAF complex on genome, which may lead to the specific phenotypes upon the deletion of that subunit. Future studies on the deletion of other specific BAF components on the binding of BRG1 will help to explain the different phenotypes in ESCs that resulted from the deletion of specific subunits of other chromatin complexes.

BAF Complex on the Balance Between Pluripotency and Differentiation

BAF complex regulates both the maintenance and differentiation of ESCs (Ho and Crabtree, 2010). Knockout of *Dpf2* does not change the level of H3K27ac around DPF2-bound lineage markers during differentiation of ESCs. Consistently, overexpression of Dpf2 in ESCs does not lead to the upregulation of endo- and mesodermal markers, supporting an idea that BAF complex regulates ESC differentiation indirectly (Zhang et al., 2019). *Tbx3* is a pluripotency TF, and the downregulation of its expression impairs ESC self-renewal (Ivanova et al., 2006). *Tbx3* also plays key roles on ESC differentiation. Deregulation of its expression impairs the differentiation of ESCs (Lu et al., 2011; Weidgang et al., 2014; Waghray et al., 2015; Zhang et al., 2019). *Dpf2* participates in the self-renewal and differentiation of ESCs via precisely regulating *Tbx3* expression in both ESCs and differentiating cells (Zhang et al., 2019). As a core pluripotency TF, *Nanog* represses expression of differentiation marker genes and maintains the self-renewal of ESCs (Niwa, 2007). *Dpf2* regulates the expression of *Nanog* with PRC2 subunit *Ezh2*, thereby controlling the proper differentiation of ESCs (Zhang et al., 2019).

BAF47 controls the differentiation of ESCs via regulating *Oct4* expression, which provides another example to demonstrate how BAF complex controls the balance between pluripotency and differentiation (You et al., 2013). The controversial result from a recent work upon the deletion of *Baf47* in ESCs indicates that more studies are required to clarify the discrepancy (Sakakura et al., 2019). Changed expression of pluripotency genes in ESCs upon the deletion of other BAF subunits has been reported (Gao et al., 2008; Ho et al., 2009b; Ho and Crabtree, 2010). It will be of interest to carry out systematic studies to determine whether and how other subunits of BAF complex regulate the expression of specific pluripotency genes and thereby control the balance between pluripotency and differentiation.

CONCLUSION

BAF complex is functionally important for the self-renewal and differentiation of ESCs. Knockout of different subunits of BAF complex changes the expression of different pluripotency TFs and impairs the differentiation of ESCs differently. Thus, it is of particular importance to explore how BAF complex regulates the balance between the maintenance of identity of ESCs and their differentiation to three germ layers. We have outlined studies that described functions of specific subunits of various BAF complexes in ESCs. Moreover, our recent study

demonstrates an attractive mechanism that distinct BAF subunit controls the integrity of only a part of BAF complex on the genome, and therefore, its deletion only affects the binding of a part of BAF complex, which directly changes the expression of distinct pluripotency TFs in both ESCs and differentiating cells with other TFs and chromatin modifiers (Zhang et al., 2019). Consistently, another recent report demonstrates that the loss of a single subunit of the BAF complex in cancer cells did not destroy the entire complex but will change the composition of the BAF complex (Schick et al., 2019). BAF complex regulates ESC differentiation via controlling the expressions of pluripotency TFs, with different subunits affecting ESC differentiation via regulating different TFs. Further systematic studies of other subunits of BAF complex are needed to warrant the mechanism, which may also explain the distinct phenotypes that resulted from the deletion of various subunits of chromatin remodeling complex.

AUTHOR CONTRIBUTIONS

YY, XC, and WZ conceived the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by a grant from the National Natural Science Foundation of China (31970812).

ACKNOWLEDGMENTS

We are indebted to all the cited authors for their work. The authors gratefully acknowledge the comments on the manuscript by Dr. Li Li.

REFERENCES

Akerberg, B. N., and Pu, W. T. (2020). Genetic and epigenetic control of heart development. *Cold Spring Harb Perspect Biol*. 12:a036756. doi: 10.1101/cshperspect.a036756

Alfert, A., Moreno, N., and Kerl, K. (2019). The BAcomplex, F. in development and disease. *Epigenetics Chromatin*. 12:19. doi: 10.1186/s13072-019-0264-y

Alpsoy, A., and Dykhuizen, E. C. (2018). Glioma tumor suppressor candidate region gene 1 (GLTSCR1) and its paralog GLTSCR1-like form SWI/SNchromatin, F. remodeling subcomplexes. *Biol. J. Chem*. 293, 3892–3903. doi: 10.1074/jbc.RA117.001065

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353. doi: 10.1038/nature04733

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., et al. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838–853. doi: 10.1016/j.stem.2014.04.003

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., et al. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* 6, 1287–1295. doi: 10.1016/S1097-2765(00)00127-1

Carlson, M., Osmond, B. C., and Botstein, D. (1981). Mutants of yeast defective in sucrose utilization. *Genetics* 98, 25–40.

Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C., and Chan, H. M. (2020). Mammalian SWI/SNF chromatin remodeling complexes: emerging mechanisms and therapeutic strategies. *Trends Genet*. 36, 936–950. doi: 10.1016/j.tig.2020.07.011

Chamberlain, S. J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* 26, 1496–1505. doi: 10.1634/stemcells.2008-0102

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., et al. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113, 643–655. doi: 10.1016/S0092-8674(03)00392-1

Clapier, C. R., and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* 78, 273–304. doi: 10.1146/annurev.biochem.77.062706.153223

Clapier, C. R., Iwasa, J., Cairns, B. R., and Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol*. 18, 407–422. doi: 10.1038/nrm.2017.26

Collinson, A., Collier, A. J., Morgan, N. P., Sienerth, A. R., Chandra, T., Andrews, S., et al. (2016). Deletion of the polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells. *Cell Rep*. 17, 2700–2714. doi: 10.1016/j.celrep.2016.11.032

Dahéron, L., Opitz, S. L., Zaehres, H., Lensch, M. W., Andrews, P. W., Itskovitz-Eldor, J., et al. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* 22, 770–778. doi: 10.1634/stemcells.22-5-770

Egli, D., and Eggan, K. (2010). Recipient cell nuclear factors are required for reprogramming by nuclear transfer. *Development* 137, 1953–1963. doi: 10.1242/dev.046151

Elfring, L. K., Deuring, R., McCallum, C. M., Peterson, C. L., and Tamkun, J. W. (1994). Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Mol. Cell Biol*. 14, 2225–2234. doi: 10.1128/MCB.14.4.2225

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156. doi: 10.1038/292154a0

Fazzio, T. G., Huff, J. T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134, 162–174. doi: 10.1016/j.cell.2008.05.031

Festuccia, N., Halbritter, F., Corsinotti, A., Gagliardi, A., Colby, D., Tomlinson, S. R., et al. (2018). Esrrb extinction triggers dismantling of naïve pluripotency and marks commitment to differentiation. *EMBO J.* 37:e95476. doi: 10.15252/embj.201695476

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., et al. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* 11, 477–490. doi: 10.1016/j.stem.2012.08.002

Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W. C., and Wang, Z. (2008). Ecell, S. pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc. Natl. Acad. Sci. U. S. A*. 105, 6656–60. doi: 10.1073/pnas.0801802105

Gatchalian, J., Malik, S., Ho, J., Lee, D. S., Kelso, T. W. R., Shokhirev, M. N., et al. (2018). A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat. Commun*. 9:5139. doi: 10.1038/s41467-018-07528-9

Guidi, C. J., Sands, A. T., Zambrowicz, B. P., Turner, T. K., Demers, D. A., Webster, W., et al. (2001). Disruption of Ini1 leads to peri-implantation lethality and tumorigenesis in mice. *Mol. Cell Biol*. 21, 3598–3603. doi: 10.1128/MCB.21.10.3598-3603.2001

Hanna, J., Cheng, A. W., Saha, K., Kim, J., Lengner, C. J., Soldner, F., et al. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl. Acad. Sci. U. S. A*. 107, 9222–9227. doi: 10.1073/pnas.1004584107

Hansis, C., Barreto, G., Maltry, N., and Niehrs, C. (2004). Nuclear reprogramming of human somatic cells by xenopus egg extract requires BRG1. *Curr. Biol*. 14, 1475–1480. doi: 10.1016/j.cub.2004.08.031

Harada, A., Ohkawa, Y., and Imbalzano, A. N. (2017). Temporal regulation of chromatin during myoblast differentiation. *Semin. Cell Dev. Biol*. 72, 77–86. doi: 10.1016/j.semcdb.2017.10.022

Ho, L., and Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature* 463, 474–484. doi: 10.1038/nature08911

Ho, L., Jothi, R., Ronan, J. L., Cui, K., Zhao, K., and Crabtree, G. R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5187–5191. doi: 10.1073/pnas.0812888106

Ho, L., Miller, E. L., Ronan, J. L., Ho, W. Q., Jothi, R., and Crabtree, G. R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat. Cell Biol.* 13, 903–913. doi: 10.1038/ncb2285

Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., et al. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5181–5186. doi: 10.1073/pnas.0812889106

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., et al. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature* 442, 533–538. doi: 10.1038/nature04915

Jiang, J., Chan, Y. S., Loh, Y. H., Cai, J., Tong, G. Q., Lim, C. A., et al. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* 10, 353–360. doi: 10.1038/ncb1698

Jiang, Z., Tang, Y., Zhao, X., Zhang, M., Donovan, D. M., and Tian, X. C. (2015). Knockdown of Brm and Baf170, components of chromatin remodeling complex, facilitates reprogramming of somatic cells. *Stem Cells Dev.* 24, 2328–2336. doi: 10.1089/scd.2015.0069

Kaeser, M. D., Aslanian, A., Dong, M. Q., Yates, J. R. 3rd, and Emerson, B. M. (2008). BRD7, a novel PBAF-specific SWI/SNF subunit is required for target gene activation and repression in embryonic stem cells. *Biol. J. Chem.* 283, 32254–32263. doi: 10.1074/jbc.M806061200

Kidder, B. L., Palmer, S., and Knott, J. G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* 27, 317–328. doi: 10.1634/stemcells.2008-0710

Kim, J. K., Huh, S. O., Choi, H., Lee, K. S., Shin, D., Lee, C., et al. (2001). Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. *Mol. Cell Biol.* 21, 7787–7795. doi: 10.1128/MCB.21.22.7787-7795.2001

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., and Yaniv, M. (2000). The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO Rep.* 1, 500–506. doi: 10.1093/embo-reports/kvd129

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression trends. *Genet.* 32, 42–56. doi: 10.1016/j.tig.2015.10.007

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* 24, 265–276. doi: 10.1101/gad.544410

Lei, I., Tian, S., Chen, V., Zhao, Y., and Wang, Z. (2020). SWI/SNF component BAF250a coordinates OCT4 and WNT signaling pathway to control cardiac lineage differentiation. *Front. Cell Dev. Biol.* 7:358. doi: 10.3389/fcell.2019.00358

Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., et al. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* 55, 201–215. doi: 10.1016/j.neuron.2007.06.019

Lickert, H., Takeuchi, J. K., Von Both, I., Walls, J. R., McAuliffe, F., Adamson, S. L., et al. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* 432, 107–112. doi: 10.1038/nature03071

Lu, R., Yang, A., and Jin, Y. (2011). Dual functions of T-box 3 (Tbx3) in the control of self-renewal and extraembryonic endoderm differentiation in mouse embryonic stem cells. *Biol. J. Chem.* 286, 8425–36. doi: 10.1074/jbc.M110.202150

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349. doi: 10.1038/nature09784

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J.* 32, 2561–2574. doi: 10.1038/emboj.2013.177

Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol.* 30, 647–675. doi: 10.1146/annurev-cellbio-100913-013116

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 78, 7634–7638. doi: 10.1073/pnas.78.12.7634

Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., et al. (2018). Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* 175, 1272–1288.e20. doi: 10.1016/j.cell.2018.09.032

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., et al., (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* 9, 625–635. doi: 10.1038/ncb1589

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., et al. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J.* 18, 4261–4269. doi: 10.1093/emboj/18.15.4261

Middeljans, E., Wan, X., Jansen, P. W., Sharma, V., Stunnenberg, H. G., and Logie, C. (2012). SS18 together with animal-specific factors defines human BAF-type SWI/SNF complexes. *PLoS ONE* 7:e33834. doi: 10.1371/journal.pone.0033834

Mitsui, K., Tokuzawa Y., Itoh, H., Segawa, H., Murakami, M., Takahashi, K., et al. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113, 631–642. doi: 10.1016/S0092-8674(03)00393-3

Neigeborn, L., and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* 108, 845–858.

Neigeborn, L., and Carlson, M. (1987). Mutations causing constitutive invertase synthesis in yeast: genetic interactions with snf mutations. *Genetics* 115, 247–253.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., et al. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor. *Cell* 95, 379–91. doi: 10.1016/S0092-8674(00)81769-9

Niwa, H. (2007). How is pluripotency determined and maintained? *Development* 134, 635–646. doi: 10.1242/dev.02787

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* 12, 2048–2060. doi: 10.1101/gad.12.13.2048

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60, 461–472. doi: 10.1016/0092-8674(90)90597-8

Papatsenko, D., Waghray, A., and Lemischka, I. R. (2018). Feedback control of pluripotency in embryonic stem cells: signaling, transcription and epigenetics. *Stem Cell Res.* 29, 180–188. doi: 10.1016/j.scr.2018.02.012

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell Biol.* 27, 3769–3779. doi: 10.1128/MCB.01432-06

Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat. Biotechnol.* 18, 399-404. doi: 10.1038/74447

Sakakura, M., Ohta, S., Yagi, M., Tanaka, A., Norihide, J., Woltjen, K., et al. (2019). Smarcb1 maintains the cellular identity and the chromatin landscapes of mouse embryonic stem cells. *Biochem. Biophys. Res. Commun.* 519, 705–713. doi: 10.1016/j.bbrc.2019.09.054

Saunders, A., Faiola, F., and Wang, J. (2013). Concise review: pursuing self-renewal and pluripotency with the stem cell factor. *Stem Cells* 31, 1227–1236. doi: 10.1002/stem.1384

Schaniel, C., Ang, Y. S., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., et al. (2009). Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells* 27, 2979–2991. doi: 10.1002/stem.223

Schick, S., Rendeiro, A. F., Runggatscher, K., Ringler, A., Boidol, B., Hinkel, M., et al. (2019). Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers. *Nat. Genet.* 51, 1399–1410. doi: 10.1038/s41588-019-0477-9

Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.* 9, 2185–2195. doi: 10.1002/j.1460-2075.1990.tb07388.x

Shen, X., Liu, Y., Hsu, Y. J., Fujiwara, Y., Kim, J., Mao, X., et al. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in

maintaining stem cell identity and executing pluripotency. *Mol. Cell*. 32, 491–502. doi: 10.1016/j.molcel.2008.10.016

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., et al. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138, 722–737. doi: 10.1016/j.cell.2009.07.039

Singhal, N., Graumann, J., Wu, G., Araúzo-Bravo, M. J., Han, D. W., Greber, B., et al. (2010). Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* 141, 943–955. doi: 10.1016/j.cell.2010.04.037

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., et al. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 336, 688–690. doi: 10.1038/336688a0

Sokpor, G., Castro-Hernandez, R., Rosenbusch, J., Staiger, J. F., and Tuoc, T. (2018). ATP-dependent chromatin remodeling during cortical neurogenesis. *Front. Neurosci*. 12:226. doi: 10.3389/fnins.2018.00226

Stern, M., Jensen, R., and Herskowitz, I. (1984). Five SWI genes are required for expression of the Hgene, O. in yeast. *Mol. J. Biol*. 178, 853–868. doi: 10.1016/0022-2836(84)90315-2

Stewart, C. L, Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Köntgen, F., et al. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* 359, 76–79. doi: 10.1038/359076a0

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., et al. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68, 561–572. doi: 10.1016/0092-8674(92)90191-E

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145–1147. doi: 10.1126/science.282.5391.1145

Vogel-Ciernia, A., Matheos, D. P., Barrett, R. M., Kramár, E. A., Azzawi, S., Chen, Y., et al. (2013). The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat. Neurosci*. 16, 552–561. doi: 10.1038/nn.3359

Waghray, A., Saiz, N., Jayaprakash, A. D., Freire, A. G., Papatsenko, D., Pereira, C. F., et al. (2015). Tbx3 controls Dppa3 levels and exit from pluripotency toward mesoderm. *Stem Cell Rep*. 5, 97–110. doi: 10.1016/j.stemcr.2015.05.009

Weidgang, C. E., Russell, R., Tata, P. R., Kühl, S. J., Illing, A., Müller, M., et al. (2014). TBX3 directs cell-fate decision toward mesendoderm. *Stem Cell Rep*. 2:747. doi: 10.1016/j.stemcr.2014.04.011

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., et al. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 336, 684–687. doi: 10.1038/336684a0

Winata, P., William, M., Keena, V., Takahashi, K., and Cheng, Y. Y. (2018). "DNA methylation in mammalian cells," in *Gene Expression and Regulation in Mammalian Cells - Transcription Toward the Establishment of Novel Therapeutics*, ed F. Uchiumi (IntechOpen). doi: 10.5772/intechopen.72991

Yan, Z., Cui, K., Murray, D. M., Ling, C., Xue, Y., Gerstein, A., et al. (2005). PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes Dev*. 19, 1662–1667. doi: 10.1101/gad.1323805

Yan, Z., Wang, Z., Sharova, L., Sharov, A. A., Ling, C., Piao, Y., et al. (2008). BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells* 26, 1155–1165. doi: 10.1634/stemcells.2007-0846

Yeo, J. C., Jiang, J., Tan, Z. Y., Yim, G. R., Ng, J. H., Göke, J., et al. (2014). Klf2 is an essential factor that sustains ground state pluripotency. *Cell Stem Cell* 14, 864–872. doi: 10.1016/j.stem.2014.04.015

Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 115, 281–92. doi: 10.1016/S0092-8674(03)00847-X

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., et al. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523. doi: 10.1038/nature06968

You, J. S., De Carvalho, D. D., Dai, C., Liu, M., Pandiyan, K., Zhou, X. J., et al. (2013). SNF5 is an essential executor of epigenetic regulation during differentiation. *PLoS Genet*. 9:e1003459. doi: 10.1371/journal.pgen.1003459

Zhang, W., Chronis, C., Chen, X., Zhang, H., Spalinskas, R., Pardo, M., et al. (2019). The BAF and PRC2 complex subunits Dpf2 and Eed antagonistically converge on Tbx3 to control ESC differentiation. *Cell Stem Cell* 24, 138–152.e8. doi: 10.1016/j.stem.2018.12.001

Zhang, X., Li, B., Li, W., Ma, L., Zheng, D., Li, L., et al. (2014). Transcriptional repression by the BRG1-SWI/SNF complex affects the pluripotency of human embryonic stem cells. *Stem Cell Rep*. 3, 460–74. doi: 10.1016/j.stemcr.2014.07.004

Zhu, B., Ueda, A., Song, X., Horike, S. I., Yokota, T., and Akagi, T. (2017). Baf53a is involved in survival of mouse ES cells, which can be compensated by Baf53b. *Sci. Rep*. 7:14059. doi: 10.1038/s41598-017-14362-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 18 February 2021
doi: 10.3389/fcell.2021.645335

TET Enzymes and 5-Hydroxymethylcytosine in Neural Progenitor Cell Biology and Neurodevelopment

Ian C. MacArthur[1,2,3] *and Meelad M. Dawlaty*[1,2,3]*

[1] Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, United States, [2] Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, NY, United States, [3] Ruth L. and David S. Gottesman Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, Bronx, NY, United States

Studies of tissue-specific epigenomes have revealed 5-hydroxymethylcytosine (5hmC) to be a highly enriched and dynamic DNA modification in the metazoan nervous system, inspiring interest in the function of this epigenetic mark in neurodevelopment and brain function. 5hmC is generated by oxidation of 5-methylcytosine (5mC), a process catalyzed by the ten–eleven translocation (TET) enzymes. 5hmC serves not only as an intermediate in DNA demethylation but also as a stable epigenetic mark. Here, we review the known functions of 5hmC and TET enzymes in neural progenitor cell biology and embryonic and postnatal neurogenesis. We also discuss how TET enzymes and 5hmC regulate neuronal activity and brain function and highlight their implications in human neurodevelopmental and neurodegenerative disorders. Finally, we present outstanding questions in the field and envision new research directions into the roles of 5hmC and TET enzymes in neurodevelopment.

Keywords: TET enzymes, 5-hydroxymethylcytosine, neural progenitor cells, neurogenesis, neurodevelopmental disorders, epigenetics

OPEN ACCESS

Edited by:
Bruno Di Stefano, Baylor College of Medicine, United States

Reviewed by:
Claudia Doege, Columbia University, United States
Ageliki Tsagaratou, University of North Carolina at Chapel Hill, United States
Gabriella Ficz, Queen Mary University of London, United Kingdom

***Correspondence:**
Meelad M. Dawlaty
meelad.dawlaty@einsteinmed.org

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *23 December 2020*
Accepted: *01 February 2021*
Published: *18 February 2021*

Citation:
MacArthur IC and Dawlaty MM (2021) TET Enzymes and 5-Hydroxymethylcytosine in Neural Progenitor Cell Biology and Neurodevelopment. Front. Cell Dev. Biol. 9:645335. doi: 10.3389/fcell.2021.645335

INTRODUCTION

Precise temporal and spatial control of gene expression is essential for metazoan neurogenesis. This is achieved, in part, by reversible covalent modifications of DNA and histones which influence the accessibility and recruitment of transcription factors to target genes. Methylation of the 5-position carbon of cytosine (5mC) is one DNA modification influencing the transcriptional state of chromatin. DNA methylation is largely believed to be a suppressive mark achieved by *de novo* methyltransferases DNMT3A/B and maintained by maintenance methyltransferase DNMT1 (Wu and Zhang, 2014). In 2009, the discovery that ten-eleven translocation (TET) proteins (TET1, TET2, and TET3) are dioxygenases capable of oxidizing 5mC to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009) ushered in interest to study this modified base not only as an intermediate in DNA demethylation but also as a novel epigenetic mark. Oxidation of 5mC to 5hmC by TETs facilitates passive and active DNA demethylation (Tahiliani et al., 2009; Ito et al., 2011; Wu and Zhang, 2014), the latter via iterative conversion of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) and subsequent removal by DNA glycosylases and the base excision repair pathway (He et al., 2011; Ito et al., 2011). In addition to being an intermediate in DNA demethylation, 5hmC has been recognized as a stable epigenetic mark. This is supported by initial

findings that 5hmC is enriched in Purkinje neurons (Kriaucionis and Heintz, 2009), and subsequent studies confirming the presence of 5hmC and the expression of TET enzymes across many neural cell types and tissues (Globisch et al., 2010; Szwagierczak et al., 2010; Ruzov et al., 2011; Mellen et al., 2012). It has been demonstrated that 5hmC may persist for months without turnover in the brain (Bachman et al., 2014), further supporting a potential role for 5hmC as a bona fide epigenetic mark with regulatory roles in the nervous system.

TET enzymes are required for mammalian development, as loss of all three enzymes in embryonic stem cells compromises differentiation (Dawlaty et al., 2014) and in mice leads to early embryonic lethality due to gastrulation arrest (Dai et al., 2016). Loss of TET3 leads to perinatal lethality (Gu et al., 2011), though individual loss of TET1 and TET2 is compatible with development of viable mice (Dawlaty et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011). Combined loss of TET1 and TET2 leads to partial perinatal lethality, with a subset of neonates exhibiting exencephaly and other developmental abnormalities (Dawlaty et al., 2013). Similarly, combined loss of TET1 and TET3 leads to early developmental arrest and holoprosencephaly (Kang et al., 2015). This phenotypic variability suggests potential compensatory roles among TET paralogs. However, owing to the early embryonic lethality of triple TET deficiency, the absolute molecular and physiological requirements of TETs and 5hmC in neurogenesis has not yet been well-defined.

Genomic Distribution of 5hmC in the Brain

To understand the roles of 5hmC and TETs in regulation of neural gene expression, several studies have mapped the genomic distribution of 5hmC in various neural cell types and tissues over the course of embryonic and postnatal development (Jin et al., 2011; Szulwach et al., 2011; Khare et al., 2012; Hahn et al., 2013; Lister et al., 2013). In the embryonic mouse cortex, 5hmC levels increase as neural progenitor cells develop into mature neurons (Hahn et al., 2013). Interestingly, this increase is not necessarily accompanied by an increase in unmodified cytosine levels, suggesting that 5hmC can be a stable epigenetic mark in neurons and not merely a DNA demethylation intermediate (Hahn et al., 2013). Consistent with the notion that 5hmC is derived from 5mC, genomic regions in the fetal mouse brain that are enriched for 5hmC also tend to be enriched for 5mC. Notably, many of these regions become depleted of both marks in the adult mouse (Lister et al., 2013), demonstrating that 5hmC facilitates DNA demethylation in the developing brain.

5hmC levels increase in various mouse and human brain tissues over the course of life (Szulwach et al., 2011), and may have implications for neurodegenerative diseases. 5hmC is enriched in gene bodies and promoters, depleted from intergenic regions and transcription start sites, and is deposited at brain-specific enhancers (Jin et al., 2011; Szulwach et al., 2011; Hahn et al., 2013; Lister et al., 2013; Cui et al., 2020). The presence of 5hmC in gene bodies is associated with increased gene expression, suggesting that TET enzymes and 5hmC contribute to a transcriptionally permissive state of chromatin in the brain (Jin et al., 2011; Szulwach et al., 2011; Hahn et al., 2013; Lister et al., 2013). 5hmC also demarcates intron-exon boundaries in human brain cells and marks constitutively expressed exons, suggesting a potential role in control of splicing (Khare et al., 2012). Genes with high levels of 5hmC, for example *Syt1* and *Nav2*, belong to functional categories critical for nervous system function, such as synaptic transmission and neurogenesis (Khare et al., 2012; Hahn et al., 2013). 5hmC is also associated with repetitive elements as it is enriched at SINE and LTR elements in the cerebellum and hippocampus, and depleted from LINE elements in the cerebellum (Szulwach et al., 2011). Enrichment at SINE and LTRs increases over postnatal life in the cerebellum (Szulwach et al., 2011), indicating a possible role in regulation of repetitive element activity in the brain. Indeed, *Tet2/3* knockdown reverses loss-of-*Uhrf1*-mediated increased DNA hydroxymethylation and activation of IAP elements in NPCs (Ramesh et al., 2016). Together, these observations support important roles for 5hmC and TETs in mammalian neurogenesis and brain function.

Regulation of Neural Progenitor Cells and Neurogenesis by TET Enzymes and 5hmC

Studies of embryonic stem cell (ESC) differentiation have suggested a critical role for TET enzymes and 5hmC in neural lineage commitment. Deficiency of all three TETs in ESCs compromises pluripotency and *Tet1/2/3* triple knockout (TKO) ESCs fail to form neural pigmented epithelium in teratoma assays, though they are able to form other neural tissue types (Dawlaty et al., 2014). These cells fail to contribute to nervous system structures when injected into wild type blastocysts to form chimeras (Dawlaty et al., 2014). Consistently, *Tet* TKO mouse embryos and ESCs differentiated toward the neural lineage have reduced neuroectodermal and increased mesodermal gene expression, in part due to failure to inhibit Wnt signaling (Li et al., 2016). Likewise, *TET* TKO human ESCs exhibit aberrant neuroectodermal gene expression when differentiated toward the neural lineage and fail to demethylate the *PAX6* promoter, a transcription factor critical for neurodevelopment (Verma et al., 2018). These studies support a requirement for TET enzymes in the commitment of ESCs to a neural fate, an idea further supported by studies of *TET* genes in ESC specification to neural progenitor cells (NPCs). TET enzymes, in particular TET2, regulate enhancer methylation during differentiation of ESCs to NPCs (Hon et al., 2014). Though *Tet2* knockout ESCs can successfully differentiate into NPCs, these cells exhibit delayed induction of neural gene expression programs accompanied by enhancer hypermethylation and reduced histone H3 lysine 27 acetylation (Hon et al., 2014). This is in line with DNA hypermethylation in the embryonic cerebral cortex of *Tet2* knockout mice (Lister et al., 2013). TET3 plays a role in the epigenetic regulation of NPC specification and maintenance of cellular identity (Montibus et al., 2020; Santiago et al., 2020). During differentiation of mouse ESCs to NPCs, the catalytic activity of TET3 promotes expression of histone demethylase *Kdm6b*, an epigenetic regulator critical for gene regulation during neurogenesis (Montibus et al., 2020), and loss of TET3 promotes

NPC apoptosis (Li T. et al., 2015). Knockdown of *Tet3* in NPCs promotes de-repression of pluripotency genes *Oct4* and *Nanog*, implicating TET3 in the maintenance of NPC identity (Santiago et al., 2020). These studies implicate TETs in the epigenetic regulation of NPC biology.

In adult NPCs, different TET paralogs have unique functions, highlighting some non-redundant and context-specific roles. Loss of TET2 increases the proliferation of adult NPCs and reduces their differentiation into neurons and astrocytes *in vivo*, indicating that TET2 promotes NPC differentiation (Li et al., 2017). Deletion of *Tet3* decreases NPC proliferation in the subventricular zone of the mouse cortex and promotes astrocytic differentiation, consistent with a role for TET3 in maintaining NPC identity (Montalban-Loro et al., 2019). *Tet1* knockout mice have fewer NPCs in the dentate gyrus of the hippocampus and conditional deletion of *Tet1* or *Tet2* in NPCs compromises hippocampal neurogenesis (Zhang et al., 2013; Gontier et al., 2018). While most functions of TETs in NPCs are attributed to their enzymatic activity (Zhang et al., 2013; Li et al., 2017; Montibus et al., 2020), some functions are independent of enzymatic activity, such as transcriptional repression of the imprinted gene *Snrnp* by TET3 (Montalban-Loro et al., 2019). Investigating these dual roles of TETs and dissecting their requirements in NPC biology and neurogenesis will be essential.

Although global or neural-specific loss of each *Tet* gene in mice influences NPC biology, it does not block neurogenesis or cause gross neuroanatomical defects (Rudenko et al., 2013; Zhang et al., 2013). However, combined loss of TET1/2 and TET1/3 causes exencephaly and holoprosencephaly in some embryos, respectively (Dawlaty et al., 2013; Kang et al., 2015) suggesting redundancy between TETs in neurogenesis that warrants further investigation. Findings from other organisms have also supported a role for TET enzymes in embryonic neurogenesis. *Xenopus laevis* embryos depleted of TET3 are microcephalic and eyeless with deregulation of neurodevelopmental programs leading to aberrant expression of neuronal, eye, neural crest, and sonic hedgehog signaling genes (Xu et al., 2012). Moreover, *tet2/3* mutant zebrafish exhibit abnormal brain and eye morphology (Li C. et al., 2015) and impaied retinal neurogenesis, partly due to overactive Notch and Wnt signaling (Seritrakul and Gross, 2017). Aberrant expression of mesodermal genes was also observed in *tet2/3* mutant retinas (Seritrakul and Gross, 2017), a finding similar to those in *Tet1/2/3* knockout embryos (Li et al., 2016). Importantly, TET enzymes mediate demethylation of conserved developmental enhancers in brain during the phylotypic stage of vertebrates, as demonstrated in zebrafish, *Xenopus tropicalis*, and mouse (Bogdanovic et al., 2016). Together, these findings support highly conserved and overlapping functions for TETs in neurodevelopment.

Role of TET Enzymes and 5hmC in Postnatal Brain and Mature Neuronal Function

In addition to roles in regulation of NPC biology, TETs and 5hmC are important in postnatal neurodevelopment and mature neurons. As previously mentioned, 5hmC accumulates over the course of life (Szulwach et al., 2011). During development of mouse olfactory bulb neurons, which occurs throughout life, 5hmC is enriched in neurons relative to immature cells and is associated with increased neurodevelopmental gene expression (Colquitt et al., 2013). Likewise, 5hmC increases over the course of postnatal retinal maturation, and is enriched at neurogenesis genes (Perera et al., 2015). In the cerebellum, 5hmC increases during an important period of neuronal circuit formation, and TET1 and TET3 are required for proper branching of granule cell dendrites (Zhu et al., 2016). Chimeric *Tet3* knockout mice generated by injection of *Tet3* sgRNAs in one cell of a two-cell-stage embryo develop histologically normal cerebral cortices composed of *Tet3* wild type and knockout cells but exhibit abnormal electrophysiology in recordings of excitatory and inhibitory neurotransmission, suggesting that TET3 is required for developmental synapse and circuit formation (Wang et al., 2017). These studies implicate TETs in shaping the epigenetic landscape during specification of mature neural cell types after birth and in the development of higher order structures, including neuronal circuits.

5hmC and TET enzymes have also been shown to be highly dynamic within post-mitotic neurons. Cortical 5hmC has cell-type specific distributions associated with differential gene expression (Kozlenkov et al., 2018), and the ability of TETs to promote active DNA demethylation and alter gene expression in response to neuronal activity and to influence behavior has been the subject of extensive study (Guo et al., 2011; Kaas et al., 2013; Rudenko et al., 2013; Zhang et al., 2013; Li et al., 2014; Yu et al., 2015). *Tet1* expression is downregulated in response to neuronal activity in hippocampus where it regulates spatial memory and fear memory extinction (Guo et al., 2011; Kaas et al., 2013; Rudenko et al., 2013; Zhang et al., 2013). Hippocampal neurons upregulate *Tet3* to initiate active DNA demethylation in response to neuronal stimulation (Yu et al., 2015). In cortical neurons, *Tet3* is upregulated during fear extinction learning and, like *Tet1* in the hippocampus, is required for fear memory extinction (Li et al., 2014). Fear extinction learning is accompanied by *Tet3*-mediated upregulation of the *Gephyrin* gene and a transcriptionally-permissive reshaping of chromatin around this locus (Li et al., 2014). Loss of *Tet3* is sufficient to produce anxiety-like behaviors in mice, partly due to increased expression of immediate early genes like *Npas4* (Antunes et al., 2020), a role that is opposite to the anxiolytic and anti-depressant effects of *Tet1* (Feng et al., 2017). In general, the mechanistic basis by which TET enzymes influence behavior is, in part, due to reshaping of neuronal 5mC and 5hmC landscapes in response to activity. This remodeling of the epigenome is required for proper expression of genes involved in memory consolidation and synaptic function, such as *Bdnf* and *Arc*, and is sufficient to alter the electrophysiological properties of neurons (Guo et al., 2011; Kaas et al., 2013; Rudenko et al., 2013; Zhang et al., 2013; Li et al., 2014; Yu et al., 2015). In post-mitotic cerebellar neurons, 5hmC dynamics can influence recruitment of key gene regulatory factors. For example, 5hmC in gene bodies is associated with reduced MeCP2 occupancy and increased gene expression, possibly due to loss of MeCP2 repression (Mellen et al., 2017). Of note, reduced MeCP2 occupancy is specifically

associated with 5hmC at gene body CpG dinucleotides, whereas 5hmC at CpA sites flanking enhancers retains MeCP2 binding. This highlights the ability of 5hmC to influence recruitment of gene regulatory factors in a sequence-dependent manner (Mellen et al., 2017). Moreover, findings that common *MECP2* mutations in Rett syndrome disrupt MeCP2 binding to 5hmC has implications for this mark in Rett syndrome pathogenesis (Mellen et al., 2012; Brown et al., 2016). Together, these studies propose crucial roles for 5hmC and TET enzymes in mature neuronal function.

Implication of TET Enzymes and 5hmC in Human Neurodevelopmental and Neurodegenerative Disorders and Addiction

Compelling evidence for the importance of TETs in neurodevelopment and brain function is the identification of *TET* gene mutations and alterations in 5hmC levels in human neurodevelopmental and neurodegenerative disorders. Mutations in *TET3* were recently identified to underlie an inherited syndrome of intellectual disability and craniofacial abnormalities (Beck et al., 2020). While most mutations are in the catalytic domain and are sufficient to impair enzymatic activity, some are outside of this domain, and one mutation does not affect catalytic activity (Beck et al., 2020), underscoring the importance of TET3 catalytic and non-catalytic functions in human neurodevelopment. Interestingly, the clinical characteristics of patients with TET3 deficiency resemble those of patients with Tatton-Brown-Rahman syndrome and Sotos syndrome, caused by mutations in *DNMT3A* and *NSD1*, respectively (Kurotaki et al., 2002; Tatton-Brown et al., 2014). This highlights the general importance of DNA and histone methylation dynamics in human craniofacial and neural development. Other *TET* mutations have been observed in individuals with intellectual disability. Mutations in *TET1* were identified in consanguineous Pakistani and Iranian families with familial intellectual disability syndromes (Harripaul et al., 2018), and a germline *TET2* variant in an individual diagnosed with intellectual disability and delayed verbal comprehension in the absence of any other known genetic causes (Kaasinen et al., 2019). Together, these findings support an important role for TETs in the etiology of neurodevelopmental disorders and intellectual disability.

In addition TET enzymes and 5hmC are recurrently dysregulated in neurodegenerative conditions and in aging brain. Induced pluripotent stem cell-derived NPCs and neurons from Alzheimer's disease (AD) patients exhibit differential hydroxymethylation at genes associated with neurodevelopment and synaptic function, including at known AD susceptibility loci, compared to cells derived from healthy controls (Fetahu et al., 2019). Consistently, presumptive loss-of-function mutations in *TET2* have been identified in patients with early onset AD and frontotemporal dementia (Cochran et al., 2020). Interestingly, TET2 promotes proinflammatory gene expression in microglia and *TET2* expression is increased in microglia associated with amyloid beta plaques in the brains of AD patients and mouse models (Carrillo-Jimenez et al., 2019). Thus, the positive and negative roles of TET2 in AD are likely specific to distinct stages in clinical course and cell types. *TET* variants or dysregulation have also been implicated in Parkinson's disease (PD). *TET1* mutations were reported in a Chinese cohort of PD patients (Shu et al., 2019). Intriguingly, increased expression of *TET2* and increased 5hmC at neural enhancers is observed in prefrontal cortex of patients with PD and *Tet2* knockout mice are protected from inflammatory damage to the substantia nigra (Marshall et al., 2020). Conversely, *Tet2* expression declines in the hippocampus of aging mice and is associated with age-related cognitive decline (Gontier et al., 2018). Remarkably, restoration of hippocampal *Tet2* expression by stereotactic lentivirus injection is sufficient to improve cognitive function in aged mice (Gontier et al., 2018). Moreover, recent findings that TET1 and TET2 are required for axonal regeneration by reprogramming factor expression highlights their potential as therapeutic targets (Lu et al., 2020). These observations in human disease and mouse models warrant further studies to clarify the discordant roles of TET enzymes in the etiology of neurodegenerative diseases and aging.

In addition to their roles in neurodevelopmental and neurodegenerative disorders, TET enzymes are associated with addictive behaviors in humans. *TET1* expression is decreased in the nucleus accumbens (NAc) of humans suffering from cocaine addiction and cocaine administration to mice is sufficient to alter 5hmC at enhancers in NAc (Feng et al., 2015). *Tet* expression in NAc is also responsive to methamphetamine administration in rats (Jayanthi et al., 2018). Further work is necessary to clarify the role of TETs in mediating addictive behaviors.

DISCUSSION

A dozen years since the discovery that TET enzymes promote DNA hydroxylation and demethylation and the first studies reporting the abundance of 5hmC in the mammalian nervous system (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009), work in the field has shed some light on their functions in neural physiology. TETs are dynamically expressed during development and in particular during embryonic and adult neurogenesis. 5hmC is a highly enriched mark in the brain and its levels increase over the course of embryonic neurogenesis and postnatal life where it is associated with neural gene expression. TET enzymes are required for various aspects of neurodevelopment, NPC biology, and neuronal activity. Findings that 5hmC and TETs are dysregulated in human neurodevelopmental and neurodegenerative disorders and addiction open new frontiers for utilizing them in clinical diagnostics and therapeutics. Despite this progress, several fundamental questions remain unanswered. These pertain to: (1) mechanisms of TET recruitment to target sites, (2) functional redundancy between TET paralogs, (3) gene activation and silencing by the dual enzymatic and non-enzymatic functions of TET enzymes, (4) relevance of 5hmC readers and interactomes of TETs in gene regulation, (5) re-establishment of 5hmC upon active DNA demethylation at activity-dependent genes in post-mitotic neurons, and (6) involvement of 5fC

and 5caC, the other oxidized derivatives of 5mC, in DNA demethylation and beyond.

It has been substantiated that increases in gene body 5hmC is associated with activation of neural genes over the course of development, but the mechanism by which 5hmC influences gene expression remains incompletely understood. One possibility is that 5hmC recruits specific factors to promote transcription, and several groups have sought to identify such readers of 5hmC (Yildirim et al., 2011; Mellen et al., 2012; Spruijt et al., 2013). For example, MeCP2 binds 5hmC at neuronal genes to facilitate transcription, a finding with implications in the pathogenesis of Rett syndrome, where *MECP2* is mutated and 5hmC levels are altered (Mellen et al., 2012). Alternatively, evidence also supports a role for MeCP2 and MBD2 in protecting 5mC from conversion to 5hmC (Szulwach et al., 2011; Ludwig et al., 2017). In addition, UHRF2 is a specific reader of 5hmC in NPCs while THAP11 interacts with 5hmC in brain tissue (Spruijt et al., 2013). The functions of these and other 5hmC readers, including WDR76 and THY96, in the nervous system remain to be further explored (Spruijt et al., 2013). In contrast, the observation that elevated 5hmC levels may persist at genes after silencing suggests that any factor recruitment by this mark is not sufficient to maintain gene expression in the presence of antagonistic or in the absence of agonistic transcriptional cues (Colquitt et al., 2013). This would be consistent with a necessary but not sufficient role for TET-mediated 5hmC deposition and DNA demethylation for potent gene transcription (Baumann et al., 2019). This could also indicate a requirement for higher-order formation of 5fC and 5caC for transcription factor recruitment, given the fact that a growing number of proteins bind 5fC more specifically than 5hmC (Iurlaro et al., 2013). Additional work is needed to substantiate a causal role for 5hmC in neural gene regulation.

Although functional studies have defined important roles for individual TETs in neural development and physiology, the absolute requirements of these enzymes and 5hmC is not fully established. This is in part due to the possible compensatory effects of TETs in studies involving deletion of individual *TET* genes. Leveraging conditional genetic systems for spatial and temporal deletion of all three *TET* genes in specific cell types may allow for identification of their novel functions in the brain. While TET enzymes certainly influence gene expression through their enzymatic functions, non-enzymatic activities of these proteins involving formation and recruitment of chromatin regulatory complexes have also been described (Chen et al., 2013; Ito et al., 2019; Montalban-Loro et al., 2019). Further work is necessary to fully elucidate protein-protein interactions by which TETs influence transcription and chromatin state, such as those between TET2 and FOXO3 in NPCs (Li et al., 2017) and TET3 and NSD3 in mature neurons (Perera et al., 2015). Comparison of the neural phenotypes associated with loss of TET enzymes vs. loss of their enzymatic activity alone will help dissect key enzymatic-dependent and independent functions of TET proteins in the brain. Use of existing and development of new *Tet* catalytic mutant mouse models will facilitate the *in vivo* study of TET catalytic-independent functions. Furthermore, the observation that loss of *Tet* in *Drosophila* results in aberrant brain development and reduced RNA hydroxymethylation warrants investigation into the role of TET-mediated RNA modifications in mammalian brains (Delatte et al., 2016). In summary, as illustrated in **Figure 1**, TET enzymes and 5hmC play crucial roles in various aspects of neurobiology, from regulation of NPCs and neurogenesis to adult brain function and human diseases. Gaining further insights into their roles will enhance our understanding of metazoan nervous system development and the etiology of human neurological disorders.

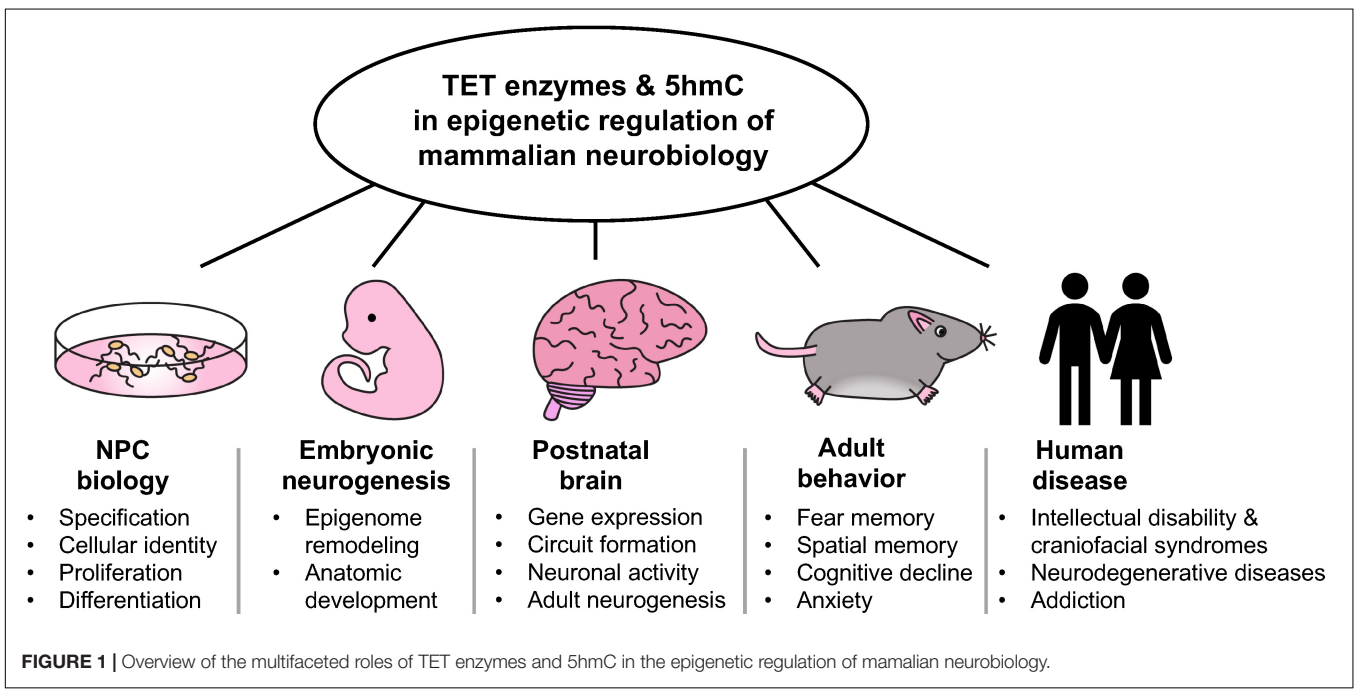


FIGURE 1 | Overview of the multifaceted roles of TET enzymes and 5hmC in the epigenetic regulation of mamalian neurobiology.

AUTHOR CONTRIBUTIONS

IM wrote the first draft and prepared the figure. MD edited and revised the draft and figure, and finalized the final version with input from IM. Both authors contributed to the article and approved the submitted version.

FUNDING

The Dawlaty laboratory was supported by the National Institutes of Health (R01GM122839, R01HL148852), the New State Department of Health (NYSTEM IIRP Award), the Feldstein Medical Foundation, and the Hirschl and Weill-Caulier Charitable Trust Funds to study the biology of Tet enzymes in stem cells and development. MD's salary was in part supported by these funds. IM was supported by the Medical Scientist Training Program training grant T32GM007288-48 from the National Institutes of Health to Albert Einstein College of Medicine.

ACKNOWLEDGMENTS

We thank Drs. Frank Soldner and Jean Hébert of the Albert Einstein College of Medicine's Neuroscience and Genetics Departments for critical reading of the manuscript, and members of the Dawlaty laboratory for helpful discussions. We apologize to colleagues whose work we could not discuss here due to space limitations.

REFERENCES

Antunes, C., Da Silva, J. D., Guerra-Gomes, S., Alves, N. D. F., Ferreira, F., Loureiro-Campos, E., et al. (2020). Tet3 ablation in adult brain neurons increases anxiety-like behavior and regulates cognitive function in mice. *Mol. Psychiatry*. doi: 10.1038/s41380-020-0695-7 [Epub ahead of print].

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* 6, 1049–1055. doi: 10.1038/nchem.2064

Baumann, V., Wiesbeck, M., Breunig, C. T., Braun, J. M., Koferle, A., Ninkovic, J., et al. (2019). Targeted removal of epigenetic barriers during transcriptional reprogramming. *Nat. Commun.* 10:2119. doi: 10.1038/s41467-019-10146-8

Beck, D. B., Petracovici, A., He, C., Moore, H. W., Louie, R. J., Ansar, M., et al. (2020). Delineation of a human mendelian disorder of the DNA demethylation machinery: TET3 deficiency. *Am. J. Hum. Genet.* 106, 234–245. doi: 10.1016/j.ajhg.2019.12.007

Bogdanovic, O., Smits, A. H., de la Calle Mustienes, E., Tena, J. J., Ford, E., and Williams, R. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* 48, 417–426. doi: 10.1038/ng.3522

Brown, K., Selfridge, J., Lagger, S., Connelly, J., De Sousa, D., Kerr, A., et al. (2016). The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* 25, 558–570. doi: 10.1093/hmg/ddv496

Carrillo-Jimenez, A., Deniz, O., Niklison-Chirou, M. V., Ruiz, R., Bezerra-Salomao, K., Stratoulias, V., et al. (2019). TET2 regulates the neuroinflammatory response in microglia. *Cell Rep.* 29, 697–713.e8. doi: 10.1016/j.celrep.2019.09.013

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561–564. doi: 10.1038/nature11742

Cochran, J. N., Geier, E. G., Bonham, L. W., Newberry, J. S., Amaral, M. D., Thompson, M. L., et al. (2020). Non-coding and loss-of-function coding variants in TET2 are associated with multiple neurodegenerative diseases. *Am. J. Hum. Genet.* 106, 632–645. doi: 10.1016/j.ajhg.2020.03.010

Colquitt, B. M., Allen, E. W., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14682–14687. doi: 10.1073/pnas.1302759110

Cui, X. L., Nie, J., Ku, J., Dougherty, U., West-Szymanski, D. C., Collin, F., et al. (2020). A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* 11:6161. doi: 10.1038/s41467-020-20001-w

Dai, H. Q., Wang, B. A., Yang, L., Chen, J. J., Zhu, G. C., Sun, M. L., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. *Nature* 538, 528–532. doi: 10.1038/nature20095

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., et al. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* 29, 102–111. doi: 10.1016/j.devcel.2014.03.003

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* 24, 310–323. doi: 10.1016/j.devcel.2012.12.015

Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9, 166–175. doi: 10.1016/j.stem.2011.07.010

Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., et al. (2016). RNA biochemistry. transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* 351, 282–285. doi: 10.1126/science.aac5253

Feng, J., Pena, C. J., Purushothaman, I., Engmann, O., Walker, D., Brown, A. N., et al. (2017). Tet1 in nucleus accumbens opposes depression- and anxiety-like behaviors. *Neuropsychopharmacology* 42, 1657–1669. doi: 10.1038/npp.2017.6

Feng, J., Shao, N., Szulwach, K. E., Vialou, V., Huynh, J., Zhong, C., et al. (2015). Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. *Nat. Neurosci.* 18, 536–544. doi: 10.1038/nn.3976

Fetahu, I. S., Ma, D., Rabidou, K., Argueta, C., Smith, M., Liu, H., et al. (2019). Epigenetic signatures of methylated DNA cytosine in Alzheimer's disease. *Sci. Adv.* 5:eaaw2880. doi: 10.1126/sciadv.aaw2880

Globisch, D., Munzel, M., Muller, M., Michalakis, S., Wagner, M., Koch, S., et al. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5:e15367. doi: 10.1371/journal.pone.0015367

Gontier, G., Iyer, M., Shea, J. M., Bieri, G., Wheatley, E. G., Ramalho-Santos, M., et al. (2018). Tet2 rescues age-related regenerative decline and enhances cognitive function in the adult mouse brain. *Cell Rep.* 22, 1974–1981. doi: 10.1016/j.celrep.2018.02.001

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610. doi: 10.1038/nature10443

Guo, J. U., Su, Y., Zhong, C., Ming, G. L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423–434. doi: 10.1016/j.cell.2011.03.022

Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.* 3, 291–300. doi: 10.1016/j.celrep.2013.01.011

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., et al. (2018). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Mol. Psychiatry* 23, 973–984. doi: 10.1038/mp.2017.60

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi: 10.1126/science.1210944

Hon, G. C., Song, C. X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., et al. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* 56, 286–297. doi: 10.1016/j.molcel.2014.08.026

Ito, K., Lee, J., Chrysanthou, S., Zhao, Y., Josephs, K., Sato, H., et al. (2019). Non-catalytic roles of Tet2 are essential to regulate hematopoietic stem and progenitor cell homeostasis. *Cell Rep.* 28, 2480–2490.e4. doi: 10.1016/j.celrep.2019.07.094

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303. doi: 10.1126/science.1210597

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., et al. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* 14:R119. doi: 10.1186/gb-2013-14-10-r119

Jayanthi, S., Gonzalez, B., McCoy, M. T., Ladenheim, B., Bisagno, V., and Cadet, J. L. (2018). Methamphetamine induces TET1- and TET3-dependent DNA hydroxymethylation of Crh and Avp genes in the rat nucleus accumbens. *Mol. Neurobiol.* 55, 5154–5166. doi: 10.1007/s12035-017-0750-9

Jin, S. G., Wu, X., Li, A. X., and Pfeifer, G. P. (2011). Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nuclei. Acid. Res.* 39, 5015–5024. doi: 10.1093/nar/gkr120

Kaas, G. A., Zhong, C., Eason, D. E., Ross, D. L., Vachhani, R. V., Ming, G. L., et al. (2013). TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* 79, 1086–1093. doi: 10.1016/j.neuron.2013.08.032

Kaasinen, E., Kuismin, O., Rajamaki, K., Ristolainen, H., Aavikko, M., Kondelin, J., et al. (2019). Impact of constitutional TET2 haploinsufficiency on molecular and clinical phenotype in humans. *Nat. Commun.* 10:1252. doi: 10.1038/s41467-019-09198-7

Kang, J., Lienhard, M., Pastor, W. A., Chawla, A., Novotny, M., Tsagaratou, A., et al. (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4236–E4245. doi: 10.1073/pnas.1510510112

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., et al. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* 19, 1037–1043. doi: 10.1038/nsmb.2372

Kozlenkov, A., Li, J., Apontes, P., Hurd, Y. L., Byne, W. M., Koonin, E. V., et al. (2018). A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Sci. Adv.* 4:eaau6190. doi: 10.1126/sciadv.aau6190

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786

Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., et al. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* 30, 365–366. doi: 10.1038/ng863

Li, C., Lan, Y., Schwartz-Orbach, L., Korol, E., Tahiliani, M., Evans, T., et al. (2015). Overlapping requirements for Tet2 and Tet3 in normal development and hematopoietic stem cell emergence. *Cell Rep.* 12, 1133–1143. doi: 10.1016/j.celrep.2015.07.025

Li, T., Yang, D., Li, J., Tang, Y., Yang, J., and Le, W. (2015). Critical role of Tet3 in neural progenitor cell maintenance and terminal differentiation. *Mol. Neurobiol.* 51, 142–154. doi: 10.1007/s12035-014-8734-5

Li, X., Wei, W., Zhao, Q. Y., Widagdo, J., Baker-Andresen, D., Flavell, C. R., et al. (2014). Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7120–7125. doi: 10.1073/pnas.1318906111

Li, X., Yao, B., Chen, L., Kang, Y., Li, Y., Cheng, Y., et al. (2017). Ten-eleven translocation 2 interacts with forkhead box O3 and regulates adult neurogenesis. *Nat. Commun.* 8:15903. doi: 10.1038/ncomms15903

Li, X., Yue, X., Pastor, W. A., Lin, L., Georges, R., Chavez, L., et al. (2016). Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc. Natl. Acad. Sci. U.S.A.* 113, E8267–E8276. doi: 10.1073/pnas.1617802113

Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., et al. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* 118, 4509–4518. doi: 10.1182/blood-2010-12-325241

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* 341:1237905. doi: 10.1126/science.1237905

Lu, Y., Brommer, B., Tian, X., Krishnan, A., Meer, M., Wang, C., et al. (2020). Reprogramming to recover youthful epigenetic information and restore vision. *Nature* 588, 124–129. doi: 10.1038/s41586-020-2975-4

Ludwig, A. K., Zhang, P., Hastert, F. D., Meyer, S., Rausch, C., Herce, H. D., et al. (2017). Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acids Res.* 45, 2438–2457. doi: 10.1093/nar/gkw1197

Marshall, L. L., Killinger, B. A., Ensink, E., Li, P., Li, K. X., Cui, W., et al. (2020). Epigenomic analysis of Parkinson's disease neurons identifies Tet2 loss as neuroprotective. *Nat. Neurosci.* 23, 1203–1214. doi: 10.1038/s41593-020-0690-y

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430. doi: 10.1016/j.cell.2012.11.022

Mellen, M., Ayata, P., and Heintz, N. (2017). 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proc. Natl. Acad. Sci. U.S.A.* 114, E7812–E7821. doi: 10.1073/pnas.1708044114

Montalban-Loro, R., Lozano-Urena, A., Ito, M., Krueger, C., Reik, W., Ferguson-Smith, A. C., et al. (2019). TET3 prevents terminal differentiation of adult NSCs by a non-catalytic action at Snrpn. *Nat. Commun.* 10:1726. doi: 10.1038/s41467-019-09665-1

Montibus, B., Cercy, J., Bouschet, T., Charras, A., Maupetit-Mehouas, S., Nury, D., et al. (2020). TET3 controls the expression of the H3K27me3 demethylase Kdm6b during neural commitment. *Cell Mol. Life Sci*. doi: 10.1007/s00018-020-03541-8 [Epub ahead of print].

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 20, 11–24. doi: 10.1016/j.ccr.2011.06.001

Perera, A., Eisen, D., Wagner, M., Laube, S. K., Kunzel, A. F., Koch, S., et al. (2015). TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell Rep.* 11, 283–294. doi: 10.1016/j.celrep.2015.03.020

Ramesh, V., Bayam, E., Cernilogar, F. M. I, Bonapace, M., Schulze, M., Riemenschneider, M. J., et al. (2016). Loss of Uhrf1 in neural stem cells leads to activation of retroviral elements and delayed neurodegeneration. *Genes Dev.* 30, 2199–2212. doi: 10.1101/gad.284992.116

Rudenko, A., Dawlaty, M. M., Seo, J., Cheng, A. W., Meng, J., Le, T., et al. (2013). Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* 79, 1109–1122. doi: 10.1016/j.neuron.2013.08.003

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., et al. (2011). Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res.* 21, 1332–1342. doi: 10.1038/cr.2011.113

Santiago, M., Antunes, C., Guedes, M., Iacovino, M., Kyba, M., Reik, W., et al. (2020). Tet3 regulates cellular identity and DNA methylation in neural progenitor cells. *Cell Mol. Life Sci.* 77, 2871–2883. doi: 10.1007/s00018-019-03335-7

Seritrakul, P., and Gross, J. M. (2017). Tet-mediated DNA hydroxymethylation regulates retinal neurogenesis by modulating cell-extrinsic signaling pathways. *PLoS Genet.* 13:e1006987. doi: 10.1371/journal.pgen.1006987

Shu, L., Qin, L., Min, S., Pan, H., Zhong, J., Guo, J., et al. (2019). Genetic analysis of DNA methylation and hydroxymethylation genes in Parkinson's disease. *Neurobiol. Aging* 84, 242.e13–242.e16. doi: 10.1016/j.neurobiolaging.2019.02.025

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159. doi: 10.1016/j.cell.2013.02.004

Szulwach, K. E., Li, X., Li, Y., Song, C. X., Wu, H., Dai, Q., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616. doi: 10.1038/nn.2959

Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 38:e181. doi: 10.1093/nar/gkq684

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Vecchio, S. Del, et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* 46, 385–388. doi: 10.1038/ng.2917

Verma, N., Pan, H., Dore, L. C., Shukla, A., Li, Q. V., Pelham-Webb, B., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* 50, 83–95. doi: 10.1038/s41588-017-0002-y

Wang, L., Li, M. Y., Qu, C., Miao, W. Y., Yin, Q., Liao, J., et al. (2017). CRISPR-Cas9-mediated genome editing in one blastomere of two-cell embryos reveals a novel Tet3 function in regulating neocortical development. *Cell Res.* 27, 815–829. doi: 10.1038/cr.2017.58

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* 156, 45–68. doi: 10.1016/j.cell.2013.12.019

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* 151, 1200–1213. doi: 10.1016/j.cell.2012.11.014

Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Ee, L. S., et al. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510. doi: 10.1016/j.cell.2011.11.054

Yu, H., Su, Y., Shin, J., Zhong, C., Guo, J. U., Weng, Y. L., et al. (2015). Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. *Nat. Neurosci.* 18, 836–843. doi: 10.1038/nn.4008

Zhang, R. R., Cui, Q. Y., Murai, K., Lim, Y. C., Smith, Z. D., Jin, S., et al. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* 13, 237–245. doi: 10.1016/j.stem.2013.05.006

Zhu, X., Girardo, D., Govek, E. E., John, K., Mellen, M., Tamayo, P., et al. (2016). Role of Tet1/3 genes and chromatin remodeling genes in cerebellar circuit formation. *Neuron* 89, 100–112. doi: 10.1016/j.neuron.2015.11.030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

REVIEW
published: 18 February 2021
doi: 10.3389/fcell.2021.637309

Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions

*Li Sun†, Xiuling Fu†, Gang Ma† and Andrew P. Hutchins**

Department of Biology, Southern University of Science and Technology, Shenzhen, China

A major event in embryonic development is the rearrangement of epigenetic information as the somatic genome is reprogrammed for a new round of organismal development. Epigenetic data are held in chemical modifications on DNA and histones, and there are dramatic and dynamic changes in these marks during embryogenesis. However, the mechanisms behind this intricate process and how it is regulating and responding to embryonic development remain unclear. As embryos develop from totipotency to pluripotency, they pass through several distinct stages that can be captured permanently or transiently *in vitro*. Pluripotent naïve cells resemble the early epiblast, primed cells resemble the late epiblast, and blastomere-like cells have been isolated, although fully totipotent cells remain elusive. Experiments using these *in vitro* model systems have led to insights into chromatin changes in embryonic development, which has informed exploration of pre-implantation embryos. Intriguingly, human and mouse cells rely on different signaling and epigenetic pathways, and it remains a mystery why this variation exists. In this review, we will summarize the chromatin rearrangements in early embryonic development, drawing from genomic data from *in vitro* cell lines, and human and mouse embryos.

Keywords: embryonic stem cells, naïve, primed, epigenetic, transposable elements, reprogramming

OPEN ACCESS

Edited by:
Justin Brumbaugh, University of Colorado Boulder, United States

Reviewed by:
Guokai Chen, University of Macau, China
Xinyi Lu, Nankai University, China

***Correspondence:**
Andrew P. Hutchins
andrewh@sustech.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *03 December 2020*
Accepted: *19 January 2021*
Published: *18 February 2021*

Citation:
Sun L, Fu X, Ma G and Hutchins AP (2021) Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions. Front. Cell Dev. Biol. 9:637309. doi: 10.3389/fcell.2021.637309

INTRODUCTION

Each cell contains the same DNA that is interpreted to provide specialized cell function, yet the interpretation of the DNA code is cell type-specific, and epigenetic barriers exist that impair and permit cell type conversions. Cell type conversions are surprisingly rare in the adult organism, outside some stem cells, which retain limited ability to derive various progeny (Avgustinova and Benitah, 2016). Despite the importance of understanding cell type control for regenerative therapies, exactly how this process is controlled remains surprisingly elusive (Hutchins et al., 2017). Epigenetic control mechanisms are a major contributor to this process; however, there are a wide array of overlapping and competing mechanisms, particularly histone and DNA chemical modifications (Ohbo and Tomizawa, 2015). Epigenetic modifications form barriers that can be permissive for some cell type transitions but intolerant for others. These epigenetic barriers can resist all cell type conversions but can also act as bidirectional valves, guiding cells toward a differentiated cell type but then blocking reversion to progenitor cells and locking cells into a fixed cell type (Arabaci et al., 2020). Epigenetic control is mediated by DNA-binding proteins, particularly transcription factors (TFs) (Fu et al., 2017), that recruit epigenetic enzymes to regulate chemical

modifications on DNA and histones and reshape or remodel chromatin structure. This process both responds to cell type changes and controls them and is critical in normal developmental processes. During embryonic development of a new organism, epigenetic information is reset (Guo et al., 2014; Xia and Xie, 2020), and many epigenetic rearrangements occur during early pre-implantation development and accompany, and in some instances drive, developmental changes. Research to understand the epigenetic reconfigurations in embryonic development has been intense, as understanding epigenetic control could give fine-grained control of cell type states, which would be useful for a wide range of medical applications, from regeneration (Avgustinova and Benitah, 2016), and cell replacement therapy, to understanding how cell type control is disrupted in cancerous cells (Meacham and Morrison, 2013). In this review, we will discuss epigenetic control of early embryonic stages, using work from embryos, *in vitro* mimics of embryonic cells, and cell type conversion systems.

NAÏVE AND PRIMED CELLS IN MICE

Amongst the first *in vitro*-derived embryonic cells were the human and mouse embryonal carcinoma (EC) cell lines, which can be maintained *in vitro* but are derived from teratocarcinomas and have multiple mutations and abnormal karyotypes (Martin, 1980). Several EC lines were derived, and while each line has a set of common embryonic features, they also have line-specific effects, such as restricted differentiation potential and different culture requirements (Andrews, 1988; Alonso et al., 1991). When ECs are injected into a mouse blastocyst, they can contribute to development but often lead to teratomas in the adult mice. Fully viable EC-derived chimeras have been reported (Hanaoka et al., 1987); however, considering what we now know about the rapid growth of normal, untransformed mouse embryonic stem cells (mESCs), it is not inconceivable that those EC cultures harbored small numbers of mESCs. Compared with ECs, mESCs are untransformed, contribute to chimeras at high frequency without generating teratomas in the adult, and can be grown *in vitro* in defined conditions indefinitely. *In vitro*, mESCs can be differentiated to all three germ lineages and have become a powerful model of the early stages of embryonic development (Rossant, 2011). mESCs are thought to most closely resemble the early epiblast (Boroviak et al., 2014). They express marker genes specific for the epiblast; and in female mESCs, both X chromosomes are active; and silencing is required to exit the pluripotent state (Schulz et al., 2014). mESCs were first derived in 1981 (Evans and Kaufman, 1981), but it was not until 1998 that human ESCs (hESCs) were reported (Thomson et al., 1998). However, there are several morphological and molecular differences between mESCs and hESCs: mESC colonies are domed, hESCs are flat, hESCs rely on glycolysis, and mESCs oxidate phosphorylation (Zhou et al., 2012; Sperber et al., 2015). There are also major differences in the ectopic signaling pathways that are required: mESCs rely on bone morphogenic protein 4 (BMP4) and leukemia inhibitory factor (LIF) signaling (Ying et al., 2003); however, hESCs rely on Activin A and fibroblast growth factors (FGFs) (Beattie et al., 2005); and not only is BMP signaling not required for hESCs, but inhibition of BMP signaling is even beneficial (Xu et al., 2005). Application of the hESC growth medium to mouse cells led to the isolation of epiblast stem cells (EpiSCs). These cells were derived from E6.5 pre-gastrulating embryos and are quite different in morphology, gene expression, and culture conditions than mESCs (Brons et al., 2007; Tesar et al., 2007). EpiSCs, instead, more closely resemble hESCs, and the pre-gastrulating epiblast, a later developmental stage than mESCs. To explain the properties of mESCs and EpiSCs, a two-phase model of embryonic development was proposed, consisting of a "naïve" (mESC) state, which is closer to the early epiblast, and a "primed" (EpiSC) state, which is closer to the late pre-gastrulating epiblast. Other cell types exist on a continuum between these two conditions and, sometimes, outside of this classification (Nichols and Smith, 2009). Several features distinguish the naïve and primed states (**Figure 1**). Primed and naïve cells have a distinctive morphology: primed cells are flat and more two-dimensional, while naïve cells have a domed shape and are more three-dimensional (3D). Primed cells are more glycolytic, while naïve cells rely more on oxidative phosphorylation (Zhou et al., 2012). In females, the X chromosomes are active in naïve but silent in primed cells. Finally, in the naïve state, the chromatin tends to be looser and, overall, less repressive, which enables more transposable element (TE) expression.

The Naïve and Ground States in Mouse Embryonic Stem Cells

Typical mESCs are grown in serum supplemented with the cytokine LIF (Smith et al., 1988; Williams et al., 1988) and adopt a "naïve" state. mESCs can also be supported in medium referred to as ground state or "2iLIF" medium, which consists of PD0325901 (a MEK inhibitor), CHIR99021 (a GSK3 inhibitor), and LIF (Ying et al., 2008). Intriguingly, growth in 2iLIF improves the cells, making them easier to maintain, less heterogeneous, and less prone to spontaneous differentiation (Kolodziejczyk et al., 2015; Takashima et al., 2015). Cells grown in 2iLIF also have an altered chromatin state and reduced levels of repressive marks; for example, H3K27me3 levels are reduced at developmental genes (Marks et al., 2012) and lower levels of other repressive histone modifications (Sim et al., 2017). The change in DNA methylation is more drastic: serum + LIF-grown mESCs have slightly reduced DNA methylation levels, compared with somatic cells, but cells grown in 2iLIF are nearly completely demethylated (Habibi et al., 2013; Leitch et al., 2013). The mechanism appears to be mediated by one of the inhibitors in the 2i cocktail, the MEK inhibitor PD0325901, which indirectly reduces the levels of H3K9me2 and so blocks the recruitment of UHRF1 to chromatin. The loss of UHRF1 then leads to a failure to recruit the methyltransferase maintainer DNMT1, and so DNA is passively demethylated as cells divide (von Meyenn et al., 2016). A complementary mechanism posits a more active process involving MEK inhibitor indirectly stabilizing the histone H3K9me3 demethylase KDM4C (JMJD2C). This leads to demethylation of H3K9me3, the active conversion of 5mC to 5hmC by TET1 hydroxylase, and eventual

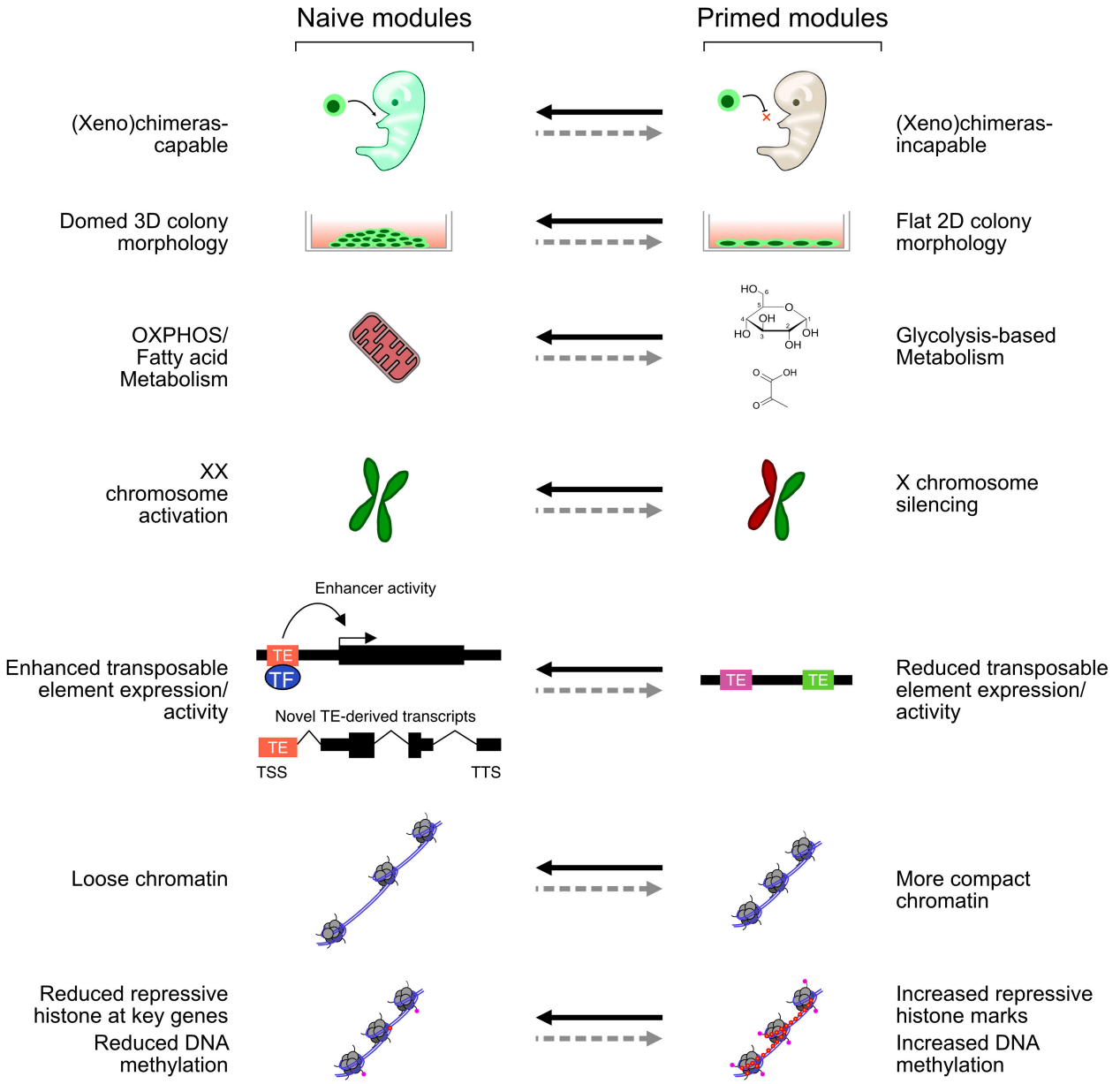


FIGURE 1 | Selected modules representing key features of naïve and primed cells. Each module can be switched on and off, perhaps independently, in naïve and primed mouse and human embryonic stem cells (hESCs). While mouse ESCs (mESCs) activate all these modules, different naïve-like hESCs may only activate some aspects.

demethylation (Sim et al., 2017). The global DNA demethylation in naïve cells is reminiscent of a similar DNA demethylation that occurs in early embryonic development between the two-cell (2C) and blastocyst stages (Guo et al., 2014; Smith et al., 2014). However, how close MEK inhibitor-driven DNA demethylation is to embryonic DNA hypomethylation is unclear. Another issue that is complex in 2iLIF-grown cells is that while H3K27me3 at gene promoters is reduced, 2iLIF cells have increased overall H3K27me3 levels and often have increased heterochromatic marks (van Mierlo et al., 2019). This suggests that 2iLIF cells have an overall more repressed chromatin state. One possibility is that the loss of DNA methylation leads to compensatory mechanisms that repress genes and generate heterochromatin by methylating histones. Indeed, H3K27me3 and H3K9me3 compensate for the loss of DNA methylation to repress TEs in mESCs (Walter et al., 2016).

The bromodomain-containing protein BRD4 is an epigenetic reader, which binds to acetylated histones. *Brd4* knockout mice lack an inner cell mass, and when *Brd4* was inhibited or knocked down in naïve cells, it led to differentiation, at least partly due to the loss of *Nanog* expression and other pluripotency genes (Di Micco et al., 2014; Liu et al., 2014; Horne et al., 2015). BRD4

also has a key role in maintaining enhancers, by recruiting CDK9 and the mediator complex (Di Micco et al., 2014). Beyond the direct regulation of pluripotent genes, BRD4 has a complex role in mediating the differences between the 2iLIF ground state and serum + LIF-grown naïve cells, as 2iLIF cells can tolerate the loss of *Brd4*, while serum + LIF cells cannot (Zhang et al., 2020b). The mechanism involves the inhibition of GSK3 in 2iLIF conditions, which leads to stabilization of beta-catenin, and the recruitment of BRD4 and a multimolecular protein complex to pluripotency genes to make 2iLIF-grown mESCs resistant to differentiation (Zhang et al., 2020b). This results in reduced pause-release of RNA polymerase II (polII) and more stable transcription, which helps explain previous observations that 2iLIF-grown cells have more homogenous gene expression than serum + LIF-grown cells (Kolodziejczyk et al., 2015). The mediator complex is a multiprotein complex that is important at integrating signals to activate nearby gene expression by recruiting polII. It has a key role in regulating super-enhancers, which are large regions of DNA with a potent enhancer activity that are important for regulating cell type-specific genes (Whyte et al., 2013). Mediator, in addition to being recruited by BRD4, also has a role in naïve and primed states, as chemical inhibition of two CDKs, CDK8/19, promotes the naïve state in both humans and mice (Lynch et al., 2020). This effect is driven by the hyperactivation of enhancers, as inhibition of CDK8/19 leads to derepression of mediator. BRD4 is not the only bromodomain-containing protein involved in naïve and primed state control. BRD9 was identified as a member of a non-canonical chromatin remodeling BAF complex, and when BRD9 was inhibited, the cells began to acquire aspects of EpiSCs, and, like BRD4, *Brd9* is dispensable in 2iLIF conditions (Gatchalian et al., 2018).

Chromatin inside the cell is tightly packed into successive 3D layers that can be broadly divided into a hierarchy of three organizational features (Rowley and Corces, 2018). The first level is the A and B compartments, which, very roughly, correspond to euchromatin (A compartment) and heterochromatin (B compartment). At the second level, topologically associated domains (TADs) are megabase domains of chromatin that extensively interact within a TAD but weakly between TADs. Finally, at the third level, individual TFs and epigenetic factors form contacts between strands of DNA to form chromatin loops, which are often responsible for bringing distal enhancers together with promoters. mESCs have unique features at all three of these levels, which are suggestive of open and relaxed chromatin. As mESCs are differentiated to neurons, the A compartments decrease and the B compartments increase in interaction frequency, indicating the loss of active chromatin and the acquisition of repressed chromatin as mESCs differentiate (Bonev et al., 2017). In human cells, the situation is similar, and a high-resolution HiC dataset in hESCs and somatic cells showed many A to B compartment switches (Dixon et al., 2015). At the second level, TAD compartment structure strengthened as mESCs differentiated, and TADs containing actively expressed genes interacted weakly, while inactive TADs increased (Bonev et al., 2017). The chromatin state can also influence the 3D genome folding, as knockout of the H3K9me1/2 methyltransferase *Ehmt2* led to reduced TAD boundary strength, although compartments were unaffected (Jiang et al., 2020).

Intriguingly, the 3D structure in developing embryos is initially undefined. From the zygote to the late 2C stage, the TADs and chromatin loops are nearly completely absent, and only compartments on the paternal genome are weakly present (Du et al., 2017; Ke et al., 2017). TADs and chromatin loops reestablish at the eight-cell to the morula stages (Du et al., 2017; Ke et al., 2017). TADs and compartments reform around the same time as zygotic genome activation (ZGA), and there is some evidence that the reestablishment of 3D structure can influence embryonic development. In somatic cell nuclear transfer (SCNT) experiments, the somatic nucleus inside the oocyte briefly retains TADs, which are relaxed at the 2C stage and match normal development. However, the brief window when TADs are erroneously present impairs minor ZGA and embryonic development at the 2C stage, which can be rescued by depleting cohesin to help disrupt TADs in the somatic nucleus (Zhang et al., 2020a). Based on these observations, the totipotent stages of embryonic development (zygote to late eight-cell stage) seem to require relaxed unstructured 3D chromatin. However, it is unclear if this is a necessary feature of totipotency or a consequence of epigenetic reprogramming in early embryogenesis or simply reflects rearrangements in chromatin that are independent of embryonic development. It would be useful to explore these issues in the more experimentally tractable mESCs, and several systems have been explored that lead to dissolution of the 3D genome. With the use of a degron system, the key cohesin complex member RAD21 was deleted in mESCs, leading to the near-complete loss of TADs but a slight strengthening of A/B compartments, and RING1B-mediated polycomb loops persisted (Rhodes et al., 2020). However, depletion of RAD21 in mESCs had surprisingly little effect on gene expression or cell phenotype (Rhodes et al., 2020). CTCF, a major 3D genome organizer, was similarly degraded in mESCs. A/B compartments were unaffected, but TADs and chromatin loops were disrupted; however, once again, the effect on the mESC phenotype was modest, although there was a proliferation defect if CTCF loss persisted (Nora et al., 2017). These results suggest that 3D structure is relatively uncoupled from cell type control, although the precise 3D structure of embryonic cells has not been fully recapitulated in mESCs and hence remains inconclusive.

A remarkable feature of mESCs is their tolerance for the loss of epigenetic regulatory enzymes with relatively few effects (He et al., 2019). For example, components of the polycomb repressor complex 1 and 2 (PRC1 and PRC2) are dispensable for mESCs (Chamberlain et al., 2008). Loss of *Rybp*, a member of an atypical PRC1 complex, also has little to no effect (Li et al., 2017b). Knockdown of *Setdb1*, a H3K9me3 methyltransferase, only predisposes cells to differentiation (Karimi et al., 2011). Co-activators also show the same pattern: *Kmt2d* (MLL2), a H3K4 methyltransferase, is dispensable (Lubitz et al., 2007). Epigenetic factor knockouts often do not substantially impact the pluripotent state, although they may make them more prone to spontaneous differentiation and alter the differentiation direction of the cells, or, as is often the case for epigenetic

factor knockouts, lead to embryonic arrest at gastrulation, for example, in the *Kdm1a* (Macfarlan et al., 2011) or *Setdb1* (Dodge et al., 2004) knockouts. A CRISPR/Cas9 screen identified around 40 epigenetic factors that, when knocked out, delayed the differentiation of mESCs, and just two epigenetic factor knockouts promoted mESC differentiation, *Cbx7* and *Sin3b* (Li et al., 2018). As their screen was set up primarily to detect improved or impaired ability to differentiate, knockouts that did not affect differentiation would not be detected. This emphasizes the remarkable ability of mESCs to tolerate the widespread loss of epigenetic regulators. This tolerance may be related to the more open and active chromatin, which appears to be a feature of early embryonic cells (Boskovic et al., 2014; Schlesinger and Meshorer, 2019).

Transcriptional Control of Primed Mouse Epiblast Stem Cells

Primed EpiSCs are a distinct cell state, compared with mESCs. EpiSCs show both molecular and phenotypic differences, particularly colony morphology and the lack of ability to form chimeras (**Figure 1**). Despite these phenotypic differences, mESCs and EpiSCs have both shared and distinct transcriptional regulation. The core pluripotent network of OCT4, SOX2, and NANOG are active in both cell types (Weinberger et al., 2016), but they bind to different genomic loci (Buecker et al., 2014; Galonska et al., 2015; Matsuda et al., 2017). The core pluripotency network is coordinated by a different set of TFs in each cell type. For example, ESRRB, NR0B1, ZFP42, and TFCP2L1 are important in mESCs (Festuccia et al., 2012; Hutchins et al., 2013; Adachi et al., 2018; Atlasi et al., 2019), but ZIC2, ZIC3, POU3F1, and OTX2 are key in EpiSCs (Acampora et al., 2013; Matsuda et al., 2017; Yang et al., 2019c). Thus, there is a core gene expression module that is common to mESCs and EpiSCs, and divergent regulatory modules specific to each cell type. Both cell types are centered around SOX2-OCT4, but mESCs use TFs that are expressed in the early blastocyst (e.g., NANOG, ESRRB, and TFCP2L1), and EpiSC-specific TFs tend to be expressed in the gastrulating blastocyst (OTX2 and POU3F1). In addition to transcriptional differences, the chromatin and epigenetic states are also altered between EpiSCs and mESCs, and enhancer usage is dramatically altered between EpiSCs and mESCs. Even though only a few hundred genes change expression between mESCs and EpiSCs, several tens of thousands of enhancer marks are differentially regulated (Factor et al., 2014). Even genes that are expressed in both cell types can utilize different enhancers (Factor et al., 2014). This effect had been seen previously at the *Pou5f1* locus, which has two enhancers, a distal enhancer that is active in preimplantation embryos and mESCs, and a proximal enhancer that is active in the epiblast and EpiSCs (Yeom et al., 1996). Yet the global profiling of chromatin highlighted how widespread this phenomenon is (Factor et al., 2014). This redistribution of enhancers is ultimately driven by changes in the TF activity, which decommission and activate panels of enhancers to control each cell state. A good example is OCT4 and SOX2, and both are expressed in mESCs and EpiSCs but are drastically altered in their binding patterns in the two-cell types (Matsuda et al., 2017). It is most likely that OCT4 and SOX2 binding is altered due to the activity of OTX2 and POU3F1, which becomes the dominant factors in EpiSCs (Matsuda et al., 2017).

Interconversion Between Mouse Naïve and Primed Cells

Mouse embryonic stem cells and EpiSCs can be interconverted *in vitro*, and while conversion of mESCs to EpiSCs is relatively efficient, the conversion of EpiSCs to ESCs remains inefficient without transgenes or epigenetic modulation (Zhou et al., 2010; Tosolini and Jouneau, 2016; Stuart et al., 2019). This indicates that EpiSCs are developmentally later, as, in general, cells can efficiently differentiate toward their progeny but are resistant to dedifferentiation to their precursor cell type. Epigenetic barriers, particularly unidirectional blocks, appear to permit the conversion of mESCs to EpiSCs but block the reverse. ZFP281 acts as just such a bidirectional valve, as it assists mESC to EpiSC conversion but blocks the reverse (Mayer et al., 2020). The mechanism involves ZFP281 co-binding with EHMT1 to methylate H3K9 and inhibit genes in the early stages of mESC to EpiSC conversion (Mayer et al., 2020), while in the reverse case ZFP281 binds the NuRD co-repressor complex to suppress *Nanog* expression and enable exit from pluripotency (Fidalgo et al., 2012). When mESCs are converted to EpiSCs, there is a global reconfiguring of chromatin (Factor et al., 2014). These properties have made the interconversion of mESCs to EpiSCs a powerful model to understand the epigenetic modulation of cell conversions.

Several factors have been identified that influence the conversion of mESCs to EpiSCs (**Table 1**). Most are TFs that promote the conversion of EpiSCs to mESCs. These TFs recruit other co-activators and co-repressors to influence the chromatin state, although their direct activity is not always clear, as TFs can often act as both activator and repressor in a context-specific manner. For example, NANOG is mainly an activator but can also work as a repressor (Heurtier et al., 2019). When NANOG binds to DNA with ZFP281, it recruits the NuRD histone deacetylase (HDAC) co-repressor complex (Fidalgo et al., 2012). *Esrrb* is a major requirement to convert EpiSCs to mESCs, as it participates in extensive chromatin remodeling (Adachi et al., 2018). Many naïve-specific enhancers are kept silent in EpiSCs by DNA methylation and inaccessible chromatin. ESRRB opens these naïve-specific enhancers by recruiting the p300 complex, displacing and phasing nucleosomes, and opening closed chromatin, making it accessible for other members of the pluripotency regulatory network, such as OCT4, SOX2, and NANOG, to bind (Adachi et al., 2018).

Epigenetic pathways have been directly implicated in the conversion of mESC to EpiSCs. Activatory epigenetic marks are also redistributed between mESCs and EpiSCs (Factor et al., 2014), and the enhancer mark H3K4me1 has been directly implicated in the conversion to mESCs. Inhibition of the histone methyltransferase that catalyzes H3K4me1, MLL1, drives EpiSCs back to a naïve state (Zhang et al., 2016). In addition to H3K4me1/MLL1 inhibition, an inhibitor of the histone H3K4/9 demethylase KDM1A (LSD1) was part of a cocktail of

TABLE 1 | Selected factors that influence the interconversions of mESCs and EpiSCs.

Molecule(s)	Function	Effect on mESCs, EpiSCs, and their interconversion	References
Brd9/BAF complex	Chromatin remodeler	BRD9, through a non-canonical BAF complex, blocks transition to a primed state	Gatchalian et al., 2018
Epiblastin A	Inhibitor	Inhibits CK1 and induces EpiSC-to-mESC conversion	Illich et al., 2016
Esrrb	Transcription factor	Overexpression sustains LIF-independent self-renewal in the absence of Nanog and reverts EpiSCs to mESCs	Festuccia et al., 2012; Adachi et al., 2018
Hif1a	Transcription factor	HIF-1α activation switches mESC metabolism and pushes cells toward an EpiSC-like state	Zhou et al., 2012
iPStat3	$Gp130^{Y118F}$ receptor	Artificially induces STAT3 signaling and converts mESCs to EpiSCs	Stuart et al., 2019
Kdm1a/Kmt2d (MLL4)	Histone methylation	Kmt2d (MLL4) is required to exit pluripotency; knockdown of Kdm1a restores the ability of mESCs to convert to EpiSCs	Cao et al., 2018
Kdm6b (JMJD3)	Histone demethylase	Demethylates H3K27me2 or H3K27me3; Kdm6b facilitates a Klf4-driven EpiSC-to-ESC conversion	Huang et al., 2020
Klf2, Klf4	Transcription factor	Overexpression sustains LIF-independent self-renewal and can convert EpiSCs to mESCs	Guo et al., 2009
Kmt2a (MLL1)	Epigenetic inhibitor	H3K4me1 methyltransferase, deletion of Kmt2a (MLL1) impairs mESC differentiation to EpiLCs; MLL1 Inhibition reprograms EpiSCs to mESCs	Zhang et al., 2016
Nanog	Transcription factor	Transient transfection of Nanog mediates reprogramming of mESCs to EpiSCs	Silva et al., 2009
Nr5a1, Nr5a2	Transcription factor	Overexpression reprograms EpiSCs to mESCs	Guo and Smith, 2010
Otx2	Transcription factor	Required to stably establish EpiSCs	Acampora et al., 2013
Prdm14	Epigenetic factor	Overexpression drives EpiLCs to mESC-like cells	Okashita et al., 2016
Sall1	Transcription factor	Promotes reprogramming EpiSCs to mESCs	Yang et al., 2019a
Setdb1	Histone methyltransferase	Methyltransferase for H3K9me3; Setdb1 loss enriches a transient 2C-like state in mESCs	Wu et al., 2020
Smarcad1	SWI/SNF helicase	Knockdown converts mESCs to EpiSC-like cells	Xiao et al., 2017
Tcf3, Etv5, Rbpj	Transcription factors	mESCs lacking Tcf3, Etv5, and Rbpj are trapped in a naïve pluripotent condition and are difficult to differentiate	Kalkan et al., 2019
Tfe3	Transcription factor	Nuclear-localized TFE3 blocks mESCs from differentiating	Betschinger et al., 2013
Wnt, Gsk3b	Transcription factor	Wnt signaling blocks the conversion of mESCs to EpiSCs and maintains mESCs in the naïve state	Ying et al., 2008; ten Berge et al., 2011
Zbtb7a/b	Transcription factor	Knockdown converts EpiSCs to naïve mESCs	Yu et al., 2020b
Zfp281	Transcription factor	Deletion of Zfp281 promotes EpiSCs reprogramming and acts downstream of Ehmt1 (G9a-like protein; a histone methyltransferase for H3K9me1/2)	Mayer et al., 2020
Zfp706	Transcription factor	Deletion of Zfp706 promotes mESC self-renewal and promotes EpiSC reprogramming	Leeb et al., 2014

mESC, mouse embryonic stem cell; EpiSC, epiblast stem cell; LIF, leukemia inhibitory factor; EpiLC, epiblast-like cell.

chemicals that could promote the conversion of EpiSCs to mESCs (Zhou et al., 2010), underlining the importance of epigenetic modulation in cell type conversions. Histone citrullination is the post-translational conversion of arginine to citrulline, and it can act as an epigenetic mark, although its functions are not well defined. In naïve mESCs, histone H1 is citrullinated and evicted from chromatin, decondensing chromatin and likely making it more accessible for TF binding (Christophorou et al., 2014). Histone H3 can also be citrullinated, and it can recruit the SWI/SNF chromatin remodeler SMARCAD1 to relax chromatin (Xiao et al., 2017). Knockdown of *Smarcad1* led to H3K9me3 deposition and heterochromatin spreading, and the cells adopted features of EpiSCs (Xiao et al., 2017). This suggests that citrullination assists in the control of heterochromatin and the maintenance of the naïve state. Overall, these observations agree with the idea that naïve cells represent an "unprogrammed" ground state with lower levels of both repressor and enhancer marks and agree with the idea that less overall epigenetic regulation is a feature of naïve mESCs (Schlesinger and Meshorer, 2019).

In summary, the naïve and primed states are relatively well described in mouse cells, and it is possible to interconvert the cell types. While conversion from mESCs to EpiSCs is relatively easy, the converse transition is difficult and often inefficient without transgenes. Indeed, there are multiple trajectories to convert EpiSCs back to mESCs, with some mechanisms passing cells through later developmental stages, such as mesoderm-like cells, and other conversion methods pass cells through earlier developmental stages (Stuart et al., 2019). It appears that there is a single main pathway in the differentiation of mESCs to EpiSCs but multiple pathways that EpiSCs must be forced along to revert to mESCs. These observations suggest the existence of epigenetic barriers, probably many, that act to impair the dedifferentiation of EpiSCs to mESCs (**Figure 2**). Ultimately, the interconversions

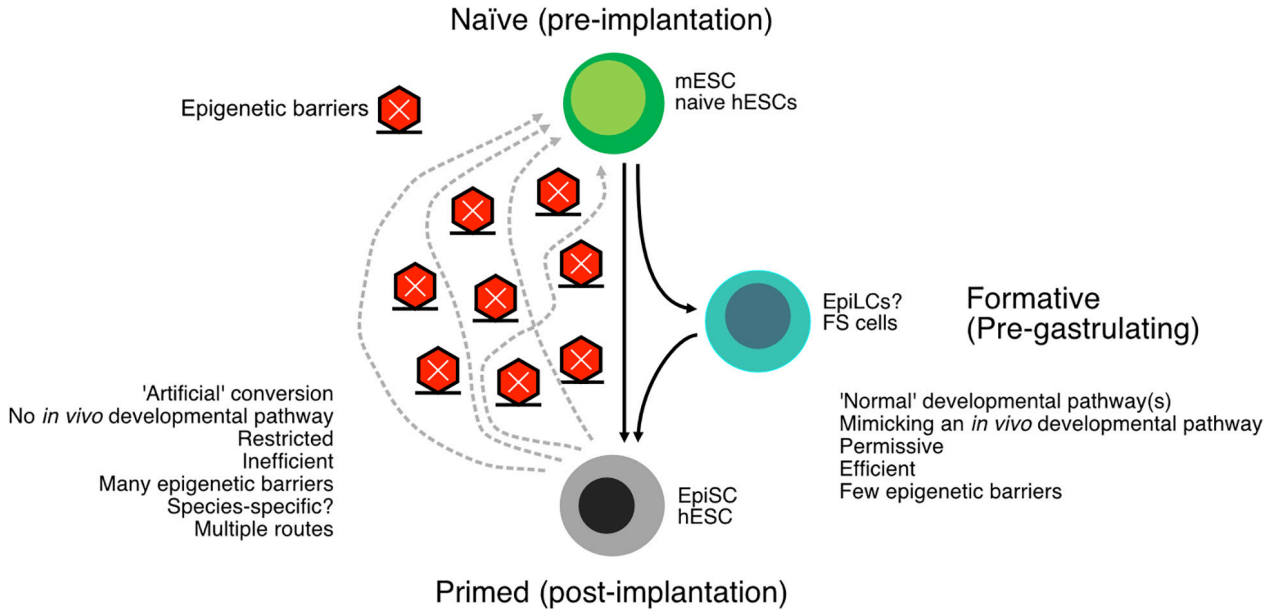


FIGURE 2 | Schematic of the conversion of naïve embryonic stem cells (ESCs) to primed ESCs. The "downstream" pathway follows a normal developmental program and tends to be relatively easy and efficient. The reverse process, reverting primed cells to naïve cells, is arguably an artificial process and consequently more difficult and inefficient. Epigenetic barriers exist between the primed and naïve cells, and there may be more than one way to proceed through these epigenetic barriers. The many epigenetic barriers and the divergent pathways have likely contributed to the difficulty in generating human naïve ESCs. In addition to the naïve and primed states, there is also a formative state that exists intermediate to naïve and primed cells but is capable of primordial germ cell (PGC) formation.

of mESCs and EpiSCs are a crucial window into epigenetic controls that underlie cell type conversions.

An Intermediate, Formative Pluripotent State

Work on the naïve state in mice and humans has led to the emergence of a more subtle conception of naïve and primed states. Instead of existing as a binary state, either naïve or primed, there is instead a spectrum of states, some stable and others unstable, that exist between naïve and primed states. A recent proposal posits the existence of a critical intermediate state, the formative state, that exists between naïve and primed mESCs (Smith, 2017). The formative state represents the loss of naïve pluripotency but precedes lineage commitment. One of the first events in the *in vitro* differentiation of mESCs is the loss of the naïve-specific TFs, such as *Tfcp2l1* and *Esrrb*. This matches a similar loss in the early post-implantation epiblast (Boroviak et al., 2014) and precedes lineage priming and acquisition of primed-specific genes. A formative state helps explain some curious phenomena that are not easily reconciled with a binary naïve and primed model. First, deriving primordial germ cells (PGCs) from both mESCs and EpiSCs is inefficient, despite both cell types being on the presumptive developmental path capable of generating PGCs (Hayashi et al., 2011). Second, a third type of *in vitro* epiblast cell, epiblast-like cells (EpiLCs), can give rise to PGCs at relatively high efficiency and can adopt a gene expression profile more reminiscent of E5.0–E6.0 pre-gastrulating blastocysts (Hayashi et al., 2011), which is similar to the timepoint for the specification of PGCs in the embryo (E5.5–E6.25) (Ohinata et al., 2009). This suggests that EpiLCs represent a transient window when PGCs are specified. An interesting aspect of the formative state is the extensive remodeling of chromatin that provides a blank slate for later lineage specification (Smith, 2017). Indeed, EpiLCs have lower levels of repressive histone marks, particularly H3K27me3 and H3K9me2, and have higher levels of bivalent promoters, marked by both H3K4me3 and H3K27me3 (Kurimoto et al., 2015; Yang et al., 2019b). This suggests the EpiLCs are poised for lineage commitment. A comprehensive multi-omic exploration of the conversion of mESCs to EpiLCs revealed a series of waves of gene expression changes that were preceded by widespread chromatin remodeling (Yang et al., 2019b). The conversion to a formative state relies on the expression and activity of *Tcf3*, *Etv5*, and *Rbpj*, as the deletion of all three impairs the differentiation of mESCs to EpiLCs (Kalkan et al., 2019). Of these three factors, their mechanism is unclear, but ETV5 binds to formative state-specific genes and promotes their expression, at least partly through histone acetylation (Kalkan et al., 2019), and all three combine to suppress the naïve pluripotency program.

A disadvantage of EpiLCs is their transient nature, and they cannot be captured *in vitro* like mESCs and EpiSCs. This makes a detailed exploration of the formative state challenging. Recently, two groups reported the isolation of cell lines that fulfill the properties of the formative state but can be maintained in culture (Kinoshita et al., 2020; Yu et al., 2020a). These cells [formative state (FS) and XPSCs] have a gene expression program distinct

from mESCs and EpiSCs, yet they can partially contribute to mouse chimeras and can contribute to the germline (Kinoshita et al., 2020) and can generate PGCs *in vitro* (Yu et al., 2020a). Both FS and XPCs show increased bivalent genes, marked by H3K4me3 and H3K27me3 (Kinoshita et al., 2020; Yu et al., 2020a). Overall, in the naïve state, chromatin is open and lacking repressive marks, while in the FS, the cells begin to acquire bivalent chromatin marks that poise the cells for later lineage commitment.

NAÏVE AND PRIMED STATES IN HUMANS

Human ESCs were first derived in 1998 from blastocyst-stage embryos (Thomson et al., 1998). hESCs are quite different from mESCs and require Activin A and FGFs, rather than serum/2i and LIF that mESCs need. Such a large difference in growth requirements led to research into why hESCs and mESCs were so different. As described above, mouse EpiSCs could be derived that more closely resemble hESCs, based upon morphology and marker gene expression (Brons et al., 2007; Tesar et al., 2007; Rossant, 2011). However, this prompted a question: If EpiSCs resembled hESCs, is it possible to generate naïve hESCs? Subsequent research began from the basis that mESCs are easiest to derive and stable in ground-state 2iLIF media (Ying et al., 2008), and most naïve hESCs began with this cocktail. However, much as the conversion of EpiSCs to mESCs is challenging, likely due to multiple routes and potent epigenetic barriers (**Figure 2**), the transition of human primed to naïve pluripotency has also been challenging. The first reported naïve conditions for hESCs involved the transfection of ectopic *OCT4*, *KLF2*, and *KLF4*, along with 2iLIF (Hanna et al., 2010). Since then, there has been a veritable explosion of competing protocols for naïve hESCs, including many that do not require transgenes (Yilmaz and Benvenisty, 2019). However, the situation remains complex, and there is considerable argument about the nature of the putative naïve hESCs.

Transcriptional and Epigenetic Control of Human Naïve and Primed Cells

The naïve and primed states are well described in mice, but in humans, the situation remains complex (Davidson et al., 2015). Research using human cells has led to the development of a different model of naïve and primed states, which suggests instead of distinct naïve and primed cell states; there are instead modules that can be switched on and off relatively independently of one another (**Figure 1**). In this model, naïve ESCs switch on a set of modules, while primed ESCs switch on a different set. This view has emerged due to the difficulty in establishing a human version of the complete mouse naïve state and the existence of naïve-like cells that only partially fulfill the naïve criteria. Human naïve ESCs cannot be derived using only 2iLIF; instead, a large number of protocols have been developed that give rise to cells that mimic several aspects of mouse naïve cells (Chan et al., 2013; Gafni et al., 2013; Wang et al., 2014; Ware et al., 2014; Duggal et al., 2015; Carter et al., 2016; Theunissen et al., 2016). However, the competing protocols have distinct transcriptional profiles, cell surface markers, and epigenome states (Yang et al., 2019b), and no comprehensive model has emerged concerning the mechanisms controlling these states in humans. Indeed, just like the conversion of primed EpiSCs to naïve mESCs, there appear to be multiple routes from primed hESCs to naïve pluripotency (Duggal et al., 2015), and potent epigenetic barriers resist the transition. Both naïve and primed hESCs are regulated through several core TFs, for example, SOX2 and OCT4, which are common to both naïve and primed hESCs in mice and humans. However, the human cells have gene regulatory networks that utilize KLF5, KLF7, TFCP2L1, FOXR1, ZIC2, and TFAP2C, for the naïve state; and OTX2 and SALL2 for the primed state (Takashima et al., 2014; Weinberger et al., 2016; Pastor et al., 2018). Some of these are active in mouse naive mESCs (KLF5 and TFCP2l1); however, several seem unique to humans (e.g., TFAP2C), and some critical regulators in mice (e.g., *Esrrb*) are not typically upregulated in human naïve cells (Kisa et al., 2017; Rostovskaya et al., 2019). Consequently, there are substantial differences in transcriptional regulation in humans (**Table 2**), and the full naïve and primed regulatory networks remain to be elucidated.

Epigenetic Influence of the Chemical Cocktails That Convert Primed to Naïve Cells

The human naïve cocktails begin with 2iLIF as a starting base, although LIF can be substituted for other molecules, and LIF is not strictly required in mouse naïve cells. Beyond 2i, many other inhibitors and signaling factors have been used, targeting JNK, MAPK, BRAF, SRC, and ROCK kinases. These inhibitors likely have widespread downstream effects on epigenetic regulation, although the pathways have not been fully explored. The two small molecules that have been directly implicated in epigenetic control are HDAC inhibitors, which are useful as a pretreatment of primed cells before conversion to naïve cells (Ware et al., 2014), and vitamin C. Vitamin C acts as a co-factor for TET and Jumonji (JMJ) domain-containing proteins. JMJ domain proteins are involved in histone demethylations, while TET domains convert 5mC to 5hmC, which is the initial step in the DNA demethylation pathway (Teslaa and Teitell, 2015). Overexpression of *TET1*, along with *MCRS1* and *THAP11*, can drive cells toward a naïve-like state (Durruthy-Durruthy et al., 2016), indicating that TET1 is important in naïve cells. Nonetheless, how vitamin C modulates TETs and JMJs to promote the formation of naïve hESCs remains unclear. Other epigenetic pathways involved include the PRC2 component EZH2, which is required to maintain primed hESCs but is dispensable for naïve cells (Shan et al., 2017). As EZH2 is a key catalytic enzyme for the repressive histone mark H3K27me3, it suggests that human naïve cells may have reduced epigenetic repression, the same as mice, although, just as in mice, human naïve cells have higher overall levels of H3K27me3 (De Clerck et al., 2019). Possibly, the situation is similar to that of the mice, and H3K27me3 is lost at the promoters of critical genes, but overall H3K27me3 is elevated in response to reduced DNA methylation.

TABLE 2 | Epigenetic factors implicated in naïve and primed hESC control.

Gene name	Function	Effect on naïve or primed hESCs	Epigenetic pathway	References
EZH2	Histone methyltransferase	Required to primed hESCs, but dispensable in naïve hESCs	H3K27me3	Shan et al., 2017
HDAC1/3	Histone deacetylase	Histone deacetylase inhibition was required to establish naïve-like hESCs	Histone deacetylation/ transcriptional regulation	Ware et al., 2014
MCRS1, *TET1*, *THAP11*	DNA demethylation (TET1) and epigenetic remodeling	Expressed in combination and can convert primed hESCs to naïve-like cells	5mC to 5hmC conversion (TET1) and unknown	Durruthy-Durruthy et al., 2016
NNMT	Nicotinamide N-methyltransferase	Knockout in primed hESCs leads to cells to acquire some naïve characteristics	Reduces histone methylation by removing the methyl-group donor SAM	Sperber et al., 2015
SIRT2	Histone/protein deacetylase	SIRT2 controls primed hESC state	Acetylation and regulation of glycolytic enzymes	Cha et al., 2017
TNKS1/2	Chromatin remodeling	Tankyrase 1/2 inhibition promotes naïve and extended pluripotency	Telomere elongation	Zimmerlin and Zambidis, 2020
TET1/2	DNA demethylation	Required for pluripotency in primed but not the naïve state	5mC to 5hmC conversion	Finley et al., 2018

hESC, human embryonic stem cell; SAM, S-adenosylmethionine.

Ultimately, there remains argument over which of these naïve cocktails captures most of the naïve state. This led to an expansion of the model that the naïve and primed cells exist on a spectrum (Weinberger et al., 2016; Smith, 2017; Yang et al., 2019b) and led to ideas of multiple interchangeable naïve and primed modules that can be switched on or off under certain conditions (**Figure 1**; Theunissen et al., 2016; Cornacchia et al., 2019). This helps explain the differences in the naïve cocktails, as each cocktail can switch on certain modules but may fail to activate them all.

Extra/Expanded-Capability Cells With Totipotent-Like Properties

An extra complication for the naïve and primed model is the existence of "extra-capability cells" that describe ESC-like cells that are pluripotent and also have some aspects of totipotency (Yang et al., 2017a,b; Gao et al., 2019). These cells are grown under culture conditions similar to naïve cells, but they drop the MEK inhibitor from 2i and include WNT pathway inhibitors and then either SRC and tankyrase inhibitors (EPSCs) or ROCK inhibitor (EPSs). First described in mice, EPS/EPSCs (extra/expanded pluripotent stem cells) can contribute to the trophoblast, which is a property that both naïve and primed mESCs and hESCs lack. However, the gold standard test of totipotency, the derivation of a complete mouse using only these cells, has not yet been reported. This suggests that their totipotent properties remain incomplete or that they lack full totipotency. Indeed, there is argument that while mouse EPS/EPSCs can occasionally localize in the trophoblast, these cells lack trophoblast-specific markers, still express epiblast markers, and do not have totipotent properties (Posfai et al., 2021). Additionally, the DNA methylation state of EPS/EPSCs is somewhat contradictory. EPSCs have high levels of methylated DNA (Yang et al., 2017a), which does not match the DNA hypomethylation of totipotent embryonic cells and 2iLIF-grown naïve mESCs, although it should be noted that it is unclear if the hypomethylation in 2iLIF cells is a cell type effect, or is a side effect of PD0325901, a component of the 2iLIF cocktail. One advantage that the EPS/EPSCs do have is in the derivation of cells from species that have been previously resistant to the isolation of ESCs, for example, deriving porcine EPSCs (Gao et al., 2019). The EPS and EPSC cocktails have also been applied to human embryos to derive putative totipotent cells (Yang et al., 2017b; Gao et al., 2019). Overall, the identity of the EPS and EPSCs remains unclear, particularly how they correspond to *in vivo* development. EPS and EPSCs may hint at a further expansion to the module concept, where EPS and EPSCs are switching off some naïve modules and activating some totipotent modules. But like the naïve/primed split, these cell types are potentially activating only some of the totipotent modules, and only partial totipotency is achieved. A fascinating study of how biological phenotypes can act independently is the reprogramming of mESCs to oocyte-like cells (Hamazaki et al., 2020). In that study, oocyte-like cells could be derived without PGC specification, meiosis, or epigenetic reprogramming of DNA demethylation. This suggests that these aspects are independent and can be switched on and off in a module-specific fashion (Hamazaki et al., 2020).

Links Between Metabolism and Epigenetic Control of the Naïve and Primed States

Pluripotent stem cell fate transitions from naïve to primed are accompanied by a metabolic switch, from oxidative phosphorylation (oxphos) to mainly glycolysis, respectively (**Figure 1**; Teslaa and Teitell, 2015). This mirrors the developing embryo, which mainly uses pyruvate and oxphos from fertilization to the blastocyst stage, before transitioning to glucose-based glycolysis and anaerobic metabolism in the late blastocyst (Devreker and Englert, 2000; Chason et al., 2011). Although it should be noted that it is unclear if glycolysis is required to produce energy for embryos, instead, it may be needed for biosynthetic pathways (Smith and Sturmey, 2013).

ESCs, conversely, are highly active cells that divide rapidly and need a lot of energy. Hence, the link between embryonic metabolism and ESC metabolism is not a complete match. There is nonetheless a close link between metabolism and epigenetic control that has not been thoroughly explored (Shyh-Chang and Ng, 2017). For example, *SIRT1* is high in hESCs and acetylates and activates glycolytic enzymes (Cha et al., 2017), and in mice, HIF1A controls glycolytic/oxphos metabolism and influences cell state (Zhou et al., 2012). Many of the reactants required for epigenetic control are metabolic products. For example, acetyl-CoA is the main acetyl donor for histone acetylation, and intracellular levels of acetyl-CoA directly regulate global histone acetylation, and so function as a signal for overall cellular energy metabolism (Cai et al., 2011). An analysis of naïve and primed hESC states revealed that naïve cells express nicotinamide N-methyltransferase (*NNMT*) at high levels (Sperber et al., 2015), which is responsible for metabolizing *S*-adenosylmethionine (SAM), the major chemical donor for histone methyltransferases. As a consequence, naïve cells have low levels of SAM and correspondingly low levels of histone methylation, while primed cells have the inverse, high SAM and high histone methylation (Sperber et al., 2015). Importantly, these metabolic changes may be required for the naïve to primed transition, as *NNMT* knockout cells transition toward a primed state (Sperber et al., 2015). Similarly, 2iLIF-grown mESCs maintain high levels alpha-ketoglutarate that biases the cells toward DNA and histone demethylation by promoting the activity of JMJ-containing demethylases (Carey et al., 2015). In addition to these direct links between metabolism and chromatin, manipulation of metabolism by altering the growth medium also has strong effects on hESCs. Lipid deprivation of primed hESCs reverts them to an intermediate naïve-like state, and the reapplication of lipids pushes the cells toward a primed state (Cornacchia et al., 2019). The exact metabolic/epigenetic pathways behind this effect remain to be elucidated, but lipid deprivation may promote glucose utilization by glycolysis, making more acetyl-CoA available and leading to increased histone acetylation and gene activation. Similarly, mouse 2iLIF-grown cells can utilize fatty acid oxidation (FAO); and inhibition of FAO leads to a reversible quiescence (Khoa et al., 2020), reminiscent of diapause in mice where embryos can be paused if the developmental environment is unfavorable. This effect is driven by the activity of MOF, a histone acetyltransferase that acetylates histones at FAO-related genes and helps activate them (Khoa et al., 2020). Ultimately, there is an intimate interdependence between metabolism, and epigenetic control in embryonic cells and the embryo (Chason et al., 2011), which remains to be comprehensively explored (Betschinger, 2017).

Epigenetic Control of the X Chromosomes in Female Cells

The epigenetic status of the X chromosomes in female ESCs is a particularly important point when discussing human naïve cells and is considered something of a hallmark for the naïve state (Nichols and Smith, 2009; Theunissen et al., 2016). In mice, the situation is relatively straightforward; in naïve mESCs, both X chromosomes are active; *Xist*, a long non-coding RNA that silences one X chromosome is not expressed; and in primed EpiSCs, *Xist* is expressed and one X chromosome is epigenetically inactivated (Bao et al., 2009). This roughly matches the *in vivo* embryonic states: in the late inner cell mass, both X chromosomes are active, and during differentiation in the late epiblast, one random X chromosome is silenced (Okamoto et al., 2004, 2011). The situation in human cells is more complex. Female hESCs (i.e., primed state) have one active and one inactive X chromosomes, which is strong evidence that hESCs are developmentally later than the early blastocyst (Nichols and Smith, 2009). Human naïve cells, depending upon the protocol used, have varying states of X chromosome inactivation, including intermediate states of *XIST* expression and epigenetic silencing (Sahakyan et al., 2017). The discrepancy between mice and humans may, at least partly, be related to the mechanism of X chromosome inactivation in humans, which appears to be more complex than in mice (Patrat et al., 2020). Briefly, in humans, *XIST* is expressed in the inner cell mass, but its expression is not correlated with epigenetic suppression (Okamoto et al., 2011; Petropoulos et al., 2016). Instead, one X chromosome is "dampened" by an unclear mechanism, before later full X chromosome inactivation by chromatin silencing (Petropoulos et al., 2016), possibly as late as post-implantation. This added complexity in humans may explain the differences in X chromosome status in naïve and primed hESCs or may reflect species-specific epigenetic regulatory differences (Okamoto et al., 2011).

Ultimately, human naïve ESCs remain surprisingly elusive to pin down (Theunissen et al., 2016). A close comparison of gene expression profiles and epigenetic states indicates that naïve hESC protocols, to date, remain distinct from naïve mESCs (Yang et al., 2019b), suggesting that the current protocols only capture aspects of the naïve state. As the conversion of primed to naïve cells is an artificial conversion, not only are potent epigenetic barriers in place, but the route to true naïve cells is unclear (**Figure 2**). Another, perhaps uncomfortable, possibility is that the naïve mouse state has no clear mimic in humans, is species-specific, or is a transitory stage in humans that cannot be stably captured *in vitro* (Rossant and Tam, 2017; Yang et al., 2019b).

EPIGENETIC CONTROL IN REPROGRAMMING SOMATIC CELLS TO PLURIPOTENT STEM CELLS

Interconversions between closely related embryonic states have helped inform our understanding of the epigenetic control of embryogenesis. Another, more drastic, cell conversion is the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) or by SCNT (Dean et al., 2003; Takahashi and Yamanaka, 2006). These two techniques have revolutionized the study of epigenetic regulation of the embryonic state, particularly in humans where early embryogenesis is hard to study. Both reprogramming techniques involve the global reconfiguration of gene expression patterns driven by epigenetic remodeling (Liu et al., 2020), and these methods have revealed

potent epigenetic barriers that restrain cell type conversions (Xu et al., 2016). Chromatin is dramatically reorganized during reprogramming (Wang et al., 2017), as enhancer–promoter interactions and active and repressive sequences make new contacts and reorder the transcriptional program (Apostolou et al., 2013; Di Stefano et al., 2020; Lu et al., 2020b). During somatic cell reprogramming, the ectopic pluripotency transgenes *Oct4*, *Sox2*, *Klf4*, and *Myc* reconnect target enhancers to promoters to induce transcriptional change (Wei et al., 2013; Beagan et al., 2016; Stadhouders et al., 2018). More importantly, during this process, chromatin reorganization occurs prior to, or independent from, gene expression changes (Wei et al., 2013; Beagan et al., 2016). Indeed, changes in chromatin accessibility often precede changes in gene expression, often by several days (Li et al., 2017a).

Broadly, somatic cells tend to have higher levels of repressive marks, which are reduced during the reprogramming to pluripotency. Vitamin C improves the reprogramming of somatic cells to pluripotency by modulating TET and JMJ domain-containing proteins (Wang et al., 2011; Chen et al., 2013a), leading to demethylation of DNA and H3K36me2/3. Other repressive epigenetic marks have been identified as major barriers for reprogramming (Arabaci et al., 2020), particularly the repressive histone modification H3K9me3 that is redistributed during iPSC reprogramming (Hawkins et al., 2010). Methyltransferases are downstream targets of BMPs and are a determinant for iPSC generation by regulating the methylation states at core pluripotency loci (Chen et al., 2013b). Similarly, the loss of the H3K9 methyltransferase *Setdb1*, or its co-factor *Trim28*, leads to improved reprogramming (Miles et al., 2017), although it may ultimately be deleterious as it causes spontaneous differentiation in the resulting iPSCs (Klimczak et al., 2017). Other H3K9me3 enzymes also impair the conversion of somatic cells to iPSCs, including *Suv39h1/2* and *Ehmt2* (G9a), along with the H3K79me3 methyltransferase *Dot1l* (Onder et al., 2012). However, the regulation of repressive histones is more subtle than just repressive mechanisms are bad for reprogramming. Reprogramming is a multi-phased program (Brambrink et al., 2008), and one of the earliest phases is the large-scale suppression of the somatic gene expression program (Chronis et al., 2017; Li et al., 2017a). Whether an epigenetic enzyme is beneficial or deleterious for reprogramming may ultimately depend upon the balance of its role in suppression or activation of the somatic and pluripotent programs (**Figure 3**). Outsize roles in suppression of the somatic or activation of the pluripotent program are likely to improve reprogramming, while the opposite is likely to shift the balance toward impairment. Consequently, epigenetic regulators have context-specific and temporal-specific effects during reprogramming. For example, knockdown of the co-repressors *Ncor1*/*Ncor2* is deleterious for the early stages of reprogramming, due to reduced somatic gene suppression, but beneficial for the late stages due to reduced pluripotent gene repression (Zhuang et al., 2018). A similar pattern was observed for the histone H3K27 demethylase *Kdm6b* (JMJD3) (Huang et al., 2020) and the H3K27me3 methyltransferase *Ezh2* (Rao et al., 2015). Other epigenetic regulators can be beneficial in both phases, although they may use different mechanisms to achieve this effect. For example, the H2AK119 ubiquitinase RYBP cooperates with PRC1 complex to suppress the somatic program via the histone demethylase KDM2B but cooperates with OCT4 to activate the pluripotent program (Li et al., 2017b). Epigenetic regulators can be something of a double-edged sword for somatic cell reprogramming (Onder et al., 2012; Li et al., 2017b; Zhuang et al., 2018).

A curious observation in studies of epigenetic reprogramming is that the loss of epigenetic regulators can have strong effects on the reprogramming process (Xu et al., 2016), yet the loss of the same factors in ESCs tends to have a much weaker or negligible effect. For example, knockdown of *Ncor1/2* improves reprogramming but does not affect mESCs (Zhuang et al., 2018). Similarly, *Sin3a/Sap30* loss impairs reprogramming (Li et al., 2017a; Saunders et al., 2017) but causes no change in mESCs. The short-term knockdown of a panel of 40 epigenetic regulators resulted in differentiation in only two knockdowns (He et al., 2019). These observations indicate that epigenetic pathways involved in the establishment of pluripotency may not always be involved in the maintenance of pluripotency or, if they are involved, can often be in contradictory ways. A good example is the knockdown of the PRC2 component *Ezh2*. Knockdown promotes reprogramming (Onder et al., 2012), but in mESCs, it affects self-renewal and makes the cells prone to differentiation. Overall, this suggests two patterns for epigenetic regulators in reprogramming and pluripotent maintenance: (1) once the reprogramming epigenetic barriers have been overcome, they are dispensable in ESCs, and (2) loss of epigenetic regulators makes ESCs unstable and more prone to differentiation. Epigenetic barriers are not always two-way and often act more like valves that can be easily traversed again if going in the opposite direction. This effect puts a limitation on screening technologies such as genome-wide knockdowns/outs and sgRNA screens. Epigenetic factors that impact the stable ESC state may not be relevant to the entirety of the reprogramming process. Consequently, challenging experiments must be performed during the reprogramming time course to understand the requirements for the temporal order of events. So far, these screens have focused on the early stages of reprogramming, which is more experimentally tractable (Borkent et al., 2016; Toh et al., 2016; Miles et al., 2017; Neganova et al., 2019). However, reprogramming is a curiously multistage-phased process, and epigenetic barriers may be transiently erected and disassembled, meaning that screening also needs to be timed to specific stages. Potentially, technologies such as Perturb-seq (Dixit et al., 2016), which provides candidate factors and phenotype readout simultaneously, may help in understanding the full range of epigenetic barriers blocking reprogramming.

TRANSPOSABLE ELEMENTS, THE EARLY EMBRYO, AND NAÏVE AND PRIMED EMBRYONIC STATES

Transposable elements are the single largest constituent of mammalian genomes (Hutchins and Pei, 2015), taking up around 40% of the total DNA sequence. They can be divided

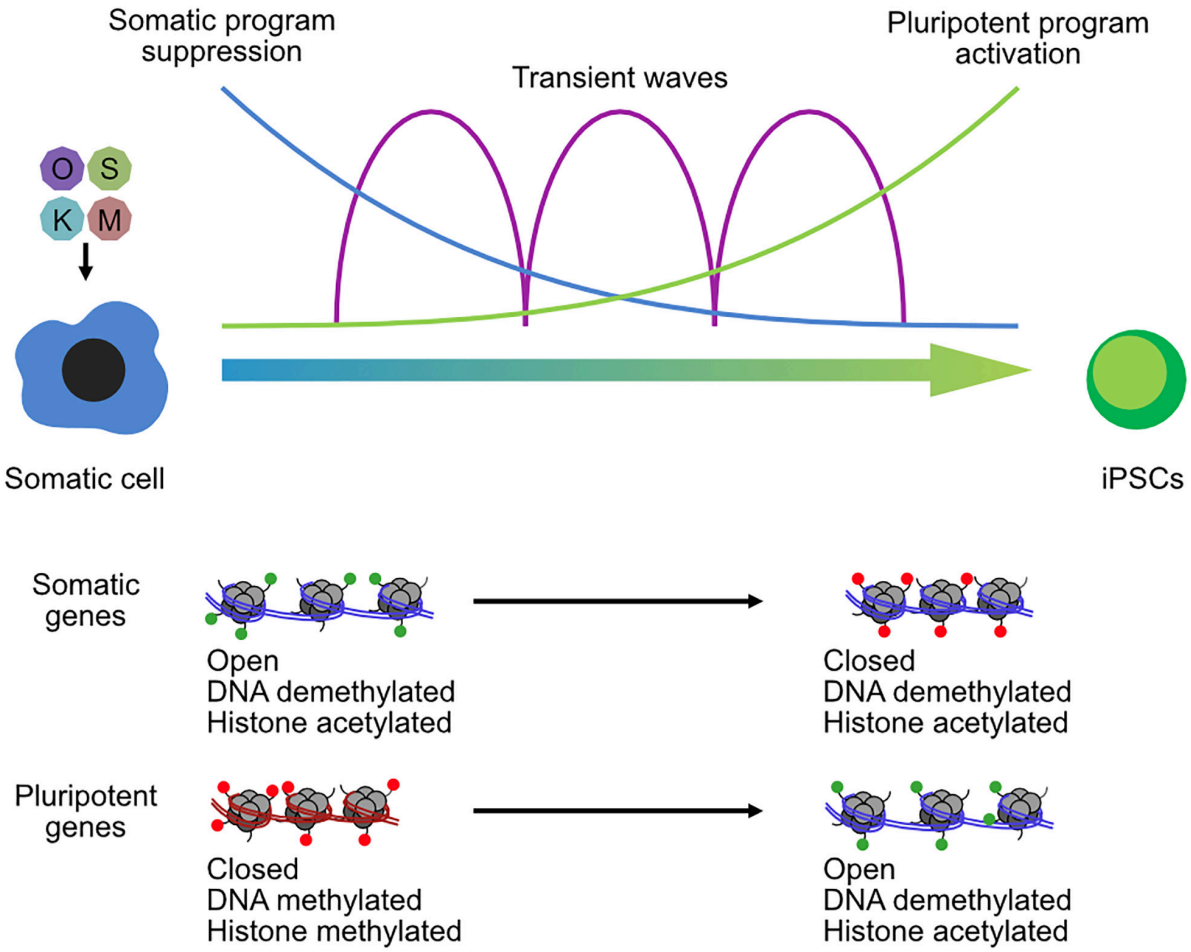


FIGURE 3 | Chromatin reconfiguration in the reprogramming of somatic cells to induced pluripotent stem cells. OSKM transgenes are transfected into somatic cells and initiate a complex series of biological programs, including the suppression of the somatic program in the early phase of reprogramming and the activation of the pluripotent program in the late phase. Transient waves of gene expression programs occur between the two-cell type states. Chromatin is remodeled throughout the reprogramming process, and especially at somatic loci and pluripotent gene loci that open and close chromatin and lose or gain histone methylation.

into four broad categories, DNA transposons, and three types of retrotransposon: long-interspersed elements (LINEs), short interspersed elements (SINEs), and the endogenous retroviruses/long-terminal repeats (ERVs/LTRs). TEs have been viewed as genetic parasites that are especially dangerous during embryogenesis when transposition duplications are capable of entering the germline. However, TEs can act as a source of evolutionary innovation, by duplicating TF binding sites, rewiring gene regulatory networks, altering splicing patterns, and many other effects on the genome and cell (Bourque et al., 2018), both beneficial and deleterious (Enriquez-Gasca et al., 2020). During early embryonic development, the genome is reprogrammed back to a naïve state. This process involves the global DNA demethylation of the genome, a process that is presumed to be a requirement for the correct execution of a new developmental program. However, DNA methylation is also one of the dominant mechanisms for the suppression of TEs in somatic tissues (Feng et al., 2010; Jonsson et al., 2019), and global DNA demethylation in the early embryo helps release waves of TE expression (Beraldi et al., 2006; Goke et al., 2015). Intriguingly, TE activity is dynamic in early embryonic cells (Wang et al., 2020) and is both stage and TE type-specific (Goke et al., 2015). In the early embryo and ESCs, instead of DNA methylation suppressing TE expression, other mechanisms take over, particularly the methylation of histone H3K9me3 by SETDB1, which is recruited to specific TEs by TRIM28 binding to KRAB-family zinc finger TFs (Ecco et al., 2016). However, there is also evidence that a wide range of epigenetic enzymes are involved in the suppression (or management) of the expression of TEs. Indeed, the early embryo can contain vast quantities of TE RNAs, a single MaLR LTR family TE can comprise up to 13% of the total oocyte RNA (Peaston et al., 2004), and SINE elements may make up a further 3% (Bachvarova, 1988). Functional roles for TEs in embryogenesis are less well explored, but LINE L1 expression is required for progression to the blastocyst stage (Percharde et al., 2018), while depletion of L1s in mESCs leads to the derepression of genes that are proximal to LINE L1s (Lu et al., 2020a). HERVKs are expressed and produce viroid-like particles

in normal human embryos (Grow et al., 2015). Multiple lines of evidence indicate that TE expression and epigenetic activity, aside from retrotransposition, are involved in the embryonic process, although their full involvement, both beneficial and deleterious, is unclear.

Transposable Elements and Two-Cell-Like Cells

TE expression has found utility as both a marker of embryonic stages and also as a tool to isolate new cell types with enhanced features. In mESC cultures, there is a small subpopulation (about 1%) that expresses a mouse-specific MERVL ERV (Macfarlan et al., 2012). Intriguingly, MERVL is expressed in the 2C stage of the mouse embryo when the cells are still totipotent. Isolation of the MERVL expressing "2C-like" cells from an mESC culture resulted in a population of cells that cycle into and out of the mESC state and have some totipotent-like properties. For example, they can partially colonize the embryonic trophoblast, a capability that normal mESCs lack. The 2C-like cells are transient and cannot be maintained, but various cocktails and protocols have been developed that improve their derivation (Iturbide and Torres-Padilla, 2020). Mechanistically, 2C-like cells rely on a transcriptional network distinct from the OCT4–SOX2 pluripotency network. The details are still being worked out, but the 2C-like state centers around several families of TFs and microRNAs, including miR344, DPPA2/4, ZSCAN4-family, ZMYM2, NELFA, and GATA2 (De Iaco et al., 2019; Fu et al., 2019; Zhang et al., 2019; Hu et al., 2020; Yang et al., 2020). These pathways ultimately center on the expression of *Dux* (**Figure 4A**; Percharde et al., 2018), a key TF required for ZGA in the developing embryo (De Iaco et al., 2017). This 2C-like transcriptional network cooperates to remodel the epigenetic state, H3K9me3 and H2AK119ub1 (ubiquitination), and chromatin assembly by CAF1, which are all particularly critical (Ishiuchi et al., 2015). MERVLs themselves tend to lack H3K9me3 and are not bound by SETDB1 (Maksakova et al., 2013) but are marked by H3K9me2 and H3K56ac (Macfarlan et al., 2011; He et al., 2019). Knockdown of *Ehmt2* (G9a) leads to the upregulation of MERVLs by a direct mechanism involving loss of H3K9me2 at MERVLs and the gain of open accessible chromatin (**Figure 4B**; Maksakova et al., 2013; Hendrickson et al., 2017). Similarly, *Kdm1a*, a histone demethylase, is important for suppressing the expression of MERVLs (Macfarlan et al., 2011), although the exact mechanism by which KDM1A suppresses MERVLs is not clear, as KDM1A can demethylate both H3K4me1 and H3K9me2. H3K4me1 marks enhancers and is generally associated with gene activation, while H3K9me2 is a repressive mark, often associated with heterochromatin and H3K9me3. H3K9me2 marks MERVLs, but H3K4me1 does not, suggesting that KDM1A may at least partially regulate MERVLs indirectly. Indeed, several direct and indirect chromatin modifiers regulate MERVLs. Both TRIM28 and RNF2 do not bind directly to MERVLs, and their corresponding marks, H3K9me3 and H2AK119ub, are not found either (Maksakova et al., 2013; He et al., 2019), but knockdown of *Trim28* or *Rnf2* leads to upregulation of MERVLs. Curiously, histone ubiquitination has two roles in both suppressing and activating MERVLs: knockdown of the histone H2A ubiquitinase *Rnf2* leads to the deubiquitination and upregulation of MERVLs (Zhang et al., 2019), while conversely, knockdown of the H2B deubiquitinase *Usp7* leads to H2B ubiquitination and upregulation of MERVLs (Chen et al., 2020). There is a similar pattern here to the conversion of primed cells to naïve cells: generally, the loss of repressive histone marks is beneficial for the conversion of cells to a 2C-like state. For example, knockdown of *Setdb1* (Wu et al., 2020), *Trim28* (KAP1) (Maksakova et al., 2013), *Dnmt1* (Fu et al., 2019), and *Rnf2* (RING1B) (He et al., 2019; Zhang et al., 2019) and inhibition of HDACs by trichostatin A (TSA) (Macfarlan et al., 2012) can all increase the number of 2C-like cells in an mESC culture.

TE activity, both as transcribed RNAs and as enhancers, is linked with the 2C-like state. In oocytes, a key factor is DPPA3 (*Dppa3*/*Stella*), a maternally inherited protein that is essential for the transition from the maternal to the zygotic gene expression program (Huang et al., 2017). When *Dppa3*/DPPA3 was removed from the maternal pool, MERVLs failed to be upregulated, and the embryos arrested at the 2C stage (Huang et al., 2017). Intriguingly, the authors found that microinjection of small interfering RNAs (siRNAs) targeting MERVLs led to a reduction in MERVL-derived Gag proteins and developmental impairment (Huang et al., 2017). These results suggest that MERVL expression is not simply a marker for the 2C stage but is also functionally relevant. MERVL sequences are spliced into other transcripts as TE–gene chimeras (Huang et al., 2017; Chen et al., 2020), and MERVL expression may also be driving transcript expression. MERVL sequences can act as an enhancer to recruit TFs to promote transcription (Huang et al., 2017; Zhang et al., 2019), and the MERVL sequence can act as a promoter (Jiang et al., 2020). In addition to MERVLs, LINE L1 RNAs silence *Dux* expression by recruiting TRIM28 to induce heterochromatin via H3K9me3 (Percharde et al., 2018). Knockdown of LINE L1 RNAs leads to reactivation of the 2C-like gene expression program and particularly reactivation of Dux (Percharde et al., 2018). This points to a surprisingly complex relationship between TE expression, transcriptional regulation, and epigenetic control of heterochromatin. LINE L1 RNAs can reactivate *Dux*, which then appears to lead to deregulation of H3K9me3, which activates MERVLs, which are spliced into key 2C-like transcripts and may also act as enhancers for genes required for the 2C-like state. Ultimately, the causal relationship between TE activation, 2C-like gene expression programs, and transcriptional and epigenetic control still needs to be unpicked, but it is a fascinating model system for the establishment of totipotency.

Transposable Elements and Human Naïve Cells

The expression of TEs has also found utility as markers of the embryonic state in hESCs (Theunissen et al., 2016). hESCs express the primate-specific HERVH ERV, and their accompanying LTR (LTR7) can act as pluripotent-specific transcription start sites (Fort et al., 2014). This pattern of TE

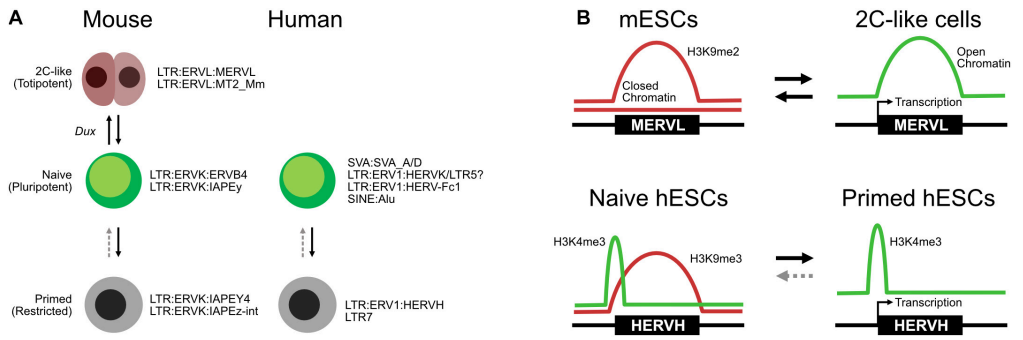


FIGURE 4 | Transposable elements (TEs) in totipotent, naïve, and primed embryonic states. **(A)** Expression of specific TEs mark cell type states in both mice and humans, although the TEs involved are species-specific. **(B)** Selected chromatin transitions at specific MERVLs in mouse embryonic stem cells (mESCs) during the transition to two-cell (2C)-like cells and at HERVHs in naïve and primed human ESCs (hESCs).

expression has been proposed as one of several criteria that define the naïve and primed hESC states (Theunissen et al., 2016). Briefly, primed hESCs express HERVH/LTR7 RNAs, while naïve cells express a more mixed set of TEs, but particularly SVAs, LTR5, and HERVK. HERVH are marked by H3K4me3 in both naïve and primed cells but are typically marked by the repressive H3K9me3 mark in primed cells (**Figure 4B**; Theunissen et al., 2016), although another study suggests that high levels of HERVH specifically mark naïve hESCs (**Figure 4A**; Wang et al., 2014). An interesting observation of that study was the splicing of HERVH directly into hESC chimeric transcripts. This is similar to the splicing of MERVLs seen in 2C-like cells, suggesting that a robust understanding of TE splicing patterns may lead to insights into embryonic cell states.

Transposable elements can be expressed as fragmentary RNA, and an area that remains poorly explored is the splicing of TEs into other transcripts. TEs can be expressed as individual units within the cell, but they can be spliced into longer transcripts, often as part of long non-coding transcripts but also into normal coding transcripts to generate novel chimeric transcripts (Bourque et al., 2018). To date, exploring the contribution of TEs to the normal transcriptome of a cell has been hampered by the use of short-reads to assemble transcripts (Babarinde et al., 2019). Nonetheless, the chimeric splicing of TEs into transcripts is a feature of pre-implantation embryonic cell types. The mouse-specific MERVLs that are transcribed in 2C-like cells are spliced into other coding and non-coding transcripts (Macfarlan et al., 2012; Huang et al., 2017; He et al., 2019). Similarly, the HERVH human-specific ERVs that are a feature of pluripotent stem cells are also spliced into other transcripts (Fort et al., 2014; Wang et al., 2014). Intriguingly, TEs are spliced into pluripotent transcripts in cancerous cells (Jang et al., 2019), the implication being that these TEs are activating pluripotent genes and converting them to oncogenes. However, HERVH activation appears not to be a general feature of cancer (Zapatka et al., 2020).

When TEs are still retrotranspositionally active, it poses a danger to the cell; however, once the coding sequences are mutated, and they are no longer functional, and epigenetic suppression mechanisms should decline due to a lack of evolutionary pressure to suppress TEs. Yet TEs maintain complex epigenetic regulatory patterns that are TE-type specific and are present long after they have stopped being capable of retrotransposition and are several million years old (Bourque et al., 2008; He et al., 2019). This suggests regulatory function and co-option for legitimate biological function. A good example is H3K9me3, a critical epigenetic mark responsible for silencing TEs in mESCs (Rowe et al., 2013; Yang et al., 2015), which is intimately involved in 2C-like cells, naïve cells, and reprogramming (Chen et al., 2013b; Bao et al., 2015; Xiao et al., 2017; Wang et al., 2018; Wu et al., 2020). H3K9me3 is remodeled during embryonic development, particularly at LTRs and ERVs (Wang et al., 2018). Knockdown of several H3K9me3-related factors, *Setdb1* and *Trim28*, impaired mouse embryonic development to the blastocyst, but *Chaf1a* (a modulator of H3K9me3 and part of the CAF1 complex) knockdown nearly completely blocked embryos from progressing past the morula stage (Wang et al., 2018). Consequently, H3K9me3 seems to be performing double duty as a major repressive mark for LTRs and ERVs, as well as erecting epigenetic barriers between cell type conversions. Ultimately, there is a tight integration between epigenetic control of TE activity and cell fate, and they should be considered as a unified mechanism with overlapping activities.

CONCLUSION

Epigenetic reconfiguration during early embryonic development is a critical process that resets the cells and makes them capable of a new round of development. The epigenetic rearrangements on chromatin are widespread and encompass changes in histone modifications, nucleosome positioning, 3D structure, and DNA modifications. A complex system of epigenetic regulators is involved in this process, and there are many distinct stages that cells transition through during normal development. Some of these states can be captured *in vitro* and have informed our understanding of the mechanisms behind embryonic development, and how autonomous and exogenous signaling, transcriptional control, and epigenetics combine to regulate development. Many mysteries remain, particularly in the role of epigenetic control in maintaining and blocking cell type conversions. Understanding this process in detail will

lead to an enhanced understanding of cell type transitions that will inform potential medical treatments, particularly cell replacement therapy.

AUTHOR CONTRIBUTIONS

LS, XF, and GM compiled the tables and prepared the figures. All authors were involved in manuscript writing. AH wrote and approved the final text, and funded the study.

FUNDING

This work was supported by the Science and Technology Planning Project of Guangdong Province (2019A050510004), the National Natural Science Foundation of China (31970589 and 31850410463), the Shenzhen Science, Technology and Innovation Commission (JCYJ20200109141018712), the Shenzhen Peacock plan (201701090668B), and the Center for Computational Science and Engineering of the Southern University of Science and Technology.

REFERENCES

Acampora, D., Di Giovannantonio, L. G., and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* 140, 43–55. doi: 10.1242/dev.085290

Adachi, K., Kopp, W., Wu, G., Heising, S., Greber, B., Stehling, M., et al. (2018). Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. *Cell Stem Cell* 23, 266–275e266. doi: 10.1016/j.stem.2018.05.020

Alonso, A., Breuer, B., Steuer, B., and Fischer, J. (1991). The F9-EC cell line as a model for the analysis of differentiation. *Int. J. Dev. Biol.* 35, 389–397.

Andrews, P. W. (1988). Human teratocarcinomas. *Biochim. Biophys. Acta* 948, 17–36. doi: 10.1016/0304-419x(88)90003-0

Apostolou, E., Ferrari, F., Walsh, R. M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., et al. (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* 12, 699–712. doi: 10.1016/j.stem.2013.04.013

Arabaci, D. H., Terzioglu, G., Bayirbasi, B., and Onder, T. T. (2020). Going up the Hill: Chromatin-based Barriers to Epigenetic Reprogramming. *FEBS J.* 2020:15628. doi: 10.1111/febs.15628

Atlasi, Y., Megchelenbrink, W., Peng, T., Habibi, E., Joshi, O., Wang, S. Y., et al. (2019). Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat. Cell Biol.* 21, 568–578. doi: 10.1038/s41556-019-0310-9

Avgustinova, A., and Benitah, S. A. (2016). Epigenetic control of adult stem cell function. *Nat. Rev. Mol. Cell Biol.* 17, 643–658. doi: 10.1038/nrm.2016.76

Babarinde, I. A., Li, Y., and Hutchins, A. P. (2019). Computational Methods for Mapping, Assembly and Quantification for Coding and Non-coding Transcripts. *Comput. Struct. Biotechnol. J.* 17, 628–637. doi: 10.1016/j.csbj.2019.04.012

Bachvarova, R. (1988). Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. *Dev. Biol.* 130, 513–523. doi: 10.1016/0012-1606(88)90346-6

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., et al. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* 461, 1292–1295. doi: 10.1038/nature08534

Bao, X., Wu, H., Zhu, X., Guo, X., Hutchins, A. P., Luo, Z., et al. (2015). The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res.* 25, 80–92. doi: 10.1038/cr.2014.165

Beagan, J. A., Gilgenast, T. G., Kim, J., Plona, Z., Norton, H. K., Hu, G., et al. (2016). Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* 18, 611–624. doi: 10.1016/j.stem.2016.04.004

Beattie, G. M., Lopez, A. D., Bucay, N., Hinton, A., Firpo, M. T., King, C. C., et al. (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem Cells* 23, 489–495. doi: 10.1634/stemcells.2004-0279

Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E., and Spadafora, C. (2006). Expression of LINE-1 retroposons is essential for murine preimplantation development. *Mol. Reprod. Dev.* 73, 279–287. doi: 10.1002/mrd.20423

Betschinger, J. (2017). Charting Developmental Dissolution of Pluripotency. *J. Mol. Biol.* 429, 1441–1458. doi: 10.1016/j.jmb.2016.12.017

Betschinger, J., Nichols, J., Dietmann, S., Corrin, P. D., Paddison, P. J., and Smith, A. (2013). Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* 153, 335–347. doi: 10.1016/j.cell.2013.03.012

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557–572e524. doi: 10.1016/j.cell.2017.09.043

Borkent, M., Bennett, B. D., Lackford, B., Bar-Nur, O., Brumbaugh, J., Wang, L., et al. (2016). A Serial shRNA Screen for Roadblocks to Reprogramming Identifies the Protein Modifier SUMO2. *Stem Cell Rep.* 6, 704–716. doi: 10.1016/j.stemcr.2016.02.004

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* 16, 516–528. doi: 10.1038/ncb2965

Boskovic, A., Eid, A., Pontabry, J., Ishiuchi, T., Spiegelhalter, C., Raghu Ram, E. V., et al. (2014). Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* 28, 1042–1047. doi: 10.1101/gad.238881.114

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19:199. doi: 10.1186/s13059-018-1577-z

Bourque, G., Leong, B., Vega, V. B., Chen, X., Lee, Y. L., Srinivasan, K. G., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762. doi: 10.1101/gr.080663.108

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., et al. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* 2, 151–159. doi: 10.1016/j.stem.2008.01.004

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B. W., Lopes, S. M. C. D., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, U191–U197. doi: 10.1038/nature05950

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., et al. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838–853. doi: 10.1016/j.stem.2014.04.003

Cai, L., Sutter, B. M., Li, B., and Tu, B. P. (2011). Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol. Cell* 42, 426–437. doi: 10.1016/j.molcel.2011.05.004

Cao, K. X., Collings, C. K., Morgan, M. A., Marshall, S. A., Rendleman, E. J., Ozark, P. A., et al. (2018). An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Sci. Adv.* 4:eaa8747. doi: 10.1126/sciadv.aap8747

Carey, B. W., Finley, L. W., Cross, J. R., Allis, C. D., and Thompson, C. B. (2015). Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* 518, 413–416. doi: 10.1038/nature13981

Carter, M. G., Smagghe, B. J., Stewart, A. K., Rapley, J. A., Lynch, E., Bernier, K. J., et al. (2016). A Primitive Growth Factor, NME7AB, Is Sufficient to Induce Stable Naive State Human Pluripotency; Reprogramming in This Novel Growth Factor Confers Superior Differentiation. *Stem Cells* 34, 847–859. doi: 10.1002/stem.2261

Cha, Y., Han, M. J., Cha, H. J., Zoldan, J., Burkart, A., Jung, J. H., et al. (2017). Metabolic control of primed human pluripotent stem cell fate and function by the miR-200c-SIRT2 axis. *Nat. Cell Biol.* 19, 445–456. doi: 10.1038/ncb3517

Chamberlain, S. J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* 26, 1496–1505. doi: 10.1634/stemcells.2008-0102

Chan, Y. S., Goke, J., Ng, J. H., Lu, X., Gonzales, K. A., Tan, C. P., et al. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* 13, 663–675. doi: 10.1016/j.stem.2013.11.015

Chason, R. J., Csokmay, J., Segars, J. H., DeCherney, A. H., and Armant, D. R. (2011). Environmental and epigenetic effects upon preimplantation embryo metabolism and development. *Trends Endocrinol. Metab.* 22, 412–420. doi: 10.1016/j.tem.2011.05.005

Chen, F., Zhang, W., Xie, D., Gao, T., Dong, Z., and Lu, X. (2020). Histone chaperone FACT represses retrotransposon MERVL and MERVL-derived cryptic promoters. *Nucleic Acids Res.* 48, 10211–10225. doi: 10.1093/nar/gkaa732

Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., et al. (2013a). Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* 45, 1504–1509. doi: 10.1038/ng.2807

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., et al. (2013b). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* 45, 34–42. doi: 10.1038/ng.2491

Christophorou, M. A., Castelo-Branco, G., Halley-Stott, R. P., Oliveira, C. S., Loos, R., Radzisheuskaya, A., et al. (2014). Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* 507, 104–108. doi: 10.1038/nature12942

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., et al. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* 168, 442–459e420. doi: 10.1016/j.cell.2016.12.016

Cornacchia, D., Zhang, C., Zimmer, B., Chung, S. Y., Fan, Y., Soliman, M. A., et al. (2019). Lipid Deprivation Induces a Stable, Naive-to-Primed Intermediate State of Pluripotency in Human PSCs. *Cell Stem Cell* 25, 120–136e110. doi: 10.1016/j.stem.2019.05.001

Davidson, K. C., Mason, E. A., and Pera, M. F. (2015). The pluripotent state in mouse and human. *Development* 142, 3090–3099. doi: 10.1242/dev.116061

De Clerck, L., Taelman, J., Popovic, M., Willems, S., Van der Jeught, M., Heindryckx, B., et al. (2019). Untargeted histone profiling during naive conversion uncovers conserved modification markers between mouse and human. *Sci. Rep.* 9:17240. doi: 10.1038/s41598-019-53681-6

De Iaco, A., Coudray, A., Duc, J., and Trono, D. (2019). DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep.* 20:201847382. doi: 10.15252/embr.201847382

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945. doi: 10.1038/ng.3858

Dean, W., Santos, F., and Reik, W. (2003). Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. *Semin. Cell Dev. Biol.* 14, 93–100. doi: 10.1016/s1084-9521(02)00141-6

Devreker, F., and Englert, Y. (2000). In vitro development and metabolism of the human embryo up to the blastocyst stage. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 92, 51–56. doi: 10.1016/s0301-2115(00)00425-5

Di Micco, R., Fontanals-Cirera, B., Low, V., Ntziachristos, P., Yuen, S. K., Lovell, C. D., et al. (2014). Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* 9, 234–247. doi: 10.1016/j.celrep.2014.08.055

Di Stefano, M., Stadhouders, R., Farabella, I., Castillo, D., Serra, F., Graf, T., et al. (2020). Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.* 11:2564. doi: 10.1038/s41467-020-16396-1

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167, 1853–1866e1817. doi: 10.1016/j.cell.2016.11.038

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336. doi: 10.1038/nature14222

Dodge, J. E., Kang, Y. K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. *Mol. Cell Biol.* 24, 2478–2486. doi: 10.1128/mcb.24.6.2478-2486.2004

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., et al. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547, 232–235. doi: 10.1038/nature23263

Duggal, G., Warrier, S., Ghimire, S., Broekaert, D., Van der Jeught, M., Lierman, S., et al. (2015). Alternative Routes to Induce Naive Pluripotency in Human Embryonic Stem Cells. *Stem Cells* 33, 2686–2698. doi: 10.1002/stem.2071

Durruthy-Durruthy, J., Wossidlo, M., Pai, S., Takahashi, Y., Kang, G., Omberg, L., et al. (2016). Spatiotemporal Reconstruction of the Human Blastocyst by Single-Cell Gene-Expression Analysis Informs Induction of Naive Pluripotency. *Dev. Cell* 38, 100–115. doi: 10.1016/j.devcel.2016.06.014

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., et al. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev. Cell* 36, 611–623. doi: 10.1016/j.devcel.2016.02.024

Enriquez-Gasca, R., Gould, P. A., and Rowe, H. M. (2020). Host Gene Regulation by Transposable Elements: The New, the Old and the Ugly. *Viruses* 12:12101089. doi: 10.3390/v12101089

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156. doi: 10.1038/292154a0

Factor, D. C., Corradin, O., Zentner, G. E., Saiakhova, A., Song, L., Chenoweth, J. G., et al. (2014). Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. *Cell Stem Cell* 14, 854–863. doi: 10.1016/j.stem.2014.05.005

Feng, S., Jacobsen, S. E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. *Science* 330, 622–627. doi: 10.1126/science.1190614

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., et al. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* 11, 477–490. doi: 10.1016/j.stem.2012.08.002

Fidalgo, M., Faiola, F., Pereira, C. F., Ding, J., Saunders, A., Gingold, J., et al. (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proc. Natl. Acad. Sci. U S A.* 109, 16202–16207. doi: 10.1073/pnas.1208533109

Finley, L. W. S., Vardhana, S. A., Carey, B. W., Alonso-Curbelo, D., Koche, R., Chen, Y., et al. (2018). Pluripotency transcription factors and Tet1/2 maintain Brd4-independent stem cell identity. *Nat. Cell Biol.* 20, 565–574. doi: 10.1038/s41556-018-0086-3

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C. A., Saxena, A., et al. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* 46, 558–566. doi: 10.1038/ng.2965

Fu, X., He, F., Li, Y., Shahveranov, A., and Hutchins, A. P. (2017). Genomic and molecular control of cell type and cell type conversions. *Cell Regen.* 6, 1–7. doi: 10.1016/j.cr.2017.09.001

Fu, X., Wu, X., Djekidel, M. N., and Zhang, Y. (2019). Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat. Cell Biol.* 21, 835–844. doi: 10.1038/s41556-019-0343-0

Gafni, O., Weinberger, L., Mansour, A. A., Manor, Y. S., Chomsky, E., Ben-Yosef, D., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature* 504, 282–286. doi: 10.1038/nature12745

Galonska, C., Ziller, M. J., Karnik, R., and Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell* 17, 462–470. doi: 10.1016/j.stem.2015.07.005

Gao, X., Nowak-Imialek, M., Chen, X., Chen, D., Herrmann, D., Ruan, D., et al. (2019). Establishment of porcine and human expanded potential stem cells. *Nat. Cell Biol.* 21, 687–699. doi: 10.1038/s41556-019-0333-2

Gatchalian, J., Malik, S., Ho, J., Lee, D. S., Kelso, T. W. R., Shokhirev, M. N., et al. (2018). A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat. Commun.* 9:5139. doi: 10.1038/s41467-018-07528-9

Goke, J., Lu, X., Chan, Y. S., Ng, H. H., Ly, L. H., Sachs, F., et al. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 16, 135–141. doi: 10.1016/j.stem.2015.01.005

Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221–225. doi: 10.1038/nature14308

Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development* 137, 3185–3192. doi: 10.1242/dev.052753

Guo, G., Yang, J., Nichols, J., Hall, J. S., Eyres, I., Mansfield, W., et al. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* 136, 1063–1069. doi: 10.1242/dev.030957

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., et al. (2014). The DNA methylation landscape of human early embryos. *Nature* 511, 606–610. doi: 10.1038/nature13544

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H., Matarese, F., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* 13, 360–369. doi: 10.1016/j.stem.2013.06.002

Hamazaki, N., Kyogoku, H., Araki, H., Miura, F., Horikawa, C., Hamada, N., et al. (2020). Reconstitution of the oocyte transcriptional network with transcription factors. *Nature* 589, 264–269. doi: 10.1038/s41586-020-3027-9

Hanaoka, K., Hayasaka, M., Noguchi, T., and Kato, Y. (1987). Viable Chimeras between Embryonal Carcinoma Cells and Mouse Embryos: Comparison of Aggregation and Injection Methods. *Dev. Growth Different.* 29, 263–270. doi: 10.1111/j.1440-169X.1987.00263.x

Hanna, J., Cheng, A. W., Saha, K., Kim, J., Lengner, C. J., Soldner, F., et al. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl. Acad. Sci. U S A.* 107, 9222–9227. doi: 10.1073/pnas.1004584107

Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6, 479–491. doi: 10.1016/j.stem.2010.03.018

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* 146, 519–532. doi: 10.1016/j.cell.2011.06.052

He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M. M., et al. (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.* 10:34. doi: 10.1038/s41467-018-08006-y

Hendrickson, P. G., Dorais, J. A., Grow, E. J., Whiddon, J. L., Lim, J. W., Wike, C. L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934. doi: 10.1038/ng.3844

Heurtier, V., Owens, N., Gonzalez, I., Mueller, F., Proux, C., Mornico, D., et al. (2019). The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. *Nat. Commun.* 10:1109. doi: 10.1038/s41467-019-09041-z

Horne, G. A., Stewart, H. J., Dickson, J., Knapp, S., Ramsahoye, B., and Chevassut, T. (2015). Nanog requires BRD4 to maintain murine embryonic stem cell pluripotency and is suppressed by bromodomain inhibitor JQ1 together with Lefty1. *Stem Cells Dev.* 24, 879–891. doi: 10.1089/scd.2014.0302

Hu, Z., Tan, D. E. K., Chia, G., Tan, H., Leong, H. F., Chen, B. J., et al. (2020). Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat. Cell Biol.* 22, 175–186. doi: 10.1038/s41556-019-0453-8

Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylicz, J. J., et al. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *Elife* 6:22345. doi: 10.7554/eLife.22345

Huang, Y., Zhang, H., Wang, L., Tang, C., Qin, X., Wu, X., et al. (2020). JMJD3 acts in tandem with KLF4 to facilitate reprogramming to pluripotency. *Nat. Commun.* 11:5061. doi: 10.1038/s41467-020-18900-z

Hutchins, A. P., and Pei, D. (2015). Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci. Bull.* 60, 1722–1733. doi: 10.1007/s11434-015-0905-x

Hutchins, A. P., Choo, S. H., Mistri, T. K., Rahmani, M., Woon, C. T., Ng, C. K., et al. (2013). Co-motif discovery identifies an Esrrb-Sox2-DNA ternary complex as a mediator of transcriptional differences between mouse embryonic and epiblast stem cells. *Stem Cells* 31, 269–281. doi: 10.1002/stem.1279

Hutchins, A. P., Yang, Z., Li, Y., He, F., Fu, X., Wang, X., et al. (2017). Models of global gene expression define major domains of cell type and tissue identity. *Nucleic Acids Res.* 45, 2354–2367. doi: 10.1093/nar/gkx054

Illich, D. J., Zhang, M., Ursu, A., Osorno, R., Kim, K. P., Yoon, J., et al. (2016). Distinct Signaling Requirements for the Establishment of ESC Pluripotency in Late-Stage EpiSCs. *Cell Rep.* 15, 787–800. doi: 10.1016/j.celrep.2016.03.073

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Boskovic, A., Ziegler-Birling, C., Rodriguez-Terrones, D., et al. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671. doi: 10.1038/nsmb.3066

Iturbide, A., and Torres-Padilla, M. E. (2020). A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming. *Curr. Opin. Genet. Dev.* 64, 26–30. doi: 10.1016/j.gde.2020.05.038

Jang, H. S., Shah, N. M., Du, A. Y., Dailey, Z. Z., Pehrsson, E. C., Godoy, P. M., et al. (2019). Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* 51, 611–617. doi: 10.1038/s41588-019-0373-3

Jiang, Q., Ang, J. Y. J., Lee, A. Y., Cao, Q., Li, K. Y., Yip, K. Y., et al. (2020). G9a Plays Distinct Roles in Maintaining DNA Methylation, Retrotransposon Silencing, and Chromatin Looping. *Cell Rep.* 33:108315. doi: 10.1016/j.celrep.2020.108315

Jonsson, M. E., Ludvik Brattas, P., Gustafsson, C., Petri, R., Yudovich, D., Pircs, K., et al. (2019). Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* 10:3182. doi: 10.1038/s41467-019-11150-8

Kalkan, T., Bornelöv, S., Mulas, C., Diamanti, E., Lohoff, T., Ralser, M., et al. (2019). Complementary Activity of ETV5, RBPJ, and TCF3 Drives Formative Transition from Naive Pluripotency. *Cell Stem Cell* 24, 785.e–801.e. doi: 10.1016/j.stem.2019.03.017

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., et al. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* 8, 676–687. doi: 10.1016/j.stem.2011.04.004

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* 170, 367–381e320. doi: 10.1016/j.cell.2017.06.029

Khoa, L. T. P., Tsan, Y. C., Mao, F., Kremer, D. M., Sajjakulnukit, P., Zhang, L., et al. (2020). Histone Acetyltransferase MOF Blocks Acquisition of Quiescence in Ground-State ESCs through Activating Fatty Acid Oxidation. *Cell Stem Cell* 27, 441–458e410. doi: 10.1016/j.stem.2020.06.005

Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G. G., et al. (2020). Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell* 2020:5. doi: 10.1016/j.stem.2020.11.005

Kisa, F., Shiozawa, S., Oda, K., Yoshimatsu, S., Nakamura, M., Koya, I., et al. (2017). Naive-like ESRRB(+) iPSCs with the Capacity for Rapid Neural Differentiation. *Stem Cell Rep.* 9, 1825–1838. doi: 10.1016/j.stemcr.2017.10.008

Klimczak, M., Czerwinska, P., Mazurek, S., Sozanska, B., Biecek, P., Mackiewicz, A., et al. (2017). TRIM28 epigenetic corepressor is indispensable for stable induced pluripotent stem cell formation. *Stem Cell Res.* 23, 163–172. doi: 10.1016/j.scr.2017.07.012

Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C., Ilicic, T., Henriksson, J., Natarajan, K. N., et al. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* 17, 471–485. doi: 10.1016/j.stem.2015.09.011

Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., et al. (2015). Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. *Cell Stem Cell* 16, 517–532. doi: 10.1016/j.stem.2015.03.002

Leeb, M., Dietmann, S., Paramor, M., Niwa, H., and Smith, A. (2014). Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* 14, 385–393. doi: 10.1016/j.stem.2013.12.008

Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* 20, 311–316. doi: 10.1038/nsmb.2510

Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., et al. (2017a). Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* 21, 819–833e816. doi: 10.1016/j.stem.2017.10.012

Li, H., Lai, P., Jia, J., Song, Y., Xia, Q., Huang, K., et al. (2017b). RNA Helicase DDX5 Inhibits Reprogramming to Pluripotency by miRNA-Based Repression of RYBP and its PRC1-Dependent and -Independent Functions. *Cell Stem Cell* 20:571. doi: 10.1016/j.stem.2017.03.014

Li, M., Yu, J. S. L., Tilgner, K., Ong, S. H., Koike-Yusa, H., and Yusa, K. (2018). Genome-wide CRISPR-KO Screen Uncovers mTORC1-Mediated Gsk3 Regulation in Naive Pluripotency Maintenance and Dissolution. *Cell Rep.* 24, 489–502. doi: 10.1016/j.celrep.2018.06.027

Liu, W., Stein, P., Cheng, X., Yang, W., Shao, N. Y., Morrisey, E. E., et al. (2014). BRD4 regulates Nanog expression in mouse embryonic stem cells and preimplantation embryos. *Cell Death Differ.* 21, 1950–1960. doi: 10.1038/cdd.2014.124

Liu, X., Ouyang, J. F., Rossello, F. J., Tan, J. P., Davidson, K. C., Valdes, D. S., et al. (2020). Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature* 586, 101–107. doi: 10.1038/s41586-020-2734-6

Lu, J. Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., et al. (2020a). Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep.* 30, 3296–3311e3295. doi: 10.1016/j.celrep.2020.02.048

Lu, L., Liu, X., Huang, W. K., Giusti-Rodriguez, P., Cui, J., Zhang, S., et al. (2020b). Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Mol. Cell* 79, 521–534e515. doi: 10.1016/j.molcel.2020.06.007

Lubitz, S., Glaser, S., Schaft, J., Stewart, A. F., and Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Mol. Biol. Cell* 18, 2356–2366. doi: 10.1091/mbc.e06-11-1060

Lynch, C. J., Bernad, R., Martinez-Val, A., Shahbazi, M. N., Nobrega-Pereira, S., Calvo, I., et al. (2020). Global hyperactivation of enhancers stabilizes human and mouse naive pluripotency through inhibition of CDK8/19 Mediator kinases. *Nat. Cell Biol.* 22, 1223–1238. doi: 10.1038/s41556-020-0573-1

Macfarlan, T. S., Gifford, W. D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., et al. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* 25, 594–607. doi: 10.1101/gad.2008511

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi: 10.1038/nature11244

Maksakova, I. A., Thompson, P. J., Goyal, P., Jones, S. J., Singh, P. B., Karimi, M. M., et al. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenet.Chromat.* 6:15. doi: 10.1186/1756-8935-6-15

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590–604. doi: 10.1016/j.cell.2012.03.026

Martin, G. R. (1980). Teratocarcinomas and mammalian embryogenesis. *Science* 209, 768–776. doi: 10.1126/science.6250214

Matsuda, K., Mikami, T., Oki, S., Iida, H., Andrabi, M., Boss, J. M., et al. (2017). ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. *Development* 144, 1948–1958. doi: 10.1242/dev.143479

Mayer, D., Stadler, M. B., Rittirsch, M., Hess, D., Lukonin, I., Winzi, M., et al. (2020). Zfp281 orchestrates interconversion of pluripotent states by engaging Ehmt1 and Zic2. *EMBO J.* 39:e102591. doi: 10.15252/embj.2019102591

Meacham, C. E., and Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature* 501, 328–337. doi: 10.1038/nature12624

Miles, D. C., de Vries, N. A., Gisler, S., Lieftink, C., Akhtar, W., Gogola, E., et al. (2017). TRIM28 is an Epigenetic Barrier to Induced Pluripotent Stem Cell Reprogramming. *Stem Cells* 35, 147–157. doi: 10.1002/stem.2453

Neganova, I., Cotts, L., Banks, P., Gassner, K., Shukurov, A., Armstrong, L., et al. (2019). Endothelial Differentiation G Protein-Coupled Receptor 5 Plays an Important Role in Induction and Maintenance of Pluripotency. *Stem Cells* 37, 318–331. doi: 10.1002/stem.2954

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* 4, 487–492. doi: 10.1016/j.stem.2009.05.015

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., et al. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930–944e922. doi: 10.1016/j.cell.2017.05.004

Ohbo, K., and Tomizawa, S. (2015). Epigenetic regulation in stem cell development, cell fate conversion, and reprogramming. *Biomol. Concepts* 6, 1–9. doi: 10.1515/bmc-2014-0036

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* 137, 571–584. doi: 10.1016/j.cell.2009.03.014

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* 303, 644–649. doi: 10.1126/science.1092727

Okamoto, I., Patrat, C., Thepot, D., Peynot, N., Fauque, P., Daniel, N., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* 472, 370–374. doi: 10.1038/nature09872

Okashita, N., Suwa, Y., Nishimura, O., Sakashita, N., Kadota, M., Nagamatsu, G., et al. (2016). PRDM14 Drives OCT3/4 Recruitment via Active Demethylation in the Transition from Primed to Naive Pluripotency. *Stem Cell Rep.* 7, 1072–1086. doi: 10.1016/j.stemcr.2016.10.007

Onder, T. T., Kara, N., Cherry, A., Sinha, A. U., Zhu, N., Bernt, K. M., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483, 598–602. doi: 10.1038/nature10953

Pastor, W. A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T. J., et al. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* 20, 553–564. doi: 10.1038/s41556-018-0089-0

Patrat, C., Ouimette, J. F., and Rougeulle, C. (2020). X chromosome inactivation in human development. *Development* 147:183095. doi: 10.1242/dev.183095

Peaston, A. E., Evsikov, A. V., Graber, J. H., de Vries, W. N., Holbrook, A. E., Solter, D., et al. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606. doi: 10.1016/j.devcel.2004.09.004

Percharde, M., Lin, C. J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., et al. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* 174, 391.e–405.e. doi: 10.1016/j.cell.2018.05.043

Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., et al. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* 167:285. doi: 10.1016/j.cell.2016.08.009

Posfai, E., Schell, J. P., Janiszewski, A., Rovic, I., Murray, A., Bradshaw, B., et al. (2021). Evaluating totipotency using criteria of increasing stringency. *Nat. Cell Biol.* 23, 49–60. doi: 10.1038/s41556-020-00609-2

Rao, R. A., Dhele, N., Cheemadan, S., Ketkar, A., Jayandharan, G. R., Palakodeti, D., et al. (2015). Ezh2 mediated H3K27me3 activity facilitates somatic transition during human pluripotent reprogramming. *Sci. Rep.* 5:8229. doi: 10.1038/srep08229

Rhodes, J. D. P., Feldmann, A., Hernandez-Rodriguez, B., Diaz, N., Brown, J. M., Fursova, N. A., et al. (2020). Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell Rep.* 30, 820–835e810. doi: 10.1016/j.celrep.2019.12.057

Rossant, J. (2011). The Impact of Developmental Biology on Pluripotent Stem Cell Research: Successes and Challenges. *Dev. Cell* 21, 20–23. doi: 10.1016/j.devcel.2011.06.010

Rossant, J., and Tam, P. P. L. (2017). New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* 20, 18–28. doi: 10.1016/j.stem.2016.12.004

Rostovskaya, M., Stirparo, G. G., and Smith, A. (2019). Capacitation of human naive pluripotent stem cells for multi-lineage differentiation. *Development* 146:172916. doi: 10.1242/dev.172916

Rowe, H. M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T. S., Tarabay, Y., et al. (2013). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* 23, 452–461. doi: 10.1101/gr.147678.112

Rowley, M. J., and Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800. doi: 10.1038/s41576-018-0060-8

Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T. W., et al. (2017). Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* 20, 87–101. doi: 10.1016/j.stem.2016.10.006

Saunders, A., Huang, X., Fidalgo, M., Reimer, M. H. Jr., Faiola, F., Ding, J., et al. (2017). The SIN3A/HDAC Corepressor Complex Functionally Cooperates with NANOG to Promote Pluripotency. *Cell Rep.* 18, 1713–1726. doi: 10.1016/j.celrep.2017.01.055

Schlesinger, S., and Meshorer, E. (2019). Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. *Dev. Cell* 48, 135–150. doi: 10.1016/j.devcel.2019.01.003

Schulz, E. G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., et al. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* 14, 203–216. doi: 10.1016/j.stem.2013.11.022

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., et al. (2017). PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naive ESCs. *Nat. Commun.* 8:672. doi: 10.1038/s41467-017-00668-4

Shyh-Chang, N., and Ng, H. H. (2017). The metabolic programming of stem cells. *Genes Dev.* 31, 336–346. doi: 10.1101/gad.293167.116

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., et al. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138, 722–737. doi: 10.1016/j.cell.2009.07.039

Sim, Y. J., Kim, M. S., Nayfeh, A., Yun, Y. J., Kim, S. J., Park, K. T., et al. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. *Stem Cell Rep.* 8, 1312–1328. doi: 10.1016/j.stemcr.2017.04.001

Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. *Development* 144, 365–373. doi: 10.1242/dev.142679

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., et al. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 336, 688–690. doi: 10.1038/336688a0

Smith, D. G., and Sturmey, R. G. (2013). Parallels between embryo and cancer cell metabolism. *Biochem. Soc. Trans.* 41, 664–669. doi: 10.1042/BST20120352

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., et al. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature* 511, 611–615. doi: 10.1038/nature13581

Sperber, H., Mathieu, J., Wang, Y., Ferreccio, A., Hesson, J., Xu, Z., et al. (2015). The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat. Cell Biol.* 17, 1523–1535. doi: 10.1038/ncb3264

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* 50, 238–249. doi: 10.1038/s41588-017-0030-7

Stuart, H. T., Stirparo, G. G., Lohoff, T., Bates, L. E., Kinoshita, M., Lim, C. Y., et al. (2019). Distinct Molecular Trajectories Converge to Induce Naive Pluripotency. *Cell Stem Cell* 25, 388–406e388. doi: 10.1016/j.stem.2019.07.009

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676. doi: 10.1016/j.cell.2006.07.024

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* 158, 1254–1269. doi: 10.1016/j.cell.2014.08.029

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., et al. (2015). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* 162, 452–453. doi: 10.1016/j.cell.2015.06.052

ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., et al. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.* 13, 1070–1075. doi: 10.1038/ncb2314

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., et al. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196–199. doi: 10.1038/nature05972

Teslaa, T., and Teitell, M. A. (2015). Pluripotent stem cell energy metabolism: an update. *EMBO J.* 34, 138–153. doi: 10.15252/embj.201490446

Theunissen, T. W., Friedli, M., He, Y., Planet, E., O'Neil, R. C., Markoulaki, S., et al. (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* 19, 502–515. doi: 10.1016/j.stem.2016.06.011

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145–1147. doi: 10.1126/science.282.5391.1145

Toh, C. X., Chan, J. W., Chong, Z. S., Wang, H. F., Guo, H. C., Satapathy, S., et al. (2016). RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming. *Cell Rep.* 15, 2597–2607. doi: 10.1016/j.celrep.2016.05.049

Tosolini, M., and Jouneau, A. (2016). From Naive to Primed Pluripotency: In Vitro Conversion of Mouse Embryonic Stem Cells in Epiblast Stem Cells. *Methods Mol. Biol.* 1341, 209–216. doi: 10.1007/7651_2015_208

van Mierlo, G., Dirks, R. A. M., De Clerck, L., Brinkman, A. B., Huth, M., Kloet, S. L., et al. (2019). Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell* 24, 123–137e128. doi: 10.1016/j.stem.2018.10.017

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., et al. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Mol. Cell* 62:983. doi: 10.1016/j.molcel.2016.06.005

Walter, M., Teissandier, A., Perez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* 5:11418. doi: 10.7554/eLife.11418

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., et al. (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* 20, 620–631. doi: 10.1038/s41556-018-0093-4

Wang, J., Huang, J., and Shi, G. (2020). Retrotransposons in pluripotent stem cells. *Cell Regen.* 9:4. doi: 10.1186/s13619-020-00046-4

Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Rasko, T., Szvetnik, A., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 516, 405–409. doi: 10.1038/nature13804

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., et al. (2011). The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* 9, 575–587. doi: 10.1016/j.stem.2011.10.005

Wang, Y., Bi, Y., and Gao, S. (2017). Epigenetic regulation of somatic cell reprogramming. *Curr. Opin. Genet. Dev.* 46, 156–163. doi: 10.1016/j.gde.2017.07.002

Ware, C. B., Nelson, A. M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E. C., et al. (2014). Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci. U S A.* 111, 4484–4489. doi: 10.1073/pnas.1319738111

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., et al. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* 13, 36–47. doi: 10.1016/j.stem.2013.05.010

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J. H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* 17, 155–169. doi: 10.1038/nrm.2015.28

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319. doi: 10.1016/j.cell.2013.03.035

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., et al. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 336, 684–687. doi: 10.1038/336684a0

Wu, K., Liu, H., Wang, Y., He, J., Xu, S., Chen, Y., et al. (2020). SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent States. *Cell Rep.* 30, 25–36e26. doi: 10.1016/j.celrep.2019.12.010

Xia, W., and Xie, W. (2020). Rebooting the Epigenomes during Mammalian Early Embryogenesis. *Stem Cell Rep.* 2020:5. doi: 10.1016/j.stemcr.2020.09.005

Xiao, S., Lu, J., Sridhar, B., Cao, X., Yu, P., Zhao, T., et al. (2017). SMARCAD1 Contributes to the Regulation of Naive Pluripotency by Interacting with Histone Citrullination. *Cell Rep.* 18, 3117–3128. doi: 10.1016/j.celrep.2017.02.070

Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T., and Thomson, J. A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat. Methods* 2, 185–190. doi: 10.1038/nmeth744

Xu, Y., Zhang, M., Li, W., Zhu, X., Bao, X., Qin, B., et al. (2016). Transcriptional Control of Somatic Cell Reprogramming. *Trends Cell Biol.* 26, 272–288. doi: 10.1016/j.tcb.2015.12.003

Yang, B. X., El Farran, C. A., Guo, H. C., Yu, T., Fang, H. T., Wang, H. F., et al. (2015). Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* 163, 230–245. doi: 10.1016/j.cell.2015.08.037

Yang, F., Huang, X., Zang, R., Chen, J., Fidalgo, M., Sanchez-Priego, C., et al. (2020). DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces

a Totipotent 2C-like State. *Cell Stem Cell* 26, 234–250e237. doi: 10.1016/j.stem.2020.01.004

Yang, J., Rajan, S. S., Friedrich, M. J., Lan, G., Zou, X., Ponstingl, H., et al. (2019a). Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming. *Stem Cell Rep.* 12, 757–771. doi: 10.1016/j.stemcr.2019.02.010

Yang, J., Ryan, D. J., Wang, W., Tsang, J. C., Lan, G., Masaki, H., et al. (2017a). Establishment of mouse expanded potential stem cells. *Nature* 550, 393–397. doi: 10.1038/nature24052

Yang, P., Humphrey, S. J., Cinghu, S., Pathania, R., Oldfield, A. J., Kumar, D., et al. (2019b). Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. *Cell Syst.* 8, 427–445e410. doi: 10.1016/j.cels.2019.03.012

Yang, S. H., Andrabi, M., Biss, R., Murtuza Baker, S., Iqbal, M., and Sharrocks, A. D. (2019c). ZIC3 Controls the Transition from Naive to Primed Pluripotency. *Cell Rep.* 27, 3215–3227e3216. doi: 10.1016/j.celrep.2019.05.026

Yang, Y., Liu, B., Xu, J., Wang, J., Wu, J., Shi, C., et al. (2017b). Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell* 169, 243–257e225. doi: 10.1016/j.cell.2017.02.005

Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., et al. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122, 881–894.

Yilmaz, A., and Benvenisty, N. (2019). Defining Human Pluripotency. *Cell Stem Cell* 25, 9–22. doi: 10.1016/j.stem.2019.06.010

Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 115, 281–292. doi: 10.1016/s0092-8674(03)00847-x

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., et al. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523. doi: 10.1038/nature06968

Yu, L., Wei, Y., Sun, H. X., Mahdi, A. K., Pinzon Arteaga, C. A., Sakurai, M., et al. (2020a). Derivation of Intermediate Pluripotent Stem Cells Amenable to Primordial Germ Cell Specification. *Cell Stem Cell* 2020:3. doi: 10.1016/j.stem.2020.11.003

Yu, S., Zhou, C., Cao, S., He, J., Cai, B., Wu, K., et al. (2020b). BMP4 resets mouse epiblast stem cells to naive pluripotency through ZBTB7A/B-mediated chromatin remodelling. *Nat. Cell Biol.* 22, 651–662. doi: 10.1038/s41556-020-0516-x

Zapatka, M., Borozan, I., Brewer, D. S., Iskar, M., Grundhoff, A., Alawi, M., et al. (2020). The landscape of viral associations in human cancers. *Nat. Genet.* 52, 320–330. doi: 10.1038/s41588-019-0558-9

Zhang, H., Gayen, S., Xiong, J., Zhou, B., Shanmugam, A. K., Sun, Y., et al. (2016). MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency. *Cell Stem Cell* 18, 481–494. doi: 10.1016/j.stem.2016.02.004

Zhang, K., Wu, D. Y., Zheng, H., Wang, Y., Sun, Q. R., Liu, X., et al. (2020a). Analysis of Genome Architecture during SCNT Reveals a Role of Cohesin in Impeding Minor ZGA. *Mol. Cell* 79, 234–250e239. doi: 10.1016/j.molcel.2020.06.001

Zhang, M., Lai, Y., Krupalnik, V., Guo, P., Guo, X., Zhou, J., et al. (2020b). beta-Catenin safeguards the ground state of mouse pluripotency by strengthening the robustness of the transcriptional apparatus. *Sci. Adv.* 6:eaba1593. doi: 10.1126/sciadv.aba1593

Zhang, W., Chen, F., Chen, R., Xie, D., Yang, J., Zhao, X., et al. (2019). Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res.* 47, 8485–8501. doi: 10.1093/nar/gkz594

Zhou, H., Li, W., Zhu, S., Joo, J. Y., Do, J. T., Xiong, W., et al. (2010). Conversion of mouse epiblast stem cells to an earlier pluripotency state by small molecules. *J. Biol. Chem.* 285, 29676–29680. doi: 10.1074/jbc.C110.150599

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., et al. (2012). HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* 31, 2103–2116. doi: 10.1038/emboj.2012.71

Zhuang, Q., Li, W., Benda, C., Huang, Z., Ahmed, T., Liu, P., et al. (2018). NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. *Nat. Cell Biol.* 20, 400–412. doi: 10.1038/s41556-018-0047-x

Zimmerlin, L., and Zambidis, E. T. (2020). Pleiotropic roles of tankyrase/PARP proteins in the establishment and maintenance of human naive pluripotency. *Exp. Cell Res.* 390:111935. doi: 10.1016/j.yexcr.2020.111935

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 23 February 2021
doi: 10.3389/fcell.2021.641410

TBL1XR1 Ensures Balanced Neural Development Through NCOR Complex-Mediated Regulation of the MAPK Pathway

Giuseppina Mastrototaro[1†‡], Mattia Zaghi[1‡], Luca Massimino[1], Matteo Moneta[1], Neda Mohammadi[2], Federica Banfi[1], Edoardo Bellini[1], Marzia Indrigo[1†], Giulia Fagnocchi[1], Anna Bagliani[3], Stefano Taverna[2], Maria Rohm[4,5,6,7], Stephan Herzig[4,5,6,7] and Alessandro Sessa[1]*

[1] Stem Cell and Neurogenesis Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy, [2] Neurimmunology Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy, [3] Medical Oncology Unit, ASST Ovest Milanese, Legnano Hospital, Legnano, Italy, [4] Institute for Diabetes and Cancer IDC, Helmholtz Center, Munich, Germany, [5] Joint Heidelberg-IDC Translational Diabetes Program, Inner Medicine 1, Heidelberg University Hospital, Heidelberg, Germany, [6] Medical Faculty, Technical University Munich, Munich, Germany, [7] German Center for Diabetes Research, Oberschleissheim, Germany

TBL1XR1 gene is associated with multiple developmental disorders presenting several neurological aspects. The relative protein is involved in the modulation of important cellular pathways and master regulators of transcriptional output, including nuclear receptor repressors, Wnt signaling, and MECP2 protein. However, TBL1XR1 mutations (including complete loss of its functions) have not been experimentally studied in a neurological context, leaving a knowledge gap in the mechanisms at the basis of the diseases. Here, we show that *Tbl1xr1* knock-out mice exhibit behavioral and neuronal abnormalities. Either the absence of TBL1XR1 or its point mutations interfering with stability/regulation of NCOR complex induced decreased proliferation and increased differentiation in neural progenitors. We suggest that this developmental unbalance is due to a failure in the regulation of the MAPK cascade. Taken together, our results broaden the molecular and functional aftermath of TBL1XR1 deficiency associated with human disorders.

Keywords: TBL1XR1, NCOR, brain development, neurodevelopmental disorders, MAPK

OPEN ACCESS

Edited by:
José Luis Sardina, Josep Carreras Leukaemia Research Institute (IJC), Spain

Reviewed by:
Cinzia Vicidomini, Massachusetts General Hospital and Harvard Medical School, United States
Ola Hermanson, Karolinska Institutet (KI), Sweden
Zheng Sun, Baylor College of Medicine, United States

***Correspondence:**
Alessandro Sessa
sessa.alessandro@hsr.it

†Present address:
Giuseppina Mastrototaro, Molmed S.p.A., Milan, Italy
Marzia Indrigo, Division of Neuroscience, RETT Research Center, IRCCS San Raffaele Scientific Institute, Milan, Italy

‡*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *14 December 2020*
Accepted: *04 February 2021*
Published: *23 February 2021*

Citation:
Mastrototaro G, Zaghi M, Massimino L, Moneta M, Mohammadi N, Banfi F, Bellini E, Indrigo M, Fagnocchi G, Bagliani A, Taverna S, Rohm M, Herzig S and Sessa A (2021) TBL1XR1 Ensures Balanced Neural Development Through NCOR Complex-Mediated Regulation of the MAPK Pathway. Front. Cell Dev. Biol. 9:641410. doi: 10.3389/fcell.2021.641410

INTRODUCTION

The building of a healthy nervous system is due to a concerted array of molecular mechanisms within and between different cell types during the embryonic life and beyond (Borrell, 2019). Not surprisingly, a huge variety of gene mutations is associated with neurodevelopmental disorders (Thapar et al., 2017). *TBL1XR1*, encoding for transducin β-like 1—related protein 1 (a member of HDAC containing NCOR/SMRT complexes) (Yoon et al., 2003), has been associated with different human developmental diseases, spanning from autism spectrum disorders (ASD) (O'Roak et al., 2012; Stessman et al., 2017), to West syndrome (Saitsu et al., 2014), schizophrenia (SCZ) (Nishi et al., 2017), intellectual disability (Pons et al., 2015; Riehmer et al., 2017). Mutations leading to the substitution Y446C of *TBL1XR1* are the only genetic cause of Pierpont syndrome

(Heinen et al., 2016). The *TBL1XR1* mutational spectrum is wide, ranging from deletions to duplications and point mutations, all inducing frameshifts or amino-acidic substitutions (Li et al., 2015). However, to our knowledge the hypothesis that mutations in *TBL1XR1* gene directly lead to neurological impairment—and whether this is due to the disruption of the function of NCOR complex—has not been investigated in high systems (e.g., mammals).

Co-repressors NCOR1 and NCOR2 (in mice NCOR and SMRT, respectively), contribute to different repressive pathways, including the inhibition of the downstream cascade of nuclear receptors such as the receptors of thyroid hormone (TRs) and retinoic acid (RARs), in the absence of their ligands (Kong et al., 2020). While the two proteins are very similar (sharing portions of sequence and domains), they play slightly different roles in different cell populations (Iemolo et al., 2020; Sun and Xu, 2020). For example, during brain development both factors are required for the proliferation capability of neural precursors, with NCOR1 loss specifically impacting on premature differentiation of astrocytes and oligondendroglia (Castelo-Branco et al., 2014), while the absence of NCOR2 leads to early differentiation to neurons and astrocytes (Hermanson et al., 2002; Jepsen et al., 2007). NCOR complexes are believed to mediate repression through the deacetylation of specific chromatin regions through a direct action of HDAC proteins, particularly HDAC3 (Zhuang et al., 2018). In this context, TBL1XR1 and its related member TBL1X—both WD40 repeat containing proteins—are important for both formation and dismantling of the NCOR complex. Indeed, they are required both for the repressive activity of the complex and the ubiquitination-dependent dismissal of the co-repressors in response to appropriate signals (e.g., the ligands of a nuclear receptor), or even when the co-repressors are missing (Perissi et al., 2004; Ishizuka and Lazar, 2005; Yoon et al., 2005; Li et al., 2015). Importantly, TBL1XR1 and TBL1X play either redundant or exclusive roles in these processes, depending on the contexts and/or the targets (Perissi et al., 2008). Both TBL1XR1, TBL1X, and the NCOR complexes are physically associated with the Rett syndrome's causative factor MECP2, further underlying the importance of these molecules for normal brain functioning (Ebert et al., 2013; Lyst et al., 2013; Kruusvee et al., 2017). Despite TBL1XR1 and the related complexes have been known for years (Perissi et al., 2010), their specific role in brain development and possible causality in the neurological aspects of human diseases have been only marginally explored.

In the present study, we describe how of the lack of TBL1XR1 affects mouse behavior, brain development and function. We show that TBL1XR1 loss of function impacts on coordination, memory, and sociability, similarly to what occurs in humans. Mutant neural progenitors proliferated less than normal, due to the lack of NCOR-mediated regulation of MAPK cascade. The reductionist loss-of-function approach was enriched by the usage of disease-specific mutant proteins for complementation experiments. This analysis indicated how point mutations are generally different from the *Tbl1xr1* knock-out (KO) condition, except for the F10L mutation that was retrieved in SCZ cases. Altogether, our data indicate a broad impact of TBL1XR1 for correct neuronal development and maturation, accounting for the complex and variegated spectrum of pathological traits associated with the different *TBL1XR1* mutations in humans.

MATERIALS AND METHODS

Mice

Tbl1xr1 floxed animals (Rohm et al., 2013) and CMV:Cre were maintained by crossing each other or with backcrossing with C57BL/6 animals at the San Raffaele Scientific Institute Institutional mouse facility. Experiments were performed in accordance with experimental protocols approved by local Institutional Animal Care and Use Committee (IACUC). Experimental subjects were sacrificed by means cervical dislocation. Both total body and brains were weighted to calculate their ratio.

Western Blot Analysis

Brain tissue and *in vitro* cells were prepared as previously indicated (Sessa et al., 2019) for western blot analysis. The following primary antibodies were used: anti-TBL1XR1 (Novus Biological NB600-270); anti-NCOR1 (Merck-Millipore ABE251); anti-HDAC3 (Abcam ab 13704); anti-GAPDH (Abcam ab8245); anti-βCATENIN (Chemicon AB19022); anti-pβCATENIN (Ser33/37/Thr41, Cell signaling #9561S); anti-ERK1 (Cell Signaling Technology 4372); anti-pERK1 (Cell Signaling Technology 5726); Anti-V5 (Thermo Fisher Scientific R960-25); Anti-Histone H3 (Abcam, ab1791).

Behavioral Testing

Animals were housed at a constant temperature of 23°C in a 12 h light/dark cycle (lights off at 7 PM), with food and water available *ad libitum*. We analyzed control and mutant littermates (males only) ranging from 2 to 4 months of age. The operator was blind to the genotype.

Rotarod

Mice were assessed for the latency to fall as previously described (Sessa et al., 2019).

Beam Walking

Beam crossings and number of paw slips were assessed as previously described (Sessa et al., 2019).

Open Field

Mice were located in a square arena (50 × 50 cm) and video-recorded for 10 min. Total distance traveled and the time spent near the walls were scored by EthoVisionXT software (Noldus Information Technology, Wageningen, Netherlands).

Catwalk

Motor function was estimated through the CatWalk system (Noldus Information Technology, Wageningen, Netherlands). The animal paw prints were recorded as the animal moved across a walkway with an illuminated glass floor fitted with high-speed video camera assembled with 8.5 mm wide-angle lenses below the floor. The day before the test, each animal was

placed on the CatWalk platform to walk freely as habituation. During the test, three uninterrupted runs (minimum of 5-step sequence patterns) were collected. Several gait parameters were calculated and analyzed by the dedicated software using the position, pressure and surface area of each paw footprint.

Marble Burying Test

Individual animals were placed in a cage (20 × 32cm) with 5 cm of bedding material and 12 marbles (12–15 mm in diameter). The marbles were placed in a 3 × 4 matrix. The number of buried marbles was counted after a 30 min session.

8-Arm Radial Maze

8-arm radial maze tests were conducted as previously described (Sessa et al., 2019). Days 4–13 are shown as experimental days 1–10 in **Figure 1F** and **Supplementary Figure S1D**. Working memory errors (**Supplementary Figure S1D**) were calculated as re-entries to arms where the pellet had already been consumed.

Elevated Plus Maze

Elevated plus maze was performed as described (Leo et al., 2014). Briefly, mice were placed in a plus-shaped maze composed by two enclosed and two open arms. The apparatus was elevated from the ground. The animals were allowed to freely explore the environment for 5 min and the time they spent in the open arms was calculated as an inverse measure of the levels of anxiety.

Social Interaction (Resident-Intruder Test)

To elicit social interactions, one mouse (experimental subject, resident) of a pair was placed in a neutral cage and allowed to freely explore for 10 min. At the end of the session, a second male mouse (intruder) of different strain was added to the cage and the behavior of the pair of animals was recorded for 20 min. The amount of time during which the resident mice engaged in social interactions (e.g., sniffing, following, grooming, biting, chasing, mounting, wrestling) was recorded on videotapes for each pair of mice. Since the test was originally designed for testing aggressiveness, we also evaluated the time that the animals spent in aggressive behavior.

3-Chamber Test

Adult mice were tested using a 3-chamber test coupled with the video tracking software Ethovision XT (Noldus) as previously described (Sessa et al., 2019).

Sniffing preference was assesses using a discrimination index (DI), i.e., the difference between the sniffing time in the occupied cage and the empty one, divided by total sniffing time.

Histology

Histological procedures to measure morphological parameters were conducted as previously described (Sessa et al., 2019).

Immunohistochemistry

Immunohistochemical analyses were performed as previously described (Colombo et al., 2006). The primary antibodies utilized were the following: anti-CTIP2 (Abcam ab18465), anti-CUX1 (santa Cruz Biotechnology sc-13024), anti-SATB2 (Abcam, ab51502), anti-PAX6 (Covance #PRB-278P), anti GFAP (chicken, 1:1,000, Abcam, ab4674), anti-DCX (rabbit, 1:1,000, Abcam, ab18723), antiKI67 (rabbit, 1:500, immunological sciences, mab-90948), anti-NEUN (rabbit, 1:500, Abcam, ab104225).

Secondary antibodies: 488-mouse (donkey, 1:2,000, Molecular Probes, A21202), 488-rabbit (donkey, 1:2,000, Molecular Probes, A21026), 594-mouse (donkey, 1:2,000, Molecular Probes, A10036), 594-rabbit (donkey, 1:2,000, Molecular Probes, A21207). DAPI (4′,6′-diamidino-2-phenylindole) was used as nuclear counterstaining.

Electrophysiology

All procedures were approved by the Italian Ministry of Health and the San Raffaele Scientific Institute Animal Care and Use Committee in accordance with the relevant guidelines and regulations. We analyzed mice of both sexes (30 days of age) as previously described (Sessa et al., 2019).

Primary Neuronal Cultures

Primary cultures of mouse embryonic hippocampal neurons were prepared from E17.5 mutant and control mice as previously described (Sessa et al., 2019).

Sholl Analysis

Neuronal cultures were transduced with lentiviral vector EF1a-GFP at a low titer the day after the plating for 1 h, in order to obtain sparse GFP cell-labeling. At the desired time points, cells were processed for IF analysis. Images of the dendritic tree of double positive GFP+/MAP2+ neurons were investigated using Sholl Analysis plug-in (Ferreira et al., 2014) in Fiji software (NIH, United States).

In vitro Spine Analysis

Mutant and control primary cultures (E17.5 murine hippocampal neurons) were infected with lentiviral vector EF1a-GFP at a low titer at DIV1 for 1 h to get few and sparse GFP labeled cells. At DIV 15 and DIV 21, the primary neurons were stained with GFP and analyzed at confocal microscope. Spine density was measured in both mutant and wild type neurons.

Golgi-Cox Staining and Spine Measurements

Golgi-Cox staining was performed as previously described (Sessa et al., 2019).

Neural Stem Cell Culture

E14.5 embryonic cortices were dissociated, fragmented and digested with papain (10 U/ml, Worthington Biochemical) and cysteine (1 mM, Sigma-Aldrich) in HBSS with 0.5 mM EDTA at 37°C. The obtained NSCs were routinely cultured in suspension as neurospheres. Cells were normally cultured in neural-inducing medium (NIM) containing: DMEM/F12 (Sigma-Aldrich) supplemented with Hormon Mix (DMEM/F12, 0.6% Glucose (Sigma-Aldrich) [30% in phosphate buffer (PBS)

(Euroclone)], Insulin (Sigma-Aldrich) 250 μg/ml, putrescine powder (Sigma-Aldrich) 97 μg/ml, apotransferrin powder (Sigma Aldrich), sodium selenite 0.3 μM, progesterone 0.2 μM), 1 mg/ml penicillin/streptomycin (Sigma-Aldrich), 2 mM glutamine (Sigma-Aldrich), 0.66% Glucose [30% in phosphate buffer salt (PBS) (Euroclone)], Heparin 4 μg/ml, 10 ng/ml basic fibroblast growth factor (bFGF) (Thermo Fisher Scientific) and 10 ng/ml epithelial growth factor (EGF) (Thermo Fisher Scientific). To assess multipotent differentiation capacity of NSCs, cells were seeded on matrigel (Corning) coated glass coverslips at a density of 3–4 $\times$ 10^5 cells per well in a 24 multi-well plate. The first day after plating, the medium was changed adding normal NIM without EGF for 2 days. The medium was then changed again adding NIM without both EGF and FGF, supplemented with 2% fetal bovine serum (FBS) (Sigma-Aldrich) for 6 days.

NSCs proliferation rate was calculated by seeding 6 $\times$ 10^5 cells in adherent conditions in a 6 multi-wells plate and counting (after detaching) the cells every 2–3 days. After the count, 6 $\times$ 10^5 cells were seeded again for 6 time points (3 replicates for each time point).

Quantitative RT-PCR

RNA was extracted using TRI Reagent (Merck) according to the manufacturer's instructions. Quantitative RT-PCR (qRT-PCR) was performed as previously described (Sessa et al., 2019) with custom designed oligos (**Supplementary Table S1**).

Immunocytochemistry

Immunocytochemicalstaining was performed as previously described (Sessa et al., 2019).

The primary antibodies utilized were the following: anti-KI67 (Immunological Science, MAB90948), anti-GFAP (Merck, AB5804), anti-TUJ1 (Covance, MRB-435P), anti-phospho histone 3 (PH3) (rabbit, 1:400, Sigma-Aldrich, H0412), anti-S100b (Dako, GA504). Anti-O4 primary antibody was produced from a hybridoma clone, using the culture media of hybridoma cells directly on living cells to perform the staining for O4 epitope. Secondary antibody used: 488-mouse (donkey, 1:2,000, Molecular Probes, A21202), 488-rabbit (donkey, 1:2,000, Molecular Probes, A21026), 594-mouse (donkey, 1:2,000, Molecular Probes, A10036), 594-rabbit (donkey, 1:2,000, Molecular Probes, A21207). DAPI (4′,6′-diamidino-2-phenylindole) was used as nuclear counterstaining.

The quantification of Radial Glia cells in the hippocampus has been performed using GFAP antibody, counting only the $GFAP^+$ cells present in the subgranular zone of the dentate gyrus.

RNA-Sequencing

RNA libraries were generated, quality-checked and sequenced as previously described (Sessa et al., 2019). Sequences (Fastq files) were aligned to the mm9 and mm10 mouse reference genomes by using the splice junction mapper TopHat (Kim et al., 2013). Differential gene expression and Functional enrichment analyses were performed with DESeq2 (Love et al., 2014) and GSEA (Huang et al., 2009), respectively. Statistical and downstream bioinformatics analyses were performed within the R environment. The software Homer was used to find *de novo*-enriched motifs in the promoters of downregulated genes with the following setting: +1,000–100 from TSS. The data were deposited in the NCBI Gene Expression Omnibus and are accessible through GSE162750. RNA-seq and ChIP-seq data from the literature were downloaded from the NCBI GEO repository with the accession codes GSM935653, GSM1817009, and GSM1817010.

Constructs and NSC Complementation

The coding sequence of *Tbl1xr1* was kindly provided by Dr. V. Perissi and cloned into pCAG-V5 vector after digestion with *Xho*I/*Not*I enzymes. Then, Tbl1xr1-V5 coding sequence was amplified using specific primers (5′-TCCCCCCGGGATGAGTATAAGCAGTGATGAGGTCAACT TCTTGG-3′; 5′-ACGCGTCGACTCACGTAGAATCGAGACC GAGGAGAGGG-3′), and inserted in the Ef1a-Setd5-V5 lentiviral construct digested with Xma1 and Sal1 to remove the Setd5 coding sequence and obtain the Ef1a-Tbl1xr1-V5 lentiviral construct. Ef1a-Tbl1xr1-V5 was subjected to PCR site-specific mutagenesis to obtain the following mutated form: (i) Ef1a- Tbl1xr1F10L-V5 (5′-CTCGAGGGATCCACCATGAGTATAAGCAGTGATGAGGT CAACTTgTTGGTATATAGGTACTTGCAAG-3′; 5′-CGGCTG CATGCTGCTGTGCAAGCTTGTCTCTGTAGGCTTGTTGTC TTGTTTGGACTACATCGGGCATAACAGC-3′): (ii) Ef1a- Tbl 1xr1G70N-V5 (5′-CTCGAGGGATCCACCATGAGTATAAG CAGTGATGAGGTCAACTTCTTGGTATATAGGTACTTGCA AG-3′; 5′-GTTGTCTTGTTTGGACTACATCGGGCATAACA GCATCTATCAGGGACAGAGACTCGATGGGTCGACCATCA AATAAGGTGtCATCCTCATTTATGCTAAC-3′); (iii) Ef1a- Tbl 1xr1L282P-V5 (5′-CTTGCCAGCACCTTGGGGCAGCATAAA GGTCCTATATTTGCATTAAAATGGAATAAGAAAGGAAAT TTCATCCCAAGTGCTGGCGTAGATAAG-3′; 5′-GCGGC CGCGGATCCTTTCCGAAGGTCTAAGACACAAACTGAACC GTCCGAAGCAC-3′); (iiii) Ef1a- Tbl1xr1Y446C-V5 (5′-CTTTGACAAAACATCAAGAGCCCGTGTGCAGTGTGGCT TTTAGTCCTGATGGC-3′; 5′-GCGGCCGCGGATCCTTTCCG AAGGTCTAAGACACAAACTGAACCGTCCGAAGCAC).

The lentiviral construct of both wildtype and mutated forms of Tbl1xr1 were used to infect NSC derived from telencephalic cortex of Tbl1xr1 KO embryos at E14.5.

Immuno-Precipitation

HEK293 cells were seeded on six 150 mm dishes and transfected with Ef1a-GFP (mock), Ef1a-Tbl1xr1-V5, Ef1a-Tbl1xr1F10L-V5, Ef1a-Tbl1xr1 G70N-V5, Ef1a-Tbl1xr1L282P-V5 and Ef1a-Tbl1xr Y446C-V5 using Ca-phosphate. 48 h after transfection, cells were harvested and suspended in 500 μl of IP buffer (20 mM Tri-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100) supplemented with complete protease inhibitors (Roche). 30′ after the incubation on ice, the lysate was centrifuged at 14,000 $\times$ *g* for 10 min at 4°C and 15% of the supernatant was added with 4X SDS protein sample buffer [100 mM Tris-HCl pH 6.8, 40% (v/v) glycerol, 312 mM SDS, 174 mM dithiothreitol (DTT), 0.04% (w/v) bromophenol blue] (input). The remaining supernatant

was immuno-precipitated with V5-antibody overnight at 4°C with agitation. 100 μl of Protein G Dynabeads (Novex, Thermo Fisher Scientific) were used per IP. After washing in IP buffer supplemented with protease inhibitors, the beads were mixed with lysate and incubated for 2 h at 4°C with agitation. At the end of the incubation, the mixed beads-lysate was washed 3 times with IP buffer and added with 2X SDS protein sample buffer for Western blot analysis.

Quantification and Statistical Analysis

Data are expressed as mean ± standard error (SEM) and significance was set at $p < 0.05$. Statistical tests are provided in the figure legends.

RESULTS

Tbl1xr1 Knock-Out Mice Display Behavioral Impairments

To analyze the importance of TBL1XR1 for neurological capabilities in living animals, we generated *Tbl1xr1* knock-out (KO) mice taking advantage of a line carrying the *Tbl1xr1* floxed allele (LoxP sequences surrounding the exon V) (Rohm et al., 2013) crossed with the CMV:Cre driver line (**Figure 1A**). Despite the full KO mouse died at an early embryonic phase (Perissi et al., 2004), CMV:Cre mediated KO animals were born nearly at mendelian ratio, successfully completed pre-natal development and reached the adulthood (**Figure 1A**), perhaps thanks to incomplete DNA recombination at some stage/cell population. However, the TBL1XR1 protein was completely absent in brain tissue (**Figure 1A**). Since mutations in the *TBL1XR1* are linked with a variety of neuro-psycho-motor disabilities, we decided to test motor, social, and cognitive performance of KO vs. control littermate mice through behavioral assessment (**Figure 1** and **Supplementary Figure S1**). Both genotypes showed comparable spontaneous activity in the open field (**Supplementary Figure S1A**), while motor coordination was impaired in KO mice as revealed by poor performance in rotarod exercise (**Figure 1B**), abnormal paw-slipping in beam walking (**Figure 1C**), and mild but appreciable differences in some catwalk parameters (**Figure 1D** and **Supplementary Figure S1B**). Interestingly, a large fraction of human *TBL1XR1* mutated patients display compromised or delayed motor skills (Vaqueiro et al., 2018).

Mutant animals tended to hide more objects compared to the WT in the marble-burying test, possibly reflecting obsessive behavior, and/or anxiety (**Figure 1E**). However, the time spent at the periphery vs. the center of an empty arena was comparable, indicating that *Tbl1xr1* KO did not show high anxiety levels (**Supplementary Figure S1C**). We also performed an elevated plus maze test to further investigate anxiety. Notably, mutant mice did not display significant differences in the time spent in the open arms, confirming that the loss of *Tbl1xr1* is not impacting on anxiety, at least in this model (**Supplementary Figure S1D**).

Next, we assessed behavior that is relevant for intellectual disability and autism spectrum disorders. To test spatial memory, we used an 8-arm radial maze test (**Figure 1F**), in which *Tbl1xr1* mutants displayed significant differences in memory performances, including working memory (**Figure 1F** and **Supplementary Figure S1E**). Sociability was assayed by evaluating the time that a test subject (either a WT or an KO animal) spent together with a host WT mouse (Silverman et al., 2010). KO animals interacted significantly less with the host, compared to the controls (**Figure 1G**). However, the two genotypes spent equal amount of time in aggressive behavior (**Supplementary Figure S1F**). The animals did not present any olfactory impairment, since both groups performed food tests (either hidden or presented) equally well (data not shown). Furthermore, since external factors may confound the assessment of sociability in freely moving subjects, we performed a three-chamber test (Papale et al., 2017) where the stranger mouse is confined in a cage (thus eliminating any possible effect of size or dominance) while the subject *Tbl1xr1* KO or WT animals was free to explore. Despite the two genotypes equally examined the arena (**Supplementary Figure S1G**), mutant animals did not discriminate between the stranger-occupied cage and the empty one (**Figure 1H**). Notably, cognitive disability is a penetrant trait in TBL1XR1 patients, while low sociality is present in TBL1XR1 patients with ASD (Laskowski et al., 2016; Vaqueiro et al., 2018).

These results indicate that our mouse model, in which TBL1XR1 is completely missing in brain tissue, is viable and presents behavioral deficits that, at least in part, resemble the impairments described in patients carrying genomic insults in the *TBL1XR1* gene, thus representing an interesting model for the related diseases.

Loss of *Tbl1xr1* Alters Morphology and Function of Neuronal Cells

Mutant brains appeared grossly normal in the adulthood, and the brain weight normalized to total body mass was conserved (**Figure 2A**). However, cortical wall thickness was reduced, particularly at medial-caudal level (**Figure 2B**). Although neuronal types were conserved across the cortical layers, we found that the number of $CUX1^+$ late-born neurons was significantly higher and $SATB2^+$ neurons were decreased in mutants, while the number of the early born $CTIP2^+$ neurons was unaltered (**Supplementary Figure S2A**). Thus, the molecular identity of upper layer neurons (where both $CUX1^+$ and $SATB2^+$ neurons are particularly abundant) is altered in these mutants.

To determine if synaptic transmission was impaired upon loss of *Tbl1xr1*, we analyzed both spontaneous glutamatergic excitatory postsynaptic currents (sEPSCs) and GABAergic inhibitory postsynaptic currents (sIPSCs) in cortical slices from postnatal day 30 (P30) KO and littermate WT mice using whole-cell voltage-clamp recordings (**Figure 2C**). While amplitude, 10–90% rise time, and decay time constant were unaltered, the sEPSC frequency was significantly increased in mutant neurons (**Figure 2C**). Conversely, we found no differences in sIPSC parameters between genotypes (data not shown). Then, we sought to further investigate neuronal cells using primary neuronal cultures from hippocampi of E17.5 mutant and WT

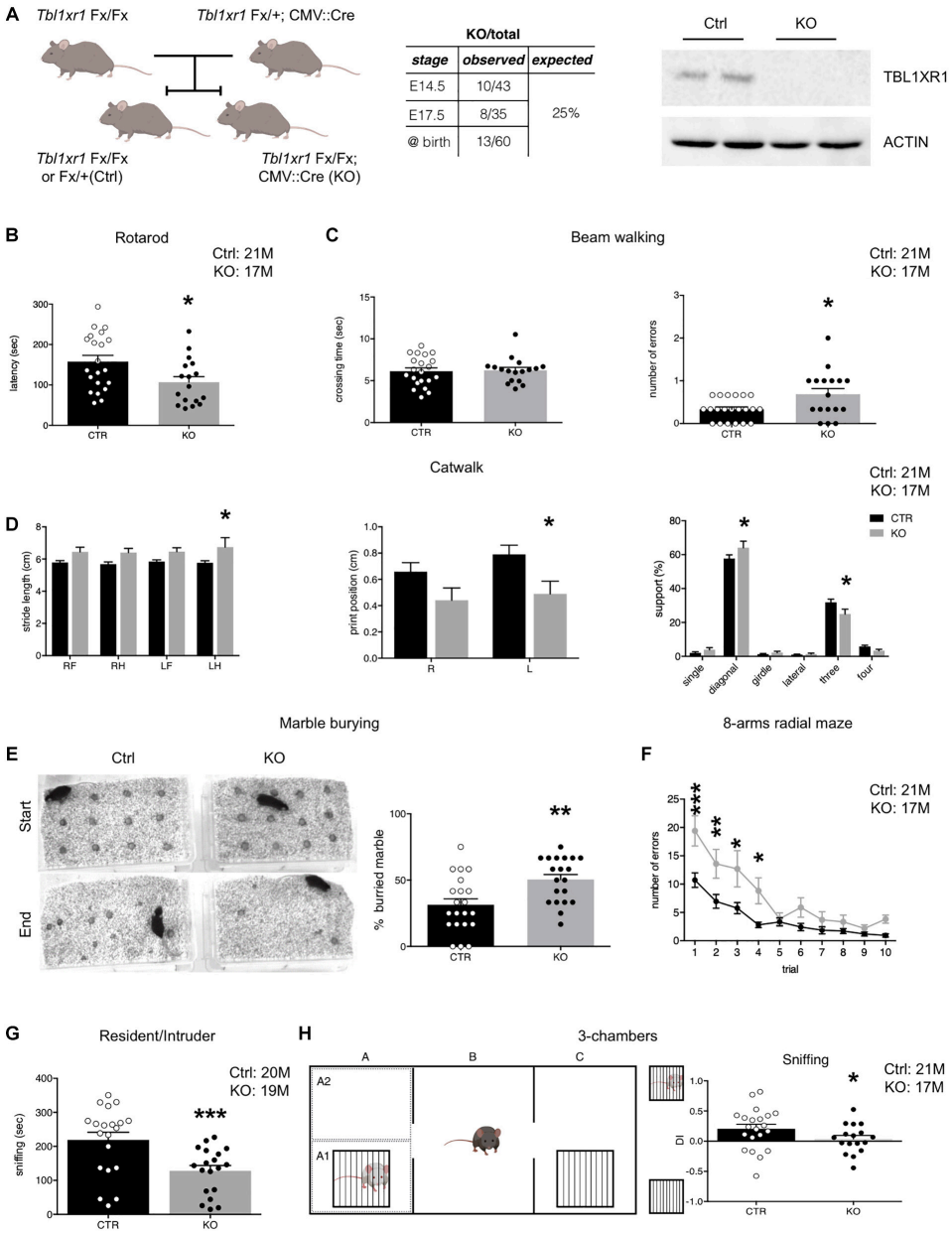


FIGURE 1 | Behavioral deficit of Tbl1xr1 mutant mouse. **(A)** Breeding strategy for obtaining both *Tbl1xr1* mutant and control (left), table summarizing mutant animals found at the indicated developmental stages (center) and Western blot for TBL1XR1 protein in control and mutant brain tissue (right). **(B)** Rotarod motor test (shown as mean + s.e.m. with dots representing individual samples): *n* (adult male mice): Ctrl = 21, KO = 17, $*p = 0.02$, statistically compared using unpaired *t*-test. **(C)** Beam walking assay in which time of crossing and number of errors (paw falls) have been quantified (shown as mean + s.e.m. with dots representing individual samples): crossing time (top): *n* (adult male mice): Ctrl = 21, KO = 17, $p = 0.9933$, statistically compared using Mann-Whitney; errors (bottom): n (adult male mice): Ctrl = 21, KO = 17, $*p = 0.0126$, statistically compared using unpaired *t*-test. **(D)** Through catwalk assay we measured stride length, position of the paw during walking (print position) and support that the animal has during walking (shown as mean + s.e.m.), n (adult male mice): Ctrl = 21, KO = 17: $*p < 0.05$. Statistically compared using 2-way ANOVA. **(E)** Marble burying test in which the percentage of buried marble spheres has been quantified (shown as mean + s.e.m. with dots representing individual samples), n (adult male mice): Ctrl = 21, KO = 20: $**p = 0.0028$, statistically compared using unpaired *t*-test. **(F)** Eight-arm radial maze test quantified in which the percentage of errors (entries in arms already visited) on total visits has been quantified (shown as means ± SEMs in each experimental day, 1 trial/day), *n* (adult male mice): Ctrl = 21, KO = 17: Multiple comparisons: day 1: $***p = 0.0001$; day 2: $**p = 0.007$; day 3: $**p = 0.0039$: day 4: $*p = 0.021$; day 5: $p > 0.9999$; day 6: $p = 0.7603$; day 7: $p > 0.9999$; day 8: $p > 0.9999$; day 9: $p > 0.9999$; day 10: $p > 0.9999$; statistically compared using 2-way ANOVA and Bonferroni's *post hoc* test. **(G)** Sociability of adult animals measured as the time spent to interact with a novel mouse in the resident/intruder test (shown at right as means ± SEMs, with dots representing individual samples), *n* (adult male mice): Ctrl = 20, KO = 19: $***p = 0.0009$; statistically compared using unpaired *t*-test. **(H)** Three-chamber test (scheme at left) quantified as percentage of the time spent sniffing the mouse enclosed in the little cage vs. the empty cage shown as discrimination index (DI) (shown as means ± SEMs, with dots representing individual samples), *n* (adult male mice): Ctrl = 21, KO = 17: sniffing: $*p = 0.0287$, statistically compared using unpaired *t*-test. See also **Supplementary Figure S1**.

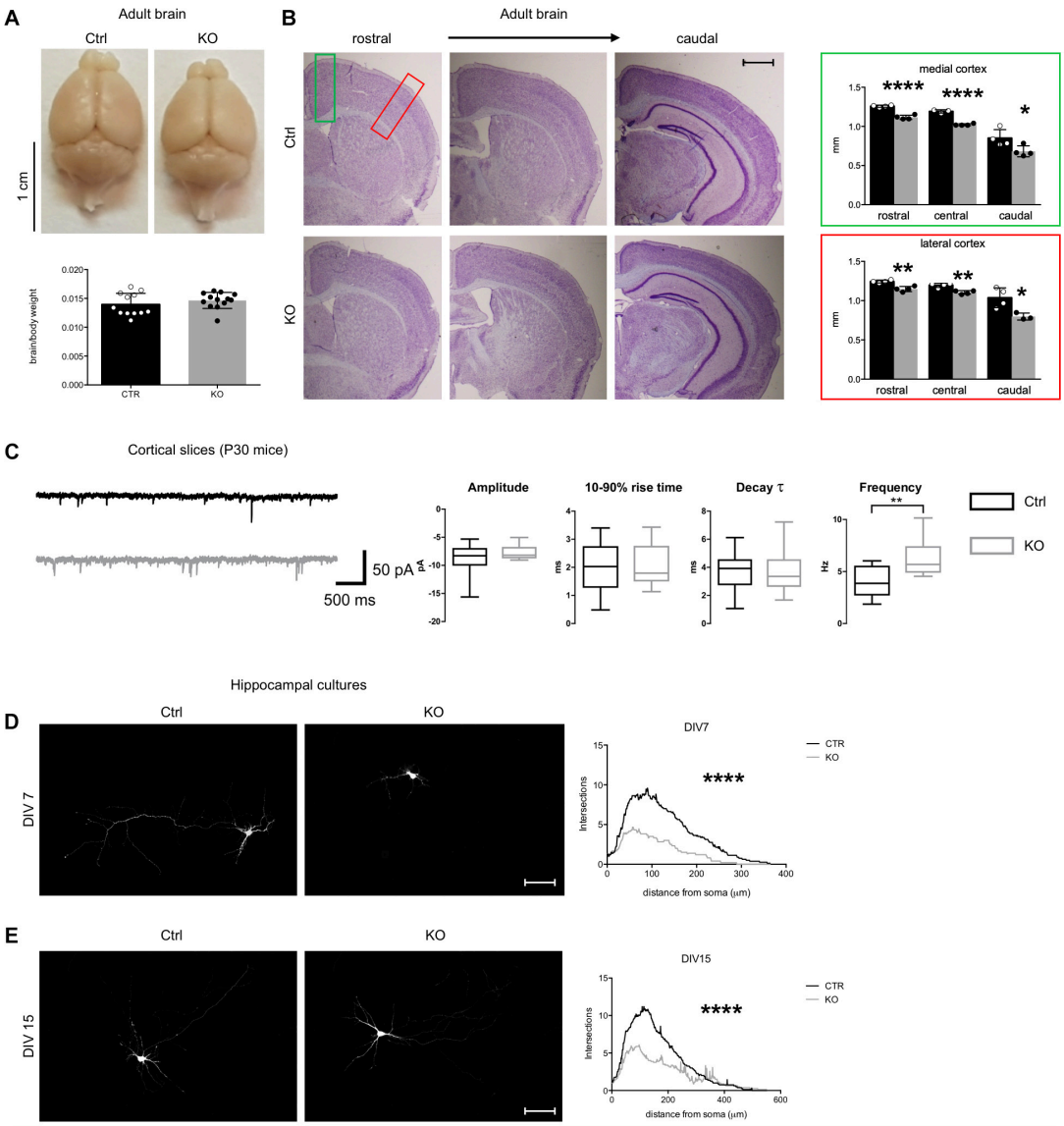


FIGURE 2 | Defects of *Tbl1xr1* mutant neurons. **(A)** Adult brains from both control and Tbl1xr1 mutant mice showed similar brain to body ratio (shown as means ± SEMs, with dots representing individual samples), *n*: Ctrl = 12, KO = 13: $p > 0.05$, unpaired *t*-test. **(B)** Left, cresyl violet staining of coronal section of both control and mutant forebrains (cerebral corte, striatum, and hippocampus from adult mice of >than 2 months of age). right quantification of the cortical thickness in medial (green square) and lateral (red) position (shown as means ± SEMs, with dots representing individual samples), *n*: Ctrl = 4, KO = 4: rostral: $^{****}p < 0.0001$; central: $^{****}p < 0.0001$; rostral: $^{*}p < 0.0415$. **(C)** Left, representative traces showing sEPSCs recorded in cortical pyramidal neurons in slices from control (black) and *Tbl1xr1* mutant (gray) mice (animals of both sexes, 30 days of age) (bars: 50 pA, 500 ms). Right, summary boxplots for EPSC parameters (shown as box for interquartile range, median line and whiskers for highest and lowest values): mean amplitude, wt -8.6 ± 0.6 pA, KO -7.8 ± 0.3 pA, $n = 19$, $p = 0.27$, unpaired *t*-test; mean 10–90% rise time, wt 1.9 ± 0.2 ms, KO 2.0 ± 0.2 ms, $n = 19$, $p = 0.78$, unpaired *t*-test; mean decay time constant, wt 3.7 ± 0.3 ms, KO 3.6 ± 0.4 pA, $n = 19$, $p = 0.87$, unpaired *t*-test; mean frequency, wt 4.0 ± 0.3 Hz, KO 6.3 ± 0.4 Hz, $n = 19$, $p < 0.001$, unpaired *t*-test. **(D)** Hippocampal cultures from control and mutant E17.5 embryos at 7 days *in vitro* (DIV) infected with low titer GFP lentivirus (at DIV0) and Sholl analysis for the quantification of the intersection of neurites accordingly with the distance to the soma, $^{****}p < 0.0001$, 2-way ANOVA. **(E)** Hippocampal cultures from control and mutant E17.5 embryos at 14 days *in vitro* (DIV) infected with low titer GFP lentivirus (at DIV0) and Sholl analysis for the quantification of the intersection of neurites accordingly with the distance to the soma, $^{****}p < 0.0001$, 2-way ANOVA. Scale bars: **(B)** 400 μm; **(D,E)** 50 μm. See also **Supplementary Figure S2**.

embryos. Mutant neurons exhibited a simpler morphology, with less neurites and arborization compared to control cells both at early and late stages of *in vitro* development (**Figures 2D,E**). In developing mutant neurons (DIV 15), dendritic spines were significantly sparser than controls (**Supplementary Figure S2B**), though their number was comparable in more mature cells *in vitro* (**Supplementary Figure S2C**). Golgi-Cox staining confirmed the normal dendritic spine number in adult KO cerebral cortex compared to controls, without evident changes in their shape (**Supplementary Figure S2D**).

Altogether, these results suggest that *Tbl1xr1* mutant mice are mildly microcephalic, as seen in particular in Pierpont syndrome patients (Heinen et al., 2016), and that their excitatory synapses are hyperactive.

Loss of *Tbl1xr1* Leads to Defective Proliferation and Differentiation of Neural Progenitors

We next sought to investigate whether the observed changes reflected any impairment in brain development during embryonic life. The embryonic cortex at E14.5 was slightly smaller in mutants, particularly in its medial region (**Supplementary Figure S3A**). Interestingly, the ventricular zone of mutant mice contained less proliferative progenitors ($PAX6^+$ cells), probably because differentiation occurred at earlier stages ($SATB2^+$ used as marker of young neuronal cells) (**Supplementary Figure S3B**). To better analyze neural progenitor proliferation/differentiation dynamics, we used an *in vitro* system constituted by NSCs (Azari et al., 2010). To avoid possible residual early stage impairments, we derived NSCs from *Tbl1xr1* floxed E14.5 embryos. After infecting them with lentivirus carrying either Cre-GFP (to generate KO's) or GFP only (as control), we sorted the GFP^+ cells and verified for correct genotypes and for Tbl1xr1 mRNA and protein absence (**Figure 3A**). When grown as neurospheres, mutant NSCs showed a smaller sphere volume compared to controls (**Figure 3B**), possibly due to a proliferative defect. When we alternatively used adherent cultures to obtain proper growth curves and immuno-labeling (using PH3 and KI67 proliferative markers), a decrease in cell cycle progression was evident (**Figures 3C,D** and **Supplementary Figure S3C**). Conversely, the differentiation capability of mutant NSCs were increased since KO cells generated more post-mitotic $TUJ1^+$ neurons, $S100b^+$ astrocytes, and $O4^+$ oligodendrocytes after applying a differentiation protocol *in vitro* (**Figure 3E**).

Finally, to investigate whether these defects are present in *Tbl1xr1* mutant postnatal progenitors, we analyzed the dentate gyrus (DG) of the hippocampus where adult neurogenesis normally occur in rodents. First, we noted that the gross morphology of the hippocampus and the DG in particular is maintained (**Figure 2B**). Immunochemistry for markers of radial glia cells (GFAP), neuroblasts/immature neurons (DCX), neurons (NEUN), and proliferative cells (KI67), indicated that the dynamics of proliferation/differentiation in both mutant and control DG is similar (**Supplementary Figure S3D**).

These data indicate that the loss of TBL1XR1 during the neural progenitor stage leads to proliferative defects and a tendency to anticipate differentiation that may eventually induce a microcephalic brain. However, adult neurogenesis in DG is preserved in *Tbl1xr1* mutants.

TBL1XR1 Ensures Correct MAPK Signaling Transduction

Together with its related protein TBL1X, TBL1XR1 is important for both gene repression mediated by NCOR/SMRT complexes, and for de-repression promoting the removal of NCOR/SMRT (Perissi et al., 2010). Thus, we investigated the level of NCOR1 using Western blot analyses on both NSCs and adult brains. The level of NCOR1 was decreased in mutants, indicating that TBL1XR1 is critical for the assembly of the co-repressor complex rather than serving as an exchanging factor (**Supplementary Figures S4A,B**). The entire complex seemed to be affected since also HDAC3, the typical histone deacetylase present in the complex, was significantly downregulated both *in vitro* and *in vivo* (**Supplementary Figures S4A,B**). The effect on both NCOR1 and HDAC3 was particularly evident in the nuclear fraction, where the epigenetic complex is normally operating (**Figure 4A**). Both TBL1R and TBL1XR1 are also required for transcriptional activity of βCATENIN on Wnt target genes (Li and Wang, 2008). Interestingly, despite the level of βCATENIN was unaffected, we detected a high level of its phosphorylated form in the mutant cells (**Figure 4B**) suggesting that, in the absence of TBL1XR1, the AXIN2/GSK3β complex is active to remove the protein and mitigate the canonical Wnt signaling (MacDonald et al., 2009).

To gain insights on the molecular aspects of the phenotype, we performed RNA-seq to analyze the global transcriptome upon *Tbl1xr1* loss. Approximately 3,500 transcripts were de-regulated (1,755 upregulated and 1,754 downregulated) (**Figure 4C** and **Supplementary Table S2**). Gene Ontology (GO) analysis showed that several biological processes were significantly altered in *Tbl1xr1* KO cells, including DNA replication and cell cycle (downregulation), as well as gliogenesis and neuron differentiation (upregulation) (**Figure 4D** and **Supplementary Table S2**), confirming our previous observations (**Figure 3**). Moreover, we found upregulated gene categories associated with important molecular pathways as the Wnt and MAPK pathways, while several categories associated with nucleic acid metabolism were downregulated (**Figure 4D** and **Supplementary Table S2**). To discriminate between possible direct and indirect effects of TBL1XR1 on gene transcription, we examined public available datasets of DNA regions bound by TBL1XR1 obtained in a lymphoblastoid cell line (GM12878 cells). Despite the cell origin was very far from NSCs, we observed that a significant percentage of both upregulated (16%) and downregulated genes (26%) are putative TBL1XR1 direct targets (**Figure 4E**). These target genes account for GO categories important for DNA/RNA catabolism, RNA processing and splicing, DNA repair, and cell cycle (putative direct targets downregulated) as well as catabolism, autophagy and cell death (putative direct targets upregulated) (**Figure 4E** and **Supplementary Table S2**). These observations support the double nature of the TBL1XR1 protein to serve both as a core member and an exchanging factor of NCOR complexes, explaining the missing repressive activity on certain loci and the transcriptional repression on others, respectively, in *Tbl1xr1* KO's. However, it has to be noted that the paralog gene *Tbl1x* was upregulated in *Tbl1xr1* NSCs (**Supplementary Table S2**), possibly vicariating the function of TBL1XR1 in certain contexts.

To understand the impact of the lack of TBL1XR1 on Wnt signaling, we analyzed direct target genes of this pathway that resulted partially deregulated. We found that transcripts changed in both directions in our dataset, that did not agree with a loss of βCATENIN activity (**Supplementary Figure S4C** and

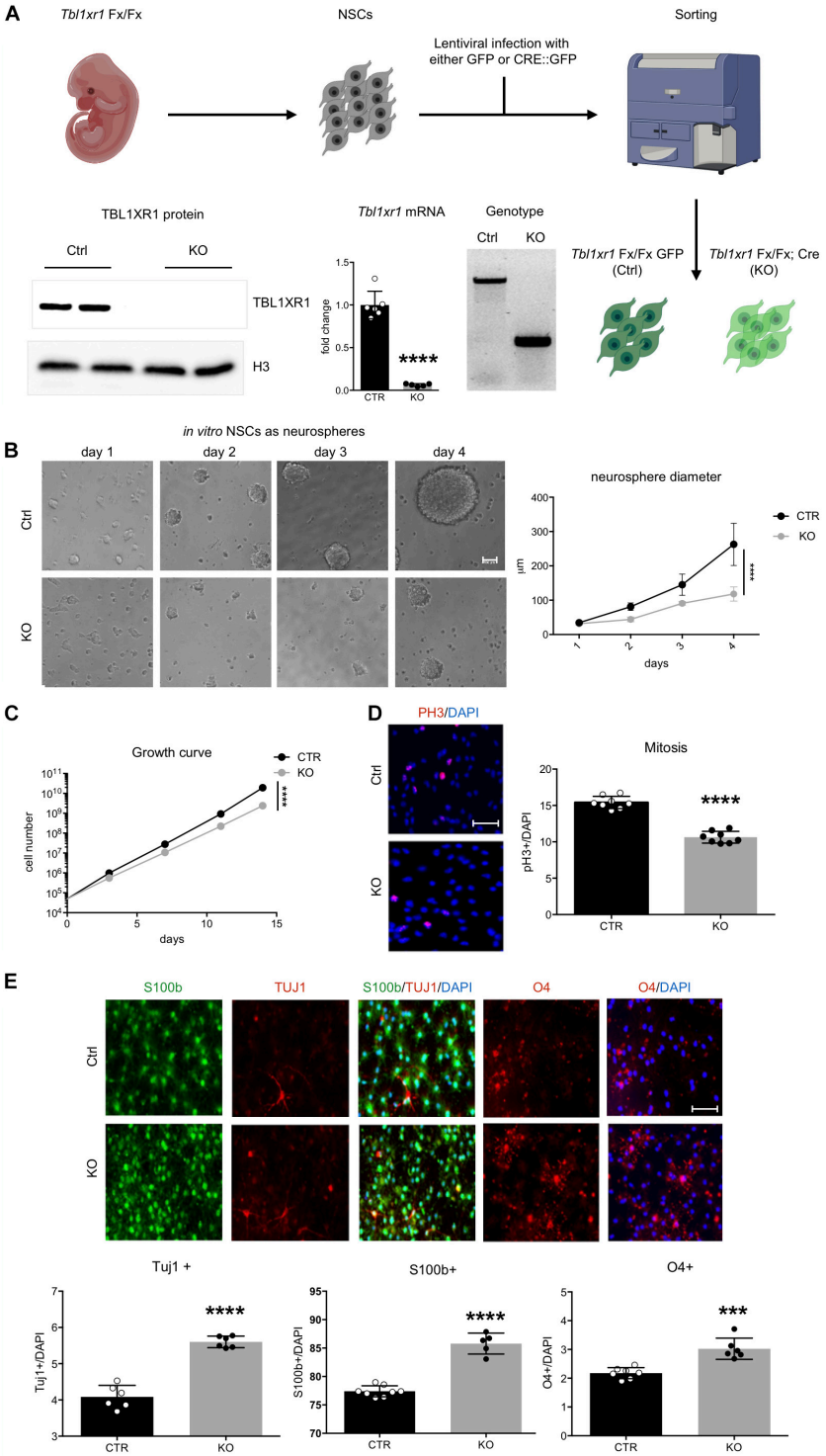


FIGURE 3 | Loss of *Tbl1xr1* affects neural stem cells dynamics. **(A)** Scheme of NSC derivation with examples of: genotyping (by PCR), expression of *Tb1xr1* (mRNA abundance by RT-qPCR, $^{****}p < 0.0001$, unpaired *t*-test), and TBL1XR1 protein level (by Western blot). **(B)** NSCs both Ctrl and KO cultured as neurospheres, on the right quantification of spheres' diameter: difference due to genotype, $F(1,40) = 63.19$, $^{****}p < 0.0001$, 2-way ANOVA. **(C)** Growth curve of adherent *in vitro* NSCs: difference due to genotype, $F(1,4) = 535.6$, $^{****}p < 0.0001$, 2-way ANOVA. **(D)** Immunocytochemistry of both Ctrl and *Tbl1xr1* KO proliferating NSCs for phosphor histone 3 (PH3) counterstained with DAPI. On the right, quantification of $PH3^{+}$ cells on total DAPI nuclei, $^{****}p < 0.0001$, unpaired *t*-test. **(E)** Up, immunocytochemistry of both Ctrl and *Tbl1xr1* KO differentiating NSCs for S100b (astrocytes), TUJ1 (neurons) and O4 (oligodendrocytes) counterstained with DAPI. Bottom, quantification: TUJ1: $^{****}p < 0.0001$, unpaired *t*-test; S100b: $^{****}p < 0.0001$, unpaired *t*-test; O4: $^{***}p = 0.0002$, unpaired *t*-test. Scale bars: **(B)** 40 µm; **(D,E)** 50 µm.

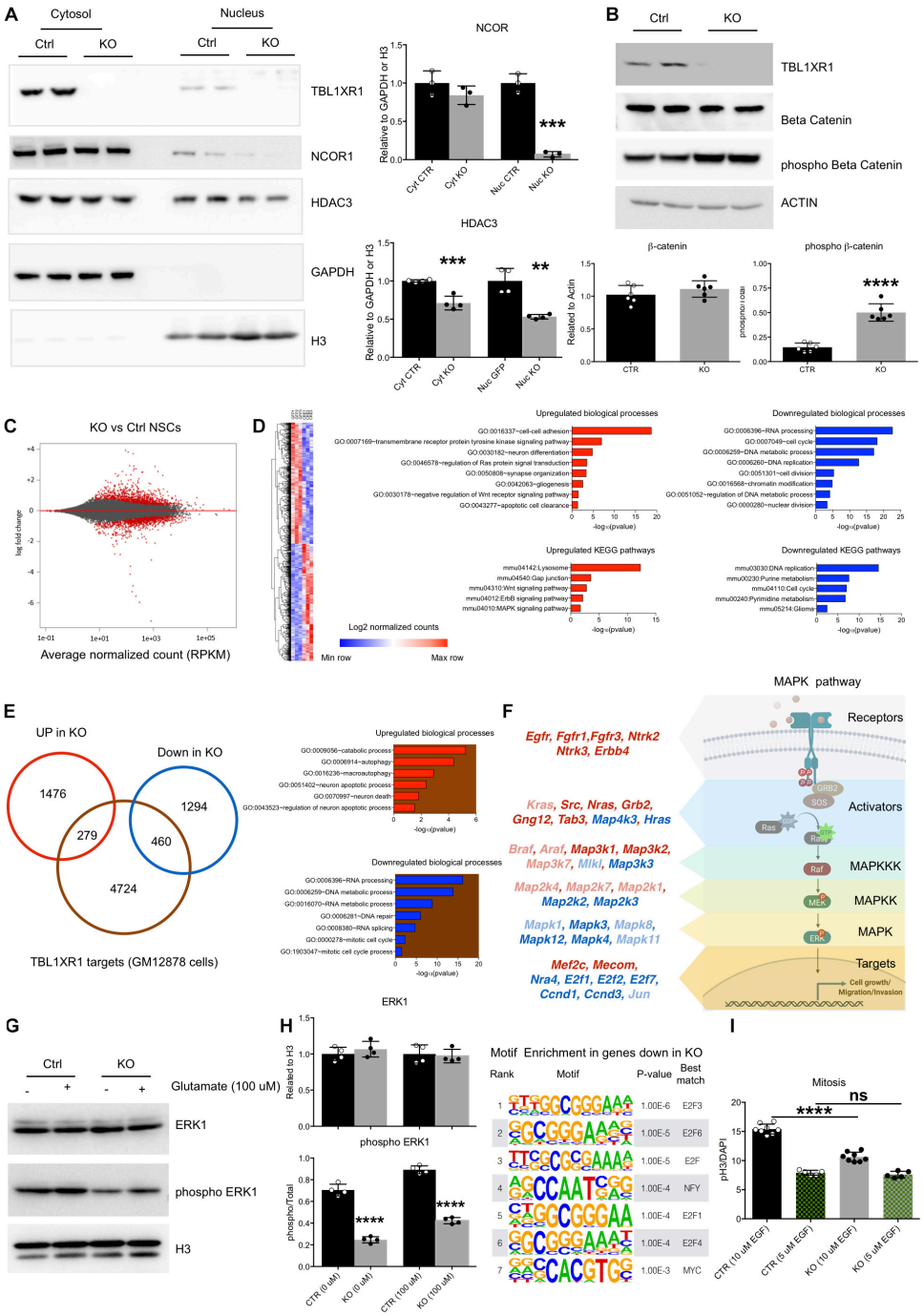


FIGURE 4 | Molecular defects of *Tbl1xr1* KO neural stem cells. **(A)** Western blot analysis of both cytosolic (enriched in GAPDH, used as normalizer) and nuclear (enriched in H3, used as normalizer) protein fractions from control and *Tbl1xr1* KO *in vitro* NSCs for the following proteins: TBL1XR1, NCOR1, and HDAC3. On the right, the quantification of the blots for NCOR1 and HDAC3 (shown as mean + s.e.m. with dots representing individual samples): cytoplasmic NCOR1: n (biological replicates): Ctrl = 3, KO = 3: $p = 0.2407$; nuclear NCOR1: *n* (biological replicates): Ctrl = 3, KO = 3: $^{***}p = 0.0002$; cytoplasmic HDAC3: *n* (biological replicates): Ctrl = 4, KO = 4: $^{***}p = 0.0007$; nuclear HDAC3: *n* (biological replicates): Ctrl = 4, KO = 4: $^{**}p = 0.0013$. Statistically compared using *t*-test. **(B)** Western blot analysis of whole protein lysates from control and *Tbl1xr1* KO *in vitro* NSCs for the following proteins: TBL1XR1, βCATENIN and its phosphorylated form (Ser33/37/Thr41). On the right, the quantification of the blots for βCATENIN and pβCATENIN: *n* (biological replicates): Ctrl = 6, KO = 6: βCATENIN $p = 0.2878$; pβCATENIN $^{****}p < 0.0001$. Statistically compared using unpaired *t*-test. **(C)** MA plot showing log2 fold changes as function of average normalized gene expression (RPKM) for control against KO *in vitro* NSCs (RNAseq data). Differentially expressed genes are highlighted in red. **(D)** Left, heat-map showing the differentially expressed genes between control and KO NSCs. Right, Gene Ontology analysis for terms indicating both the biological processes (up) and KEGG pathways (bottom), for upregulated (red, left) and downregulated (blue, right) genes in *Tbl1xr1* KO NSCs compared to control. **(E)** Left, Venn diagram showing the overlap between TBL1XR1 gene

(Continued)

FIGURE 4 | Continued
targets (brown, publicly available ChIP-seq data obtained from GM12878 cells) and genes up (red) and down-regulated (blue) in *Tbl1xr1* KO NSCs. Right, Gene Ontology analysis for terms indicating the biological processes for TBL1XR1 targets that are also upregulated (the red/brown intersection) and downregulated (the blue/brown intersection) genes in *Tbl1xr1* KO NSCs compared to control. **(F)** MAPK pathways found altered between control and *Tbl1xr1* KO NSCs ordinated accordingly their role within the pathway (scheme on the right). Red = upregulated in mutant; blue = downregulated in mutant; dark color = statistically significant; light color = trend. **(G)** Western blot analysis of whole protein lysates from control and *Tbl1xr1* KO NSCs treated or not with 100 μM of glutamate for the following proteins: ERK1 and its phosphorylated form (Thr202/Tyr204). On the right, the quantification: *n*: Ctrl = 4, KO = 4, 0 mM of glutamate: ERK1 $p = 0.3898$; pERK1 $^{****}p < 0.0001$; 100 μM of glutamate: ERK1 $p = 0.7170$; pERK1 $^{****}p < 0.0001$; Statistically compared using unpaired *t*-test. **(H)** Motif enrichment analysis using Homer on promoters (–1,000 + 100 from TSS) og genese downregulated in Tbl1xr1 KO NSCs ($n = 1{,}754$). **(I)** Quantification of the immunocytochemistry for PH3 in both Ctrl and *Tbl1xr1* KO proliferating NSCs in a medium with either normal (10 μM) or low levels (5 μM) of EGF, counterstained with DAPI. 10 μM: n: Ctrl = 8, KO = 8, $^{****}p < 0.0001$; *n*: Ctrl = 5, KO = 5, $p = 0.9055$; one-way ANOVA. See also **Supplementary Tables S2, S5**.

Supplementary Table S3). Next, we treated both control and KO NSCs with either an activator (LiCl) (Clément-Lacroix et al., 2005) or an inhibitor (IWR-1) (Martins-Neves et al., 2018) of the Wnt pathway. Interestingly, both genotypes were competent to increase cell proliferation, reaching the same level upon the activation of the pathway (**Supplementary Figure S4D**). However, the inhibition of βCATENIN, although affecting the proliferation of both lines, did not blunt the difference between control and KO NSCs (**Supplementary Figure S4E**). These data suggest that other pathways beyond Wnt contribute to the proliferative defects of *Tbl1xr1* KO NSCs.

The NCOR complex intersects multiple signaling cascades (Perissi et al., 2010). Our transcriptomic data showed a heavy de-regulation of the mitogen-activated protein kinase pathway. GO analysis indicated an upregulation of the MAPK pathway (**Figure 4D**), which are particularly relevant for the upper part of the cascade (e.g., receptors and MAP3K1 and 2), while other genes belonging to downstream nodes (e.g., transcription factors) were mostly downregulated (**Figure 4F**, **Supplementary Figures S4F,G**, and **Supplementary Table S4**). Of note, both NCOR1 and NCOR2 appear to directly target several MAPK-related genes (**Supplementary Figures S4H,I** and **Supplementary Table S4**). Thus, we decided to investigate whether this pathway is affected in our system. First, we biochemically evaluated the phosphorylated form of the extracellular signal-regulated kinase 1 (ERK1) as a proxy for its general activation, showing a strong decrease in *Tbl1xr1* KO NSCs both in basal condition and upon exogenous stimulation (100 μM glutamate) (**Figure 4G**). Notably, the general level of the protein was maintained (while the mRNA level was downregulated, **Figure 4F**), indicating that the alteration can be either subtle (and thus not detected in whole protein lysates) or recovered at post-transcriptional levels.

Then, we found that the genes downregulated in *Tbl1xr1* KO NSCs were enriched in E2F binding motifs near their promoters (−1,000 + 100 bp from the TSS) (**Figure 4H**, **Supplementary Figure S4J**, and **Supplementary Table S5**). Importantly, genes encoding for these factors such as *E2f1* and *E2f2,* known as MAPK downstream effectors (Nikolai et al., 2016), were downregulated upon *Tbl1xr1* loss (**Supplementary Figure S4K**).

NSCs *in vitro* proliferation relies on the external supply of mitogens, which eventually converge on MAPK cascade (Galli et al., 2003). Thus, we decreased the concentration of one of these factors (EGF) by 50% in the culture medium of both WT and KO cells. The reduced availability of EGF blunted the proliferative difference between the two genotypes, suggesting that the transduction of mitogen-activated cascade is affected by the loss of TBL1XR1 in NSCs (**Figure 4I**).

Altogether, these results suggest that depletion of TBL1XR1 affects directly and indirectly a number of important intracellular pathways. Among these, a prominent role is played by the MAPK pathway branch that is activated by external stimuli like mitogens. In fact, despite a de-repression of the upstream portion, the MAPK cascade results partially blocked leading to proliferation defects through the impairment of E2F transcription factors target network.

MAPK Signaling Alteration Is a Differential Feature Between *TBL1XR1*-Associated Disorders

To investigate the effects of different disease-inducing *TBL1XR1* mutations, we designed a series of complementation experiments using *Tbl1xr1* KO NSCs and lentiviral constructs carrying either *Tbl1xr1* WT coding sequence or its mutated versions corresponding to the following human mutations: F10L (case of schizophrenia) (Nishi et al., 2017), G70N (West syndrome-like) (Saitsu et al., 2014), L282P (ASD) (O'Roak et al., 2012) and Y446C (Pierpont syndrome) (Pons et al., 2015; **Figure 5A**). All versions were expressed at supra-physiological levels in NSCs (**Figure 5A**). The *Tbl1xr1* WT form was able to restore the proliferation defects of KO NSCs as shown by (i) the growth curve (**Supplementary Figures S5A,B**), (ii) the dimension of the neurospheres, when the cells were cultures as free-floating in the media (**Supplementary Figure S5B**) and (iii) the quantification of the immunostaining for M-phase marker (**Figure 5C** and **Supplementary Figure S5C**). In addition, the tendency to differentiate faster in post-mitotic derivatives was counteracted by lentiviral mediated *Tbl1xr1* expression (**Figure 5D** and **Supplementary Figure S5D**). This suggests that the defects we reported are specifically due to TBL1XR1 absence and that a fast re-introduction of the factor completely reverts the phenotype. The different mutations we tested were all equally able to rescue both the proliferative capacity and the normal differentiation, except for F10L, which failed any recovery (**Figures 5B–D** and **Supplementary Figure S5**). Also, G70N displayed an incomplete reversion of neurosphere diameter (**Supplementary Figure S5B**) but showed a normal rate of proliferation and differentiation (**Figures 5B–D** and **Supplementary Figure S5**). This suggests that all the missense mutations considered, except

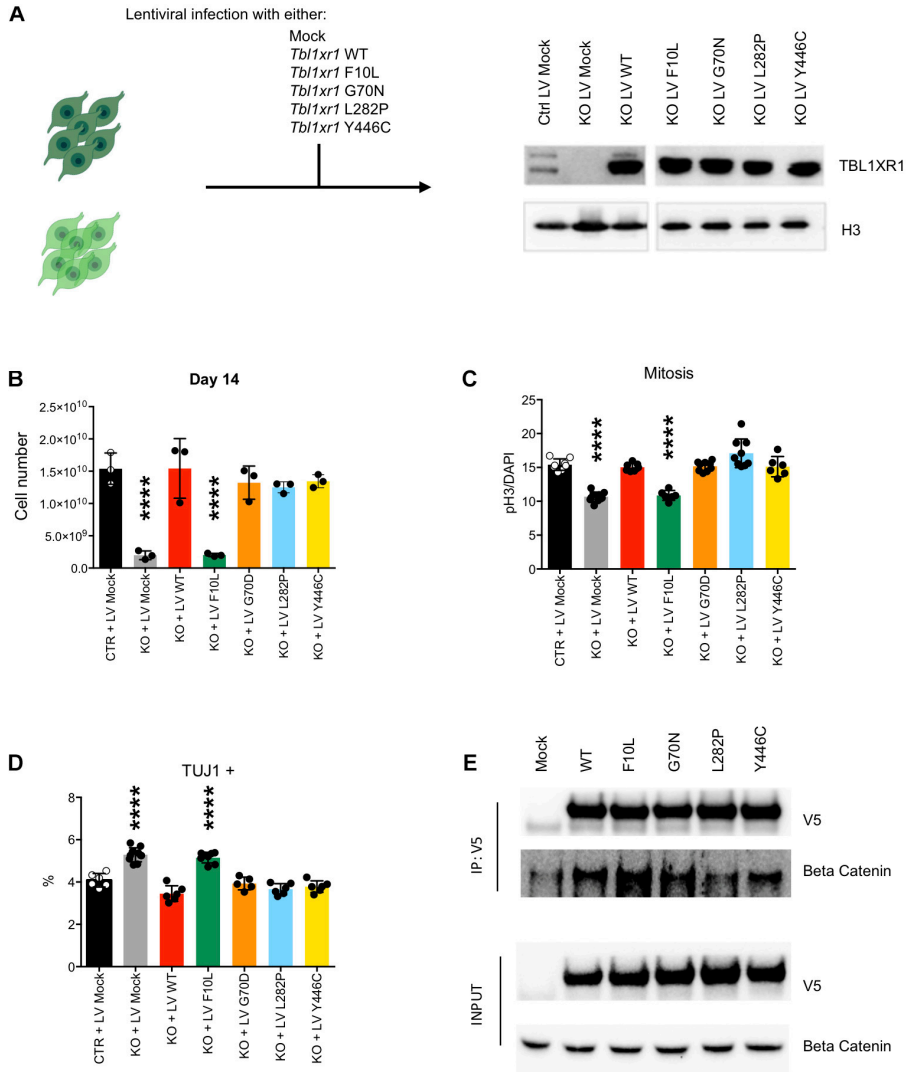


FIGURE 5 | Complementation of *Tbl1xr1* KO neural stem cells. **(A)** Experimental design and Western blot analysis of whole protein lysates from the indicated NSCs for TBL1XR1 (H3 was used for normalization). **(B)** Final point (day 14) of growth curves of the indicated NSCs (see also **Supplementary Figure S5A**) (shown as mean + s.e.m. with dots representing individual samples): Ctrl + LV Mock vs.: KO + LV Mock ****$p < 0.0001$; KO + LV WT $p > 0.9999$; KO + LV F10L ****$p < 0.0001$; KO + LV G70D $p = 0.7111$; KO + LV L282P $p = 0.4644$; KO + LV Y446C $p = 0.7894$; one-way ANOVA with Dunnett's multiple comparisons test. **(C)** Quantification of the immunocytochemistry for PH3 in the indicated NSCs (see also **Supplementary Figure S5C**) (shown as mean + s.e.m. with dots representing individual samples): Ctrl + LV Mock vs.: KO + LV Mock ****$p < 0.0001$; KO + LV WT $p = 0.9710$; KO + LV F10L ****$p < 0.0001$; KO + LV G70D $p = 0.9960$; KO + LV L282P $p = 0.0556$; KO + LV Y446C $p = 0.9924$; one-way ANOVA with Dunnett's multiple comparisons test. **(D)** Quantification of the immunocytochemistry for TUJ1 in the indicated NSCs (shown as mean + s.e.m. with dots representing individual samples): Ctrl + LV Mock vs.: KO + LV Mock ****$p < 0.0001$; KO + LV WT $p = 0.1075$; KO + LV F10L ****$p < 0.0001$; KO + LV G70D $p = 0.8997$; KO + LV L282P $p = 0.0789$; KO + LV Y446C $p = 0.3241$; one-way ANOVA with Dunnett's multiple comparisons test. **(E)** Western blot analysis of precipitated fraction and input of immune-precipitation experiment using the antibody against V5 tag that is fused to each Tbl1xr1 isoforms, for V5 and βCATENIN.

F10L, do not rely on a loss-of-function mechanism. Conversely, the schizophrenia phenotype due to F10L mutation may be due, at least in part, by endophenotypes similar to those shown by the KO. It has been reported that F10L substitution increased the ability of TBL1XR1 in the transduction of Wnt signaling, while the association with NCOR complex seemed negatively affected (Nishi et al., 2017). We independently confirmed that, among the tested TBL1XR1 mutant forms, the F10L was the one with the greater affinity for βCATENIN (**Figure 5E**). Thus, our complementation experiment supports the hypothesis that the phenotype we observed in KO cells is due to an impairment of NCOR-related functions, rather than Wnt signaling.

Collectively, these data suggest that within the complex mutational spectrum of the *TBL1XR1* gene, some genetic conditions (deletions and F10L) cause early neural defects due to dysregulation of MAPK cascade through the missing interaction with NCOR complex, while for other

mutations (e.g., G70N, L282, and Y446C) other mechanisms must be considered.

DISCUSSION

In this study, we show that the genetic knock-out of *Tbl1xr1* in mice leads to behavioral impairments that, to a certain extent, are similar to those typical of human patients mutated in the same gene. We demonstrated that *Tbl1xr1* mutant brain exhibits morphological and functional alterations originated by an increased proliferation and delayed differentiation of neural stem cells of the embryonic brain. We found impairments in the NCOR1/HDAC containing complex, transcriptional gene alterations, and a deregulation in the MAPK signaling, which are all plausible mechanisms for the disorders associated with TBL1XR1 loss of function.

TBL1XR1 mutations have been associated with neurological disorders presenting variegated manifestations that may include brain malformations, social difficulties, intellectual disability, developmental delay, learning disability, hearing loss, schizophrenia, and seizures (Kong et al., 2020). This large spectrum of phenotypes is somehow expected since TBL1XR1 may serve as cofactor for multiple functions, including the regulation of chromatin occupancy by NCOR complexes (Perissi et al., 2010), WNT signaling (Choi et al., 2011) and interaction between NCOR1/2 and MECP2 (Kruusvee et al., 2017). However, the precise function of TBL1XR1 in the processes at the basis of these phenotypes has been neither clarified nor experimentally modeled.

Here we show that animals with *Tbl1xr1* loss-of-function displayed behavioral abnormalities. These include motor coordination, memory skills, and social impairments, which are described at different levels in patients with TBL1XR1 deletions, missense mutation, and Pierpont syndrome (Vaqueiro et al., 2018; Kong et al., 2020). Of note, these phenotypes are also overlapping with some disorders due to either *NCOR1/2* or *HDAC3* mutations (Wang et al., 2016; Sajan et al., 2017; Sakaguchi et al., 2018; Iwama et al., 2019; Zhou et al., 2019). An interesting opportunity of comparison is also given by the existing animal models of TBL1XR1 interactors, namely mutants in *Ncor1/2*, *Hdac3,* and *Mecp2*. It is important to stress how neurological functions, including social and intellectual abilities, resulted indeed altered in these animals (Kong et al., 2020). In particular, our model presents memory and social impairment similar to the knock-in murine model carrying mutations in the deacetylase-activating domain of both NCOR1 and 2, that hinder their binding to HDAC3 (Zhou et al., 2019), supporting similarities between TBL1XR1, NCOR1, and NCOR2 loss-of-function. On the contrary, the two models present opposite results regarding anxiety and locomotor coordination (Zhou et al., 2019). However, our model appears different from *Ncor1/2* and *Hdac3 KO*. In particular, given the multifaceted nature of the TBL1XR1 factor, our model is not reflecting only the consequences of the loss of function of NCOR complexes. Indeed, TBL1XR1 is important for both NCOR1 and NCOR2 activity (leading to epigenetic repression) and for its dismissal (leading to gene activation) and consequently for the fine balance between these two processes. Therefore, the impairments found in our model likely depend from the different contexts of TBL1XR1 action (e.g., either the absence or presence of the ligand for nuclear receptors) in a time- and stage-dependent manner.

Here, we show that *Tbl1xr1* KO mice are microcephalic with hyperactive excitatory synapses. This may be due to different dynamics in neuronal maturation and circuit formation and/or homeostasis, even though a normal number of glutamatergic spines are present in the adult mutant neurons. Since we identified a peculiar pattern of spine development (low number at early stage, physiological number in the adult), it remains possible that the normal wiring processes of the cortical areas are altered during development of *Tbl1xr1* mutants, leading to the hyper-excitability once that the correct number of spines is restored. Further functional investigation will clarify in the future the exact role of TBL1XR1 in this respect.

We demonstrated a reduced proliferation of NSCs upon *Tbl1xr1* KO. The TBL1XR1-mediated regulation of both Wnt signaling and NCOR1 complex may contribute to the observed phenotype. Indeed, βCATENIN, of which TBL1XR1 is an interactor complementing its function as a transcription factor, positively regulates proliferation of neural precursors, particularly in embryonic stages (Oliva et al., 2018). Its upregulation leads to a dramatic tangential enlargement of the neural ectoderm (Chenn and Walsh, 2002; Marinaro et al., 2012). On the other hand, the inhibition of the canonical Wnt signaling leads to a diminished proliferation of early neural progenitors and premature differentiation/apoptosis (Holowacz et al., 2011; Bem et al., 2019), while in both late and adult neural stem cells the effect was the opposite (Hirabayashi et al., 2004; Kuwabara et al., 2009). However, our data obtained with embryonic mid-gestation NSCs do not univocally indicate an upregulation of the Wnt pathway. The NCOR1/2 complexes have been studied in the context of neural progenitor proliferation too (Jepsen et al., 2000, 2007; Hermanson et al., 2002). In *Ncor1* KO mice the cerebral cortex is mildly affected, with an apparent increase in neuronal differentiation at E14.5 (Jepsen et al., 2000). NSCs derived from $Ncor1^{-/-}$ embryonic brain failed to form colonies, proliferated less *in vitro,* and differentiated under normal culture conditions (Hermanson et al., 2002). The same study suggested that the nuclear localization (hence the possibility to work as epigenetic modifier) is lost upon differentiation, once again supporting the importance of NCOR1 for the proliferative stage (Hermanson et al., 2002). *Ncor1* knock-down has been also used to decrease the proliferative capacity of GBM cell models, which are relatively similar to NSCs (Heldring et al., 2014). $Ncor2^{-/-}$ suggested a role for NCOR2 as a regulator of NSC state counteracting Notch activity and retinoic acid-dependent differentiation (Jepsen et al., 2007). Our data support a role of NCOR1 activity for correct NSC proliferation, implying that *Tbl1xr1* and *Ncor1* loss-of-functions lead to convergent molecular dysfunctions and cellular phenotypes. This is further supported by the fact that F10L is the only *Tbl1xr1* mutation unable to restore the proliferative capability in KO NSCs. TBL1XR1 carrying the F10L substitution is known to have a reduced interaction with the NCOR

complex, in favor of an enhanced interaction with βCATENIN (Nishi et al., 2017).

We propose that the TBL1XR1-mediated control of proliferation/differentiation dynamics is through NCOR1-mediated regulation of MAPK signaling. This is supported by the fact that *Ncor1* KO NSCs were irresponsive to FGF2, one of the external mitogens able to stimulate the MAPK cascade (Hermanson et al., 2002; Mossahebi-Mohammadi et al., 2020). It has also been shown that the loss of ERK1/2 activity is detrimental for hippocampal neural progenitors *in vivo* (Vithayathil et al., 2015), as occurs in mutant of FGF receptor mutants (Ohkubo et al., 2004). This mitogen-activated pathway is involved in cancer, including the proliferative capabilities of the most aggressive brain cancers as grade IV glioblastomas (Jimenez-Pascual and Siebzehnrubl, 2019). Gain-of-function mutations in the components of RAS/MAPK pathway have been causally connected with developmental disorders (RASopathies) that include Noonan, Cardio-facio-cutaneous, and Costello syndromes (Kim and Baek, 2019). Several neurological abnormalities including neuro-cognitive impairment, macrocephaly, and seizures characterize these diseases to different levels (Rauen, 2013). At least part of these phenotypes may be due to increased proliferation of mutant neural progenitors (Rooney et al., 2016; Pfeiffer et al., 2018; Kim and Baek, 2019). Conversely, the decreased MAPK signaling in *Tbl1xr1* KO cells may explain the tendency of these cells to differentiate, mimicking a condition of mitogen withdrawal in NSCs *in vitro* (Chojnacki and Weiss, 2008), explaining the microcephaly in both our model and human patients mutated in *TBL1XR1*. Notably, MAPK activity has also been associated with neurite outgrowth and, accordingly, *Tbl1xr1* KO neurons displayed a poorly developed dendritic tree.

We suggest that the F10L substitution in TBL1XR1 is the only mutation, among those studied here, to be similar to the TBL1XR1 loss-of-function. This is probably due to a decreased ability of TBL1XR1^{F10L} to interact with the NCOR complex (Nishi et al., 2017), leading to an aberrant MAPK regulation. This mutation was found as *de novo* in Japanese patients affected by sporadic cases of schizophrenia. Interestingly, albeit schizophrenia is a complex and heterogeneous disorder, analyses of expression variance within patients have associated this condition with the MAPK pathway (Igolkina et al., 2018). Moreover, several components of the cascade have been implicated in enhanced risk of SCZ (Xu et al., 2008).

In conclusion, we report for the first time both animal and cellular models to investigate the wide spectrum of neurological manifestations in TBL1XR1-associated disorders. We found behavioral, functional, and developmental impairments that are associated with deregulation of MAPK signaling in *Tbl1xr1* mutant mice. While further analyses will help clarify the exact impact of TBL1XR1 on adult brain functionality, our data shed new light on the molecular pathways at the crossroad of nuclear receptor activity, fundamental cellular and developmental programs, and neurological aspects of TBL1XR1-related human disorders.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the OSR Institutional Animal Care and Use Committee, Ministerial Authorization IACUC#820.

AUTHOR CONTRIBUTIONS

GM, MZ, MM, FB, and GF performed and analyzed the experiments. LM and EB performed bioinformatic analyses. MI performed behavioral tests. AB helped in the data analysis and in manuscript preparation. ST supervised and performed, with NM, the electrophysiological analyses. MR and SH generated the *Tbl1xr1* flox mouse line. AS supervised the project, conceived, and designed the experiments, analyzed the data, provided with funding, and wrote the manuscript with input from the co-authors. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Telethon (GGP15096), the Italian Ministry of Health (GR-2013-02355540 and GR-2016-02362536) and the Italian Ministry of University and Research (PRIN2017 # 2017M95WBA). SH was supported by the Helmholtz Future Topic AMPro.

ACKNOWLEDGMENTS

We thank Dr. V. Perissi for providing the coding sequence of *Tbl1xr1*. We also thank the members of the Broccoli lab for helpful discussions. The drawings were created using elements from Biorender.com.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.641410/full#supplementary-material

Supplementary Figure 1 | Behavioral deficit of Tbl1xr1 mutant mouse. **(A)** Open field arena to test spontaneous activity as total distance traveled (shown as mean + s.e.m. with dots representing individual samples, examples of traces on the left), n (adult male mice): Ctrl = 21, KO = 17: $p = 0.1392$; statistically compared using Mann-Whitney test. **(B)** Through catwalk assay we measured the duration (left) and the speed (right) of the run (shown as mean + s.e.m.), *n* (adult male mice): Ctrl = 21, KO = 17: duration $p = 0.1514$, statistically compared using *t*-test; speed $p = 0.1386$, statistically compared using *t*-test. **(C)** Thigmotaxis, *n* (adult male mice): Ctrl = 21, KO = 17: $p = 0.7853$; statistically compared using

Mann-Whitney test. **(D)** Time in the open arms during elevated plus maze test (shown as mean + s.e.m.), *n* (adult male mice): Ctrl = 10, KO = 10: $p = 0.2395$; statistically compared using Mann-Whitney test. **(E)** Working memory errors (re-entries to arms where the pellet has already been consumed) during the eight-arm radial maze test (shown as means ± SEMs in each experimental day, 1 trial/day), *n* (adult male mice): Ctrl = 21, KO = 17: Multiple comparisons: day 1: $^{***}p < 0.0001$; day 2: $p = 0.1408$; day 3: $^{****}p < 0.0001$; day 4: $^{***}p = 0.0007$; day 5: $p > 0.9999$; day 6: $p = 0.1250$; day 7: $p > 0.9999$; day 8: $p > 0.9999$; day 9: $p > 0.9999$; day 10: $p = 0.9367$; statistically compared using 2-way ANOVA and Bonferroni's *post hoc* test. **(F)** Aggressiveness of adult animals measured as the time spent to fight or bite in the resident/intruder test (shown at right as means ± SEMs, with dots representing individual samples), *n*: Ctrl = 20, KO = 19: $p = 0.1842$; statistically compared using unpaired *t*-test. **(G)** Three-chamber test quantified as the time that mouse used to stay in both the different chambers and the sections of chamber A, as well as the time the tested animals spent to sniff both cages (shown as means ± SEMs, with dots representing individual samples), *n* (adult male mice): Ctrl = 21, KO = 17: chamber preference: **(A)** $p = 0.4776$, **(B)** $p = 0.9723$, **(C)** $p = 0.1780$, two-way ANOVA and Sidak's *post hoc* test; chamber A preference: A1 $p = 0.2870$, A2 $p = 0.2230$, two-way ANOVA and Sidak's *post hoc* test; sniffing: $p = 0.3869$, unpaired *t*-test.

Supplementary Figure 2 | Defects of *Tbl1xr1* mutant neurons. **(A)** Example images and quantification of immunohistochemistry for CTIP2 (left), CUX1 (middle) and SATB2 (right) in cerebral cortex of adult mice (>2 months of age) (shown as mean + s.e.m. with dots representing individual samples): CTIP2: *n*: Ctrl = 10, KO = 10 $p = 0.1135$; CUX1: *n*: Ctrl = 10, KO = 10 $^{***}p = 0.0002$; SATB2: *n*: Ctrl = 9, KO = 9 $^{***}p = 0.0001$; statistically compared using unpaired *t*-test. **(B)** Spine quantification of DIV 15 hippocampal neurons infected with GFP lentivirus at DIV0 (shown as means ± SEMs, with dots representing individual samples), *n* (replicates): Ctrl = 7, KO = 7: $^{****}p < 0.0001$, unpaired *t*-test. **(C)** Spine quantification of DIV 21 hippocampal neurons infected with GFP lentivirus at DIV0 (shown as means ± SEMs, with dots representing individual samples), *n* (replicates): Ctrl = 9, KO = 9: $p = 0.6245$, unpaired *t*-test. **(D)** Golgi-Cox staining of brains of control and mutant adult mice (>2 months of age). Top, magnification of analyzed dendrite portions. Bottom, quantification of spine density (shown as means ± SEMs, with dots representing individual samples): *n*: Ctrl = 14 (from 2 mice), KO = 14 mice (from 2 mice), $p = 0.4076$; unpaired *t*-test. Scale bars: **(A)** 100 μm; **(B–D)** 10 μm.

Supplementary Figure 3 | Defective neural stem cells in *Tbl1xr1* mutant embryos. **(A)** Left, DAPI staining of coronal section of both control and mutant embryonic forebrains (at the stage of E14.5). Right quantification of the cortical thickness in medial (green square) and lateral (red) position (shown as means ± SEMs, with dots representing individual samples), *n* (different animals): Ctrl = 6, KO = 6: medial, $^{**}p = 0.0014$, unpaired *t*-test; lateral $p = 0.3129$, unpaired *t*-test. **(B)** Left, immunohistochemistry for PAX6 (red) and SATB2 (green) on coronal section of E14.5 mouse cortices. Right, quantification (shown as means ± SEMs, with dots representing individual samples), *n* (different animals): Ctrl = 8, KO = 8: PAX6: $^{****}p < 0.0001$; SATB2: $^{****}p < 0.0001$; statistically compared unpaired *t*-test. **(C)** Left, immunocytochemistry of both Ctrl and *Tbl1xr1* KO proliferating *in vitro* NSCs for KI67 counterstained with DAPI. Right (shown as means ± SEMs, with dots representing individual samples), *n* (biological replicates): Ctrl = 8, KO = 8: $^{****}p < 0.0001$, unpaired *t*-test. **(D)** Immunochemistry of both Ctrl and *Tbl1xr1* KO dentate gyrus for GFAP (white), DCX (red), NEUN (green), and KI67 (red) counterstained with DAPI (blue). Below, quantification (shown as means ± SEMs, with dots representing individual samples), n: (different animals) Ctrl = 3, KO = 3: GFAP: $p = 0.4436$; DCX: $p = 4346$; NEUN: $p = 0.5593$; KI67: $p = 2{,}810$; statistically compared unpaired *t*-test. Scale bars: **(A,C)** 50 μm; **(B)** 25 μm; **(D)** 100 μm.

Supplementary Figure 4 | Additional molecular defects of *Tbl1xr1* KO neural stem cells. **(A)** Western blot analysis of whole protein lysates from control and *Tbl1xr1* KO NSCs for the following proteins: NCOR1 and HDAC3 (H3 was used for normalization). On the right, the quantification of the blots for NCOR1 and HDAC3 (shown as mean + s.e.m. with dots representing individual samples): NCOR1: n (biological replicates): Ctrl = 4, KO = 5: $^{**}p = 0.0047$; HDAC3: *n*: Ctrl = 4, KO = 4: $^{**}p = 0.0020$. Statistically compared using unpaired *t*-test. **(B)** Western blot analysis of whole protein lysates from control and *Tbl1xr1* KO adult brains for the following proteins: NCOR1 and HDAC3 (H3 was used for normalization). On the right, the quantification of the blots for NCOR1 and HDAC3 (shown as mean + s.e.m. with dots representing individual samples): NCOR1: *n* (biological replicates): Ctrl = 4, KO = 4: $^{**}p = 0.0019$; HDAC3: *n* (biological replicates): Ctrl = 5, KO = 5: $^{***}p = 0.0006$. Statistically compared using unpaired *t*-test. **(C)** Heat map showing genes associated with Wnt pathway according to KEGG. **(D)** Quantification of the immunocytochemistry for PH3 in both Ctrl and *Tbl1xr1* KO proliferating NSCs in a medium either with or without LiCl in the medium (40 mM), counterstained with DAPI. *n* (biological replicates): Ctrl = 6, KO = 6: no LiCl: $^{****}p < 0.0001$; 40 mM LiCl: $p = 0.9982$; one-way ANOVA. **(E)** Quantification of the immunocytochemistry for PH3 in both Ctrl and *Tbl1xr1* KO proliferating NSCs in a medium either with or without IWR1 in the medium (10 mM), counterstained with DAPI. *n*: Ctrl = 6, KO = 6: no IWR1: $^{****}p < 0.0001$; 10 μM IWR1: $^{\#\#\#\#}p < 0.0001$; one-way ANOVA. **(F)** Validation of genes related with MAPK deregulated in mutant NSCs by RT-qPCR. Quantification (shown as mean + s.e.m. with dots representing individual samples): *Fgfr1* $^{***}p = 0.0002$, *Egfr* $^{**}p = 0.0012$, *Map3k1* $^{****}p < 0.0001$, *Map3k2* $^{***}p = 0.0001$, *Map2k2* $^{***}p = 0.0002$, *Map2k3* $^{**}p = 0.0015$, *Mapk3* $^{****}p < 0.0001$, *Mapk4* $^{**}p = 0.0015$. Statistically compared using unpaired *t*-test. **(G)** Heat map showing genes associated with MAPK1 as shown in **Figure 4F**. **(H)** Venn diagram showing the overlap between a curated list of MAPK related genes (black), publicly available ChIP-seq data for NCOR1 binding in IPSCs (GSE1817009, yellow) and publicly available ChIP-seq data for NCOR2 (a.k.a. SMRT) binding in IPSCs (GSE1817010, orange). **(I)** IGV snapshot for ChIP tracks of NCOR1 (GSE1817009, yellow) and NCOR2 (a.k.a. SMRT) (GSE1817010, orange) showing peaks near the *Map2k2* gene (in IPSC). **(J)** Scatter plot showing Homer results for motif enrichment analysis in promoters (–1,000 + 100 bp from TSS) of the downregulated genes in *Tbl1xr1* KO compared to control NSCs. **(K)** Validation of genes encoding for E2F transcription factors deregulated in mutant NSCs by RT-qPCR. Quantification (shown as mean + s.e.m. with dots representing individual samples): *E2f1* $^{****}p < 0.0001$, *E2f2* $^{****}p < 0.0001$. Statistically compared using unpaired *t*-test. See also **Supplementary Tables S3, S4**).

Supplementary Figure 5 | Additional characterization of the complementation of *Tbl1xr1* KO neural stem cells. **(A)** Growth curve of the indicated adherent *in vitro* NSCs, see also **Figure 5B**. **(B)** Left, histograms showing sphere's diameters of indicated NSCs cultured as neurospheres at the indicated days *in vitro* after disaggregation. STATS. Right, examples of microphotographs of the neurospheres of the indicated NSCs cultured at the indicated days *in vitro* after disaggregation. **(C)** Immunocytochemistry of the indicated proliferating NSCs for phosphor histone 3 (PH3) counterstained with DAPI. Quantification is shown in **Figure 5C**. **(D)** Immunocytochemistry of the indicated proliferating NSCs for TUJ1 (red) and S100b (green) counterstained with DAPI (left) and O4 counterstained with DAPI (right). Quantification (shown as mean + s.e.m. with dots representing individual samples): TUJ1: is shown in **Figure 5C**; S100b: Ctrl + LV Mock vs.: KO + LV Mock $^{****}p < 0.0001$; KO + LV WT $p = 0.5496$; KO + LV F10L $^{****}p < 0.0001$; KO + LV G70D $p = 0.2906$; KO + LV L282P $p = 0.8682$; KO + LV Y446C $p = 0.5420$; one-way ANOVA with Dunnett's multiple comparisons test. O4: Ctrl + LV Mock vs.: KO + LV Mock $^{***}p = 0.0002$; KO + LV WT $p = 0.3449$; KO + LV F10L $^{****}p < 0.0001$; KO + LV G70D $p = 0.3247$; KO + LV L282P $p = 0.6425$; KO + LV Y446C $p = 0.7304$; one-way ANOVA with Dunnett's multiple comparisons test. Scale bars: **(A)** 60 μm; **(C,D)** 50 μm.

Supplementary Table 1 | List of oligonucleotides used as primers for RT-qPCR in this work.

Supplementary Table 2 | Sheet 1: Results of DESeq2 analysis of RNAseq data between Tbl1xr1 KO and control (GFP) NSCs. Sheet 2: Lists of differentially expressed genes (DEGs) divided in down and upregulated. Sheet 3: Complete lists of Gene Ontology and KEEG pathway analyses performed on DEGs. Sheet 4: Lists as in Venn diagram in **Figure 4E**. Sheet 5: GO analysis of genes that are both upregulated and TBL1XR1 targets. Sheet 6: GO analysis of genes that are both downregulated and TBL1XR1 targets.

Supplementary Table 3 | List of Wnt signaling related genes (curated using KEGG) as in heatmap in **Supplementary Figure S4C**.

Supplementary Table 4 | Sheet 1: List of MAPK signaling related genes (curated using different databases) used in Venn diagram in **Supplementary Figure S4H**. Sheet 2: List of MAPK signaling related genes downregulated in Tbl1xr1 NSCs

NSCs as in scheme of the **Figure 4F**. Sheet 3: Lists as in Venn diagram in **Supplementary Figure S4H**.

Supplementary Table 5 | Full list of Motif Enrichment Analysis by Homer using −1,000 + 100 bp from TSS as parameters.

REFERENCES

Azari, H., Rahman, M., Sharififar, S., and Reynolds, B. A. (2010). Isolation and expansion of the adult mouse neural stem cells using the neurosphere assay. *J. Vis. Exp.* 45:2393. doi: 10.3791/2393

Bem, J., Brożko, N., Chakraborty, C., Lipiec, M. A., Koziński, K., Nagalski, A., et al. (2019). Wnt/β-catenin signaling in brain development and mental disorders: keeping TCF7L2 in mind. *FEBS Lett.* 593, 1654–1674. doi: 10.1002/1873-3468.13502

Borrell, V. (2019). Recent advances in understanding neocortical development. *F1000Res* 8:F1000FacultyRev–1791. doi: 10.12688/f1000research.20332.1

Castelo-Branco, G., Lilja, T., Wallenborg, K., Falcao, A. M., Marques, S. C., Gracias, A., et al. (2014). Neural stem cell differentiation is dictated by distinct actions of nuclear receptor corepressors and histone deacetylases. *Stem Cell Reports* 3, 502–515. doi: 10.1016/j.stemcr.2014.07.008

Chenn, A., and Walsh, C. A. (2002). Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* 297, 365–369. doi: 10.1126/science.1074192

Choi, H. K., Choi, K. C., Yoo, J. Y., Song, M., Ko, S. J., Kim, C. H., et al. (2011). Reversible SUMOylation of TBL1-TBLR1 regulates beta-catenin-mediated Wnt signaling. *Mol. Cell* 43, 203–216. doi: 10.1016/j.molcel.2011.05.027

Chojnacki, A., and Weiss, S. (2008). Production of neurons, astrocytes and oligodendrocytes from mammalian CNS stem cells. *Nat. Protoc.* 3, 935–940. doi: 10.1038/nprot.2008.55

Clément-Lacroix, P., Ai, M., Morvan, F., Roman-Roman, S., Vayssière, B., Belleville, C., et al. (2005). Lrp5-independent activation of Wnt signaling by lithium chloride increases bone formation and bone mass in mice. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17406–17411. doi: 10.1073/pnas.0505259102

Colombo, E., Giannelli, S. G., Galli, R., Tagliafico, E., Foroni, C., Tenedini, E., et al. (2006). Embryonic stem-derived versus somatic neural stem cells: a comparative analysis of their developmental potential and molecular phenotype. *Stem Cells* 24, 825–834. doi: 10.1634/stemcells.2005-0313

Ebert, D. H., Gabel, H. W., Robinson, N. D., Kastan, N. R., Hu, L. S., Cohen, S., et al. (2013). Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* 499, 341–345. doi: 10.1038/nature12348

Ferreira, T. A., Blackman, A. V., Oyrer, J., Jayabal, S., Chung, A. J., Watt, A. J., et al. (2014). Neuronal morphometry directly from bitmap images. *Nat. Methods* 11, 982–984. doi: 10.1038/nmeth.3125

Galli, R., Gritti, A., Bonfanti, L., and Vescovi, A. L. (2003). Neural stem cells: an overview. *Circ. Res.* 92, 598–608. doi: 10.1161/01.RES.0000065580.02404.F4

Heinen, C. A., Jongejan, A., Watson, P. J., Redeker, B., Boelen, A., Boudzovitch-Surovtseva, O., et al. (2016). A specific mutation in TBL1XR1 causes Pierpont syndrome. *J. Med. Genet.* 53, 330–337. doi: 10.1136/jmedgenet-2015-103233

Heldring, N., Nyman, U., Lönnerberg, P., Onnestam, S., Herland, A., Holmberg, J., et al. (2014). NCoR controls glioblastoma tumor cell characteristics. *Neuro Oncol.* 16, 241–249. doi: 10.1093/neuonc/not214

Hermanson, O., Jepsen, K., and Rosenfeld, M. G. (2002). N-CoR controls differentiation of neural stem cells into astrocytes. *Nature* 419, 934–939. doi: 10.1038/nature01156

Hirabayashi, Y., Itoh, Y., Tabata, H., Nakajima, K., Akiyama, T., Masuyama, N., et al. (2004). The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells. *Development* 131, 2791–2801. doi: 10.1242/dev.01165

Holowacz, T., Huelsken, J., Dufort, D., and van der Kooy, D. (2011). Neural stem cells are increased after loss of β-catenin, but neural progenitors undergo cell death. *Eur. J. Neurosci.* 33, 1366–1375. doi: 10.1111/j.1460-9568.2011.07632.x

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Iemolo, A., Montilla-Perez, P., Lai, I.-C., Meng, Y., Nolan, S., Wen, J., et al. (2020). A cell type-specific expression map of NCoR1 and SMRT transcriptional co-repressors in the mouse brain. *J. Comp. Neurol.* 528, 2218–2238. doi: 10.1002/cne.24886

Igolkina, A. A., Armoskus, C., Newman, J. R. B., Evgrafov, O. V., McIntyre, L. M., Nuzhdin, S. V., et al. (2018). Analysis of gene expression variance in schizophrenia using structural equation modeling. *Front. Mol. Neurosci.* 11:192. doi: 10.3389/fnmol.2018.00192

Ishizuka, T., and Lazar, M. A. (2005). The nuclear receptor corepressor deacetylase activating domain is essential for repression by thyroid hormone receptor. *Mol. Endocrinol.* 19, 1443–1451. doi: 10.1210/me.2005-0009

Iwama, K., Mizuguchi, T., Takeshita, E., Nakagawa, E., Okazaki, T., Nomura, Y., et al. (2019). Genetic landscape of Rett syndrome-like phenotypes revealed by whole exome sequencing. *J. Med. Genet.* 56, 396–407. doi: 10.1136/jmedgenet-2018-105775

Jepsen, K., Hermanson, O., Onami, T. M., Gleiberman, A. S., Lunyak, V., McEvilly, R. J., et al. (2000). Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* 102, 753–763. doi: 10.1016/s0092-8674(00)00064-7

Jepsen, K., Solum, D., Zhou, T., McEvilly, R. J., Kim, H. J., Glass, C. K., et al. (2007). SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. *Nature* 450, 415–419. doi: 10.1038/nature06270

Jimenez-Pascual, A., and Siebzehnrubl, F. A. (2019). Fibroblast growth factor receptor functions in glioblastoma. *Cells* 8:715. doi: 10.3390/cells8070715

Kim, Y. E., and Baek, S. T. (2019). Neurodevelopmental aspects of RASopathies. *Mol. Cells* 42, 441–447. doi: 10.14348/molcells.2019.0037

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.

Kong, Y., Zhou, W., and Sun, Z. (2020). Nuclear receptor corepressors in intellectual disability and autism. *Mol. Psychiatry* 25, 2220–2236. doi: 10.1038/s41380-020-0667-y

Kruusvee, V., Lyst, M. J., Taylor, C., Tarnauskaitė, Ž, Bird, A. P., and Cook, A. G. (2017). Structure of the MeCP2-TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. *Proc. Natl. Acad. Sci. U.S.A.* 114, E3243–E3250. doi: 10.1073/pnas.1700731114

Kuwabara, T., Hsieh, J., Muotri, A., Yeo, G., Warashina, M., Lie, D. C., et al. (2009). Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat. Neurosci.* 12, 1097–1105. doi: 10.1038/nn.2360

Laskowski, R. A., Tyagi, N., Johnson, D., Joss, S., Kinning, E., McWilliam, C., et al. (2016). Integrating population variation and protein structural analysis to improve clinical interpretation of missense variation: application to the WD40 domain. *Hum. Mol. Genet.* 25, 927–935. doi: 10.1093/hmg/ddv625

Leo, L. M., Almeida-Corrêa, S., Canetti, C. A., Amaral, O. B., Bozza, F. A., and Pamplona, F. A. (2014). Age-dependent relevance of endogenous 5-lipoxygenase derivatives in anxiety-like behavior in mice. *PLoS One* 9:e85009. doi: 10.1371/journal.pone.0085009

Li, J., and Wang, C. Y. (2008). TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis. *Nat. Cell Biol.* 10, 160–169. doi: 10.1038/ncb1684

Li, J. Y., Daniels, G., Wang, J., and Zhang, X. (2015). TBL1XR1 in physiological and pathological states. *Am. J. Clin. Exp. Urol.* 3, 13–23.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Lyst, M. J., Ekiert, R., Ebert, D. H., Merusi, C., Nowak, J., Selfridge, J., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* 16, 898–902. doi: 10.1038/nn.3434

MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* 17, 9–26. doi: 10.1016/j.devcel.2009.06.016

Marinaro, C., Pannese, M., Weinandy, F., Sessa, A., Bergamaschi, A., Taketo, M. M., et al. (2012). Wnt signaling has opposing roles in the developing and the adult brain that are modulated by Hipk1. *Cereb. Cortex* 22, 2415–2427. doi: 10.1093/cercor/bhr320

Martins-Neves, S. R., Paiva-Oliveira, D. I., Fontes-Ribeiro, C., Bovée, J. V. M. G., Cleton-Jansen, A. M., and Gomes, C. M. F. (2018). IWR-1, a tankyrase inhibitor,

attenuates Wnt/β-catenin signaling in cancer stem-like cells and inhibits in vivo the growth of a subcutaneous human osteosarcoma xenograft. *Cancer Lett.* 414, 1–15. doi: 10.1016/j.canlet.2017.11.004

Mossahebi-Mohammadi, M., Quan, M., Zhang, J. S., and Li, X. (2020). FGF signaling pathway: a Key regulator of stem cell pluripotency. *Front. Cell Dev. Biol.* 8:79. doi: 10.3389/fcell.2020.00079

Nikolai, B. C., Lanz, R. B., York, B., Dasgupta, S., Mitsiades, N., Creighton, C. J., et al. (2016). HER2 signaling drives DNA anabolism and proliferation through SRC-3 phosphorylation and E2F1-regulated genes. *Cancer Res.* 76, 1463–1475. doi: 10.1158/0008-5472.CAN-15-2383

Nishi, A., Numata, S., Tajima, A., Zhu, X., Ito, K., Saito, A., et al. (2017). De novo non-synonymous TBL1XR1 mutation alters Wnt signaling activity. *Sci. Rep.* 7:2887. doi: 10.1038/s41598-017-02792-z

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619–1622. doi: 10.1126/science.1227764

Ohkubo, Y., Uchida, A. O., Shin, D., Partanen, J., and Vaccarino, F. M. (2004). Fibroblast growth factor receptor 1 is required for the proliferation of hippocampal progenitor cells and for hippocampal growth in mouse. *J. Neurosci.* 24, 6057–6069. doi: 10.1523/JNEUROSCI.1140-04.2004

Oliva, C. A., Montecinos-Oliva, C., and Inestrosa, N. C. (2018). Wnt signaling in the central nervous system: new insights in health and disease. *Prog. Mol. Biol. Transl. Sci.* 153, 81–130. doi: 10.1016/bs.pmbts.2017.11.018

Papale, A., d'Isa, R., Menna, E., Cerovic, M., Solari, N., Hardingham, N., et al. (2017). Severe intellectual disability and enhanced gamma-aminobutyric acidergic synaptogenesis in a novel model of rare RASopathies. *Biol. Psychiatry* 81, 179–192. doi: 10.1016/j.biopsych.2016.06.016

Perissi, V., Aggarwal, A., Glass, C. K., Rose, D. W., and Rosenfeld, M. G. (2004). A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell* 116, 511–526. doi: 10.1016/s0092-8674(04)00133-3

Perissi, V., Scafoglio, C., Zhang, J., Ohgi, K. A., Rose, D. W., Glass, C. K., et al. (2008). TBL1 and TBLR1 phosphorylation on regulated gene promoters overcomes dual CtBP and NCoR/SMRT transcriptional repression checkpoints. *Mol. Cell* 29, 755–766. doi: 10.1016/j.molcel.2008.01.020

Perissi, V., Jepsen, K., Glass, C. K., and Rosenfeld, M. G. (2010). Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* 11, 109–123. doi: 10.1038/nrg2736

Pfeiffer, V., Götz, R., Camarero, G., Heinsen, H., Blum, R., and Rapp, U. R. (2018). Impaired neuronal maturation of hippocampal neural progenitor cells in mice lacking CRAF. *PLoS One* 13:e0192067. doi: 10.1371/journal.pone.0192067

Pons, L., Cordier, M. P., Labalme, A., Till, M., Louvrier, C., Schluth-Bolard, C., et al. (2015). A new syndrome of intellectual disability with dysmorphism due to TBL1XR1 deletion. *Am. J. Med. Genet. A.* 167A, 164–168. doi: 10.1002/ajmg.a.36759

Rauen, K. A. (2013). The RASopathies. *Annu. Rev. Genomics Hum. Genet.* 14, 355–369. doi: 10.1146/annurev-genom-091212-153523

Riehmer, V., Erger, F., Herkenrath, P., Seland, S., Jackels, M., Wiater, A., et al. (2017). A heritable microduplication encompassing TBL1XR1 causes a genomic sister-disorder for the 3q26.32 microdeletion syndrome. *Am. J. Med. Genet. A.* 173, 2132–2138. doi: 10.1002/ajmg.a.38285

Rohm, M., Sommerfeld, A., Strzoda, D., Jones, A., Sijmonsma, T. P., Rudofsky, G., et al. (2013). Transcriptional cofactor TBLR1 controls lipid mobilization in white adipose tissue. *Cell Metab.* 17, 575–585. doi: 10.1016/j.cmet.2013.02.010

Rooney, G. E., Goodwin, A. F., Depeille, P., Sharir, A., Schofield, C. M., Yeh, E., et al. (2016). Human iPS cell-derived neurons uncover the impact of increased Ras signaling in costello syndrome. *J. Neurosci.* 36, 142–152. doi: 10.1523/JNEUROSCI.1547-15.2016

Saitsu, H., Tohyama, J., Walsh, T., Kato, M., Kobayashi, Y., Lee, M., et al. (2014). A girl with West syndrome and autistic features harboring a de novo TBL1XR1 mutation. *J. Hum. Genet.* 59, 581–583. doi: 10.1038/jhg.2014.71

Sajan, S. A., Jhangiani, S. N., Muzny, D. M., Gibbs, R. A., Lupski, J. R., Glaze, D. G., et al. (2017). Enrichment of mutations in chromatin regulators in people with Rett syndrome lacking mutations in MECP2. *Genet. Med.* 19, 13–19. doi: 10.1038/gim.2016.42

Sakaguchi, Y., Uehara, T., Suzuki, H., Sakamoto, Y., Fujiwara, M., Kosaki, K., et al. (2018). Haploinsufficiency of NCOR1 associated with autism spectrum disorder, scoliosis, and abnormal palatogenesis. *Am. J. Med. Genet. A.* 176, 2466–2469. doi: 10.1002/ajmg.a.40354

Silverman, J. L., Yang, M., Lord, C., and Crawley, J. N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nat. Rev. Neurosci.* 11, 490–502. doi: 10.1038/nrn2851

Sessa, A., Fagnocchi, L., Mastrototaro, G., Massimino, L., Zaghi, M., Indrigo, M., et al. (2019). SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron*, 104, 271–289. doi: 10.1016/j.neuron.2019.07.013

Stessman, H. A., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* 49, 515–526. doi: 10.1038/ng.3792

Sun, Z., and Xu, Y. (2020). Nuclear Receptor Coactivators (NCOAs) and Corepressors (NCORs) in the Brain. *Endocrinology* 161, 1–12. doi: 10.1210/endocr/bqaa083

Thapar, A., Cooper, M., and Rutter, M. (2017). Neurodevelopmental disorders. *Lancet Psychiatry* 4, 339–346. doi: 10.1016/S2215-0366(16)30376-5

Vaqueiro, A. C., de Oliveira, C. P., Cordoba, M. S., Versiani, B. R., de Carvalho, C. X., Alves Rodrigues, P. G., et al. (2018). Expanding the spectrum of TBL1XR1 deletion: report of a patient with brain and cardiac malformations. *Eur. J. Med. Genet.* 61, 29–33. doi: 10.1016/j.ejmg.2017.10.008

Vithayathil, J., Pucilowska, J., Goodnough, L. H., Atit, R. P., and Landreth, G. E. (2015). Dentate Gyrus development requires ERK activity to maintain progenitor population and MAPK pathway feedback regulation. *J. Neurosci.* 35, 6836–6848. doi: 10.1523/JNEUROSCI.4196-14.2015

Wang, T., Guo, H., Xiong, B., Stessman, H. A., Wu, H., Coe, B. P., et al. (2016). De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nat Commun.* 7:13316. doi: 10.1038/ncomms13316

Yoon, H. G., Chan, D. W., Huang, Z. Q., Li, J., Fondell, J. D., Qin, J., et al. (2003). Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1. *EMBO J.* 22, 1336–1346. doi: 10.1093/emboj/cdg120

Yoon, H. G., Choi, Y., Cole, P. A., and Wong, J. (2005). Reading and function of a histone code involved in targeting corepressor complexes for repression. *Mol. Cell Biol.* 25, 324–335. doi: 10.1128/MCB.25.1.324-335.2005

Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat. Genet.* 40, 880–885. doi: 10.1038/ng.162

Zhou, W., He, Y., Rehman, A. U., Kong, Y., Hong, S., Ding, G., et al. (2019). Loss of function of NCOR1 and NCOR2 impairs memory through a novel GABAergic hypothalamus-CA3 projection. *Nat. Neurosci.* 22, 205–217. doi: 10.1038/s41593-018-0311-1

Zhuang, Q., Li, W., Benda, C., Huang, Z., Ahmed, T., Liu, P., et al. (2018). NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. *Nat. Cell Biol.* 20, 400–412. doi: 10.1038/s41556-018-0047-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 01 March 2021
doi: 10.3389/fcell.2021.643055

Epigenetic Regulation of Cell-Fate Changes That Determine Adult Liver Regeneration After Injury

*Luigi Aloia**

MRC Laboratory for Molecular Cell Biology, University College London, London, United Kingdom

The adult liver has excellent regenerative potential following injury. In contrast to other organs of the body that have high cellular turnover during homeostasis (e.g., intestine, stomach, and skin), the adult liver is a slowly self-renewing organ and does not contain a defined stem-cell compartment that maintains homeostasis. However, tissue damage induces significant proliferation across the liver and can trigger cell-fate changes, such as trans-differentiation and de-differentiation into liver progenitors, which contribute to efficient tissue regeneration and restoration of liver functions. Epigenetic mechanisms have been shown to regulate cell-fate decisions in both embryonic and adult tissues in response to environmental cues. Underlying their relevance in liver biology, expression levels and epigenetic activity of chromatin modifiers are often altered in chronic liver disease and liver cancer. In this review, I examine the role of several chromatin modifiers in the regulation of cell-fate changes that determine efficient adult liver epithelial regeneration in response to tissue injury in mouse models. Specifically, I focus on epigenetic mechanisms such as chromatin remodelling, DNA methylation and hydroxymethylation, and histone methylation and deacetylation. Finally, I address how altered epigenetic mechanisms and the interplay between epigenetics and metabolism may contribute to the initiation and progression of liver disease and cancer.

Keywords: plasticity, cell-fate change, epigenetics, liver cancer, chronic liver disease, liver regeneration, liver progenitor, metabolism

OPEN ACCESS

Edited by:
Justin Brumbaugh,
University of Colorado Boulder,
United States

Reviewed by:
Kirsten Sadler,
New York University Abu Dhabi,
United Arab Emirates
Wei-Yu Lu,
University of Birmingham,
United Kingdom

***Correspondence:**
Luigi Aloia
l.aloia@ucl.ac.uk

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *17 December 2020*
Accepted: *05 February 2021*
Published: *01 March 2021*

Citation:
Aloia L (2021) Epigenetic Regulation of Cell-Fate Changes That Determine Adult Liver Regeneration After Injury. Front. Cell Dev. Biol. 9:643055. doi: 10.3389/fcell.2021.643055

INTRODUCTION

The adult liver has extraordinary regenerative potential. This knowledge dates back to the myth of Prometheus, who was condemned to an eternal punishment caused by an eagle eating his liver, which would replenish itself every day. Indeed, after partial hepatectomy, which leads to removal of up to 70% of the liver, liver mass and functions become significantly restored within a few days. Mechanisms underlying restoration of liver mass and function in response to hepatectomy have been elegantly reviewed elsewhere (Michalopoulos and DeFrances, 1997; Michalopoulos, 2007). In this review, I focus on adult liver regeneration in response to tissue damage that can normally challenge the liver, in line with its central role in nutrient metabolism and detoxification in the adult body.

The adult liver is a slowly self-renewing organ and does not exhibit a defined stem-cell compartment that maintains homeostasis, opposite to other organs that have excellent regenerative capacity, including intestine, stomach, and skin. A great number of experimental models indicate

that cellular plasticity, intended as the capacity of liver cells to adopt alternative fates and functions in response to environmental changes, determines efficient adult liver regeneration following damage (Tetteh et al., 2015; Aloia et al., 2016; Kopp et al., 2016; Gadd et al., 2020). A key question is: what molecular mechanisms regulate cellular plasticity and cell-fate changes to achieve efficient liver regeneration upon injury? In 1957, Conrad Hal Waddington postulated that the epigenetic landscape contributes to determine cell-fate decisions and acquisition of cell-identity during embryo development (Waddington, 1957). Increasing evidence indicates that the epigenetic landscape is dynamically regulated in the adulthood and allows cell-fate changes and cellular plasticity in adult tissues in response to environmental cues and tissue injury (Takahashi and Yamanaka, 2015; Rajagopal and Stanger, 2016; Flavahan et al., 2017; Paksa and Rajagopal, 2017). Of note, here I refer to epigenetic mechanisms as modifications of the chromatin (the ensemble of DNA and histones), which are involved in the regulation of gene transcription. Importantly, epigenetic mechanisms are inheritable by the cell progeny and reversible in response to environmental changes. One intriguing hypothesis is that dynamic regulation of epigenetic mechanisms might represent an efficient way for liver cells to respond to injury, repair the damaged tissue, and return to the homeostatic state once the injury is resolved. Supporting this hypothesis, our data indicated that transient epigenetic remodelling, mediated by the methylcytosine dioxygenase TET1, triggers liver cellular plasticity and contributes to liver regeneration following injury (Aloia et al., 2019). Here, I review the role of several chromatin modifiers in the regulation of cell-fate changes required for liver epithelial regeneration after injury in mouse models (**Table 1**). In addition, I address the interplay between epigenetics and metabolism in liver regeneration and chronic liver disease and discuss how epigenetic alterations may contribute to liver disease and cancer.

HEPATOCYTES AND CHOLANGIOCYTES EXHIBIT REMARKABLE PLASTICITY IN RESPONSE TO LIVER INJURY

Two epithelial cell types are present in the adult liver: (i) hepatocytes, which represent >60% of total liver cells and perform most of the metabolic and detoxification functions of the organ and (ii) cholangiocytes, which represent 3–5% of total liver cells and form biliary ducts that collect the bile produced by the hepatocytes and export it to the intestine for digestive purposes (Ober and Lemaigre, 2018). Both hepatocytes and cholangiocytes exhibit remarkable plasticity in response to injury. The role of epithelial plasticity in liver regeneration has been elegantly reviewed by Stuart Forbes and colleagues (Gadd et al., 2020).

Increasing evidence indicates that the vast majority of the hepatocytes, if not all, can acquire proliferation to ensure efficient restoration of liver mass and functions (Chen et al., 2020; Matsumoto et al., 2020; Monga, 2020; Sun et al., 2020). Upon liver injury induced by chemicals or toxins, hepatocytes can trans-differentiate into cholangiocytes (Yanger et al., 2013) or de-differentiate into bipotent liver progenitors, capable of generating both liver epithelial cell types (Tarlow et al., 2014; Yimlamai et al., 2014; Han et al., 2019). Liver injury can also trigger cholangiocyte proliferation and expansion in the liver parenchyma, a process known as ductular reaction (Roskams et al., 2004). Importantly, cholangiocytes can de-differentiate *in vivo* into bipotent liver progenitors that give rise to both hepatocytes and cholangiocytes (Itoh and Miyajima, 2014; Ko et al., 2020; So et al., 2020). The role of cholangiocytes in liver regeneration has been extensively debated. Several experimental models indicated that hepatocytes are solely responsible for restoration of the epithelial compartment after liver injury and the contribution of cholangiocytes is neglectable (Malato et al., 2011; Schaub et al., 2014; Yanger et al., 2014; Font-Burgada et al., 2015; Lin et al., 2018). However, different mouse models demonstrated that cholangiocytes significantly contribute to liver regeneration when hepatocytes are compromised. Data from Lu et al. (2015) indicated that cholangiocytes have liver repopulation capacity upon transplantation into adult mouse liver that shows extensive hepatocyte senescence and necrosis induced by genetic deletion of the E3 ubiquitin ligase *Mdm2*. Moreover, lineage-tracing experiments demonstrated that cholangiocytes can restore from 10 to 70% of total liver hepatocytes following diet-induced liver injury in mouse models exhibiting genetic deletion of β*1-integrin* (Raven et al., 2017) or β-*catenin* (Russell et al., 2018) or over-expression of p21 (Raven et al., 2017) in hepatocytes or after prolonged liver injury in the absence of genetic alterations that might compromise hepatocyte functions (Deng et al., 2018; Manco et al., 2019).

The capacity of both hepatocytes and cholangiocytes to acquire a bipotent progenitor state relies on their origin from a common embryonic progenitor, the hepatoblast (Zaret and Grompe, 2008; Ober and Lemaigre, 2018). Of note, chromatin modifications such as trimethylation of lysine 9 and 27 of histone H3 (H3K9me3 and H3K27me3) play a crucial role in the establishment of hepatoblast identity (Nicetto et al., 2019), supporting the relevance of epigenetic mechanisms in liver cell-identity and cell-fate decisions. In the following sections, I examine the role of several epigenetic modifiers in liver regeneration mediated by adult hepatocytes and cholangiocytes following tissue damage.

EPIGENETIC MECHANISMS REGULATING HEPATOCYTE-MEDIATED LIVER REGENERATION

The hepatocytes play a crucial role in adult liver regeneration after injury (Malato et al., 2011; Schaub et al., 2014; Yanger et al., 2014; Font-Burgada et al., 2015; Lin et al., 2018). Arid1a, a DNA interacting subunit of the SWI/SNF chromatin remodelling complex, has been shown to regulate hepatocyte plasticity in mouse models carrying liver-specific deletion of *Arid1a* (by using an *Albumin*-Cre/*Arid1a* flx/flx mouse line) or conditional deletion of *Arid1a* in hepatocytes (mediated by infection with an AAV-Cre in *Arid1a* flx/flx mice) (Sun et al., 2016; Li et al., 2019). Sun et al. (2016) found that *Arid1a* deletion increased hepatocyte proliferation and reduced necrosis, inflammation, and fibrosis

TABLE 1 | Role of selected chromatin modifiers in liver epithelial regeneration following tissue injury.

Name	Function	Liver cell type	Role in liver epithelial regeneration upon injury
ARID1A *AT-Rich Interaction Domain 1A*	DNA interacting subunit of the chromatin remodelling complex SWI/SNF	Hepatocyte	• Blocks proliferation and stabilises the differentiated state (Sun et al., 2016). • Facilitates YAP-dependent transcriptional activity and induction of bipotent liver progenitors (Li et al., 2019).
DNMT1 *DNA (cytosine-5)-methyltransferase 1*	DNA methyltransferase (maintenance of 5-methycytosine during replication)	Hepatocyte	• Protects from DNA damage and senescence. Deletion causes differentiation defects (Kaji et al., 2016).
EZH 1/2 *Enhancer of zeste homolog 1/2*	Enzymatic subunits of the Polycomb Repressive Complex 2, PRC2	Hepatocyte	• Repress cell-cycle inhibitors (Bae et al., 2015).
HDAC1 *Histone* deacetylase 1	Histone deacetylase	Cholangiocyte	• Promotes differentiation of bipotent liver progenitors into hepatocytes, repressing the expression of Sox9 (Ko et al., 2019).
TET1 *Ten-eleven translocation methylcytosine dioxygenase 1*	Oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)	Cholangiocyte	• Enables de-differentiation into bipotent liver progenitors, promoting the expression of stem-cell and regenerative genes, including members and targets of the YAP signalling pathway (Aloia et al., 2019).
UHRF1 *Ubiquitin Like With PHD And Ring Finger Domains 1*	Recruitment of DNMT1 to chromatin; Binding to H3K9me3	Hepatocyte	• Depletion results in epigenetic compensation that facilitates the induced expression of cell-cycle and regenerative genes (Wang et al., 2019).

in response to liver damage induced by carbon tetrachloride, CCl4. Moreover, *Arid1a* deletion resulted in accelerated weight recovery and extended survival following treatment with the toxic diet 3,5-diethoxycarbonyl-1,4-dihydrocollidine, DDC (6 weeks). Depletion of Arid1a following damage, in the recovering liver after a 12-week treatment with CCl4 or a 4-week treatment with DDC, reduced fibrosis and accelerated regain of normal weight. Thus, these data indicated that Arid1a deletion facilitated liver recovery from damage. At the molecular level, Sun et al. (2016) showed that Arid1a genomic binding might facilitate chromatin accessibility to key transcription factors that repress proliferation (e.g., E2F4) and maintain the hepatocyte differentiated cell state, such as C/EBPa and Hnf4a (Sun et al., 2016). Together, these findings suggested that Arid1a might impair liver regeneration by stabilising the hepatocyte differentiated state and reducing hepatocyte plasticity (Sun et al., 2016). This hypothesis is challenged by the findings of Li et al. (2019), who showed that conditional *Arid1a* deletion resulted in a significant reduction of $Hnf4a^{+}/Sox9^{+}$ bipotent liver progenitors in response to DDC treatment (2 weeks). At the molecular level, they found that Arid1a determined increased chromatin accessibility in the homeostatic liver, which facilitates YAP genomic binding after liver injury. YAP transcriptional activity was previously shown to promote hepatocyte de-differentiation into liver progenitors (Yimlamai et al., 2014). Consistently, Li et al. (2019) found that the number of $YAP^{+}/Sox9^{+}$ liver progenitors was significantly reduced in *Arid1a*-depleted mice upon over-expression of a constitutively active YAPS127A. Of note, they found that *Arid1a* deletion resulted in hepatomegaly, hepatocyte hypertrophy, and impaired liver functions upon DDC-mediated liver damage, whereas its depletion following DDC-induced damage did not affect liver recovery (Li et al., 2019). The different role of Arid1a in liver regeneration and recovery after injury reported by Sun et al. (2016) and Li et al. (2019) could be explained by the different length of DDC treatment and recovery, which might result in different levels of tissue injury at the time of the analyses performed (4–6 weeks DCC followed by 0–10 days recovery vs 2 weeks DDC followed by 8 weeks recovery, respectively). The different length of injury/recovery might result in a different timing of Arid1a binding to DNA and SWI/SNF-mediated chromatin remodelling activity. This, in turn, might influence the induction of hepatocyte de-differentiation into progenitors and their return to the homeostatic state. Of note, in their experimental conditions, Sun et al. (2016) reported increased hepatocyte proliferation following *Arid1a* depletion, whereas Li et al. (2019) did not observe a significant difference. This has important consequences on hepatocyte plasticity, since increased proliferation of mature hepatocytes is likely to determine faster regeneration, thus reducing the requirement of hepatocyte de-differentiation into progenitors.

To achieve efficient liver regeneration, it is crucial that adult hepatocytes maintain their functions during proliferation and cell division. Bae et al. (2015) found that EZH1 and EZH2 regulated hepatocyte proliferation and promoted increased mouse survival following CCl4. EZH1/2 are the catalytic subunits of the Polycomb repressive complex 2 (PRC2), which result in trimethylation of lysine 27 of histone 3 (H3K27me3), an epigenetic mark mainly associated to gene repression. Bae et al. (2015) found that key target genes of EZH1/2 are the cell-cycle inhibitors *Cdkn2a* and *Cdkn2b*, which became up-regulated in *Ezh1/2* knock-out (KO) livers. Consistent with the importance of hepatocyte proliferation and division in liver regeneration, epigenetic regulation mediated by the maintenance DNA methyltransferase DNMT1 and its regulator UHRF1 has been shown to play an important role in hepatocyte viability and regenerative capacity (Kaji et al., 2016; Wang et al., 2019). Kaji et al. (2016) and Wang et al. (2019) performed liver-specific deletion of *Dnmt1* and *Uhrf1*, respectively, crossing *Albumin*-Cre with specific flx/flx mouse lines. Wang et al. (2019) showed that *Uhrf1* deletion resulted in global hypomethylation, which induced redistribution of repressive H3K27me3 from promoters to transposable elements to block their aberrant

expression. This has implications in liver regeneration, since reduced H3K27me3 levels at promoter facilitated the induced expression of cell-cyle and regenerative genes (Wang et al., 2019). Kaji et al. (2016) showed that *Dnmt1* deletion induced liver fibrosis and inflammation and resulted in hepatocyte senescence. Intriguingly, *Dnmt1* deletion resulted in maturation defects and lack of expression of the cytochrome *Cyp2e1*, which is responsible for the bioactivation of CCl4, thus making *Dnmt1* KO mice resistant to liver injury induced by CCl4 (Kaji et al., 2016). Of note, Bae et al. (2015) and Kaji et al. (2016) showed that impaired hepatocyte regenerative capacity or tissue damage caused by hepatocyte senescence and necrosis triggered ductular reaction.

These findings highlight the relevance of epigenetic mechanisms in the regulation of hepatocyte viability, proliferation, and plasticity. Further work will determine how dynamic epigenetic regulation influences the balance between proliferation of mature hepatocytes and hepatocyte de-differentiation into progenitors and how those two processes act to coordinate liver response to injury to achieve efficient regeneration.

EPIGENETIC MECHANISMS REGULATING CHOLANGIOCYTE-MEDIATED LIVER REGENERATION

Cholangiocyte expansion and de-differentiation into liver progenitors have been extensively reported in a variety of injury models in mammals (Okabe et al., 2009; Sackett et al., 2009; Dorrell et al., 2011; Shin et al., 2011; Espanol-Suner et al., 2012; Huch et al., 2013). Lineage-tracing experiments in mouse models demonstrated a crucial role for cholangiocytes in adult liver regeneration when hepatocyte response to injury is compromised by genetic alterations or following severe and prolonged liver damage (Raven et al., 2017; Deng et al., 2018; Russell et al., 2018; Manco et al., 2019).

What mechanisms regulate cholangiocyte plasticity? Our data demonstrated that upon liver injury, cholangiocytes undergo epigenetic remodelling mediated by the methylcytosine dioxygenase TET1 (Aloia et al., 2019). TET1/2/3 oxidise the repressive DNA mark 5-methylcytosine into 5-hydroxymethylcytosine (5hmC). 5hmC can act as a stable epigenetic mark associated with gene activation or represent an intermediate of complete de-methylation after further oxidation catalysed by TET proteins (Rasmussen and Helin, 2016). We found that TET1 was lowly expressed in homeostatic adult cholangiocytes and was up-regulated following DDC-mediated liver injury *in vivo* vs healthy liver, opposite to *Tet2/3*, which became down-regulated. By using a *Tet1* hypomorphic mouse, we found that TET1 reduced levels impaired cholangiocyte proliferation after acute DDC treatment (5 days) and resulted in liver fibrosis after prolonged DDC-mediated liver injury (∼8 weeks including off-treatment intervals) (Aloia et al., 2019). Lineage-tracing experiments demonstrated that TET1 was required for the formation of cholangiocyte-derived hepatocyte regenerative clusters following DDC-mediated liver injury (Aloia et al., 2019), using a liver injury model developed by Raven et al. (2017). Together, our findings indicated that TET1 triggers cholangiocyte plasticity in response to liver injury and is required for efficient liver epithelial regeneration. At the molecular level, we found that TET1 regulates the expression of stem-cell genes that identify hepatoblasts (e.g., *Lgr5*; Prior et al., 2019) and adult human and mouse bipotent liver progenitors (e.g., *Trop2*; Okabe et al., 2009; Aizarani et al., 2019). Consistently, TET1 was required for the establishment of mouse intrahepatic cholangiocyte organoids, which exhibit induced expression of stem-cell genes (e.g., *Lgr5* and *Trop2*) and can differentiate into hepatocytes (Huch et al., 2013). Moreover, genome-wide TET1 DamID-sequencing (van den Ameele et al., 2019) in intrahepatic cholangiocyte organoids revealed that TET1 regulates the expression of components and targets of signalling pathways required for liver regeneration, including YAP (Yimlamai et al., 2014; Pepe-Mooney et al., 2019; Planas-Paz et al., 2019). Together, our data suggested that TET1 enables de-differentiation of cholangiocytes into bipotent liver progenitors both *in vitro and in vivo* to ensure efficient liver regeneration and organoid formation (Aloia et al., 2019).

Of note, the increase in *Tet1* levels after liver injury *in vivo* (day 3 following DDC treatment) was only transient, since *Tet1* expression decreased at the peak of DDC-mediated liver damage (day 5). In line with this, we found that 5hmC levels were transiently increased in cholangiocytes *in vivo* at day 3 at the transcriptional start site of >3000 genes, which are mainly involved in biological processes such as development, chromatin modifications, and cell-cycle (Aloia et al., 2019). Therefore, this suggested that dynamic epigenetic remodelling *in vivo* allows cholangiocytes to both respond to liver injury and return to the homeostatic state to avoid the detrimental effects of an aberrant progenitor response. Supporting this hypothesis, $Lgr5^{+}$ cells have been identified as tumour initiating cells in mouse liver cancer (Cao et al., 2020) and YAP can induce liver cancer (Dong et al., 2007).

What epigenetic mechanisms regulate cholangiocyte differentiation into hepatocytes? Following experiments in zebrafish that led to the identification of a regenerative role for Hdac1 by repressing the expression of Sox9 and modulating the Notch signalling, Ko et al. (2019) showed that the HDAC inhibitor MS-275 impaired cholangiocyte-mediated hepatocyte regeneration in a mouse model developed by Russell et al. (2018). Of note, MS-275 treatment did not affect cholangiocyte proliferation and ductular reaction (Ko et al., 2019), thus suggesting that Hdac1 specifically regulates cholangiocyte differentiation into hepatocytes without affecting cholangiocyte response to liver injury.

These findings highlight the relevance of dynamic epigenetic regulation of cholangiocyte plasticity for liver regeneration in response to severe or persistent liver injury that compromises the hepatocytes. Of note, ductular reaction is often observed in human chronic liver disease, concomitantly with massive hepatocyte alterations and necrosis (Ko et al., 2020). Therefore, understanding the epigenetic mechanisms that regulate

cholangiocyte plasticity might indicate novel therapeutic strategies aimed at stimulating cholangiocyte regenerative capacity to ameliorate human chronic liver disease.

THE INTERPLAY BETWEEN EPIGENETICS AND METABOLISM IN LIVER REGENERATION AND DISEASE

Epigenetic mechanisms are often regulated by metabolic inputs (Etchegaray and Mostoslavsky, 2016). For example, TET1 epigenetic activity is dependent on 2-oxoglutarate, which is an intermediate of the Krebs cycle (Rasmussen and Helin, 2016). The adult liver is the central organ of the body for nutrient metabolism and presents metabolic zonation with specialised metabolic functions. Liver zonation follows decreasing oxygen and nutrient gradient across the liver lobule from the portal triad (formed by the hepatic artery, portal vein, and biliary ducts), to the central vein, where the blood meets the systemic circulation (Kietzmann, 2017). This suggests that different types and durations of liver injuries may affect different metabolic zones and trigger specific epigenetic mechanisms, which, in turn, promote distinct cellular response to damage. Importantly, metabolic alterations might cause chronic liver disease. For example, metabolic disorders are associated with the most common form of human chronic liver disease, non-alcoholic fatty liver disease (NAFLD). Patients with NAFLD exhibit increased fat accumulation and can present significant inflammation, named as non-alcoholic steatohepatitis (NASH) (Benedict and Zhang, 2017). At late stages, chronic liver disease is characterised by cirrhosis, which accounts for >1 million deaths per year worldwide and can predispose to liver cancer (Asrani et al., 2019). Therefore, it is reasonable to speculate that metabolic dysfunctions induce epigenetic alterations that, in turn, might impair liver regeneration capacity and exacerbate liver disease and favour its progression towards liver cirrhosis. The interplay between epigenetics and metabolism is bidirectional, since altered epigenetic mechanisms may also cause metabolic dysfunctions and trigger liver disease. A circadian rhythm was shown to determine chromatin recruitment of the histone deacetylase Hdac3, which in turn controls the expression of genes related to lipid metabolism, and determines the balance between gluconeogenesis and lipid synthesis (Feng et al., 2011; Sun et al., 2012). Consistently, *Hdac3* deletion in mouse resulted in hepatic steatosis (Feng et al., 2011). In addition, Arid1a was found to regulate lipogenesis and fatty acid oxidation and its

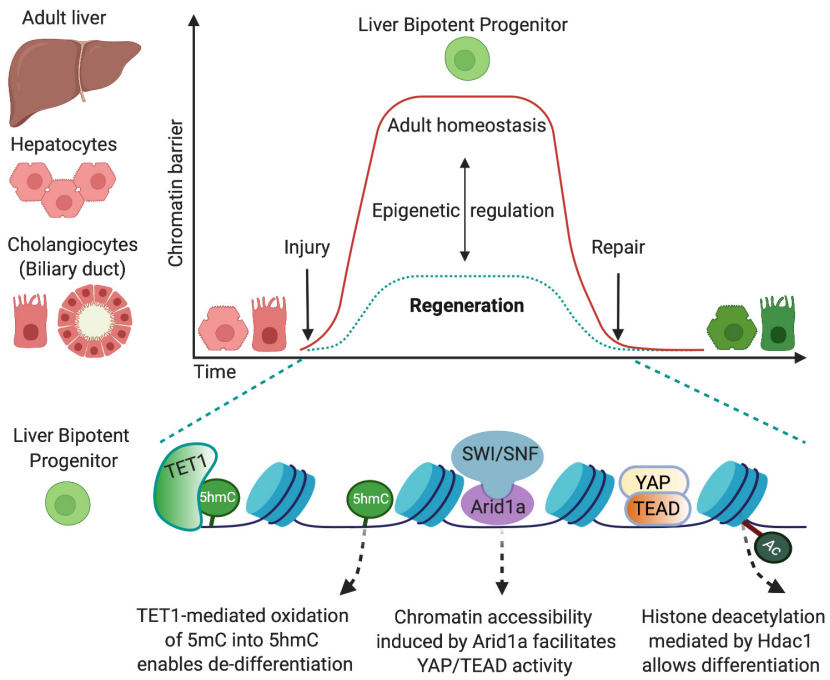


FIGURE 1 | Epigenetic mechanisms that allow cell-fate changes into bipotent liver progenitors. The adult liver is formed by two epithelial cell types, hepatocytes and cholangiocytes, which derive from a common bipotent embryonic progenitor, the hepatoblast. In the adulthood, in homeostatic conditions, the epigenetic landscape preserves cell-identity by maintaining chromatin conformations that stabilise the differentiated state (red). Upon injury, both adult hepatocytes and cholangiocytes can de-differentiate into bipotent liver progenitors that give rise to both liver epithelial cell types (green) and restore the liver epithelial compartment. This cell-fate change into progenitors is enabled by epigenetic mechanisms that determine permissive chromatin states, including increased chromatin accessibility mediated by Arid1a and the chromatin remodelling complex SWI/SNF (Li et al., 2019) and oxidation of 5-methycytosine (5mC) into 5-hydromethylcytosyne (5hmC) mediated by the methylcytosine dioxygenase TET1 (Aloia et al., 2019). Such permissive states facilitate the chromatin binding of the transcriptional machinery (e.g., YAP/TEAD) that promotes the establishment of liver progenitor identity. Importantly, dynamic epigenetic regulation allows liver epithelial cells to return to the homeostatic state once the injury is resolved and the tissue is repaired. In this regard, TET1 and 5hmC levels are only transiently enriched in cholangiocytes at early stages upon liver injury (Aloia et al., 2019), and histone deacetylation mediated by Hdac1 is required for differentiation of cholangiocyte progenitors into hepatocytes (Ko et al., 2019).

deletion was implicated in NASH in mouse (Fang et al., 2015; Moore et al., 2019).

Together, this suggests that investigating the complex interplay between epigenetics and metabolism will represent an important step to understand how the liver achieves efficient regeneration and identify the molecular mechanisms that influence initiation and progression of chronic liver disease.

DISCUSSION

In this review, I have examined the role of several chromatin modifiers that regulate epithelial cell-fate changes to achieve efficient liver regeneration after injury in mouse models (**Figure 1**). It is worth mentioning that whether mouse models allow recapitulating several aspects of human liver regeneration and disease, species-specific difference should be taken in account, and the epigenetic mechanisms identified in mouse models should be validated also in human models such as cell lines, organoid systems, or primary specimens.

The expression levels of chromatin modifiers and genomic distribution of their epigenetic marks are often altered in human chronic liver disease and liver cancer (Hardy and Mann, 2016). For example, loss of function mutations in *ARID1A* have been reported in >16% of human hepatocellular carcinoma (Guichard et al., 2012). Arid1a was shown to have a dual role as both oncogene and tumour suppressor in mouse, supporting the hypothesis that epigenetic mechanisms determine specific cellular phenotypes according to the context of the surrounding tissue. Specifically, Arid1a triggered liver cancer initiation by increasing reactive oxygen species, whereas its loss in pre-existing tumours promoted cancer growth and metastasis (Sun et al., 2017). UHRF1 is often over-expressed in human liver cancer and high UHRF1 levels are associated with a poor prognosis. Interestingly, UHRF1 over-expression was shown to induce hypomethylation in human liver cancer (Mudbhary et al., 2014), as observed upon its depletion in mouse, where it induced epigenetic compensation (Wang et al., 2019). Further studies will determine the relevance of dynamic epigenetic remodelling in human liver cancer.

Pharmacological interventions based on drugs targeting epigenetic activity have a huge potential for the development of therapeutic strategies in human disease and cancer (Cheng et al., 2019). Therefore, advances in our knowledge of the epigenetic mechanisms that regulate liver plasticity might open new avenues for the treatment of chronic liver disease and liver cancer. In this regard, 3D liver organoid cultures derived from adult liver biopsies or induced pluripotent stem cells have proven reliable models of homeostatic and regenerative cholangiocytes (Huch et al., 2013, 2015; Sampaziotis et al., 2017; Aloia et al., 2019; Rimland et al., 2020) and hepatocytes (Hu et al., 2018; Peng et al., 2018), steatohepatitis (Ouchi et al., 2019; Ramli et al., 2020), and primary liver cancer (Broutier et al., 2017; Nuciforo et al., 2018), thus representing promising platforms to uncover the molecular mechanisms underlying liver regeneration, disease, and cancer.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by MRC core funding to the MRC Laboratory for Molecular Cell Biology at University College London, award code (MC_U12266B).

ACKNOWLEDGMENTS

LA thanks Laura Broutier and Serena Mirra for critical reading of the manuscript. The figure was created with BioRender.com.

REFERENCES

Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J. S., et al. (2019). A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* 572, 199–204. doi: 10.1038/s41586-019-1373-2

Aloia, L., McKie, M. A., and Huch, M. (2016). Cellular plasticity in the adult liver and stomach. *J. Physiol.* 594, 4815–4825. doi: 10.1113/jp271769

Aloia, L., McKie, M. A., Vernaz, G., Cordero-Espinoza, L., Aleksieva, N., van den Ameele, J., et al. (2019). Epigenetic remodelling licences adult cholangiocytes for organoid formation and liver regeneration. *Nat. Cell Biol.* 21, 1321–1333. doi: 10.1038/s41556-019-0402-6

Asrani, S. K., Devarbhavi, H., Eaton, J., and Kamath, P. S. (2019). Burden of liver diseases in the world. *J. Hepatol.* 70, 151–171. doi: 10.1016/j.jhep.2018.09.014

Bae, W. K., Kang, K., Yu, J. H., Yoo, K. H., Factor, V. M., Kaji, K., et al. (2015). The methyltransferases enhancer of zeste homolog (EZH) 1 and EZH2 control hepatocyte homeostasis and regeneration. *FASEB J.* 29, 1653–1662. doi: 10.1096/fj.14-261537

Benedict, M., and Zhang, X. (2017). Non-alcoholic fatty liver disease: an expanded review. *World J. Hepatol.* 9, 715–732. doi: 10.4254/wjh.v9.i16.715

Broutier, L., Mastrogiovanni, G., Verstegen, M. M., Francies, H. E., Gavarro, L. M., Bradshaw, C. R., et al. (2017). Human primary liver cancer-derived organoid cultures for disease modeling and drug screening. *Nat. Med.* 23, 1424–1435. doi: 10.1038/nm.4438

Cao, W., Li, M., Liu, J., Zhang, S., Noordam, L., Verstegen, M. M. A., et al. (2020). LGR5 marks targetable tumor-initiating cells in mouse liver cancer. *Nat. Commun.* 11:1961.

Chen, F., Jimenez, R. J., Sharma, K., Luu, H. Y., Hsu, B. Y., Ravindranathan, A., et al. (2020). Broad distribution of hepatocyte proliferation in liver homeostasis and regeneration. *Cell Stem Cell* 26, 27–33.e4.

Cheng, Y., He, C., Wang, M., Ma, X., Mo, F., Yang, S., et al. (2019). Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct. Target. Ther.* 4:62.

Deng, X., Zhang, X., Li, W., Feng, R. X., Li, L., Yi, G. R., et al. (2018). Chronic liver injury induces conversion of biliary epithelial cells into hepatocytes. *Cell Stem Cell* 23, 114–122.e3.

Dong, J., Feldmann, G., Huang, J., Wu, S., Zhang, N., Comerford, S. A., et al. (2007). Elucidation of a universal size-control mechanism in *Drosophila* and mammals. *Cell* 130, 1120–1133. doi: 10.1016/j.cell.2007.07.019

Dorrell, C., Erker, L., Schug, J., Kopp, J. L., Canaday, P. S., Fox, A. J., et al. (2011). Prospective isolation of a bipotential clonogenic liver progenitor cell in adult mice. *Genes Dev.* 25, 1193–1203. doi: 10.1101/gad.2029411

Espanol-Suner, R., Carpentier, R., Van Hul, N., Legry, V., Achouri, Y., Cordi, S., et al. (2012). Liver progenitor cells yield functional hepatocytes in response to chronic liver injury in mice. *Gastroenterology* 143, 1564–1575.e7.

Etchegaray, J. P., and Mostoslavsky, R. (2016). Interplay between metabolism and epigenetics: a nuclear adaptation to environmental changes. *Mol. Cell* 62, 695–711. doi: 10.1016/j.molcel.2016.05.029

Fang, J. Z., Li, C., Liu, X. Y., Hu, T. T., Fan, Z. S., and Han, Z. G. (2015). Hepatocyte-specific arid1a deficiency initiates mouse steatohepatitis and hepatocellular carcinoma. *PLoS One.* 10:e0143042. doi: 10.1371/journal.pone.0143042

Feng, D., Liu, T., Sun, Z., Bugge, A., Mullican, S. E., Alenghat, T., et al. (2011). A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* 331, 1315–1319. doi: 10.1126/science.1198125

Flavahan, W. A., Gaskell, E., and Bernstein, B. E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science* 357:eaal2380. doi: 10.1126/science.aal2380

Font-Burgada, J., Shalapour, S., Ramaswamy, S., Hsueh, B., Rossell, D., Umemura, A., et al. (2015). Hybrid periportal hepatocytes regenerate the injured liver without giving rise to cancer. *Cell* 162, 766–779. doi: 10.1016/j.cell.2015.07.026

Gadd, V. L., Aleksieva, N., and Forbes, S. J. (2020). Epithelial Plasticity during liver injury and regeneration. *Cell Stem Cell* 27, 557–573. doi: 10.1016/j.stem.2020.08.016

Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I. B., et al. (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* 44, 694–698. doi: 10.1038/ng.2256

Han, X., Wang, Y., Pu, W., Huang, X., Qiu, L., Li, Y., et al. (2019). Lineage tracing reveals the bipotency of SOX9(+) hepatocytes during liver regeneration. *Stem Cell Rep.* 12, 624–638. doi: 10.1016/j.stemcr.2019.01.010

Hardy, T., and Mann, D. A. (2016). Epigenetics in liver disease: from biology to therapeutics. *Gut* 65, 1895–1905. doi: 10.1136/gutjnl-2015-311292

Hu, H., Gehart, H., Artegiani, B., LÖpez-Iglesias, C., Dekkers, F., Basak, O., et al. (2018). Long-term expansion of functional mouse and human hepatocytes as 3D organoids. *Cell* 175, 1591–1606.e19.

Huch, M., Dorrell, C., Boj, S. F., van Es, J. H., Li, V. S., van de Wetering, M., et al. (2013). In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature* 494, 247–250. doi: 10.1038/nature11826

Huch, M., Gehart, H., van Boxtel, R., Hamer, K., Blokzijl, F., Verstegen, M. M., et al. (2015). Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* 160, 299–312. doi: 10.1016/j.cell.2014.11.050

Itoh, T., and Miyajima, A. (2014). Liver regeneration by stem/progenitor cells. *Hepatology* 59, 1617–1626.

Kaji, K., Factor, V. M., Andersen, J. B., Durkin, M. E., Tomokuni, A., Marquardt, J. U., et al. (2016). DNMT1 is a required genomic regulator for murine liver histogenesis and regeneration. *Hepatology* 64, 582–598. doi: 10.1002/hep.28563

Kietzmann, T. (2017). Metabolic zonation of the liver: the oxygen gradient revisited. *Redox Biol.* 11, 622–630. doi: 10.1016/j.redox.2017.01.012

Ko, S., Russell, J. O., Molina, L. M., and Monga, S. P. (2020). Liver progenitors and adult cell plasticity in hepatic injury and repair: knowns and unknowns. *Annu. Rev. Pathol.* 15, 23–50. doi: 10.1146/annurev-pathmechdis-012419-032824

Ko, S., Russell, J. O., Tian, J., Gao, C., Kobayashi, M., Feng, R., et al. (2019). Hdac1 regulates differentiation of bipotent liver progenitor cells during regeneration via Sox9b and Cdk8. *Gastroenterology* 156, 187–202.e14.

Kopp, J. L., Grompe, M., and Sander, M. (2016). Stem cells versus plasticity in liver and pancreas regeneration. *Nat. cell Biol.* 18, 238–245. doi: 10.1038/ncb3309

Li, W., Yang, L., He, Q., Hu, C., Zhu, L., Ma, X., et al. (2019). A homeostatic arid1a-dependent permissive chromatin state licenses hepatocyte responsiveness to liver-injury-associated YAP signaling. *Cell Stem Cell* 25, 54–68.e5.

Lin, S., Nascimento, E. M., Gajera, C. R., Chen, L., Neuhofer, P., Garbuzov, A., et al. (2018). Distributed hepatocytes expressing telomerase repopulate the liver in homeostasis and injury. *Nature* 556, 244–248. doi: 10.1038/s41586-018-0004-7

Lu, W. Y., Bird, T. G., Boulter, L., Tsuchiya, A., Cole, A. M., Hay, T., et al. (2015). Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* 17, 971–983. doi: 10.1038/ncb3203

Malato, Y., Naqvi, S., Schurmann, N., Ng, R., Wang, B., Zape, J., et al. (2011). Fate tracing of mature hepatocytes in mouse liver homeostasis and regeneration. *J. Clin. Invest.* 121, 4850–4860. doi: 10.1172/jci59261

Manco, R., Clerbaux, L. A., Verhulst, S., Bou Nader, M., Sempoux, C., Ambroise, J., et al. (2019). Reactive cholangiocytes differentiate into proliferative hepatocytes with efficient DNA repair in mice with chronic liver injury. *J. Hepatol.* 70, 1180–1191. doi: 10.1016/j.jhep.2019.02.003

Matsumoto, T., Wakefield, L., Tarlow, B. D., and Grompe, M. (2020). *In Vivo* lineage tracing of polyploid hepatocytes reveals extensive proliferation during liver regeneration. *Cell Stem Cell* 26, 34–47.e3.

Michalopoulos, G. K. (2007). Liver regeneration. *J Cell Physiol.* 213, 286–300.

Michalopoulos, G. K., and DeFrances, M. C. (1997). Liver regeneration. *Science* 276, 60–66.

Monga, S. P. (2020). No zones left behind: democratic hepatocytes contribute to liver homeostasis and repair. *Cell Stem Cell* 26, 2–3. doi: 10.1016/j.stem.2019.12.002

Moore, A., Wu, L., Chuang, J. C., Sun, X., Luo, X., Gopal, P., et al. (2019). Arid1a loss drives nonalcoholic steatohepatitis in mice through epigenetic dysregulation of hepatic lipogenesis and fatty acid oxidation. *Hepatology* 69, 1931–1945. doi: 10.1002/hep.30487

Mudbhary, R., Hoshida, Y., Chernyavskaya, Y., Jacob, V., Villanueva, A., Fiel, M. I., et al. (2014). UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. *Cancer Cell* 25, 196–209. doi: 10.1016/j.ccr.2014.01.003

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* 363, 294–297. doi: 10.1126/science.aau0583

Nuciforo, S., Fofana, I., Matter, M. S., Blumer, T., Calabrese, D., Boldanova, T., et al. (2018). Organoid models of human liver cancers derived from tumor needle biopsies. *Cell Rep.* 24, 1363–1376. doi: 10.1016/j.celrep.2018.07.001

Ober, E. A., and Lemaigre, F. P. (2018). Development of the liver: insights into organ and tissue morphogenesis. *J. Hepatol.* 68, 1049–1062. doi: 10.1016/j.jhep.2018.01.005

Okabe, M., Tsukahara, Y., Tanaka, M., Suzuki, K., Saito, S., Kamiya, Y., et al. (2009). Potential hepatic stem cells reside in EpCAM+ cells of normal and injured mouse liver. *Development* 136, 1951–1960. doi: 10.1242/dev.031369

Ouchi, R., Togo, S., Kimura, M., Shinozawa, T., Koido, M., Koike, H., et al. (2019). Modeling steatohepatitis in humans with pluripotent stem cell-derived organoids. *Cell Metab.* 30, 374–384.e6.

Paksa, A., and Rajagopal, J. (2017). The epigenetic basis of cellular plasticity. *Curr. Opin. Cell Biol.* 49, 116–122. doi: 10.1016/j.ceb.2018.01.003

Peng, W. C., Logan, C. Y., Fish, M., Anbarchian, T., Aguisanda, F., Alvarez-Varela, A., et al. (2018). Inflammatory cytokine TNFalpha promotes the long-term expansion of primary hepatocytes in 3D culture. *Cell* 175, 1607–1619.e15.

Pepe-Mooney, B. J., Dill, M. T., Alemany, A., Ordovas-Montanes, J., Matsushita, Y., Rao, A., et al. (2019). Single-cell analysis of the liver epithelium reveals dynamic heterogeneity and an essential role for YAP in homeostasis and regeneration. *Cell Stem Cell* 25, 23–38.e8.

Planas-Paz, L., Sun, T., Pikiolek, M., Cochran, N. R., Bergling, S., Orsini, V., et al. (2019). YAP, but Not RSPO-LGR4/5, signaling in biliary epithelial cells promotes a ductular reaction in response to liver injury. *Cell Stem Cell* 25, 39–53.e10.

Prior, N., Hindley, C. J., Rost, F., Melendez, E., Lau, W. W. Y., Gottgens, B., et al. (2019). Lgr5(+) stem and progenitor cells reside at the apex of a heterogeneous embryonic hepatoblast pool. *Development* 146:dev174557. doi: 10.1242/dev.174557

Rajagopal, J., and Stanger, B. Z. (2016). Plasticity in the adult: how should the waddington diagram be applied to regenerating tissues? *Dev. Cell* 36, 133–137. doi: 10.1016/j.devcel.2015.12.021

Ramli, M. N. B., Lim, Y. S., Koe, C. T., Demircioglu, D., Tng, W., Gonzales, K. A. U., et al. (2020). Human pluripotent stem cell-derived organoids as models of liver disease. *Gastroenterology* 159, 1471–1486.e12.

Rasmussen, K. D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* 30, 733–750. doi: 10.1101/gad.276568.115

Raven, A., Lu, W. Y., Man, T. Y., Ferreira-Gonzalez, S., O'Duibhir, E., Dwyer, B. J., et al. (2017). Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. *Nature* 547, 350–354. doi: 10.1038/nature23015

Rimland, C. A., Tilson, S. G., Morell, C. M., Tomaz, R. A., Lu, W. Y., Adams, S. E., et al. (2020). Regional differences in human biliary tissues and corresponding *in vitro* derived organoids. *Hepatology* doi: 10.1002/hep.31252 [Epub ahead of print].

Roskams, T. A., Theise, N. D., Balabaud, C., Bhagat, G., Bhathal, P. S., Bioulac-Sage, P., et al. (2004). Nomenclature of the finer branches of the biliary tree: canals,

ductules, and ductular reactions in human livers. *Hepatology* 39, 1739–1745. doi: 10.1002/hep.20130

Russell, J. O., Lu, W. Y., Okabe, H., Abrams, M., Oertel, M., Poddar, M., et al. (2018). Hepatocyte-specific beta-catenin deletion during severe liver injury provokes cholangiocytes to differentiate into hepatocytes. *Hepatology* 69, 742–759. doi: 10.1002/hep.30270

Sackett, S. D., Li, Z., Hurtt, R., Gao, Y., Wells, R. G., Brondell, K., et al. (2009). Foxl1 is a marker of bipotential hepatic progenitor cells in mice. *Hepatology* 49, 920–929. doi: 10.1002/hep.22705

Sampaziotis, F., Justin, A. W., Tysoe, O. C., Sawiak, S., Godfrey, E. M., Upponi, S. S., et al. (2017). Reconstruction of the mouse extrahepatic biliary tree using primary human extrahepatic cholangiocyte organoids. *Nat. Med.* 23, 954–963.

Schaub, J. R., Malato, Y., Gormond, C., and Willenbring, H. (2014). Evidence against a stem cell origin of new hepatocytes in a common mouse model of chronic liver injury. *Cell Rep.* 8, 933–939. doi: 10.1016/j.celrep.2014.07.003

Shin, S., Walton, G., Aoki, R., Brondell, K., Schug, J., Fox, A., et al. (2011). Foxl1-Cre-marked adult hepatic progenitors have clonogenic and bilineage differentiation potential. *Genes Dev.* 25, 1185–1192. doi: 10.1101/gad.2027811

So, J., Kim, A., Lee, S. H., and Shin, D. (2020). Liver progenitor cell-driven liver regeneration. *Exp. Mol. Med.* 52, 1230–1238. doi: 10.1038/s12276-020-0483-0

Sun, T., Pikiolek, M., Orsini, V., Bergling, S., Holwerda, S., Morelli, L., et al. (2020). AXIN2(+) pericentral hepatocytes have limited contributions to liver homeostasis and regeneration. *Cell Stem Cell* 26, 97–107.e6.

Sun, X., Chuang, J. C., Kanchwala, M., Wu, L., Celen, C., Li, L., et al. (2016). Suppression of the SWI/SNF component arid1a promotes mammalian regeneration. *Cell Stem Cell* 18, 456–466. doi: 10.1016/j.stem.2016.03.001

Sun, X., Wang, S. C., Wei, Y., Luo, X., Jia, Y., Li, L., et al. (2017). Arid1a has context-dependent oncogenic and tumor suppressor functions in liver cancer. *Cancer Cell* 32, 574–589.e6.

Sun, Z., Miller, R. A., Patel, R. T., Chen, J., Dhir, R., Wang, H., et al. (2012). Hepatic Hdac3 promotes gluconeogenesis by repressing lipid synthesis and sequestration. *Nat. Med.* 18, 934–942. doi: 10.1038/nm.2744

Takahashi, K., and Yamanaka, S. (2015). A developmental framework for induced pluripotency. *Development* 142, 3274–3285. doi: 10.1242/dev.114249

Tarlow, B. D., Pelz, C., Naugler, W. E., Wakefield, L., Wilson, E. M., Finegold, M. J., et al. (2014). Bipotential adult liver progenitors are derived from chronically injured mature hepatocytes. *Cell Stem Cell* 15, 605–618. doi: 10.1016/j.stem.2014.09.008

Tetteh, P. W., Farin, H. F., and Clevers, H. (2015). Plasticity within stem cell hierarchies in mammalian epithelia. *Trends Cell Biol.* 25, 100–108. doi: 10.1016/j.tcb.2014.09.003

van den Ameele, J., Krautz, R., and Brand, A. H. (2019). TaDa! analysing cell type-specific chromatin *in vivo* with Targeted DamID. *Curr. Opin. Neurobiol.* 56, 160–166. doi: 10.1016/j.conb.2019.01.021

Waddington, C. (1957). "A discussion of some aspects of theoretical biology," in *The Strategy of the Genes*, ed. C. H. Waddington (London: London George Allen and Unwin).

Wang, S., Zhang, C., Hasson, D., Desai, A., SenBanerjee, S., Magnani, E., et al. (2019). Epigenetic compensation promotes liver regeneration. *Dev. Cell* 50, 43–56.e6.

Yanger, K., Knigin, D., Zong, Y., Maggs, L., Gu, G., Akiyama, H., et al. (2014). Adult hepatocytes are generated by self-duplication rather than stem cell differentiation. *Cell Stem Cell* 15, 340–349. doi: 10.1016/j.stem.2014.06.003

Yanger, K., Zong, Y., Maggs, L. R., Shapira, S. N., Maddipati, R., Aiello, N. M., et al. (2013). Robust cellular reprogramming occurs spontaneously during liver regeneration. *Genes Dev.* 27, 719–724. doi: 10.1101/gad.207803.112

Yimlamai, D., Christodoulou, C., Galli, G. G., Yanger, K., Pepe-Mooney, B., Gurung, B., et al. (2014). Hippo pathway activity influences liver cell fate. *Cell* 157, 1324–1338. doi: 10.1016/j.cell.2014.03.060

Zaret, K. S., and Grompe, M. (2008). Generation and regeneration of cells of the liver and pancreas. *Science* 322, 1490–1494. doi: 10.1126/science.1161431

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 04 March 2021
doi: 10.3389/fcell.2021.643361

The Evolutionary Conserved SWI/SNF Subunits ARID1A and ARID1B Are Key Modulators of Pluripotency and Cell-Fate Determination

*Luca Pagliaroli and Marco Trizzino**

Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, United States

OPEN ACCESS

Edited by:
Bruno Di Stefano, Baylor College of Medicine, United States

Reviewed by:
Michael Hoetker, Harvard Medical School, United States
Srinivas Vinod Saladi, Massachusetts Eye & Ear Infirmary, Harvard Medical School, United States

***Correspondence:**
Marco Trizzino
marco.trizzino@jefferson.edu

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *18 December 2020*
Accepted: *15 February 2021*
Published: *04 March 2021*

Citation:
Pagliaroli L and Trizzino M (2021) The Evolutionary Conserved SWI/SNF Subunits ARID1A and ARID1B Are Key Modulators of Pluripotency and Cell-Fate Determination. Front. Cell Dev. Biol. 9:643361. doi: 10.3389/fcell.2021.643361

Organismal development is a process that requires a fine-tuned control of cell fate and identity, through timely regulation of lineage-specific genes. These processes are mediated by the concerted action of transcription factors and protein complexes that orchestrate the interaction between *cis*-regulatory elements (enhancers, promoters) and RNA Polymerase II to elicit transcription. A proper understanding of these dynamics is essential to elucidate the mechanisms underlying developmental diseases. Many developmental disorders, such as Coffin-Siris Syndrome, characterized by growth impairment and intellectual disability are associated with mutations in subunits of the SWI/SNF chromatin remodeler complex, which is an essential regulator of transcription. *ARID1B* and its paralog *ARID1A* encode for the two largest, mutually exclusive, subunits of the complex. Mutations in *ARID1A* and, especially, *ARID1B* are recurrently associated with a very wide array of developmental disorders, suggesting that these two SWI/SNF subunits play an important role in cell fate decision. In this mini-review we therefore discuss the available scientific literature linking *ARID1A* and *ARID1B* to cell fate determination, pluripotency maintenance, and organismal development.

Keywords: chromatin remodeling, SWI/SNF, ARID1A, ARID1B, development, pluripotency, cell identity

INTRODUCTION

The SWI/SNF Complex

The SWI/SNF (SWItch/Sucrose Non-Fermentable) chromatin remodeling complex leverages an ATP-dependent mechanism to modify the structure of the chromatin and modulate its accessibility to transcriptional regulators (**Figure 1**). It was first discovered in yeast (SWI/SNF) (Stern et al., 1984), later in *Drosophila* (Brm-associated protein, BAP) (Kennison and Tamkun, 1988; Tamkun et al., 1992) and finally in mammals (Brg/Brahma-associated factors, BAF) (Wang et al., 1996).

In mammals, the different subunits comprise eight different bromodomains, two PHD finger proteins, two chromodomain and multiple proteins with DNA binding domains (Wang et al., 1996a,b; Wang et al., 1998; Lessard et al., 2007). These various subunits are not always present at the same time.

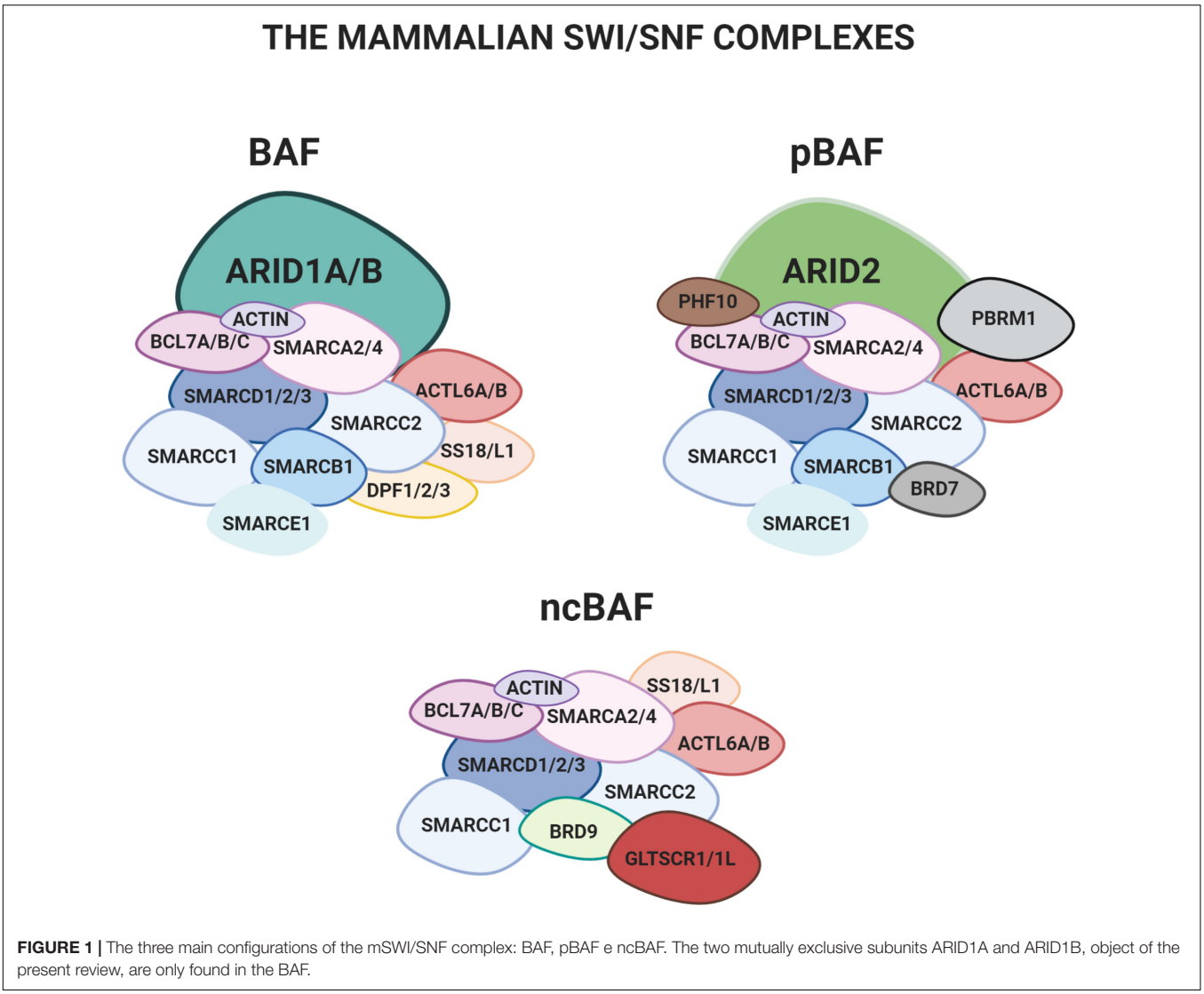


FIGURE 1 | The three main configurations of the mSWI/SNF complex: BAF, pBAF e ncBAF. The two mutually exclusive subunits ARID1A and ARID1B, object of the present review, are only found in the BAF.

Mammalian SWI/SNF (mSWI/SNF) complexes are assembled from subunits encoded by 29 genes, including multiple paralogs, which generate an extensive diversity in composition. Three versions of the mSWI/SNF were recently characterized in detail: 1) BRG1/BRM-associated factor complex (BAF), 2) polybromo containing complex (pBAF), and 3) a non-canonical version of the complex (ncBAF) (**Figure 1**) (Mashtalir et al., 2018).

The SWI/SNF complex is able to modify the structure of the chromatin leveraging the energy generated by the hydrolysis of ATP. SWI/SNF binds to the nucleosome in a central cavity where the DNA is exposed (Havas et al., 2000; Saha et al., 2005). Once bound, the complex uses the energy derived by the ATP hydrolysis to break the binding between histones and DNA, promoting the formation of a transient DNA loop that spreads around the nucleosome, ultimately orchestrating the changes in chromatin accessibility (Dechassa et al., 2008; Tang et al., 2010; Tyagi et al., 2016). As a consequence of this process, the chromatin becomes more accessible and permissive to the binding of transcription factors (Kadoch et al., 2017).

All the existing mammalian configurations of the complex contain an ATPase subunit, either SMARCA4 (BRG1) or SMARCA2 (BRM), which catalyzes the hydrolysis of ATP. Several other "core-subunits" are shared by all the different configurations (e.g., SMARCD1/2/3, SMARCC1/2 and a few others; Mashtalir et al., 2018). Finally, some of the subunits are only present in specific configurations. Among these, the mutually exclusive AT-rich interactive domain proteins ARID1A or ARID1B are only found in the BAF (Wilsker et al., 2004; Patsialou et al., 2005; Raab et al., 2015; Mashtalir et al., 2018).

ARID1A and ARID1B are conserved throughout metazoans and expressed across most human cells and tissues. Mutations in the genes encoding for these two subunits are associated with a wide array of developmental disorders and cancers, suggesting that they are implicated in the maintenance of cell identity and in the determination of cell fate. Based on this premise, in the

present review, we discuss the role of ARID1A and ARID1B in the maintenance of pluripotency, in the determination of cell fate, and, more broadly, in organismal development.

Mutations in *ARID1A*

ARID1A encodes for the AT-Rich Interactive Domain-containing protein 1A (ARID1A/BAF250a). It is the most frequently mutated member of the SWI/SNF family in cancer. *ARID1A* mutations are associated with a wide range of cancers, including ovarian endometrioid/clear-cell carcinomas, pancreatic cancer, gastric carcinoma, esophageal adenocarcinoma, renal carcinoma and breast tumors (Biegel et al., 2014; Wu et al., 2014; Kadoch and Crabtree, 2015; Masliah-Planchon et al., 2015; Takeda et al., 2016). Despite this evidence, the role of ARID1A in cancer is still not fully understood, with some studies suggesting a tumor suppression role, while a few others indicate an oncogenic function (Fang et al., 2015; Gibson et al., 2016; Zhai et al., 2016; Zhao et al., 2016; Mathur et al., 2017). In the case of endometrial and ovarian cancers, these mutations might either hamper the nuclear import of ARID1A, or affect the ability of ARID1A to interact with the subunits of SWI/SNF complex (Guan et al., 2012). The most frequently dysregulated pathway is PI3K/AKT, along with the downstream signaling cascades PTEN and PIKC3A (Takeda et al., 2016).

Mutations in *ARID1A* also lead to Coffin-Siris Syndrome, a neurodevelopmental disorder which will be further discussed in this review (Santen et al., 2013; Wieczorek et al., 2013; Kosho et al., 2014; Tsurusaki et al., 2014; Lee and Ki, 2021). *ARID1A* mutations are typically frame-shift or nonsense, spread across the gene with no specific hotspots, resulting in loss of protein level.

Mutations in *ARID1B*

ARID1B encodes for the AT-rich interactive domain-containing protein 1B (ARID1B/BAF250b). *ARID1B* mutations are commonly associated with neurodevelopmental disorders. Most frequently, *ARID1B* mutations are *de novo* haploinsufficient mutations, with no specific gene hotspots.

To date, the majority of the reported mutations are either nonsense or frameshift. These mutations result in non-functional truncated proteins, triggering *ARID1B*-haploinsufficiency and associated pathologies (Schweingruber et al., 2013; Sim et al., 2015). Mutations in *ARID1B* may disrupt the ability of the SWI/SNF complex to bind the chromatin (Sim et al., 2015).

Often, *ARID1B* mutations result in Coffin-Siris Syndrome, a relatively rare genetic disorder that manifests at birth and is characterized by both intellectual disabilities and physical phenotypes (Coffin and Siris, 1970; van der Sluijs et al., 2019). Coffin-Siris patients usually show coarse facial features, impaired craniofacial development, and hypoplastic fifth finger nails (Schrier et al., 2012). Further, most individuals present mild to severe intellectual disability, speech impairment and impaired motor skills (Vergano and Deardorff, 2014). Other characteristics of this disorder include respiratory infections, feeding issues, hearing loss, sparse scalp hair and hypermobility of joints (Vergano and Deardorff, 2014). Mutations in *ARID1B* have also been linked to Autism Spectrum Disorder (ASD), Intellectual Disabilities (ID), epilepsy and neuroblastoma (Vergano et al., 1993; Halgren et al., 2012; Hoyer et al., 2012; Santen et al., 2012; Vals et al., 2014; Yu et al., 2015; Ben-Salem et al., 2016; Sonmez et al., 2016; Jung et al., 2017; Lee et al., 2017; Shibutani et al., 2017; Yu et al., 2018; Demily et al., 2019; Filatova et al., 2019; Pranckeniene et al., 2019; Sekiguchi et al., 2019; van der Sluijs et al., 2019; Curcio et al., 2020; Fujita et al., 2020; Lian et al., 2020; Pascolini et al., 2020; Smith et al., 2020). *ARID1B* mutations can be associated with both syndromic and non-syndromic forms of ID (van der Sluijs et al., 2019). In this context, Coffin-Siris patients almost always show some degree of ID, and often present some characteristics that can be associated to Autism Spectrum Disorder. Recently, van der Sluijs et al. (2019) sought to determine genotypic and phenotypic differences between ARID1B-ID and ARID1B-CSS. They found only minor differences between ARID1B-ID and ARID1B-CSS patients, and suggested that ARID1B-related disorders seem to consist of a spectrum, and patients should be managed similarly.

Several studies tried to uncover genes and pathways most commonly dysregulated in ARID1B-ID and ARID1B-CSS. A recent paper looked at gene expression in monocytes of CSS patients. The study identified few differentially expressed genes (*CRYZ*, *TRGV5*, *TSPAN33*, *TPPP3*, *SAMD9L*, *DDX60*, *FMN1*, *PER1*, *MIR3648*, and *GSTM1*) (Kalmbach et al., 2019) and the pathway analysis did not reveal any statistically significant network. A previous study investigated gene expression in a single CSS patient carrying a novel microduplication of *ARID1B*, and identified EIF2 signaling and the regulation of eIF4 and p70S6K signaling as top canonical pathways (Seabra et al., 2017). Using an *ARID1B*-haploinsufficient mouse model, Celen et al. (2017) detected dysregulations in the Ephrin, nNOS, axonal guidance and glutamate receptor signaling pathways. Gene expression profile performed by Shibutani et al. (2017) suggested that $ARID1B^{+/-}$ mice exhibit a pattern very similar to autistic brains centered on immature fast spiking cells. Amongst the several differentially expressed genes, *HOXB2*, *PRL*, *PODNL1*, and *PTH2* were the most downregulated, whereas *AREG*, *GBP8*, *KLR2*, and *ZP2* were the most upregulated (Shibutani et al., 2017).

The Role of ARID1A and ARID1B in Pluripotency and Cell Fate Determination

The contribution of ARID1A and ARID1B to cell pluripotency has been predominantly investigated in mouse embryos and in embryonic stem cells (ESCs). These cells are distinguished by their ability to differentiate into almost any cell lineage.

Gao et al. (2008) demonstrated that embryos carrying a homozygous *ARID1A* knockout are able to differentiate in primitive endoderm and epiblast layers but are unable to generate the mesodermal layer. Moreover, $ARID1A^{-/-}$ mouse ESCs fail to maintain a normal stem cell phenotype in culture and spontaneously differentiate (Gao et al., 2008). These pluripotency anomalies seem to be lineage specific, since the ESCs cannot differentiate into cardiomyocytes or adipocytes, but can differentiate into ectoderm-derived neurons (Gao et al., 2008). Consistent with this, Lei et al. (2015) observed dysregulated expression of key developmental and pluripotency genes in

TABLE 1 | Biological processes and phenotypes associated to ARID1A and ARID1B.

STEM CELL PLURIPOTENCY		
ARID1A$^{-/-}$ mouse and human ESCs	Failure to maintain pluripotency, spontaneous differentiation, dysregulated expression of pluripotency genes	Gao et al. (2008); Lei et al. (2015), Liu J. et al. (2020)
ARID1B$^{-/-}$ mouse ESCs	Failure to maintain pluripotency, dysregulated expression of pluripotency genes	Yan et al. (2008)
CELL DIFFERENTIATION AND PROLIFERATION		
ARID1A$^{-/-}$ mouse and human ESCs	Bias toward neuronal differentiation	Gao et al. (2008), Liu J. et al. (2020)
	Failure to differentiate into cardiomyocytes or adipocytes	Gao et al. (2008); Liu J. et al. (2020)
ARID1A$^{-/-}$ hematopoietic stem cells	Impaired differentiation into myeloid and lymphoid lineages	Han et al. (2019)
ARID1B$^{-/-}$ Zebrafish	Reduced body length due to dysregulated Wnt/β-catenin signaling	Liu X. et al. (2020)
ARID1A$^{+/-}$ Zebrafish	Excessive cell proliferation in the sympathoadrenal lineage	Shi et al. (2020)
ARID1A$^{-/-}$ Mouse (liver)	Impaired liver regeneration, increased vacuole accumulation, liver dysfunction	Li et al. (2019)
ARID1A$^{-/-}$ Mouse (liver)	Increased liver regeneration	Sun et al. (2016)
ARID1A$^{-/-}$ Mouse (preosteoblasts)	Dysregulated cell cycle	NagJr., Patsialou et al. (2005)
DEVELOPMENTAL PHENOTYPES		
ARID1A$^{+/-}$ Mouse (neural crest)	Craniofacial defects, shortened snouts, low ears, defects in developing cardiac neural crest	Chandler and Magnuson (2016)
ARID1B$^{+/-}$ Mouse	Impaired maturation of dendritic spines, reduced dendritic innervation, lack of arborization and dendrite growth in cortical and hippocampal pyramidal neurons	Ka et al. (2016)
	Social and emotional impairments (parvalbumin neurons), learning and memory dysfunction (somatostatin neurons)	Smith et al. (2020)
	Reduced number of cortical GABAergic interneurons, decreased proliferation of interneuron progenitors in the ganglionic eminence	Jung et al. (2017)
	Hydrocephalus, reduced size of the corpus callosum and dentate gyrus, impairment in social behavior, growth deficit	Celen et al. (2017)
	Hydrocephalus	Shibutani et al. (2017)

ARID1A$^{-/-}$ mouse ESCs. In particular, the most frequently affected genes were associated with the generation of the mesodermal and endodermal layers (Lei et al., 2015).

Similar results were published by Liu J. et al. (2020), who investigated the role of ARID1A in early human cardiac development and neurogenesis. The study demonstrated that homozygous deletion of *ARID1A* in human ESCs results in spontaneous neuronal differentiation due to increased expression of several genes associated with neurodevelopment. Simultaneously, the same cells displayed downregulation of genes associated with cardiomyocyte differentiation (Liu J. et al., 2020).

Han et al. (2019) studied the function of ARID1A in hematopoietic stem cells (HSC), and uncovered that this SWI/SNF subunit is important for the generation of myeloid colonies, for normal T cell maturation, and for the differentiation of both myeloid and lymphoid lineages.

ARID1A loss/gain of function are thought to have context-dependent effects. For instance, *ARID1A* deletion is lethal in early embryonic mouse development (Gao et al., 2008). On the other hand, the depletion of this SWI/SNF subunit induces proliferation of ovarian clear cell carcinoma cells (Xiao et al., 2012; Yamamoto et al., 2012; Lakshminarasimhan et al., 2017). In contrast, another study leveraged a mouse ovarian cancer model and demonstrated that ARID1A loss enhances epithelial differentiation and prolongs survival (Chandler et al., 2015; Zhai et al., 2016). *ARID1A* is instead overexpressed in many hepatocellular carcinomas (Zhao et al., 2016), while the expression of this gene is reduced or lost in colorectal cancer (Mathur et al., 2017).

While there is extensive research investigating the role of ARID1A in cell differentiation, the work performed on ARID1B is thus far limited to a few studies. In this context, Yan et al. (2008) demonstrated that *ARID1B*$^{-/-}$ mouse ESCs are viable but exhibit a slower proliferation rate and tend to spontaneously differentiate. Consistent with this observation, ESCs with homozygous *ARID1B* deletion displayed reduced

expression of several pluripotency markers, including *OCT4* and *NANOG*. This suggests that ARID1B may be required to regulate stem cell pluripotency. Recently, Boerstler et al. (2020) leveraged CRISPR/Cas9 to generate a human *ARID1B*-haploinsufficient ESC line with an in-frame deletion of exons 5 and 6 of the gene. Future studies leveraging this cell line may help clarifying the role of ARID1B in pluripotency (Boerstler et al., 2020).

Shi et al. (2020) established *ARID1A* and *ARID1B* deletion mutant lines in zebrafish to investigate the effect of these subunits in neuroblastoma. The authors observed that depletion of ARID1A or ARID1B results in an increased rate of cell proliferation in the sympathoadrenal lineage, which ultimately leads to higher tumor penetrance (Shi et al., 2020).

A zebrafish model was also used to elucidate how ARID1B regulates organismal development (Liu X. et al., 2020). In this study, the authors demonstrated that *ARID1B* haploinsufficiency results in reduced body length due to dysregulated Wnt/β-catenin signaling pathway (Liu X. et al., 2020). An association between ARID1B and the Wnt/β-catenin signaling pathway had already been proposed by Vasileiou et al. (2015).

ARID1A and ARID1B in Neurodevelopment

ARID1A was associated with neural crest differentiation and craniofacial development (Chandler and Magnuson, 2016). Neural crest cells are a transient, ectoderm-derived, cell population that can migrate throughout the embryo to give origin to craniofacial bone and cartilage, peripheral neurons and glia, melanocytes, and smooth muscle cells (Shakhova and Sommer, 2008). Chandler and Magnuson (2016) generated mice with a conditional, neural crest specific, heterozygous deletion of *ARID1A*. The *ARID1A*-depleted mice displayed craniofacial defects, including shortened snouts and low ears. Additionally, most of the bones involved in the ventral cranial skeleton were greatly reduced in size, leading to abnormal facial features (Chandler and Magnuson, 2016). The study also revealed that conditional haploinsufficiency of *ARID1A* results in defects in developing cardiac neural crest due to an incomplete colonization of the outflow tract and septation of the arterial trunk, ultimately producing defects in the pharyngeal arch arteries. Consistently, homozygous *ARID1A* mutants did not survive *in utero* (Chandler and Magnuson, 2016).

As mentioned, mutations in *ARID1B* often lead to a wide array of neurodevelopmental disorders, including Autism Spectrum Disorders, Coffin-Siris Syndrome, and other forms of Intellectual Disabilities (Vergano et al., 1993; Halgren et al., 2012; Hoyer et al., 2012; Santen et al., 2012; Vals et al., 2014; Yu et al., 2015; Ben-Salem et al., 2016; Sonmez et al., 2016; Jung et al., 2017; Lee et al., 2017; Shibutani et al., 2017; Yu et al., 2018; Demily et al., 2019; Filatova et al., 2019; Pranckeniene et al., 2019; Sekiguchi et al., 2019; van der Sluijs et al., 2019; Curcio et al., 2020; Fujita et al., 2020; Lian et al., 2020; Pascolini et al., 2020; Smith et al., 2020).

Based on these lines of evidence, several studies investigated the role of *ARID1B* in neurodevelopment (Ka et al., 2016; Celen et al., 2017; Jung et al., 2017; Shibutani et al., 2017; Smith et al., 2020). Ka et al. (2016) demonstrated that ARID1B is required for arborization and dendrite growth in cortical and hippocampal pyramidal neurons. *ARID1B* haploinsufficiency resulted in reduced dendritic innervation as well as diminished attachment of dendrites to the pial surface (Ka et al., 2016). In the same study, Ka et al. (2016) found that *ARID1B* mono-allelic loss impairs the formation and maturation of dendritic spines, generating malformations that morphologically resemble those reported in animal models of multiple neuropsychiatric disorders such as ID, ASD, Rett-Syndrome, Down-Syndrome and Fragile-X-Syndrome (Irwin et al., 2002; McKinney et al., 2005; Jentarra et al., 2010; Moffat et al., 2019).

Recently, Smith et al. (2020) leveraged a mouse model to elucidate the consequences of *ARID1B*-haploinsufficiency on the development and function of parvalbumin (PV) and somatostatin (SST) neurons, two of the most prevalent interneuron subtypes. Briefly, the authors discovered that *ARID1B*-haploinsufficiency in PV neurons leads to social and emotional impairments, which are key features of ASD, while *ARID1B* deficiency in the SST population results in learning and memory dysfunction (Smith et al., 2020).

In a similar study, Jung et al. (2017) demonstrated that *ARID1B*-haploinsufficient mice present a reduced number of cortical GABAergic interneurons and decreased proliferation of interneuron progenitors in the ganglionic eminence. These neurological phenotypes are often recovered in autism and schizophrenia patients (Benes and Berretta, 2001; Pizzarelli and Cherubini, 2011). Additionally, in a third mouse model study, Celen et al. (2017) showed that *ARID1B*-haploinsufficient mice are characterized by hydrocephalus, a condition frequently reported also in Coffin-Siris patients (Schrier Vergano et al., 1993). Brain abnormalities were detected in *ARID1B*-haploinsufficient mice also by Shibutani et al. (2017). The *ARID1B*-haploinsufficient mice also exhibited reduced size of the corpus callosum and dentate gyrus, along with impairment in social behavior, altered vocalization, presence of anxiety-like behavior, and growth deficit (Celen et al., 2017).

ARID1A and ARID1B in Cell Proliferation and Tissue Regeneration

Recent studies conducted on mouse liver linked ARID1A to tissue regeneration (Sun et al., 2016; Li et al., 2019). Specifically, Li et al. (2019) demonstrated that ARID1A is required for the generation of liver-progenitor-like cells (LPLCs) in different types of periportal injuries. In detail, mice with conditional *ARID1A* knockout in the liver displayed impaired LPLCs formation and reduced regeneration of damaged liver tissue (Li et al., 2019). Moreover, *ARID1A*-knockout livers were characterized by significantly increased accumulation of fatty vacuoles and impaired liver function. Conversely, a prior study also performed in the mouse liver demonstrated that suppression of ARID1A is sufficient to promote liver regeneration (Sun et al., 2016).

These two studies suggest a dual role for ARID1A in the hepatic context. Sun et al. (2016) proposed a mechanism focused on CYP-metabolism. On the other hand, Li et al. (2019) suggested the presence of a hepatocyte plasticity network where ARID1A promotes the formation of LPLC during injury, while hampering cell proliferation during the recovery stage.

The role of ARID1A and ARID1B in cell proliferation was also investigated using mouse derived preosteoblasts (NaglJr., Patsialou et al., 2005; Flores-Alcantar et al., 2011). Notably, NaglJr., Patsialou et al. (2005) showed that deletion of *ARID1A* leads to failure in cell cycle arrest. Conversely, the same study demonstrated that loss of ARID1B has not significant impact on the cell cycle (Nagl et al., 2005).

Molecular Processes Modulated by ARID1A and ARID1B

The SWI/SNF complex is mainly considered as a transcriptional activator, which antagonizes the Polycomb Repressor Complexes (PRC1 and PRC2) in the modulation of gene expression (Kadoch and Crabtree, 2015; Alfert et al., 2019). Nonetheless, repressing activity for the SWI/SNF has also been reported. For instance, a recent study performed on HepG2 cells (hepatocellular carcinoma line) uncovered that ARID1A-containing BAF activates and represses roughly equal numbers of genes (Raab et al., 2015). The same study also demonstrated that ARID1B-containing BAF is primarily a repressor of enhancer activity (Raab et al., 2015). More specifically, Raab et al. (2015) investigated the localization of ARID1A and ARID1B binding sites in HepG2 cells via chromatin immunoprecipitation followed by sequencing (ChIP-seq). The authors observed binding of ARID1A at most enhancers and promoters, while ARID1B was predominantly located at enhancers. Loss of ARID1A from HepG2 cells resulted in a roughly equal number of activated and repressed genes, whereas loss of ARID1B predominantly resulted in transcriptional activation (Raab et al., 2015).

Consistently, ARID1A and ARID1B have been recently associated with acetylation of histone tails at both enhancers and promoters (Chandler et al., 2013; Lei et al., 2015; Raab et al., 2015; Alver et al., 2017; Kelso et al., 2017; Trizzino et al., 2018). For example, Mathur et al. (2017) observed that human *ARID1A*$^{-/-}$ colorectal cancer cells display dampened acetylation levels at Histone H3 lysine 27 (H3K27ac), which is usually associated with transcriptional activity at enhancers and promoters. Correlation between ARID1A loss and attenuation of enhancer acetylation was also observed in zebrafish models (Shi et al., 2020).

Liu X. et al. (2020) profiled chromatin accessibility in wild-type and ARID1A-deleted human ES cells. With these experiments, the authors discovered that loss of ARID1A generated a loss in accessibility at cardiogenic genes, as well as an increase in accessibility at neurogenic genes (Liu X. et al., 2020). These data are thus consistent with a dual (activator/repressor) role of ARID1A in the transcriptional regulation of ESCs. An additional study on human ES cells also revealed that acute depletion of ARID1A increases nucleosome occupancy, and therefore repression, at a set of H3K4me3- and/or H3K27me3-associated promoters (Lei et al., 2015).

The consequences of ARID1A loss on chromatin accessibility were further investigated by Kelso et al. (2017) in colorectal carcinoma lines. The authors demonstrated that loss of ARID1A and ARID1B correlates with global dampening of chromatin accessibility, along with a significant decrease of histone modifications normally associated with transcriptional activation at enhancers (Kelso et al., 2017).

Recently, Wu et al. (2019) linked ARID1A-containing BAF to Condensin, a protein complex involved in the regulation of genomic organization and chromatin looping. The study demonstrated that ovarian cancer cell lines depleted of ARID1A exhibit decreased binding of Condensing-II at active enhancers. Further, they illustrated that ARID1A-loss leads to improper genome compartmentalization (Wu et al., 2019).

Finally, in a recent study conducted in ovarian cancer cell lines, Trizzino et al. (2018) demonstrated that ARID1A and ARID1B play a role in the regulation of RNA Polymerase II promoter-proximal pausing, a widespread mechanism that controls the timing of expression of developmental genes genome-wide.

CONCLUSION

In conclusion, multiple lines of evidence point toward a model in which the ARID1A- and ARID1B-containing configurations of the SWI/SNF complex (i.e., the BAF) play an important role in the regulation of pluripotency, as well as in cell fate determination and development (**Table 1**). Multiple molecular and genomic functions were ascribed to these two SWI/SNF subunits. However, the mechanisms by which ARID1A and ARID1B regulate pluripotency and cell fate are still not fully understood and are likely context-specific. The discovery of such mechanisms, along with the transcription factors and the molecular pathways involved, may open new roads for the diagnosis and the treatment of developmental disorders and cancer.

AUTHOR CONTRIBUTIONS

MT and LP designed and wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

Funding for this review were provided to MT by the G. Harold and Leila Y. Mathers Foundation. LP and MT were also supported by the Thomas Jefferson University Open Access Fund.

ACKNOWLEDGMENTS

We are thankful to Dr. Sruti Patoori (TJU) for providing insightful comments on the manuscript.

REFERENCES

Alfert, A., Moreno, N., and Kerl, K. (2019). The BAF complex in development and disease. *Epigenetics Chromatin* 12:19.

Alver, B. H., Kim, K. H., Lu, P., Wang, X., Manchester, H. E., Wang, W., et al. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* 8:14648.

Benes, F. M., and Berretta, S. (2001). GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. *Neuropsychopharmacology* 25, 1–27. doi: 10.1016/s0893-133x(01)00225-1

Ben-Salem, S., Sobreira, N., Akawi, N. A., Al-Shamsi, A. M., John, A., Pramathan, T., et al. (2016). Gonadal mosaicism in ARID1B gene causes intellectual disability and dysmorphic features in three siblings. *Am. J. Med. Genet. A* 170A, 156–161. doi: 10.1002/ajmg.a.37405

Biegel, J. A., Busse, T. M., and Weissman, B. E. (2014). SWI/SNF chromatin remodeling complexes and cancer. *Am. J. Med. Genet. C Semin. Med. Genet.* 166C, 350–366.

Boerstler, T., Wend, H., Krumbiegel, M., Kavyanifar, A., Reis, A., Lie, D. C., et al. (2020). CRISPR/Cas9 mediated generation of human ARID1B heterozygous knockout hESC lines to model Coffin-Siris syndrome. *Stem. Cell Res.* 47:101889. doi: 10.1016/j.scr.2020.101889

Celen, C., Chuang, J. C., Luo, X., Nijem, N., Walker, A. K., Chen, F., et al. (2017). Arid1b haploinsufficient mice reveal neuropsychiatric phenotypes and reversible causes of growth impairment. *Elife* 6:e25730.

Chandler, R. L., Brennan, J., Schisler, J. C., Serber, D., Patterson, C., and Magnuson, T. (2013). ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol. Cell. Biol.* 33, 265–280. doi: 10.1128/mcb.01008-12

Chandler, R. L., Damrauer, J. S., Raab, J. R., Schisler, J. C., Wilkerson, M. D., Didion, J. P., et al. (2015). Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat. Commun.* 6:6118.

Chandler, R. L., and Magnuson, T. (2016). The SWI/SNF BAF-A complex is essential for neural crest development. *Dev. Biol.* 411, 15–24. doi: 10.1016/j.ydbio.2016.01.015

Coffin, G. S., and Siris, E. (1970). Mental retardation with absent fifth fingernail and terminal phalanx. *Am. J. Dis. Child.* 119, 433–439. doi: 10.1001/archpedi.1970.02100050435009

Curcio, M. R., Ferranti, S., Lotti, F., and Grosso, S. (2020). Coffin-Siris syndrome and epilepsy. *Neurol. Sci.* 42, 727–729. doi: 10.1007/s10072-020-04782-y

Dechassa, M. L., Zhang, B., Horowitz-Scherer, R., Persinger, J., Woodcock, C. L., Peterson, C. L., et al. (2008). Architecture of the SWI/SNF-nucleosome complex. *Mol. Cell. Biol.* 28, 6010–6021. doi: 10.1128/mcb.00693-08

Demily, C., Duwime, C., Lopez, C., Hemimou, C., Poisson, A., Plasse, J., et al. (2019). Corpus callosum metrics predict severity of visuospatial and neuromotor dysfunctions in ARID1B mutations with Coffin-Siris syndrome. *Psychiatr. Genet.* 29, 237–242. doi: 10.1097/ypg.0000000000000225

Fang, J. Z., Li, C., Liu, X. Y., Hu, T. T., Fan, Z. S., and Han, Z. G. (2015). Hepatocyte-specific arid1a deficiency initiates mouse steatohepatitis and hepatocellular carcinoma. *PLoS One* 10:e0143042. doi: 10.1371/journal.pone.0143042

Filatova, A., Rey, L. K., Lechler, M. B., Schaper, J., Hempel, M., Posmyk, R., et al. (2019). Mutations in SMARCB1 and in other Coffin-Siris syndrome genes lead to various brain midline defects. *Nat. Commun.* 10:2966.

Flores-Alcantar, A., Gonzalez-Sandoval, A., Escalante-Alcalde, D., and Lomeli, H. (2011). Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. *Cell Tissue Res.* 345, 137–148. doi: 10.1007/s00441-011-1182-x

Fujita, T., Ihara, Y., Hayashi, H., Ishii, A., Ideguchi, H., Inoue, T., et al. (2020). Coffin-Siris syndrome with bilateral macular dysplasia caused by a novel exonic deletion in ARID1B. *Congenit. Anom. (Kyoto)* 60, 189–193. doi: 10.1111/cga.12383

Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W. C., and Wang, Z. (2008). ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6656–6661. doi: 10.1073/pnas.0801802105

Gibson, W. J., Hoivik, E. A., Halle, M. K., Taylor-Weiner, A., Cherniack, A. D., Berg, A., et al. (2016). The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat. Genet.* 48, 848–855. doi: 10.1038/ng.3602

Guan, B., Gao, M., Wu, C. H., Wang, T. L., and Shih Ie, M. (2012). Functional analysis of in-frame indel ARID1A mutations reveals new regulatory mechanisms of its tumor suppressor functions. *Neoplasia* 14, 986–993. doi: 10.1593/neo.121218

Halgren, C., Kjaergaard, S., Bak, M., Hansen, C., El-Schich, Z., Anderson, C. M., et al. (2012). Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B. *Clin. Genet.* 82, 248–255. doi: 10.1111/j.1399-0004.2011.01755.x

Han, L., Madan, V., Mayakonda, A., Dakle, P., Woon, T. W., Shyamsunder, P., et al. (2019). Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice. *Leukemia* 33, 2291–2305. doi: 10.1038/s41375-019-0438-4

Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P. A., Lilley, D. M., et al. (2000). Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. *Cell* 103, 1133–1142. doi: 10.1016/s0092-8674(00)00215-4

Hoyer, J., Ekici, A. B., Endele, S., Popp, B., Zweier, C., Wiesener, A., et al. (2012). Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* 90, 565–572. doi: 10.1016/j.ajhg.2012.02.007

Irwin, S. A., Idupulapati, M., Gilbert, M. E., Harris, J. B., Chakravarti, A. B., Rogers, E. J., et al. (2002). Dendritic spine and dendritic field characteristics of layer V pyramidal neurons in the visual cortex of fragile-X knockout mice. *Am. J. Med. Genet.* 111, 140–146. doi: 10.1002/ajmg.10500

Jentarra, G. M., Olfers, S. L., Rice, S. G., Srivastava, N., Homanics, G. E., Blue, M., et al. (2010). Abnormalities of cell packing density and dendritic complexity in the MeCP2 A140V mouse model of Rett syndrome/X-linked mental retardation. *BMC Neurosci.* 11:19. doi: 10.1186/1471-2202-11-19

Jung, E. M., Moffat, J. J., Liu, J., Dravid, S. M., Gurumurthy, C. B., and Kim, W. Y. (2017). Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat. Neurosci.* 20, 1694–1707. doi: 10.1038/s41593-017-0013-0

Ka, M., Chopra, D. A., Dravid, S. M., and Kim, W. Y. (2016). Essential roles for ARID1B in dendritic arborization and spine morphology of developing pyramidal neurons. *J. Neurosci.* 36, 2723–2742. doi: 10.1523/jneurosci.2321-15.2016

Kadoch, C., and Crabtree, G. R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. *Sci. Adv.* 1:e1500447. doi: 10.1126/sciadv.1500447

Kadoch, C., Williams, R. T., Calarco, J. P., Miller, E. L., Weber, C. M., Braun, S. M., et al. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* 49, 213–222. doi: 10.1038/ng.3734

Kalmbach, A., Schroder, C., Klein-Hitpass, L., Nordstrom, K., Ulz, P., Heitzer, E., et al. (2019). Genome-wide analysis of the nucleosome landscape in individuals with Coffin-Siris Syndrome. *Cytogenet. Genome Res.* 159, 1–11. doi: 10.1159/000503266

Kelso, T. W. R., Porter, D. K., Amaral, M. L., Shokhirev, M. N., Benner, C., and Hargreaves, D. C. (2017). Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *Elife* 6:e30506.

Kennison, J. A., and Tamkun, J. W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8136–8140. doi: 10.1073/pnas.85.21.8136

Kosho, T., Okamoto, N., and Coffin-Siris Syndrome International Collaborators. (2014). Genotype-phenotype correlation of Coffin-Siris syndrome caused by mutations in SMARCB1, SMARCA4, SMARCE1, and ARID1A. *Am. J. Med. Genet. C Semin. Med. Genet.* 166C, 262–275. doi: 10.1002/ajmg.c.31407

Lakshminarasimhan, R., Andreu-Vieyra, C., Lawrenson, K., Duymich, C. E., Gayther, S. A., Liang, G., et al. (2017). Down-regulation of ARID1A is sufficient to initiate neoplastic transformation along with epigenetic reprogramming in non-tumorigenic endometriotic cells. *Cancer Lett.* 401, 11–19. doi: 10.1016/j.canlet.2017.04.040

Lee, C. G., and Ki, C. S. (2021). A novel de novo heterozygous ARID1A missense variant cluster in cis c.[5954C>G;6314C>T;6334C>T;6843G>C] causes a

Coffin-Siris syndrome. *Ann. Lab. Med.* 41, 350–353. doi: 10.3343/alm.2021.41.3.350

Lee, S. H., Kim, J. S., Zheng, S., Huse, J. T., Bae, J. S., Lee, J. W., et al. (2017). ARID1B alterations identify aggressive tumors in neuroblastoma. *Oncotarget* 8, 45943–45950. doi: 10.18632/oncotarget.17500

Lei, I., West, J., Yan, Z., Gao, X., Fang, P., Dennis, J. H., et al. (2015). BAF250a protein regulates nucleosome occupancy and histone modifications in priming embryonic stem cell differentiation. *J. Biol. Chem.* 290, 19343–19352. doi: 10.1074/jbc.m115.637389

Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., et al. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* 55, 201–215. doi: 10.1016/j.neuron.2007.06.019

Li, W., Yang, L., He, Q., Hu, C., Zhu, L., Ma, X., et al. (2019). A homeostatic arid1a-dependent permissive chromatin state licenses hepatocyte responsiveness to liver-injury-associated YAP signaling. *Cell Stem Cell* 25, 54–68.e5.

Lian, S., Ting, T. W., Lai, A. H. M., Tan, E. S., Wei, H., Cham, B., et al. (2020). Coffin-Siris syndrome-1: report of five cases from Asian populations with truncating mutations in the ARID1B gene. *J. Neurol. Sci.* 414, 116819. doi: 10.1016/j.jns.2020.116819

Liu, J., Liu, S., Gao, H., Han, L., Chu, X., Sheng, Y., et al. (2020). Genome-wide studies reveal the essential and opposite roles of ARID1A in controlling human cardiogenesis and neurogenesis from pluripotent stem cells. *Genome Biol.* 21:169.

Liu, X., Hu, G., Ye, J., Ye, B., Shen, N., Tao, Y., et al. (2020). De Novo ARID1B mutations cause growth delay associated with aberrant Wnt/beta-catenin signaling. *Hum. Mutat.* 41, 1012–1024. doi: 10.1002/humu.23990

Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., et al. (2018). Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* 175, 1272–1288e20.

Masliah-Planchon, J., Bieche, I., Guinebretiere, J. M., Bourdeaut, F., and Delattre, O. (2015). SWI/SNF chromatin remodeling and human malignancies. *Annu. Rev. Pathol.* 10, 145–171.

Mathur, R., Alver, B. H., San Roman, A. K., Wilson, B. G., Wang, X., Agoston, A. T., et al. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat. Genet.* 49, 296–302. doi: 10.1038/ng.3744

McKinney, B. C., Grossman, A. W., Elisseou, N. M., and Greenough, W. T. (2005). Dendritic spine abnormalities in the occipital cortex of C57BL/6 Fmr1 knockout mice. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 136B, 98–102. doi: 10.1002/ajmg.b.30183

Moffat, J. J., Jung, E. M., Ka, M., Smith, A. L., Jeon, B. T., Santen, G. W. E., et al. (2019). The role of ARID1B, a BAF chromatin remodeling complex subunit, in neural development and behavior. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 89, 30–38. doi: 10.1016/j.pnpbp.2018.08.021

Nagl, N. G. Jr., Patsialou, A., Haines, D. S., Dallas, P. B., Beck, G. R. Jr., and Moran, E. (2005). The p270 (ARID1A/SMARCF1) subunit of mammalian SWI/SNF-related complexes is essential for normal cell cycle arrest. *Cancer Res.* 65, 9236–9244. doi: 10.1158/0008-5472.can-05-1225

Pascolini, G., Valiante, M., Bottillo, I., Laino, L., Fleischer, N., Ferraris, A., et al. (2020). Striking phenotypic overlap between Nicolaides-Baraitser and Coffin-Siris syndromes in monozygotic twins with ARID1B intragenic deletion. *Eur. J. Med. Genet.* 63:103739. doi: 10.1016/j.ejmg.2019.103739

Patsialou, A., Wilsker, D., and Moran, E. (2005). DNA-binding properties of ARID family proteins. *Nucleic Acids Res.* 33, 66–80. doi: 10.1093/nar/gki145

Pizzarelli, R., and Cherubini, E. (2011). Alterations of GABAergic signaling in autism spectrum disorders. *Neural Plast.* 2011:297153.

Pranckeniene, L., Siavriene, E., Gueneau, L., Preiksaitiene, E., Mikstiene, V., Reymond, A., et al. (2019). De novo splice site variant of ARID1B associated with pathogenesis of Coffin-Siris syndrome. *Mol. Genet. Genomic. Med.* 7:e1006.

Raab, J. R., Resnick, S., and Magnuson, T. (2015). Genome-wide transcriptional regulation mediated by biochemically distinct SWI/SNF complexes. *PLoS Genet.* 11:e1005748. doi: 10.1371/journal.pgen.1005748

Saha, A., Wittmeyer, J., and Cairns, B. R. (2005). Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat. Struct. Mol. Biol.* 12, 747–755. doi: 10.1038/nsmb973

Santen, G. W., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., et al. (2012). Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat. Genet.* 44, 379–380. doi: 10.1038/ng.2217

Santen, G. W., Aten, E., Vulto-van Silfhout, A. T., Pottinger, C., van Bon, B. W. I, and van Minderhout, J. (2013). Coffin-Siris syndrome and the BAF complex: genotype-phenotype study in 63 patients. *Hum. Mutat.* 34, 1519–1528.

Schrier, S. A., Bodurtha, J. N., Burton, B., Chudley, A. E., Chiong, M. A., D'Avanzo, G. M., et al. (2012). The Coffin-Siris syndrome: a proposed diagnostic approach and assessment of 15 overlapping cases. *Am. J. Med. Genet. A* 158A, 1865–1876. doi: 10.1002/ajmg.a.35415

Schrier Vergano, S., Santen, G., Wieczorek, D., Wollnik, B., Matsumoto, N., and Deardorff, M. A. (1993). "Coffin-Siris Syndrome," in *GeneReviews((R))*, eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, et al. (Seattle, WA: University of Washington).

Schweingruber, C., Rufener, S. C., Zund, D., Yamashita, A., and Muhlemann, O. (2013). Nonsense-mediated mRNA decay – mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim. Biophys. Acta* 1829, 612–623. doi: 10.1016/j.bbagrm.2013.02.005

Seabra, C. M., Szoko, N., Erdin, S., Ragavendran, A., Stortchevoi, A., Maciel, P., et al. (2017). A novel microduplication of ARID1B: clinical, genetic, and proteomic findings. *Am. J. Med. Genet A* 173, 2478–2484. doi: 10.1002/ajmg.a.38327

Sekiguchi, F., Tsurusaki, Y., Okamoto, N., Teik, K. W., Mizuno, S., Suzumura, H., et al. (2019). Genetic abnormalities in a large cohort of Coffin-Siris syndrome patients. *J. Hum. Genet.* 64, 1173–1186.

Shakhova, O., and Sommer, L. (2008). "Neural crest-derived stem cells," in *StemBook [Internet]*, (Cambridge, MA: Harvard Stem Cell Institute).

Shi, H., Tao, T., Abraham, B. J., Durbin, A. D., Zimmerman, M. W., Kadoch, C., et al. (2020). ARID1A loss in neuroblastoma promotes the adrenergic-to-mesenchymal transition by regulating enhancer-mediated gene expression. *Sci. Adv.* 6:eaaz3440. doi: 10.1126/sciadv.aaz3440

Shibutani, M., Horii, T., Shoji, H., Morita, S., Kimura, M., Terawaki, N., et al. (2017). Arid1b haploinsufficiency causes abnormal brain gene expression and autism-related behaviors in mice. *Int. J. Mol. Sci.* 18:1872.

Sim, J. C., White, S. M., and Lockhart, P. J. (2015). ARID1B-mediated disorders: mutations and possible mechanisms. *Intractable Rare Dis. Res.* 4, 17–23.

Smith, A. L., Jung, E. M., Jeon, B. T., and Kim, W. Y. (2020). Arid1b haploinsufficiency in parvalbumin- or somatostatin-expressing interneurons leads to distinct ASD-like and ID-like behavior. *Sci. Rep.* 10:7834.

Sonmez, F. M., Uctepe, E., Gunduz, M., Gormez, Z., Erpolat, S., Oznur, M., et al. (2016). Coffin-Siris syndrome with cafe-au-lait spots, obesity and hyperinsulinism caused by a mutation in the ARID1B gene. *Intractable Rare Dis. Res.* 5, 222–226.

Stern, M., Jensen, R., and Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. *J. Mol Biol* 178, 853–868.

Sun, X., Chuang, J. C., Kanchwala, M., Wu, L., Celen, C., Li, L., et al. (2016). Suppression of the SWI/SNF component arid1a promotes mammalian regeneration. *Cell Stem Cell* 18, 456–466.

Takeda, T., Banno, K., Okawa, R., Yanokura, M., Iijima, M., Irie-Kunitomi, H., et al. (2016). ARID1A gene mutation in ovarian and endometrial cancers (Review). *Oncol. Rep.* 35, 607–613.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., et al. (1992). brahma: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68, 561–572.

Tang, L., Nogales, E., and Ciferri, C. (2010). Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* 102, 122–128.

Trizzino, M., Barbieri, E., Petracovici, A., Wu, S., Welsh, S. A., Owens, T. A., et al. (2018). The tumor suppressor ARID1A controls global transcription via pausing of RNA polymerase II. *Cell Rep.* 23, 3933–3945.

Tsurusaki, Y., Okamoto, N., Ohashi, H., Mizuno, S., Matsumoto, N., Makita, Y., et al. (2014). Coffin-Siris syndrome is a SWI/SNF complex disorder. *Clin. Genet.* 85, 548–554.

Tyagi, M., Imam, N., Verma, K., and Patel, A. K. (2016). Chromatin remodelers: we are the drivers!! *Nucleus* 7, 388–404.

Vals, M. A., Oiglane-Shlik, E., Noukas, M., Shor, R., Peet, A., Kals, M., et al. (2014). Coffin-Siris syndrome with obesity, macrocephaly, hepatomegaly and hyperinsulinism caused by a mutation in the ARID1B gene. *Eur. J. Hum. Genet.* 22, 1327–1329.

van der Sluijs, P. J., Jansen, S., Vergano, S. A., Adachi-Fukuda, M., Alanay, Y., AlKindy, A., et al. (2019). The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to Coffin-Siris syndrome. *Genet. Med.* 21, 1295–1307.

Vasileiou, G., Ekici, A. B., Uebe, S., Zweier, C., Hoyer, J., Engels, H., et al. (2015). Chromatin-remodeling-factor ARID1B represses Wnt/beta-catenin signaling. *Am. J. Hum. Genet.* 97, 445–456.

Vergano, S. A., van der Sluijs, P. J., and Santen, G. (1993). "ARID1B-Related Disorder," *GeneReviews((R))*, eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, G. Mirzaa, et al. (Seattle WA): University of Washington.

Vergano, S. S., and Deardorff, M. A. (2014). Clinical features, diagnostic criteria, and management of Coffin-Siris syndrome. *Am. J. Med. Genet. C Semin. Med. Genet.* 166C, 252–256.

Wang, W., Chi, T., Xue, Y., Zhou, S., Kuo, A., and Crabtree, G. R. (1998). Architectural DNA binding by a high-mobility-group/kinesin-like subunit in mammalian SWI/SNF-related complexes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 492–498.

Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., et al. (1996a). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J.* 15, 5370–5382.

Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R., and Crabtree, G. R. (1996b). Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev.* 10, 2117–2130.

Wieczorek, D., Bogershausen, N., Beleggia, F., Steiner-Haldenstatt, S., Pohl, E., Li, Y., et al. (2013). A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum. Mol. Genet.* 22, 5121–5135.

Wilsker, D., Patsialou, A., Zumbrun, S. D., Kim, S., Chen, Y., Dallas, P. B., et al. (2004). The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. *Nucleic Acids Res.* 32, 1345–1353.

Wu, R. C., Wang, T. L., and Shih, Ie.-M. (2014). The emerging roles of ARID1A in tumor suppression. *Cancer Biol. Ther.* 15, 655–664.

Wu, S., Fatkhutdinov, N., Rosin, L., Luppino, J. M., Iwasaki, O., Tanizawa, H., et al. (2019). ARID1A spatially partitions interphase chromosomes. *Sci. Adv.* 5:eaaw5294.

Xiao, W., Awadallah, A., and Xin, W. (2012). Loss of ARID1A/BAF250a expression in ovarian endometriosis and clear cell carcinoma. *Int. J. Clin. Exp. Pathol.* 5, 642–650.

Yamamoto, S., Tsuda, H., Takano, M., Tamai, S., and Matsubara, O. (2012). Loss of ARID1A protein expression occurs as an early event in ovarian clear-cell carcinoma development and frequently coexists with PIK3CA mutations. *Mod. Pathol.* 25, 615–624.

Yan, Z., Wang, Z., Sharova, L., Sharov, A. A., Ling, C., Piao, Y., et al. (2008). BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells* 26, 1155–1165.

Yu, D., Jiao, X., Cao, T., and Huang, F. (2018). Serum miRNA expression profiling reveals miR-486-3p may play a significant role in the development of autism by targeting ARID1B. *Neuroreport* 29, 1431–1436.

Yu, Y., Yao, R., Wang, L., Fan, Y., Huang, X., Hirschhorn, J., et al. (2015). De novo mutations in ARID1B associated with both syndromic and non-syndromic short stature. *BMC Genomics* 16:701. doi: 10.1186/s12864-015-1898-1

Zhai, Y., Kuick, R., Tipton, C., Wu, R., Sessine, M., Wang, Z., et al. (2016). Arid1a inactivation in an Apc- and Pten-defective mouse ovarian cancer model enhances epithelial differentiation and prolongs survival. *J. Pathol.* 238, 21–30.

Zhao, J., Chen, J., Lin, H., Jin, R., Liu, J., Liu, X., et al. (2016). The clinicopathologic significance of BAF250a (ARID1A) expression in hepatocellular carcinoma. *Pathol. Oncol. Res.* 22, 453–459.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

BRIEF RESEARCH REPORT
published: 04 March 2021
doi: 10.3389/fcell.2021.643603

Developmental Accumulation of Gene Body and Transposon Non-CpG Methylation in the Zebrafish Brain

Samuel E. Ross[1,2], *Daniel Hesselson*[3,4] *and Ozren Bogdanovic*[1,5]*

[1] Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, NSW, Australia, [2] Faculty of Medicine, St Vincent's Clinical School, University of New South Wales, Sydney, NSW, Australia, [3] Centenary Institute, The University of Sydney, Sydney, NSW, Australia, [4] Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia, [5] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
José Luis Sardina,
Josep Carreras Leukaemia Research Institute (IJC), Spain

Reviewed by:
Luo Chongyuan,
University of California, Los Angeles, United States
Michael Imbeault,
University of Cambridge, United Kingdom
Joan Barau,
Institute of Molecular Biology, Germany

***Correspondence:**
Ozren Bogdanovic
o.bogdanovic@garvan.org.au;
o.bogdanovic@gmail.com

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: 18 December 2020
Accepted: 08 February 2021
Published: 04 March 2021

Citation:
Ross SE, Hesselson D and Bogdanovic O (2021) Developmental Accumulation of Gene Body and Transposon Non-CpG Methylation in the Zebrafish Brain. Front. Cell Dev. Biol. 9:643603. doi: 10.3389/fcell.2021.643603

DNA methylation predominantly occurs at CG dinucleotides in vertebrate genomes; however, non-CG methylation (mCH) is also detectable in vertebrate tissues, most notably in the nervous system. In mammals it is well established that mCH is targeted to CAC trinucleotides by DNMT3A during nervous system development where it is enriched in gene bodies and associated with transcriptional repression. Nevertheless, the conservation of developmental mCH accumulation and its deposition by DNMT3A is largely unexplored and has yet to be functionally demonstrated in other vertebrates. In this study, by analyzing DNA methylomes and transcriptomes of zebrafish brains, we identified enrichment of mCH at CAC trinucleotides (mCAC) at defined transposon motifs as well as in developmentally downregulated genes associated with developmental and neural functions. We further generated and analyzed DNA methylomes and transcriptomes of developing zebrafish larvae and demonstrated that, like in mammals, mCH accumulates during post-embryonic brain development. Finally, by employing CRISPR/Cas9 technology, we unraveled a conserved role for Dnmt3a enzymes in developmental mCAC deposition. Overall, this work demonstrates the evolutionary conservation of developmental mCH dynamics and highlights the potential of zebrafish as a model to study mCH regulation and function during normal and perturbed development.

Keywords: DNA methylation, brain, nervous system, zebrafish, repetitive elements

BACKGROUND

In genomes of vertebrate adult somatic cells, the majority of CpG sites are methylated (>80%) with the exception of CpG-rich promoters and distal regulatory elements (Bird, 2002; Jones, 2012; Schübeler, 2015). While otherwise ubiquitous, CpG methylation (mCG) at regulatory elements is known to participate in long-term gene silencing processes (de Mendoza et al., 2019). In mammals,

albeit at much lower levels, methylation of cytosines outside the CpG context (mCH, H = T,C,A) has also been reported in the majority of tissues (Schultz et al., 2015). mCH, or more particularly methylation of CA dinucleotides (mCA), occurs most commonly in mammalian embryonic stem cells (ESCs) and in the brain (Lister et al., 2009, 2013; Schultz et al., 2015). In ESCs, mCH is enriched at CAG trinucleotides in gene bodies and is positively correlated with gene expression. Additionally, increased levels of mCH were observed at repetitive elements in ESCs (Ziller et al., 2011; Arand et al., 2012; Guo et al., 2014b). In mammalian brains, mCH rivals the levels of mCG and is enriched at CAC trinucleotides (mCAC) in gene bodies where it negatively correlates with expression and is deposited *de novo* by DNMT3A (Lister et al., 2013). In line with its repressive role in the nervous system, mCH is depleted at open chromatin regions (Lister et al., 2013). mCH in the postnatal mammalian brain displays a rapid increase during initial phases of synaptogenesis, which corresponds to 2–4 weeks in mouse, and first 2 years of life in humans. This is followed by a longer period of slower accumulation (Lister et al., 2013). While mCH is found at high levels and studied extensively in plants (Zhang et al., 2018), the function of mCH in vertebrates is less well known. Several studies, however, have demonstrated that Methyl-CpG Binding Protein 2 (MeCP2) is able to bind to and regulate genes marked by mCH, which was particularly evident at long genes (Guo et al., 2014a; Chen et al., 2015; Gabel et al., 2015; Boxer et al., 2020; Clemens et al., 2020). Whether this is due to biological or technical reasons is currently debated (Raman et al., 2018). Mutations in MeCP2 are the most prevalent cause of Rett syndrome, and interestingly, altered readout of mCH deposited by DNMT3A appears to play a central role in Rett syndrome pathogenesis (Chen et al., 2015; Lavery et al., 2020). MeCP2 is conserved across vertebrates, such as zebrafish, where depletion of MeCP2 results in similar pathologies to Rett syndrome including altered motor behavior, improper synapse formation and acute inflammation (Pietri et al., 2013; Gao et al., 2015; Nozawa et al., 2017; van der Vaart et al., 2017).

A recent report described the conserved enrichment of mCH in vertebrate brains, which originated alongside MeCP2 and DNMT3A enzymes at the root of the vertebrate lineage (de Mendoza et al., 2021). This study also highlighted the anti-correlation between gene body mCH and expression in some, but not all, vertebrate brains. In our previous work, we found highly specific mCH enrichment at TGCT tetranucleotides within zebrafish mosaic satellite repeats in embryonic and adult tissues, deposited by the teleost specific Dnmt3ba enzyme (Ross et al., 2020). However, the developmental dynamics and distribution of neural-specific mCH, and a functional role for DNMT3A or MeCP2 in relation to mCH, has yet to be demonstrated outside of mammalian brains. Here we expand upon the utility of the zebrafish model in the study of mCH and reveal that like in mammals, mCH accumulates during brain development via Dnmt3a enzymes and becomes enriched at downregulated genes and Tc1-like transposable elements. This study thus extends our knowledge of vertebrate mCH conservation and lays the foundation for future work that will allow for the precise dissection of mCH regulatory functions during zebrafish embryogenesis and nervous system formation.

MATERIALS AND METHODS

Zebrafish Usage and Ethics

Zebrafish work was conducted at the Garvan Institute of Medical Research in accordance with the Animal Ethics Committee AEC approval and with the Australian Code of Practice for Care and Use of Animals for Scientific Purposes. Adult wild type (AB/Tübingen) *Danio rerio* (zebrafish) were bred in an equal ratio of males and females. Embryos were collected 0 h post-fertilization (hpf) and incubated in 1X E3 medium (5 mM NaCl, 0.33 mM CaCl2, 0.17 mM KCl, 0.33mM H14MgO11S) for 4 days at 28.5°C before being transferred onto a filtered system.

Genomic DNA and RNA Extraction

Whole brains were dissected from zebrafish larvae and adults before being snap-frozen in liquid nitrogen and stored at −80°C. Genomic DNA (gDNA) was extracted from brains using the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, Chadstone, VIC, Australia) according to manufacturer's instructions. For RNA extraction, half of the lysate from the first step of DNA extraction from the QIAGEN DNeasy Blood & Tissue Kit was added to TRIsure (Bioline) and purified following manufacturer's instructions. All experiments in this study were performed in two biological replicates.

CRISPR/Cas9 Zebrafish Knockouts

Guide RNAs (gRNA) targeting *dnmt3aa* and *dnmt3ab* loci were designed with CRISPRscan (Moreno-Mateos et al., 2015). gRNAs for both loci were synthesized and co-injected into 1-cell stage embryos as previously described (Ross et al., 2020). CRISPR/Cas9 knockouts (cKO) fish were grown to 4 weeks of age before their brains were harvested for DNA and RNA extraction. Amplicons surrounding the CRISPR/Cas9 cut sites were PCR-amplified from genomic DNA, ligated to NEXTFLEX Bisulfite-Seq barcodes (PerkinElmer, Waltham, MA, United States), and spiked into libraries that were sequenced on the Illumina HiSeq X platform. Knockout efficiencies were calculated from the sequenced amplicons using CRISPResso (Pinello et al., 2016). RNA was reverse transcribed to cDNA using the SensiFAST^{TM} cDNA Synthesis Kit (Bioline), following the manufacturer's protocol and subjected to qPCR analysis. Relative expression levels were calculated using the $2-\Delta\Delta CT$ method and *bactin* gene as the control transcript. Two sample *t*-tests were performed on CT values. All oligos used in this study can be found in **Supplementary Table 1**.

Whole Genome Bisulfite Sequencing (WGBS)

WGBS libraries were prepared from 500 ng of zebrafish brain gDNA, spiked with 0.025 ng of unmethylated lambda phage DNA (Promega, Madison, WI, United States) and sequenced on the Illumina HiSeq X platform (2 × 150 bp) as previously described (Ross et al., 2020).

Reduced Representation Bisulfite Sequencing (RRBS)

RRBS libraries were prepared from 500 ng of zebrafish brain gDNA spiked with 0.025 ng of unmethylated lambda phage DNA (Promega, Madison, WI, United States). gDNA was digested with 10 U *Bcc*I (CCATC(N)$_4$) and 10 U *Ssp*I (AATATT) for 2 h. A separate aliquot was digested with 20 U *Msp*I (New England BioLabs, Ipswich, MA, United States). RRBS libraries were constructed as previously described (Ross et al., 2020), sequenced, and the BAM files corresponding to different digestion reactions (*Bcc*I/*Ssp*I or *Msp*I) were merged before downstream analysis.

WGBS and RRBS Data Analyses

WGBS reads were trimmed with Trimmomatic: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW: 5:20 LEADING:3 TRAILING:3 MINLEN:60 (Bolger et al., 2014), and mapped using WALT (-m 5 -t 20 -N 10000000) (Chen et al., 2016) onto the GRCz11 reference genome (UCSC), containing the λ genome. BAM files, containing only uniquely mapped reads, were deduplicated using sambamba markdup (Tarasov et al., 2015). RRBS data were processed as above with the additional option of: HEADCROP:5 CROP:140 added during trimming and without deduplication. BAM files were made FLAG-compatible and processed with CGmapTools (Chen et al., 2016; Guo et al., 2018) (convert bam2cgmap) to obtain ATCGmap files, which were corrected for CH positions that showed evidence of CG SNPs (de Mendoza et al., 2021). A summary of library statistics can be found in **Supplementary Table 2**. Genomic data were visualized in UCSC (Kent et al., 2002) and IGV (Robinson et al., 2011) browsers.

DNA Sequence Motif Analyses

BED file coordinates of the 10,000 most highly methylated mCH and mCAC sites, with a minimal depth of 10, were extended by 4 base pairs upstream and downstream. The resulting files were used as input for HOMER "findMotifsGenome.pl" function (Heinz et al., 2010), establishing the search for *de novo* motifs of length 9 (-len 9 -size given) with the GRCz11 genome used as the background sequence. Motifs were visualized using the "ggseqlogo" package in R (Wagih, 2017) and motif positions in the genome were called using the scanMotifGenomeWide.pl function (with and without -mask option checked).

mCH Level Calculation and Plotting

Bedgraphs were generated from corrected CGmapTools outputs and converted to bigWig using bedGraphToBigwig script from Kent utils. Average mCH levels were determined from bedGraph files and calculated by dividing the sum of reads supporting a methylated cytosine by the sum of all reads mapping to that position. mCH levels in genomic features and gene bodies were calculated using BEDtools map (Quinlan and Hall, 2010). mCH levels, TPMs, and gene length were plotted using the boxplot function in R (outline = FALSE). Heatmaps were generated using deepTools (Ramirez et al., 2014) computeMatrix with the following parameters: "computeMatrix scale-regions -m 650 -b 500 -a 500 -bs 25." NAN values were replaced with 0. Heatmaps were plotted with the plotHeatmap function, sorted, and clustered based on methylation levels. Scatterplots were generated using the *geom_bin2d* function in *ggplot2* [(bins = 50) + geom_smooth(method = l m)]. Pearson correlations were calculated using the *rcorr* function in R.

DMR Calling

Differentially methylated regions (DMRs) were called using DSS (delta = 0.1, p.threshold = 0.05, minlen = 100, minCG = 5, dis.merge = 500, pct.sig = 0.5) (Feng et al., 2014).

Repeatmasker Track Analyses

Repeatmsker tracks were obtained from UCSC. The percentage of repeat subfamilies overlapping the top-methylated CAC motifs was determined with BEDtools (intersectBed).

Correlation of mCH With Genomic Features

Sequenced ChIP-seq reads (Kaaij et al., 2016; Yang et al., 2020) were trimmed with Trimmomatic (ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 SLIDINGWINDOW:5:20 LEADING:3 TRAILING:3 MINLEN:20) (Bolger et al., 2014) before being mapped to the GRCz11 genome using Bowtie2 with default settings (Langmead and Salzberg, 2012). BAM files were deduplicated using sambamba markdup (Tarasov et al., 2015). Peaks were called using MACS2 (Zhang et al., 2008). bigWigs were generated using deepTools (Ramirez et al., 2014) bamCompare (-e 300 -p 20 –normalizeUsing RPKM –centerReads), and bedGraphs using UCSC bigWigToBedGraph. Spearman correlations were calculated using BEDtools map (Quinlan and Hall, 2010) and rcorr(type = c("spearman") from the *Hmisc* package in R.

RNA-Seq

RNA-seq libraries were prepared with 1,000 ng of input RNA using the KAPA mRNA HyperPrep Kit, according to manufacturer's instructions and sequenced on the Illumina HiSeq X platform (2 × 150 bp).

RNA-Seq Analyses

RNA-seq reads were trimmed using Trimmomatic: ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 SLIDING-WINDOW:5:20 LEADING:3 TRAILING:3 MINLEN:60 (Bolger et al., 2014) and aligned to the GRCz11 genome using STAR (Dobin et al., 2013). Differential gene expression analysis was performed using edgeR (Robinson et al., 2010). Robinson et al., 2010 with genes selected based on a minimum ± 1.5 logFC (FDR < 0.05) between any of the analyzed time points. Z-scores were calculated based on the log2 transformations of TPM values and plotted with the *pheatmap* package in R. Analysis of published RNA-seq data was performed based on the provided read count tables with TPM values calculated from the average of 5–6 month old brain datasets (Aramillo Irizar et al., 2018) or from collated counts of 30 neurons (Lange et al., 2020).

RESULTS AND DISCUSSION

mCH Is Enriched at Defined CAC-Containing Motifs in Zebrafish Brains

To investigate mCH in the zebrafish nervous system, we analyzed WGBS data (bisulfite conversion > 99.5%) of adult brain (mean coverage = 9.9X), as well as of adult liver (mean coverage = 7.6X), to use as a non-neural control tissue (Bogdanovic et al., 2016; Skvortsova et al., 2019). We employed stringent genotype correction (de Mendoza et al., 2021) to allow for more sensitive interrogation of mCH patterns. To better understand the sequence context of mCH deposition in the zebrafish brain and how it compares to non-neural tissues (liver), we performed a motif search on the 10,000 most highly methylated CH sites. Expectedly, we recovered the TGCT motif associated with mosaic satellite repeats in both brain and liver (Ross et al., 2020), and the previously characterized TACAC-containing motif (Lister et al., 2013; de Mendoza et al., 2021), which was specific to the brain sample (**Supplementary Figure 1A**). Furthermore, the brain sample showed a notable enrichment in CAC trinucleotide methylation (∼1.5%) with a threefold increase compared to the unmethylated lambda genome spike-in control (∼0.5%) (**Supplementary Figure 1B**). This mCAC enrichment was not evident in the liver sample, thus eliminating the possibility of sequence-specific biases or artifacts pertaining to bisulfite conversion (Olova et al., 2018). We next investigated the genomic distribution of mCAC in zebrafish brains to assess if the depletion in regulatory elements and enrichment in gene bodies previously described in mammals (Lister et al., 2013; Guo et al., 2014a; He and Ecker, 2015) is evolutionarily conserved. To achieve this, we annotated transcription start sites (TSS), exons, introns, 5′UTRs, 3′UTRs, intergenic regions, and sites of H3K27ac enrichment, which correspond to active gene-regulatory elements (Kaaij et al., 2016). We found that intronic and intergenic regions are the only regions enriched in mCAC and mCH (**Supplementary Figures 1C–E**), whereas a notable depletion, similar to the one described in mammals, was observed at H3K27ac peaks (**Supplementary Figure 1F**). These results support the previous observations of mCA presence in gene bodies of vertebrate brains (de Mendoza et al., 2021) and demonstrate a conserved depletion of mCH in active regulatory regions.

Further analyses of sequence motifs associated with the mCAC context unraveled two novel sequences in addition to the previously described vertebrate-conserved TACAC motif (**Figure 1A**). Due to previous associations of mCH with repetitive DNA in zebrafish (Ross et al., 2020), we wanted to assess if any of the most significantly methylated CAC motifs were enriched in repetitive elements. The top TACAC motif displayed comparable methylation levels in repetitive elements and in the repeat-masked genomic fraction, with an average methylation level nearly three-fold higher (∼4%) than the average global mCAC levels (**Figure 1B**). However, for the remaining two motifs we found a robust increase in average methylation levels at repetitive elements when compared to the repeat-masked genome (∼6.5% and 6%, **Figure 1B**). Further analysis revealed that the TACAC motif is broadly distributed in the genome whereas the second and third motif were mainly located in TDR and TC1DR3 repetitive elements, respectively (**Figure 1C**). Both TDR and TC1DR3 elements belong to the Tc1-mariner superfamily, which is found across eukaryotes. These elements are characterized by two inverted terminal repeats, and an open reading frame (ORF) (Wicker et al., 2007). The identified motifs are found on average once per element at positions 205/541 and 1005/1071 in TDR18 and TC1DR3 model sequences, respectively (Storer et al., 2021). To further validate these associations, we interrogated previously published heart and forebrain WGBS DNA methylomes (Ross et al., 2020; de Mendoza et al., 2021). These analyses again revealed brain-specific enrichment of mCH specifically at these elements (**Supplementary Figure 1G**), thus excluding the possibility of library preparation and sequencing artifacts causing the observed enrichment.

We also found that average methylation of the top three methylated CAC motifs correlated strongly with overall mCH in gene bodies (R = 0.52) (**Figure 1D**), suggestive of a significant contribution of CAC methylation to gene body mCH. This correlation was stronger than the one observed between average gene mCH and gene length (R = 0.36), which was previously described in mammals (Gabel et al., 2015; Boxer et al., 2020; **Supplementary Figure 1H**). Additionally, when Gene ontology (GO) analysis (Raudvere et al., 2019) of genes containing methylated TDR and TC1DR3 CAC motifs was performed, we found overrepresentation in developmental and neural development terms (**Figure 1E**). Due to its widespread genomic abundance the TACAC motif was omitted from the GO analysis. Overall, zebrafish brain is enriched in mCH, particularly in the CAC trinucleotide context, predominantly in introns and intergenic regions, as well as in members of the Tc1-mariner transposon family.

mCH Is Targeted to Genes Downregulated in the Nervous System

To explore the relationship between mCH and gene expression in the zebrafish brain, we plotted average gene mCH and mCAC values against gene expression (transcripts per million-TPM) levels from available datasets (Bogdanovic et al., 2016; Aramillo Irizar et al., 2018; **Figure 2A**). We revealed a gene cluster (n = 1,860) with higher-than-average mCH levels, which displayed lower expression than the bulk transcriptome (**Figures 2A,B**). To provide more genomic context to these findings, we first interrogated whether this elevated mCH was driven by a higher proportion of intron sequence in these genes. To that end, we plotted gene length, gene body mCAC, intron mCAC and exon mCAC for mCH-enriched genes as well as for genes that did not display any notable mCH enrichment (**Figure 2C**). While the mCH-enriched genes were significantly longer (Wilcoxon test, *** $P < 0.001$), in line with observations in mammals (Gabel et al., 2015; Boxer et al., 2020), the elevation in mCH was not driven exclusively by intron

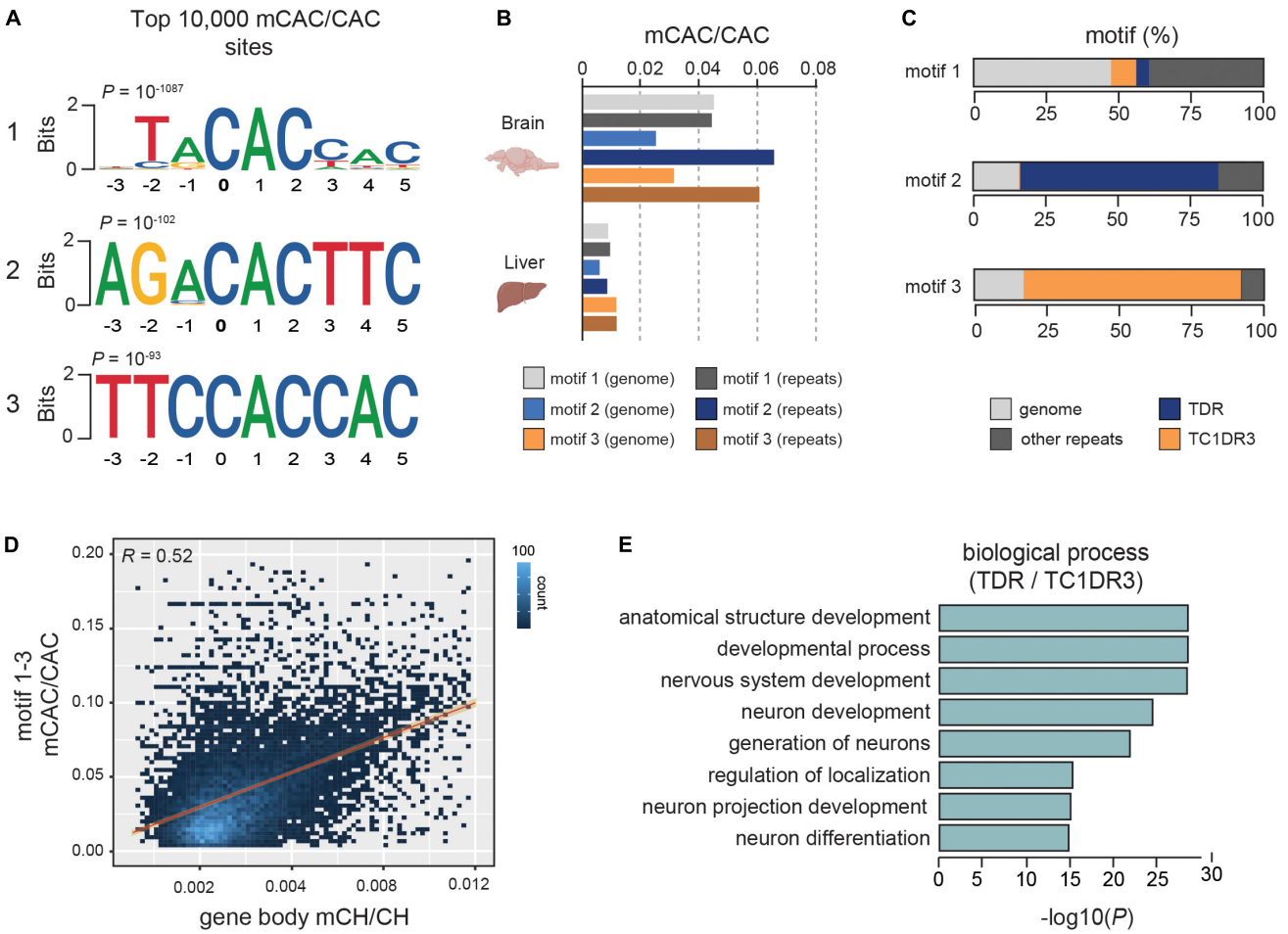


FIGURE 1 | mCH is enriched at defined CAC-containing motifs in zebrafish brains. **(A)** Top three motifs called from the 10,000 most methylated CAC trinucleotides in the zebrafish brain. **(B)** Average mCAC/CAC methylation levels of the top three mCAC motifs in the bulk genome and repetitive elements of zebrafish brain and liver. **(C)** Genomic annotation of top ranked CAC motifs. **(D)** Genomic correlation between average gene body mCH/CH and mCAC/CAC at top three most methylated CAC motifs. **(E)** Gene ontology enrichment of genes containing methylated CAC motifs in TDR and TC1DR3 elements.

contribution as both introns and exons located within these genes had significantly higher levels of mCAC (Wilcoxon test, $^{***}P < 0.001$) (**Figure 2C**). Interestingly, genes with higher levels of mCAC also contained a higher percentage of Tc1-like elements in relation to total gene length (**Figure 2D**). To confirm the observation of poorly expressed genes being marked by mCH, we plotted average TPM levels from total RNA-seq data from adult brains (Aramillo Irizar et al., 2018) and combined single cell data from neurons (Lange et al., 2020; **Figure 2E**). Genes with higher levels of mCH (cluster 1) had significantly lower average TPM when compared to genes with moderate/low levels of mCH (cluster 2) (Wilcoxon test, $^{***}P < 0.001$), or to a randomly selected subset of genes (n = 1,860, **Figure 2E**) (Wilcoxon test, $^{***}P < 0.001$). This difference in expression levels between the two clusters was even more pronounced in neurons where mCH is expected to be the highest based on mammalian data (Lister et al., 2013; **Figure 2E**). GO analysis of these highly mCH-methylated genes again revealed enrichment for terms associated with embryonic and neural development (**Supplementary Figure 2A**). This enrichment of developmental genes for mCH is also in line with observations in other vertebrate brains (de Mendoza et al., 2021).

As mCH-enriched genes are on average poorly expressed, longer, and associated with neuronal and development terms, it is yet unclear which features are most important or predictive for mCH enrichment. This is further complicated by neuronal terms and gene lengths being tightly associated in zebrafish, as discussed in our previous study (Ross et al., 2020). To further explore and rank features that may be associated with genomic mCH, we analyzed histone modification ChIP-seq data from the zebrafish brain (Yang et al., 2020) and assessed the correlations between these diverse histone marks and mCH levels in gene bodies (**Supplementary Figure 2B**). We also performed correlation analyses for mCH and various other genomic features (**Supplementary Figure 2B**). These analyses revealed that gene length has the strongest positive correlation with gene body mCH levels and that H3K4me3 at transcription start sites has the strongest

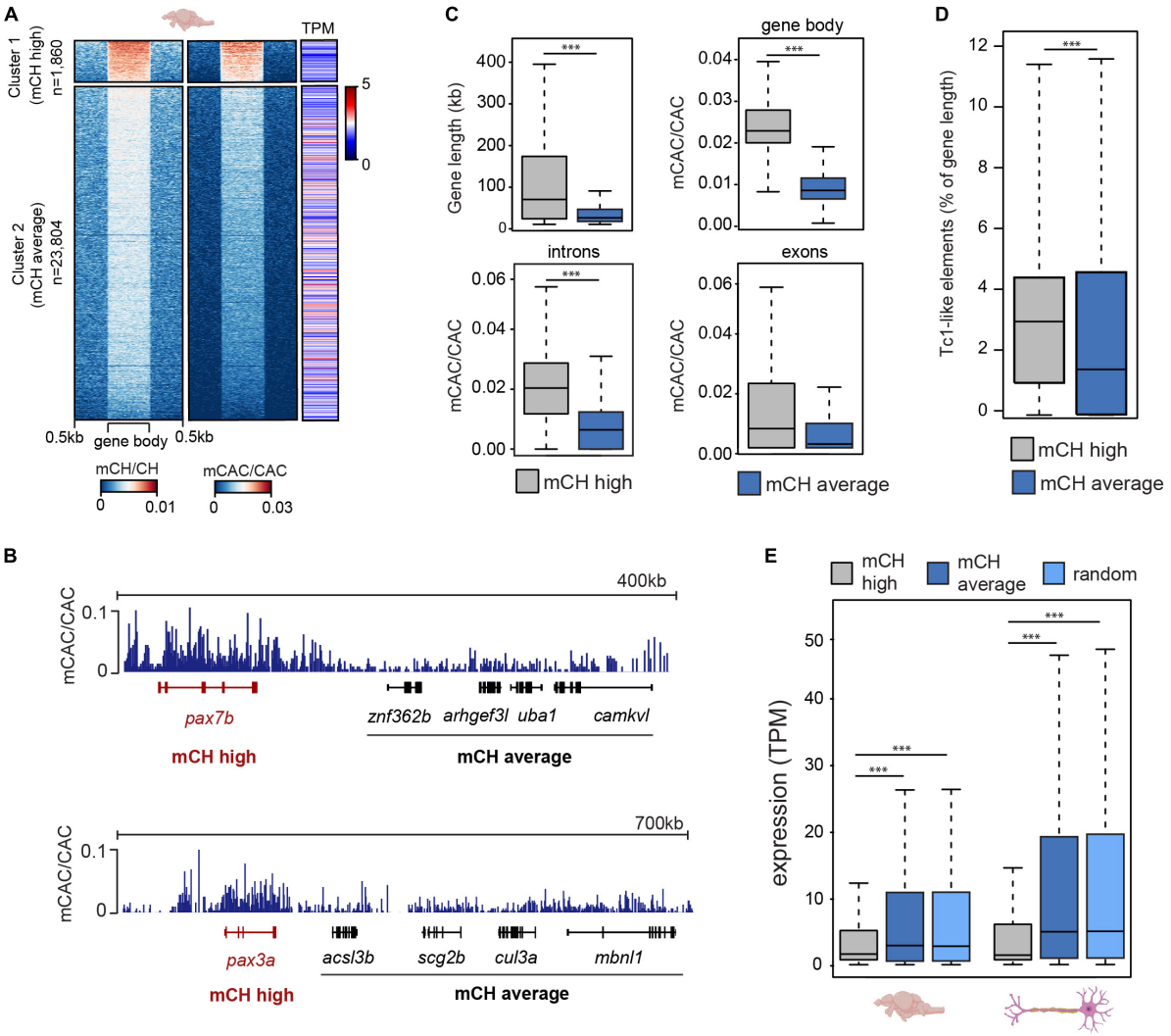


FIGURE 2 | mCH is present at long genes with low expression levels. **(A)** mCH/CH levels, mCAC/CAC levels, and gene expression represented as transcripts per million (TPM) in adult zebrafish brains. **(B)** Genome browser snapshot of mCAC/CAC levels in cluster 1 (red) and cluster 2 (black) genes in adult brains. **(C)** Comparisons of gene length (top left), average gene mCAC/CAC (top right), average exon mCAC/CAC (bottom left) and average intron mCAC/CAC (bottom right) in genes with high (cluster 1) and moderate/low levels of mCAC/CAC (cluster 2) (Wilcoxon test, $^{***}P < 0.001$). **(D)** Tc1-like element percentage of total gene length in genes with high (cluster 1) and moderate/low levels of mCAC/CAC (cluster 2) (Wilcoxon test, $^{***}P < 0.001$). **(E)** Average expression levels (TPM) in 6-month old brains ($n = 5$) and neurons ($n = 30$) at genes with high (cluster 1), or moderate/low (cluster 2) mCH, and a random subset of genes ($n = 1,860$) sampled from cluster 2 (Wilcoxon test, $^{***}P < 0.001$).

negative correlation with gene body mCH. Overall, these results further demonstrate that mCH-enriched genes are longer and associated with a repressive chromatin environment. Therefore, like in mammals, mCH in the zebrafish adult brain is associated with transcriptional repression which is particularly evident in long genes.

mCH Accumulates During Zebrafish Brain Development

As mCH has previously been shown to accumulate during mammalian brain development (Lister et al., 2013), we next investigated whether comparable mCH dynamics could be observed in zebrafish. We generated RRBS libraries (Meissner et al., 2005) using a combination of enzymes, which were selected based on virtual digestion, to enrich for regions containing highly methylated CAC motifs (**Figure 1A**) identified in adult brains. We assayed zebrafish brains starting from 1 week (1W), where brain structures such as the cerebrum are not identifiable, up until 6 weeks (6W) and adulthood, where all structures are discernable (Maeyama and Nakayasu, 2000). This analysis revealed a gradual increase in mCAC in the brains of 1-week to 6-week-old zebrafish followed by a more notable increase in adult brains (**Figure 3A**). RNA-seq analysis across this period also revealed a gradual decrease in the expression of components of DNA methylation machinery as cells presumably become

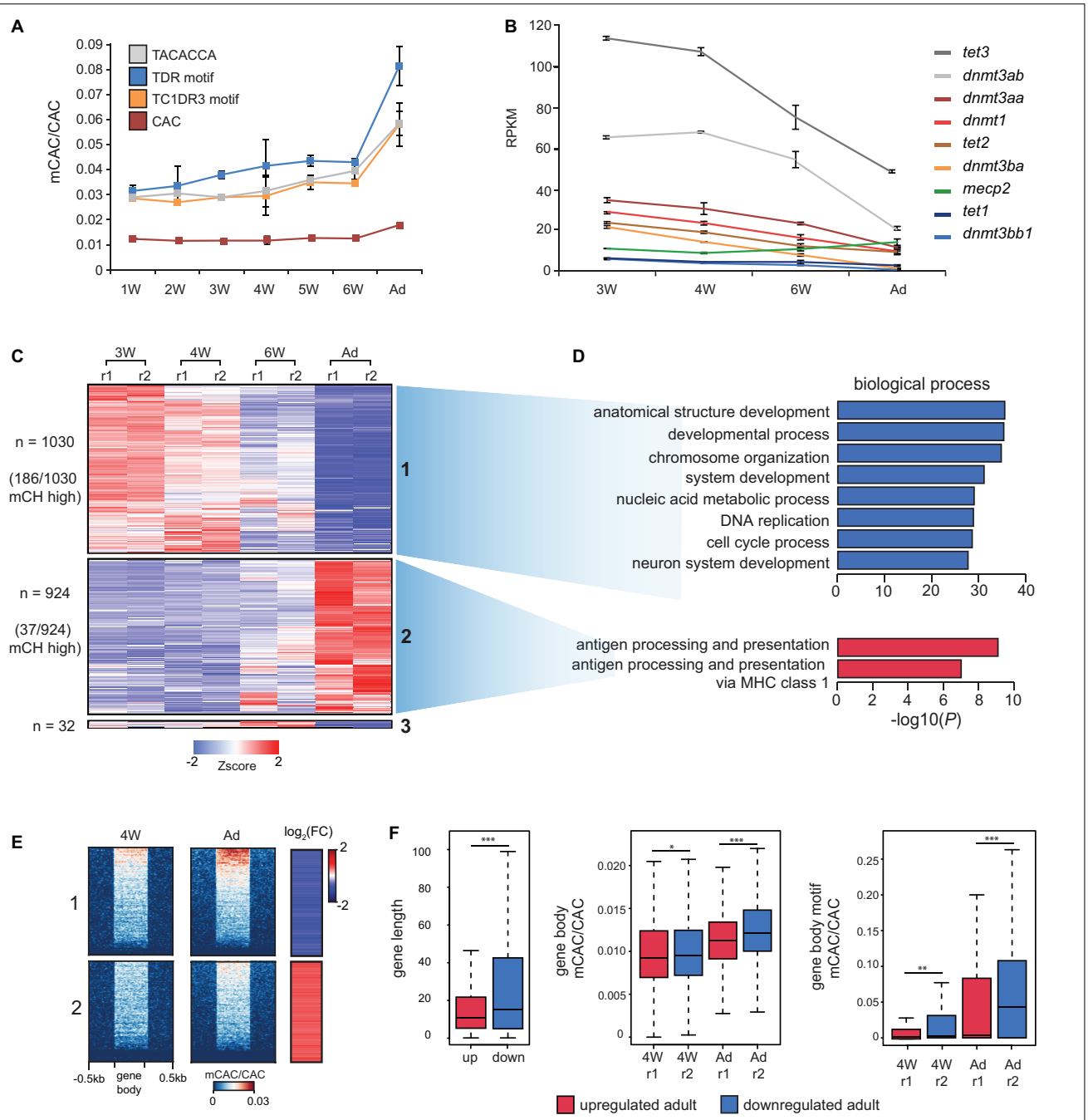


FIGURE 3 | mCH accumulates in the developing nervous system. **(A)** mCAC/CAC levels at all CAC trinucleotides and top methylated CAC motifs in larval (W = weeks old) and adult (Ad) brains, as determined by RRBS. Data is represented as the average of two biological replicates with error bars (standard deviation). **(B)** RPKM (reads per kilobase per million) values of *dnmt*, *tet*, and *mecp2* transcripts in larval and adult brains determined by RNA-seq. Data is represented as the average of two biological replicates with error bars (standard deviation). **(C)** Transcription intensities of a merged collection of differentially expressed genes called between all pairwise comparisons of larval and adult stages (r1—replicate 1, *r2* = replicate 2). **(D)** Gene ontology enrichment of differentially expressed genes in larval and adult brains. **(E)** mCAC levels and relative RNA expression levels (log2 fold change 4W/Ad) at differentially expressed genes. **(F)** Comparisons of gene length (left), gene body mCAC/CAC (middle), and gene body mCAC/CAC at top methylated motifs (right) in genes that are either upregulated or downregulated in the adult brain (Wilcoxon test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

more differentiated (**Figure 3B**). Moreover, *mecp2* expression increased during brain development coinciding with increase in mCH. This observation supports a conserved role for MeCP2 in the regulation of genes marked by mCH in vertebrates (Guo et al., 2014a; Chen et al., 2015; Gabel et al., 2015; de Mendoza et al., 2021). Differential expression analysis of all genes

across brain development revealed two major gene clusters which were either consistently upregulated (cluster 1) or downregulated (cluster 2) during brain development (**Figure 3C**). Interestingly, 18% of the downregulated genes belonged to the mCH-enriched gene cluster, compared to only 4% of the upregulated genes (χ^2-test $^{***}P < 0.001$) (**Figure 3C**). Downregulated genes were also associated with terms related to cell division and development while genes that were upregulated were enriched in terms associated with adaptive immunity (**Figure 3D**). This are consistent with ongoing developmental processes in the larval brain and with the notion that the adaptive immune system of zebrafish does not develop until 4 to 6 weeks post fertilization (Lam et al., 2004).

Since we determined that mCH dynamics in the developing zebrafish brain are comparable to developmental mCH dynamics in the postnatal mouse frontal cortex (Lister et al., 2013), we next wanted to explore if mCG dynamics similar to the one observed in mice could also be detected in zebrafish. Pairwise analyses of differentially methylated regions (DMRs) from RRBS data identified 1723 DMRs. The strongest difference in mCG was observed between 1-week-old brain and adult brain (**Supplementary Figure 3A**). This analysis also revealed that the 1-week-old zebrafish brain is most similar to the fetal mouse sample, as both the 1-week-old zebrafish brain and fetal mouse brain displayed most obvious differences when compared to other time points (Lister et al., 2013). The number of DMRs identified in zebrafish (n = 1,723) significantly differs from the number of DMRs identified in developing mouse brains (n > 142,835). This discrepancy is likely caused by different 5mC detection approaches (RRBS instead of WGBS), however other contributing factors could include: the use of whole brains instead of sorted neuron and glial populations, developmental stages not being directly comparable between zebrafish and mouse, adult zebrafish brains retaining more "juvenile features" such as radial glial cells and neurogenic capabilities (Schmidt et al., 2013), and different glia/neuron percentages of the samples, as mCG levels in glia have been reported to be more stable during development (Lister et al., 2013).

To understand better how methylation and gene expression dynamics track over developmental time, we generated WGBS datasets for 4-week-old brain tissue and compared these data against adult brain WGBS and RNA-seq data. Analysis of mCAC levels of differentially expressed genes revealed that developmentally downregulated genes accumulate more mCH when compared to upregulated ones (**Figure 3E**). This trend was also observed when visualizing mCAC levels and gene expression levels across all genes (**Supplementary Figures 3B,C**). Furthermore, quantification of mCAC levels at all mCAC trinucleotides and at the highly methylated motifs (all three combined), confirmed a significant increase in the methylation of developmentally downregulated genes (Wilcoxon test, $^{***}P < 0.001$) (**Figure 3F**). This increase in mCH in adult brains is uncoupled from global mCG changes, as DMR analysis of these WGBS datasets revealed 23,992 DMRs with the majority (n = 18,522) becoming hypomethylated during nervous system development (**Supplementary Figures 3D,E**). Altogether, these results demonstrate robust anticorrelation between mCH and gene expression during brain development in zebrafish, as well as developmental mCH accumulation, in line with observations in mammals.

Dnmt3a Enzymes Are Required for Methylation of CAC Trinucleotides in the Zebrafish Brain

Finally, to investigate if Dnmt3a-dependent methylation of CAC trinucleotides is evolutionarily conserved in zebrafish, we generated *dnmt3aa*/*dnmt3ab* CRISPR/Cas9 double knockouts (cKO). qPCR analysis of cDNA extracted from brains of 4-week-old cKOs revealed their partial knockout status with a 50% reduction in expression levels for both *dnmt3aa* and *dnmt3ab* (**Figure 4A**). Sequencing of amplicons surrounding the CRISPR/Cas9 cut sites demonstrated comparable estimates of genome editing efficiency (**Figure 4B**). Finally, WGBS analysis of these samples demonstrated that depletion of *dnmt3aa*/*dnmt3ab* resulted in a significant ($P < 0.05$, Wilcoxon test) reduction in global mCAC levels as well as in a notable reduction (43%) of mCH at top mCAC motifs ($P < 0.01$, Wilcoxon test) (**Figure 4C**). This significant reduction in mCH observed already in these partial cKOs, which could have undergone possible selection for more wild type cells, suggests that Dnmt3aa and Dnmt3ab are the major enzymes responsible for neural mCAC deposition. The reduction in mCAC levels in these cKOs can be observed across the majority of gene bodies (**Figure 4D**) and on a genome-wide and locus-specific scale (**Figure 4E**). Notably, these perturbations in mCH did not result in any obvious morphological defects in the cKO fish (data not shown). Finer analysis of brain morphology, behavior, and potential inflammatory processes in these animals will be a focus of future studies. Additionally, no changes in global CpG methylation levels or significant DMRs could be detected (**Supplementary Figure 3F**) between the cKO and WT brain, suggestive of mCH deposition being a major function of Dnmt3a enzymes in zebrafish. This is contrary to what has been observed in mouse neurons, where Dnmt3a KO resulted in ∼10% decrease in global mCG/CG levels (Lavery et al., 2020). However, given the incomplete KO, the possible redundancies with other zebrafish DNMTs (Goll and Halpern, 2011), the propagation by Dnmt1 once established, and the use of bulk cell DNA methylome data, we cannot completely rule out a role for these enzymes in neuronal CpG methylation. Nevertheless, the clear role in mCH deposition, and the fact that Dnmt3a enzymes can be traced back to the root of vertebrates (de Mendoza et al., 2021), suggests conserved functional importance of mCH in the vertebrate nervous system.

DISCUSSION

mCH has been established as an important base modification with likely biological functions during mammalian brain development (Lister et al., 2013), and links to Rett syndrome pathogenesis (Chen et al., 2015; Gabel et al., 2015; Boxer et al., 2020; Lavery et al., 2020), there are still many unknowns related to its regulation and function. Furthermore,

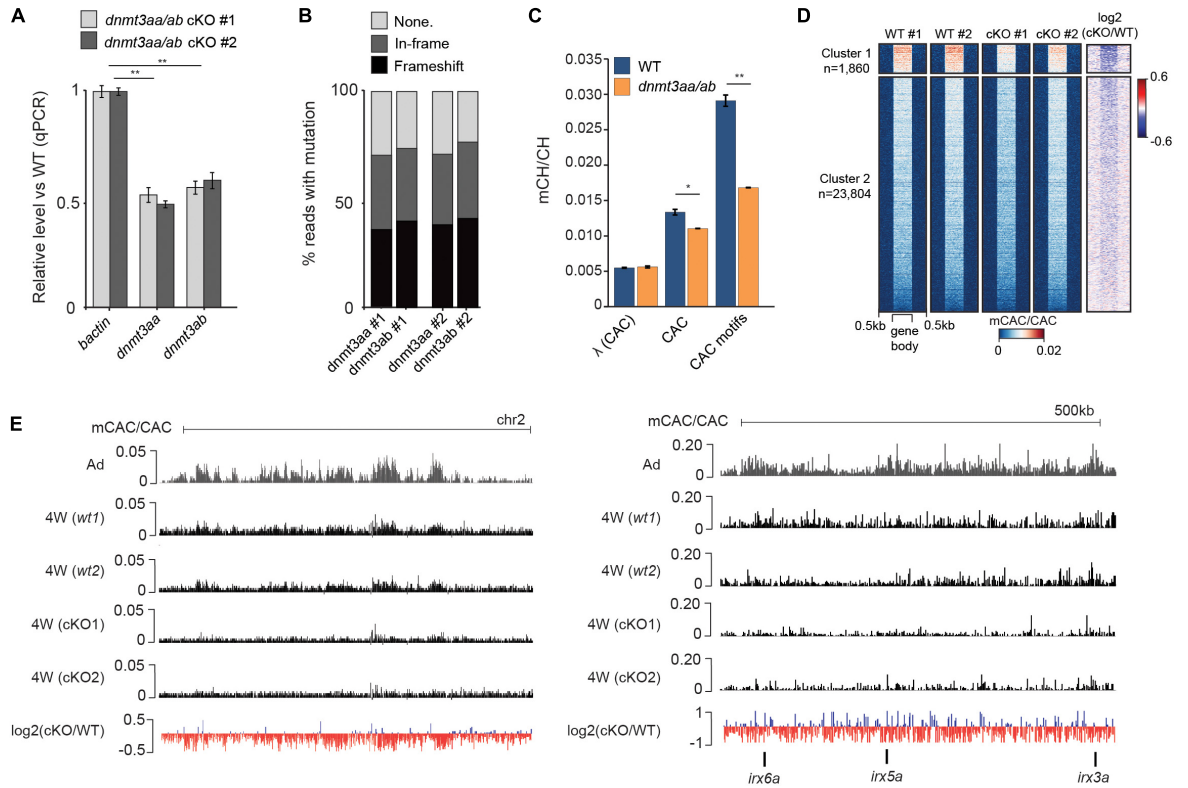


FIGURE 4 | Dnmt3a enzymes are required for methylation of CAC trinucleotides in the zebrafish brain. **(A)** Transcript levels of *dnmt3aa* and *dnmt3ab* in 4-week-old *dnmt3aa/ab* CRISPR/Cas9 KO (cKO) zebrafish brains relative to wild type (WT). The data is represented as the mean of technical replicates with error bars showing the standard error (two sample *t*-test, $^{**}P < 0.01$). **(B)** Percentage of reads with no mutation, in-frame mutations, or frameshift mutations, which map to *dnmt3aa* and *dnmt3ab* loci in *dnmt3aa/ab* cKOs. **(C)** Average mCAC/CAC levels at all CAC trinucleotides and the top methylated CAC motifs in 4-week-old WT and *dnmt3aa/ab* cKO brain. The data is represented as the average of two WGBS biological replicates (Wilcoxon test, $^{*}P < 0.05$, $^{**}P < 0.01$, λ = unmethylated lambda spike in control). **(D)** mCAC/CAC levels of all gene bodies in 4-week-old WT and *dnmt3aa/ab* cKO brains. **(E)** Genome browser snapshot of mCAC/CAC levels in adult brains, 4-week-old WT brains and 4-week-old *dnmt3aa/ab* cKO brains.

the evolutionary conservation of mCH, the mCH "writer"—DNM3TA, and the mCH "reader"—MeCP2 in vertebrates suggests that these regulatory pathways could have an ancestral role in vertebrate neurobiology (de Mendoza et al., 2021). In the current manuscript, we describe the evolutionary conservation of developmental mCH dynamics in the zebrafish nervous system. In zebrafish, like in mammals, mCH is enriched at CAC trinucleotides in gene bodies where it accumulates during brain development. Also, in line with observations in mammals, mCH depletion is evident at H3K4me3- and H3K27ac-marked regulatory regions. Similarly to our recent work on TGCT methylation of mosaic satellite repeats in zebrafish (Ross et al., 2020), and previous reports of mCH enrichment at repetitive elements in mammals (Ziller et al., 2011; Arand et al., 2012; Guo et al., 2014b), we find high levels of mCH at defined motifs associated with Tc1-like transposons (~6%). This recurring observation of mCH enrichment at repeats in multiple species supports a possible role for mCH, or DNMT3A recruitment, in the regulation of repetitive elements. In mammalian brains, active transposition of repeat elements has been shown to drive mosaicism in neuronal genomes (Muotri et al., 2005; Macia et al., 2017; Bodea et al., 2018), while MeCP2 was described as a repressor of LINE-1 elements in mouse neurons (Yu et al., 2001; Muotri et al., 2010). These data thus suggest that mCH could play an important role in regulating repetitive elements in the vertebrate brain, particularly at CG-sparse regions or active repeats such as Tc1-like transposons. These observations are also reminiscent of mCH targeting by DNMT3-related plant enzymes (Law and Jacobsen, 2010) and the findings that mCH plays vital roles in transposon silencing in plants (Domb et al., 2020).

Finally, in the current study, we generate transient CRISRPR/Cas9 KOs for *dnmt3aa*/*dnmt3ab* and demonstrate a conserved role for these enzymes in the deposition of mCH, and mCAC in particular, in the zebrafish brain. These partial KOs only have an obvious effect on mCH but not mCG levels, suggestive of a direct conservation for mCH in the vertebrate nervous system. Overall, this work provides novel insight into the evolutionary conservation of vertebrate mCH patterning and highlights the utility of the zebrafish model system, which is amenable to CRISPR/Cas9 screens, drug screens and developmental imaging, for the studies of mCH and brain development *in vivo*.

DATA AVAILABILITY STATEMENT

Data generated for this submission have been uploaded to ArrayExpress https://www.ebi.ac.uk/arrayexpress/ under the accession number E-MTAB-9924.

ETHICS STATEMENT

The animal study was reviewed and approved by the Garvan Institute of Medical Research Animal Ethics Committee (AEC approval 20/09).

AUTHOR CONTRIBUTIONS

OB conceived the study. SR performed bioinformatic analyses and CRISPR/Cas9 experiments. DH extracted brain samples. SR and OB wrote the manuscript. All authors contributed to read, and approved the final manuscript.

FUNDING

Australian Research Council (ARC) Discovery Project (DP190103852) to OB supported this work. OB was supported by NHMRC (R.D. Wright Biomedical CDF APP1162993) and CINSW (Career Development Fellowship CDF181229).

ACKNOWLEDGMENTS

Images of zebrafish brain, liver, and neurons were created with Biorender. We thank Alex de Mendoza for critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.643603/full#supplementary-material

REFERENCES

Aramillo Irizar, P., Schäuble, S., Esser, D., Groth, M., Frahm, C., Priebe, S., et al. (2018). Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat. Commun.* 9:327. doi: 10.1038/s41467-017-02395-2

Arand, J., Spieler, D., Karius, T., Branco, M. R., Meilinger, D., Meissner, A., et al. (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* 8:e1002750. doi: 10.1371/journal.pgen.1002750

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102

Bodea, G. O., McKelvey, E. G. Z., and Faulkner, G. J. (2018). Retrotransposon-induced mosaicism in the neural genome. *Open Biol.* 8:180074. doi: 10.1098/rsob.180074

Bogdanovic, O., Smits, A. H., de la Calle Mustienes, E., Tena, J. J., Ford, E., Williams, R., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* 48, 417–426. doi: 10.1038/ng.3522

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Boxer, L. D., Renthal, W., Greben, A. W., Whitwam, T., Silberfeld, A., Stroud, H., et al. (2020). MeCP2 Represses the Rate of Transcriptional Initiation of Highly Methylated Long Genes. *Mol. Cell* 77, 294.e–309.e. doi: 10.1016/j.molcel.2019.10.032

Chen, H., Smith, A. D., and Chen, T. (2016). WALT: fast and accurate read mapping for bisulfite sequencing. *Bioinformatics* 32, 3507–3509. doi: 10.1093/bioinformatics/btw490

Chen, L., Chen, K., Lavery, L. A., Baker, S. A., Shaw, C. A., Li, W., et al. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5509–5514. doi: 10.1073/pnas.1505909112

Clemens, A. W., Wu, D. Y., Moore, J. R., Christian, D. L., Zhao, G., and Gabel, H. W. (2020). MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. *Mol. Cell* 77, 279.e–293.e. doi: 10.1016/j.molcel.2019.10.033

de Mendoza, A., Lister, R., and Bogdanovic, O. (2019). Evolution of DNA Methylome Diversity in Eukaryotes. *J. Mol. Biol.* 2019:003. doi: 10.1016/j.jmb.2019.11.003

de Mendoza, A., Poppe, D., Buckberry, S., Pflueger, J., Albertin, C., Daish, T., et al. (2021). The emergence of neural non-CpG methylation system in vertebrates. *Nat. Ecol. Evol.* 2021, 1371–1372. doi: 10.1038/s41559-020-01371-2

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Domb, K., Katz, A., Harris, K., Yaari, R., Kaisler, E., Nguyen, V., et al. (2020). DNA methylation mutants in *Physcomitrella patens* elucidate individual roles of CG and non-CG methylation in genome regulation. *PNAS.* 117, 33700–33710. doi: 10.1073/pnas.2011361117

Feng, H., Conneely, K., and Wu, H. (2014). A bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* 2014:154. doi: 10.1093/nar/gku154

Gabel, H. W., Kinde, B., Stroud, H., Gilbert, C. S., Harmin, D. A., Kastan, N. R., et al. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* 522, 89–93. doi: 10.1038/nature14319

Gao, H., Bu, Y., Wu, Q., Wang, X., Chang, N., Lei, L., et al. (2015). Mecp2 regulates neural cell differentiation by suppressing the Id1 to Her2 axis in zebrafish. *J. Cell Sci.* 128, 2340–2350. doi: 10.1242/jcs.167874

Goll, M. G., and Halpern, M. E. (2011). DNA methylation in zebrafish. *Prog. Mol. Biol. Transl. Sci.* 101, 193–218. doi: 10.1016/B978-0-12-387685-0.00005-6

Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., et al. (2014a). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* 17, 215–222. doi: 10.1038/nn.3607

Guo, W., Chung, W.-Y., Qian, M., Pellegrini, M., and Zhang, M. Q. (2014b). Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic Acids Res.* 42, 3009–3016. doi: 10.1093/nar/gkt1306

Guo, W., Zhu, P., Pellegrini, M., Zhang, M. Q., Wang, X., and Ni, Z. (2018). CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* 34, 381–387. doi: 10.1093/bioinformatics/btx595

He, Y., and Ecker, J. R. (2015). Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* 16, 55–77. doi: 10.1146/annurev-genom-090413-025437

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mole. Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi: 10.1038/nrg3230

Kaaij, L. J. T., Mokry, M., Zhou, M., Musheev, M., Geeven, G., Melquiond, A. S. J., et al. (2016). Enhancers reside in a unique epigenetic environment during early zebrafish development. *Genome Biol.* 17:146. doi: 10.1186/s13059-016-1013-1

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006. doi: 10.1101/gr.229102

Lam, S. H., Chua, H. L., Gong, Z., Lam, T. J., and Sin, Y. M. (2004). Development and maturation of the immune system in zebrafish, Danio rerio: a gene expression profiling, in situ hybridization and immunological study. *Dev. Comp. Immunol.* 28, 9–28. doi: 10.1016/s0145-305x(03)00103-4

Lange, C., Rost, F., Machate, A., Reinhardt, S., Lesche, M., Weber, A., et al. (2020). Single cell sequencing of radial glia progeny reveals the diversity of newborn neurons in the adult zebrafish brain. *Development* 147:185595. doi: 10.1242/dev.185595

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lavery, L. A., Ure, K., Wan, Y.-W., Luo, C., Trostle, A. J., Wang, W., et al. (2020). Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome. *Elife* 9:52981. doi: 10.7554/eLife.52981

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* 341:1237905. doi: 10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. doi: 10.1038/nature08514

Macia, A., Widmann, T. J., Heras, S. R., Ayllon, V., Sanchez, L., Benkaddour-Boumzaouad, M., et al. (2017). Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* 27, 335–348.

Maeyama, K., and Nakayasu, H. (2000). Postembryonic Neurogenesis in Zebrafish (Danio rerio) Brain: Presence of Two Different Systems. J*zoo* 17, 959–966. doi: 10.2108/zsj.17.959

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877. doi: 10.1093/nar/gki901

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., Fernandez, J. P., Mis, E. K., Khokha, M. K., et al. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* 12, 982–988.

Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910. doi: 10.1038/nature03663

Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446. doi: 10.1038/nature09544

Nozawa, K., Lin, Y., Kubodera, R., Shimizu, Y., Tanaka, H., and Ohshima, T. (2017). Zebrafish Mecp2 is required for proper axonal elongation of motor neurons and synapse formation. *Dev. Neurobiol.* 77, 1101–1113. doi: 10.1002/dneu.22498

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R., et al. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.* 19:33. doi: 10.1186/s13059-018-1408-2

Pietri, T., Roman, A.-C., Guyon, N., Romano, S. A., Washbourne, P., Moens, C. B., et al. (2013). The first mecp2-null zebrafish model shows altered motor behaviors. *Front. Neural Circuits* 7:118. doi: 10.3389/fncir.2013.00118

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., et al. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* 34, 695–697. doi: 10.1038/nbt.3583

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Raman, A. T., Pohodich, A. E., Wan, Y.-W., Yalamanchili, H. K., Lowry, W. E., Zoghbi, H. Y., et al. (2018). Apparent bias toward long gene misregulation in MeCP2 syndromes disappears after controlling for baseline variations. *Nat. Commun.* 9:3225. doi: 10.1038/s41467-018-05627-1

Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. doi: 10.1093/nar/gku365

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Ross, S. E., Angeloni, A., Geng, F.-S., de Mendoza, A., and Bogdanovic, O. (2020). Developmental remodelling of non-CG methylation at satellite DNA repeats. *Nucleic Acids Res.* 2020, 1135. doi: 10.1093/nar/gkaa1135

Schmidt, R., Strähle, U., and Scholpp, S. (2013). Neurogenesis in zebrafish – from embryo to adult. *Neural. Dev.*. 2013:3. doi: 10.1186/1749-8104-8-3

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326. doi: 10.1038/nature14192

Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523, 212–216. doi: 10.1038/nature14465

Skvortsova, K., Tarbashevich, K., Stehling, M., Lister, R., Irimia, M., Raz, E., et al. (2019). Retention of paternal DNA methylome in the developing zebrafish germline. *Nat. Commun.* 10, 10895–10896. doi: 10.1038/s41467-019-10895-6

Storer, J., Hubley, R., Rosen, J., et al. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* 2021:00230. doi: 10.1186/s13100-020-00230-y

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034. doi: 10.1093/bioinformatics/btv098

van der Vaart, M., Svoboda, O., Weijts, B. G., Espín-Palazón, R., Sapp, V., Pietri, T., et al. (2017). Mecp2 regulates tnfa during zebrafish embryonic development and acute inflammation. *Dis. Model. Mech.* 10, 1439–1451. doi: 10.1242/dmm.026922

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647. doi: 10.1093/bioinformatics/btx469

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165

Yang, H., Luan, Y., Liu, T., Lee, H., Fang, L., Wang, Y., et al. (2020). A map of *cis*-regulatory elements and 3D genome structures in zebrafish. *Nature* 588, 337–343. doi: 10.1038/s41586-020-2962-9

Yu, F., Zingler, N., Schumann, G., and Strätling, W. H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* 29, 4493–4501. doi: 10.1093/nar/29.21.4493

Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. doi: 10.1186/gb-2008-9-9-r137

Ziller, M. J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., et al. (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* 7:e1002389. doi: 10.1371/journal.pgen.1002389

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 25 March 2021
doi: 10.3389/fcell.2021.653669

Super-Enhancers and CTCF in Early Embryonic Cell Fate Decisions

Puja Agrawal[1,2] *and Sridhar Rao*[1,2,3]*

[1] Department of Cell Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, WI, United States, [2] Versiti Blood Research Institute, Milwaukee, WI, United States, [3] Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, United States

Cell fate decisions are the backbone of many developmental and disease processes. In early mammalian development, precise gene expression changes underly the rapid division of a single cell that leads to the embryo and are critically dependent on autonomous cell changes in gene expression. To understand how these lineage specifications events are mediated, scientists have had to look past protein coding genes to the *cis* regulatory elements (CREs), including enhancers and insulators, that modulate gene expression. One class of enhancers, termed super-enhancers, is highly active and cell-type specific, implying their critical role in modulating cell-type specific gene expression. Deletion or mutations within these CREs adversely affect gene expression and development and can cause disease. In this mini-review we discuss recent studies describing the potential roles of two CREs, enhancers and binding sites for CTCF, in early mammalian development.

Keywords: CTCF, enhancer, nanog, pluripotency, embryonic stem cell

OPEN ACCESS

Edited by:
Justin Brumbaugh, University of Colorado Boulder, United States

Reviewed by:
Konstantinos Chronis, University of Illinois at Chicago, United States
Aaron Huebner, Massachusetts General Hospital, United States

***Correspondence:**
Sridhar Rao
sridhar.rao@versiti.org

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *14 January 2021*
Accepted: *18 February 2021*
Published: *25 March 2021*

Citation:
Agrawal P and Rao S (2021) Super-Enhancers and CTCF in Early Embryonic Cell Fate Decisions. Front. Cell Dev. Biol. 9:653669. doi: 10.3389/fcell.2021.653669

INTRODUCTION

Understanding the molecular underpinnings of cell fate decisions is a central question within developmental biology. For decades it has been assumed that these decisions are mediated almost exclusively by the 2% of DNA that contains protein-coding genes, while the remaining 98% was considered "junk." Over time, scientists have realized that this 2% of DNA could not exclusively explain all the developmental and disease-related decisions made during early mammalian development. This has given rise to a renewed focus on the remaining 98% of the genome, including non-coding DNA sequences and nuclear architecture and their role(s) in driving cell fate decisions by regulating gene expression. Nuclear architecture is the dynamic, three-dimensional organization of chromatin and its interaction with itself and regulatory proteins within the nucleus. The development of high-resolution imaging and chromosome conformation capture technologies, which permit a direct interrogation of long-range DNA:DNA interactions, has facilitated these important aspects, which regulate gene expression. Two types of *cis*-regulatory elements (CREs), which play a critical role in modulating gene expression, are enhancers and DNA-binding sites for CTCFs. In this mini-review we will focus on both classes of CREs specifically, as well as their mechanisms and roles in driving early cell fate decisions during early mammalian development. The examples discussed below do not encompass all the literature available by any measure but reflect selected key studies in the field focusing on embryonic stem cells (ESCs).

For further information on enhancer and CTCF functions, we suggest these reviews (Ong and Corces, 2014; Shlyueva et al., 2014; Li et al., 2016; Chen et al., 2018; Rowley and Corces, 2018; Agrawal et al., 2019; Braccioli and De Wit, 2019).

EARLY MAMMALIAN DEVELOPMENT

Early placental mammalian development is characterized by numerous rapid, distinct cell fate decisions (Arnold and Robertson, 2009). Upon fertilization, an embryo is totipotent: it can form any cell type, both embryonic and extraembryonic tissues. After a few cell divisions the cells become pluripotent: they can form the three germ layers (mesoderm, endoderm, and ectoderm) but not extraembryonic tissue. At the blastocyst stage two distinct populations are formed: trophectoderm and the pluripotent inner cell mass (ICM). The trophectoderm, which eventually gives rise to the placenta, expresses *CDX2*, which turns off the pluripotency gene *Pou5f1,* which encodes the pluripotency critical transcription factor Oct3/4 (Strumpf et al., 2005; Rayon et al., 2016). The ICM maintains pluripotency by continuing expression of Oct3/4, which subsequently differentiates into the epiblast (EPI) and the primitive endoderm (PrE). The EPI eventually gives rise to all fetal tissues and maintains pluripotency by expressing *Nanog*, while the PrE, which will form the yolk sac, differentiates by down-regulating *Nanog* and expressing *Gata6* (Messerschmidt and Kemler, 2010; Schrode et al., 2014). Development of the embryo is thus dependent on the proper temporal and spatial gene expression changes to drive lineage commitment. Experimentally, embryonic stem cells (ESCs) are derived from the early embryo and exist in multiple states, naïve (ICM) and primed (epiblast; Ghimire et al., 2018). These states are distinct with respect to transcription, where primed cells show a relative down-regulation of pluripotency genes and an up-regulation of lineage-specific genes, compared with the naïve embryos, with additional differences, indicating that pluripotency is not a single-cell state.

ENHANCERS

Enhancers are CREs that contain DNA-binding motifs for transcription factors (TFs) distal to the gene promoter, which are located immediately adjacent to the transcriptional start site (TSS). Binding of these TFs drives enhancers to promote or repress gene expression through multiple mechanisms that are beyond the scope of this review (Li et al., 2016; Chen et al., 2018; Agrawal et al., 2019). Next-generation sequencing-based (NGS) methods have revolutionized the genome-wide identification of enhancers. The combination of chromatin immunoprecipitation (ChIP) with NGS (ChIP-Seq) initially permitted the identification of potential enhancer regions based on p300 and H3K4me1 enrichment (Visel et al., 2009; Rada-Iglesias et al., 2011). These regions can be further subdivided based upon the enrichment of the mutually exclusive marks H3K27me3 and H3K27Ac, indicating whether an enhancer is "poised" or "active," respectively. Genome-wide study of histone and TF enrichment has led to the identification of a distinct class of enhancers termed super-enhancers. Super-enhancers (SEs, Whyte et al., 2013) are a highly active minority of enhancers defined by their high density of binding lineage-specific TFs and associated cofactors such as the Mediator complex and p300. SEs also exhibit relatively higher levels of H3K27Ac and are larger in size overall compared with enhancers (Parker et al., 2013; Pulakanti et al., 2013). Several studies show that SE differ from classical enhancers due to their stronger ability to drive gene expression than classical enhancers (Hnisz et al., 2015; Huang et al., 2016; Shin et al., 2016; Thomas et al., 2021). SEs are also highly cell-type specific and are often found near key lineage determining genes, implying they are critical to the establishment and/or maintenance of cell identity. For example, in ESCs there are 231 super-enhancers (Whyte et al., 2013), which are defined by their enrichment for the pluripotency TFs Nanog, Oct4, and Sox2, as well as Mediator and high levels of H3K27Ac, and many are in close proximity to pluripotency-promoting genes, including *Nanog.*

The master pluripotency TF *Nanog* is near three SEs (−5SE, −45SE, +60SE, named based on distance from *Nanog* promoter in kb; **Figure 1A**; Pulakanti et al., 2013; Whyte et al., 2013; Blinka et al., 2016; Agrawal et al., 2021). These SEs and the numerous CTCF sites at the locus make it an ideal model system to understand how CREs regulate cell fate decisions by modulating *Nanog* expression (Blinka and Rao, 2017). Mouse-derived embryonic stem cells (mESCs) are an ideal model for studying enhancer-promoter interactions, which promote *Nanog* expression. Due to their pluripotent nature, they are sensitive to changes to gene expression with a simple observation of cell fate, differentiation. Tracking *eGFP*-tagged *Nanog* and isolating GFP^- cells revealed that cells not expressing *Nanog* could give rise to *Nanog* expressing cells and continue to self-renew (Chambers et al., 2007). These cells do, however, give rise to fewer undifferentiated colonies, indicating that down-regulating *Nanog* predisposes but does not proscribe cells to differentiation. At the single-cell level, *Nanog*-low cells begin to down-regulate pluripotency-associated genes and up-regulate lineage-specific gene markers, while maintaining pluripotency (Abranches et al., 2014). Blastocyst division into the EPI and PrE is dependent on lineage specific gene expression and the location of the cells (Xenopoulos et al., 2015). One of the first markers of this division of cell fates is *Nanog,* wherein the presence of *Nanog* denotes EPI cell fate, while the loss of *Nanog* marks the development of PrE. As the ICM differentiates, there are a few PrE to EPI conversions, but no EPI to PrE conversions, indicating that the EPI state is a distinct form of pluripotency from the ICM. In culture, *Nanog* is expressed highly in naïve pluripotent ESCs but fluctuates in primed ESCs (Barral et al., 2019). Thus, the modulation of *Nanog* is important to understand, as the delicate regulation of the gene's expression is critical to defining pluripotent states. Circular Chromosome Capture (3C) of the three SEs at the *Nanog* locus demonstrates that they physically interact with the *Nanog* promoter, but CRISPR-based deletion of each SE differentially alters expression (Blinka et al., 2016). Deletion of the −45SE causes an approximately 50% decrease in *Nanog* expression, deletion of the −5SE causes a nearly 90%

decrease, but deletion of the +60SE causes no change (**Figure 1B**; Blinka et al., 2016; Agrawal et al., 2021). This leads to the important question: through what mechanism(s) do the −5SE and −45SE regulate *Nanog,* expression? Interestingly, single-cell RT-qPCR showed that deleting the −5SE altered *Nanog* expression in all cells demonstrates that it is functional in all cells. Additionally, analysis of RNA Polymerase II (RNAPII) dynamics upon deletion of the −5SE shows a complete loss of RNAPII at the *Nanog* promoter, indicating that the −5SE modulates *Nanog* expression by recruiting RNAPII or by promoting its conversion to its initiating form through phosphorylation of Ser 5 of the C-terminal domain (CTD, **Figure 1C**). Further analysis of the −45SE and its role in regulating *Nanog* gene expression is necessary to determine whether the two enhancers utilize the same or distinct mechanism(s) from the −5SE (**Figure 1D**; Henriques et al., 2018; Chen et al., 2018; Bartman et al., 2019; Sheridan et al., 2019; Zatreanu et al., 2019; Noe Gonzalez et al., 2020).

Other lineage-critical genes are also regulated by enhancers. The expression of *Shh,* a gene critical at multiple stages of development in many tissues, is critically regulated by enhancers (Anderson and Hill, 2014). In the zone of polarizing activity (ZPA) in the developing limb, mutations in the ZPA regulatory sequence (ZRS) cause polydactyly due to misregulation of *Shh* in early limb development (Lettice et al., 2003). The ZRS is interesting as it is one megabase away from *Shh,* displaying just how distant the regulation of some genes can be. The interaction of the ZRS and *Shh* is specific to E10.5–11.5 mouse embryos, limiting the ZRS activity to limb development stages (Williamson et al., 2016). Interestingly, the ZRS, a highly conserved regulatory region in vertebrates, consists of multiple discrete enhancer elements that regulate gene expression and long-range interactions, emphasizing the multiple and varied roles of enhancers in regulating development (Lettice et al., 2017). In the brain, numerous Shh-Brain-Enhancers (SBEs) modulate the spatiotemporal expression of *Shh* in the developing midbrain (Jeong et al., 2006; Yao et al., 2016). Curiously, deletion of the SBEs can alter *Shh* expression in neural progenitor cells and brain development (Jeong et al., 2008; Benabdallah et al., 2016), but 3D-FISH analysis of *Shh* and SBEs shows an increase in distance between the gene promoter and regulatory element, implying a mechanism independent of looping (Benabdallah et al., 2019).

Sox2, another pluripotency gene, has multiple enhancer regions, some of which have no effect on pluripotency when deleted but have a potential role in neural cells (Ferri et al., 2004). Other regions, specifically the *Sox2* Control Regions (SCR) disrupt maintenance of pluripotency in ESCs when deleted (Zhou et al., 2014). Similarly, to *Shh*, although the SCR is critical to *Sox2* transcription, recent live cell imaging shows that in ESCs, the SCR and *Sox2* are not in close physical proximity within the nucleus, indicating they operate at a distance (Alexander et al., 2019).

As an example of the interface between development and disease, enhancers play a critical role in the neural-crest-derived craniofacial disorder, the Pierre Robin sequence (PRS), which is classically characterized by mandibular underdevelopment along with other characteristics (Long et al., 2020). PRS is linked to mutations in a Topologically Associated Domain (TAD, described below), containing a single protein coding gene, *Sox9* (Gordon et al., 2014). *Sox9* plays a critical role during neural crest differentiation and craniofacial development (Lee and Saint-Jeannet, 2011; Dash and Trainor, 2020; Nagakura et al., 2020) and enhancers near *Sox9* have been implicated in craniofacial and chondrocyte development (Liu and Lefebvre, 2015; Yao et al., 2015). PRS associated mutations are >1Mb from *Sox9* and near a cluster of three enhancers the deletion of which leads to a 50% allele-specific decrease in *Sox9* expression (Long et al., 2020). These three clusters are enriched for activating chromatin marks exclusively in the developing human neural crest cells (hNCC), indicating a cell-type specific role for these enhancers. In a mouse model, deletion of one of these enhancer clusters causes a modest 13% decrease in *Sox9* expression that is sufficient to cause mandibular developmental changes and adversely effect survival, indicating that a dose-dependent change in *Sox9* expression due to enhancer cluster deletion alters mandibular development. Enhancer control of the *Shh, Sox2* and *Sox9* loci demonstrate how CRE modulation of gene expression allows one gene to play multiple roles at various points in development.

Although all the nuances of how enhancers determine cell fate are still being investigated, we do understand that their role is critical to a range of developmental processes. Two mechanistic insights not discussed here but also under investigation are whether and how enhancers modulate gene expression by regulating transcriptional bursting (Bartman et al., 2016, 2019; Fukaya et al., 2016) and/or phase separation (Cho et al., 2018; Sabari et al., 2018; Zamudio et al., 2019). Analysis of *Shh* and *Sox2* enhancers has shown that physical proximity may not be necessary, further expanding the question of the mechanisms that enhancers use to modulate gene expression.

CTCF

DNA is organized within the nucleus on multiple levels (Ong and Corces, 2014; Rowley and Corces, 2018). Double-helix DNA is packaged into chromatin, which is further folded into nucleosomes that can be regulated by histone modifications. The development of 3C-based techniques has allowed the study of chromatin-chromatin interactions across large genomic distances (Dekker et al., 2002; de Laat and Dekker, 2012). Hi-C, a technique that studies long range chromatin-chromatin interactions, has led to the definition of Topologically Associated Domains (**Figure 2A**, TADs). TADs are large regions of chromatin, bounded by CCCCTC binding factor (CTCF) binding sites, in which the majority of interactions remain within the TAD. CTCF is a DNA-binding insulator protein and has binding sites present throughout the genome, within and at the boundaries of TADs. TADs can be further broken down to sub-TADs which are cell-type specific insulated regions within a TAD bounded by CTCF and cohesin (**Figure 2A**, Phillips-Cremins et al., 2013; Dowen et al., 2014; Hnisz et al., 2016). These sub-TADs are found to contain tissue or development-specific genes. Insulated neighborhoods (INs) are among the

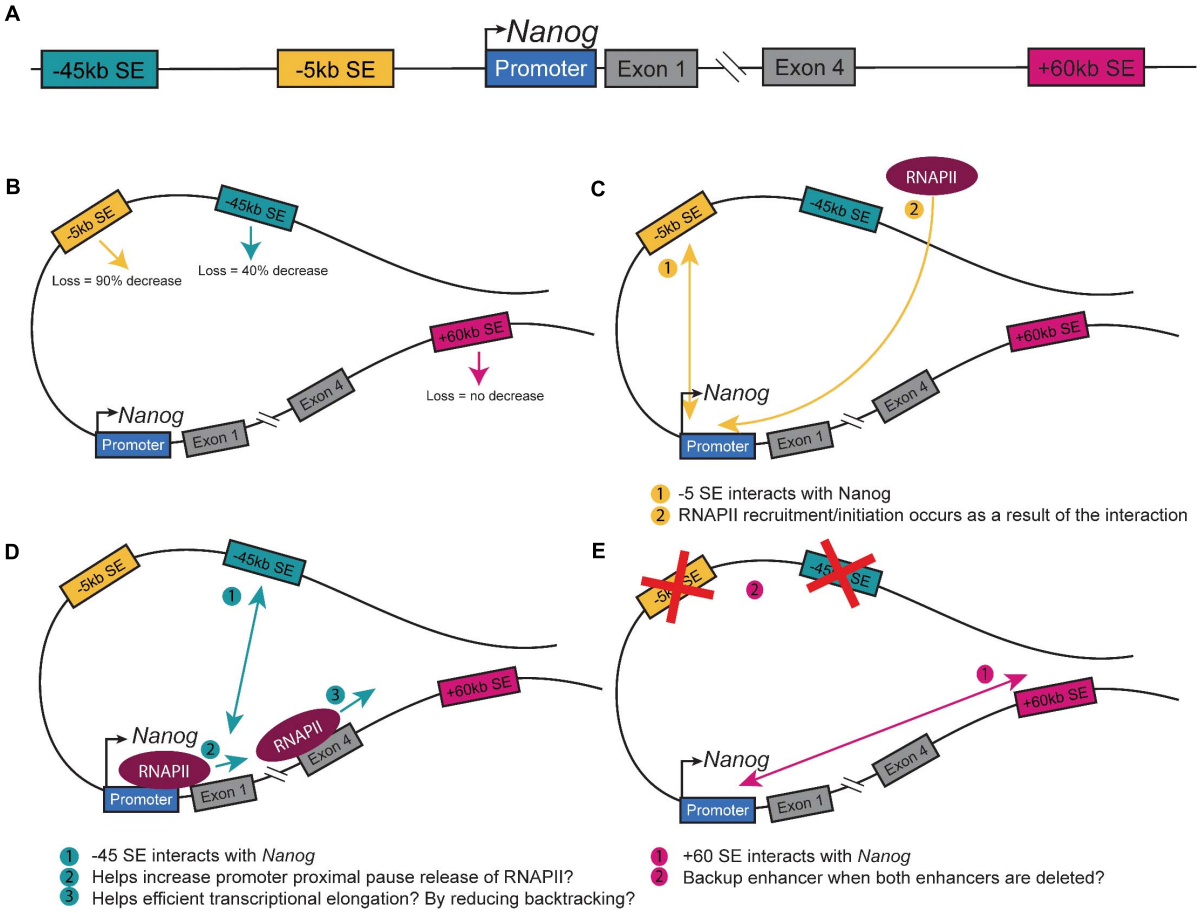


FIGURE 1 | Nanog Super-enhancers. **(A)** Schematic of the *Nanog* extended locus with the three SEs. Schematic of a theoretical loop containing all three SEs and Nanog depicting. **(B)** The changes in *Nanog* expression upon deletion (Blinka et al., 2016; Agrawal et al., 2021). **(C)** The interaction between the −5SE and *Nanog* and how it alters RNAPII dynamics (Agrawal et al., 2021). **(D)** The interaction between the −45SE and *Nanog* and its potential mechanisms. **(E)** The interaction between the +60SE and *Nanog* and the possible backup role it has in regulating the gene.

types of sub-TAD specifically identified in ESCs described as a CTCF-CTCF loop containing an SE and its target gene (Dowen et al., 2014). Deletion of one boundary CTCF site at key pluripotency genes (i.e., *Nanog, Pouf51*) is sufficient to alter gene expression, indicating these INs facilitate proper gene expression regulation, perhaps by creating and/or preventing enhancer-promoter interactions. A key point to consider is that INs and TADs were identified using different techniques and therefore may be identifying the same phenomenon but at different scales due to methodological differences.

The insulator protein CTCF contains 11 highly-conserved zinc-finger DNA binding domains (Phillips and Corces, 2009; Ong and Corces, 2014; Rowley and Corces, 2018). Insulator proteins have numerous functions, including regulating intrachromosomal interactions and the spread of heterochromatin. Knockout of CTCF is embryonic lethal at E5.5, and depletion of maternal CTCF in oocytes adversely impacts blastocyst stage embryos (Moore et al., 2012). CTCF has pleiotropic effects on gene expression, including altering DNA methylation and affecting splicing, as well as the regulation of enhancer-promoter interactions through chromatin looping (Ong and Corces, 2014). Current dogma supports an extrusion model of chromatin looping between two CTCF-bound sites that are in convergent orientation in collaboration with the cohesin complex (Rao et al., 2014; de Wit et al., 2015). The cohesin complex is a ring structure made up of four protein components, Rad21, Smc1, Smc3, and Stag1/2 (Fisher et al., 2017). While its primary function is to stabilize sister chromatids during cell division, cohesin also plays a critical role in maintaining chromatin loops, often in conjunction with CTCF. But 50–80% of CTCF binding sites are co-localized with cohesin. Although depletion of cohesin does not alter CTCF occupancy, there is a loss of looped domains, indicating that CTCF-dependent looping relies on cohesin (Rao et al., 2017). Further analysis of the CTCF protein revealed that mutating the N-terminus reduces TAD insulation (Nora et al., 2019). The orientation of the CTCF binding site positions the N-terminus to interact with cohesin, providing a molecular explanation of the necessity of convergent CTCF sites to create a chromatin loop. The necessity of cohesin and CTCF binding for chromatin looping, however,

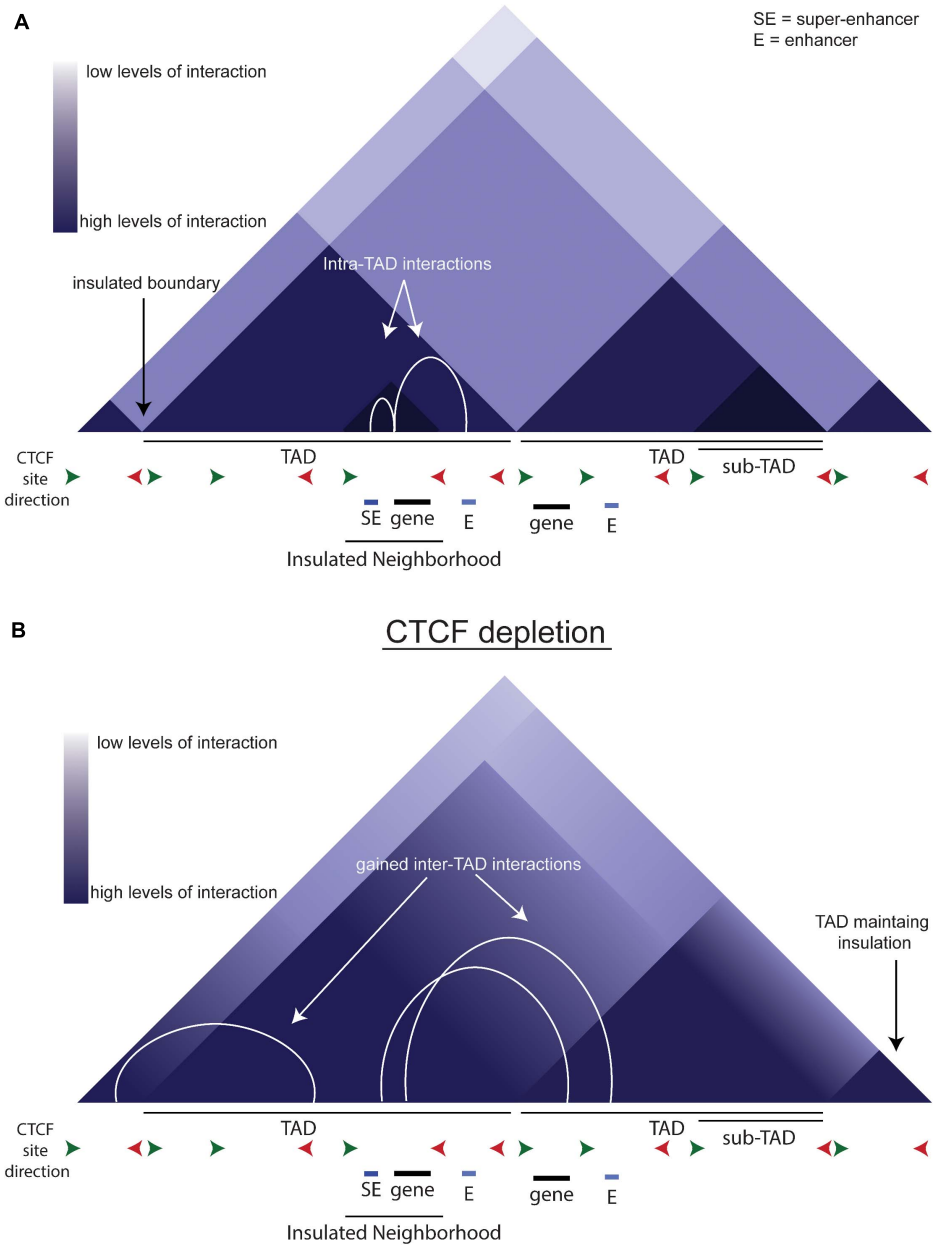


FIGURE 2 | TADs and CTCF. **(A)** Schematic of chromatin organization. The triangles depict a schematic of a Hi-C contact map. White arcs show the intra-TAD interactions that are insulated by CTCF. **(B)** Schematic of what a Hi-C contact map may look like upon CTCF depletion. White arcs show the gained interactions across previous TAD boundaries.

is currently being questioned, as two recent students have shown that depletion of CTCF and cohesin does not affect a majority of enhancer:promoter interactions (Thiecke et al., 2020; Kubo et al., 2021).

Many studies have shown that CTCF plays a role during development (reviewed in Arzate-Mejía et al., 2018). At the *HoxA* locus, deletion of CTCF sites within the locus alters the enrichment of the repressive H3K27me3 mark, permitting aberrant *HoxA* gene expression during the differentiation of ESCs to motor neurons (Narendra et al., 2015). In the developing heart, deletion of the CTCF gene in cardiac progenitor cells causes embryonic lethality at E12.5, through loss of chromatin interactions and, potentially, enhancer:promoter interactions at key developmental genes (Gomez-Velazquez et al., 2017). As a distinct example, inhibiting CTCF function prevents Schwann cell differentiation and causes hypomyelination *in vivo* (Wang et al., 2020). These are a few examples of the variety of roles CTCF plays in gene expression regulation, but little is known on how to identify what role CTCF is playing modulating nearby genes and

whether there are biological signs that differentiate one CTCF site's function from another.

In an auxin-inducible depletion system in mESCs, loss of CTCF reduces proliferation, indicating that CTCF is integral to the maintenance of pluripotency (Nora et al., 2017). The acute depletion of CTCF causes an overall loss of looping anchored by CTCF and cohesin (**Figure 2B**). There is also an increase in inter-TAD interactions that is reversed upon auxin wash-off, implying that CTCF normally prevents these interactions. Interestingly, while CTCF depletion causes loss of insulation on most boundaries, some boundaries remain intact (**Figure 2B**), indicating that the role of CTCF is not universal across the genome. CTCF sites are more likely to be near genes up-regulated by CTCF depletion, and these genes are more likely to be near active enhancers, so they are normally separated by a TAD boundary. Thus, the loss of the TAD boundary likely permits aberrant enhancer-promoter interaction(s) causing the up-regulation of these genes. In conclusion, CTCF is critical to the precise regulation of gene expression by modulating enhancer:promoter contacts.

SUPER-ENHANCERS AND CTCF

Several studies have focused on the interaction between CTCF and enhancers, specifically SEs. CTCF has been found to be bound near most SEs and within some SEs (Dowen et al., 2014; Ing-Simmons et al., 2015; Huang et al., 2017). One study finds that a mammary gland specific SE modulates not only the mammary gland specific gene *Wap*, but also the non-mammary-specific *Ramp3*. *Ramp3* lies outside of the chromatin loop formed by two CTCF sites flanking *Wap* and the associated SE. Deletion of the separating CTCF sites greatly increases the interaction between the SE and *Ramp3* and *Ramp3* expression, thus showing that the CTCF sites insulate against this interaction to maintain proper expression (Willi et al., 2017). In a variety of cancers, SEs are gained near *MYC* to up-regulate *MYC* and drive oncogenic transformation. These SEs lay between two CTCF sites that form an IN neighborhood and modulate expression (Schuijers et al., 2018). These studies, and others (Hay et al., 2016; Hanssen et al., 2017; Shin, 2019), support the theory that CTCF is critical to SE-mediated gene expression regulation.

DISCUSSION

The full potential of CTCF sites and enhancers in regulating gene expression is still being investigated and a number of questions remain. As an example, in view of the evidence that CTCF sites modulate SE function, at the *Nanog* locus, what is CTCF's role? Is the IN truly the two sites just 5′ and 3′ of the gene, as implied in Dowen et al., and can other nearby CTCF sites compensate? And are these CTCF sites encouraging and/or preventing enhancer:promoter interactions? Beyond *Nanog*, the question still remains whether there are different classes of CTCF binding sites that can be identified through a biological signature. Given that loss of CTCF can cause an up-regulation and down-regulation of gene expression, how can we identify which role it plays for any given gene? Similarly, there are still questions regarding enhancers and their role in cell fate decisions: specifically are all super-enhancers and the same? It is clear from the data regarding the *Nanog* +60E that this cannot be true, because although it has all the defining features of a super-enhancer, in ESCs, at least, it does not play a significant role. Perhaps it is a redundant enhancer, becoming fully functional if the −5SE and −45SE are out of commission or is utilized in other cell types such as primordial germ cells (**Figure 1E**). Can we identify different features of SEs and assign them to different categories? The field of nuclear architecture remains open to investigation, and as we further study chromatin-chromatin interactions, the minute control of gene expression control will clarify how cell fate decisions are made and how this process breaks down in disease pathology.

AUTHOR CONTRIBUTIONS

PA wrote the first draft of the manuscript. SR and PA edited the manuscript together. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by funding from NIDDK (DK120152) to PA and the MCW MSTP T32 (NIGMS, GM080202) to PA. Additional support came from NIH (CA204231) to SR.

REFERENCES

Abranches, E., Guedes, A. M. V., Moravec, M., Maamar, H., Svoboda, P., Raj, A., et al. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* 141, 2770–2779. doi: 10.1242/dev.108910

Agrawal, P., Blinka, S., Pulakanti, K., Reimer, M. H., Stelloh, C., Meyer, A. E., et al. (2021). Genome editing demonstrates that the -5 kb Nanog enhancer regulates Nanog expression by modulating RNAPII initiation and/or recruitment. *JBC* 296:100189. doi: 10.1074/jbc.RA120.015152

Agrawal, P., Heimbruch, K. E., and Rao, S. (2019). Genome-wide maps of transcription regulatory elements and transcription enhancers in development and disease. *Compr. Physiol.* 9, 439–455. doi: 10.1002/cphy.c180028

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., et al. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8:e41769. doi: 10.7554/eLife.41769

Anderson, E., and Hill, R. E. (2014). Long range regulation of the sonic hedgehog gene. *Curr. Opin. Genet. Dev.* 27, 54–59. doi: 10.1016/j.gde.2014.03.011

Arnold, S. J., and Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* 10, 91–103. doi: 10.1038/nrm2618

Arzate-Mejía, R. G., Recillas-Targa, F., and Corces, V. G. (2018). Developing in 3D: the role of CTCF in cell differentiation. *Development* 145:dev137729. doi: 10.1242/dev.137729

Barral, A., Rollan, I., Sanchez-Iranzo, H., Jawaid, W., Badia-Careaga, C., Menchero, S., et al. (2019). Nanog regulates Pou3f1 expression at the exit from pluripotency during gastrulation. *Biol. Open* 8:bio046367. doi: 10.1242/bio.046367

Bartman, C. R., Hamagami, N., Keller, C. A., Giardine, B., Hardison, R. C., Blobel, G. A., et al. (2019). Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol. Cell* 73, 519–532.e4. doi: 10.1016/j.molcel.2018.11.004

Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A., and Blobel, G. A. (2016). Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* 62, 237–247. doi: 10.1016/j.molcel.2016.03.007

Benabdallah, N. S., Gautier, P., Hekimoglu-Balkan, B., Lettice, L. A., Bhatia, S., and Bickmore, W. A. (2016). SBE6: a novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells. *Open Biol.* 6:160197. doi: 10.1098/rsob.160197

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., et al. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell* 76, 473–484.e7. doi: 10.1016/j.molcel.2019.07.038

Blinka, S., and Rao, S. (2017). Nanog expression in embryonic stem cells - an ideal model system to dissect enhancer function. *BioEssays* 39:1700086. doi: 10.1002/bies.201700086

Blinka, S., Reimer, M. H., Pulakanti, K., and Rao, S. (2016). Super-Enhancers at the nanog locus differentially regulate neighboring pluripotency-associated genes. *Cell Rep.* 17, 19–28. doi: 10.1016/j.celrep.2016.09.002

Braccioli, L., and De Wit, E. (2019). CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* 63, 157–165. doi: 10.1042/EBC20180069

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., et al. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* 450, 1230–1234. doi: 10.1038/nature06403

Chen, F. X., Smith, E. R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 19, 464–478. doi: 10.1038/s41580-018-0010-5

Cho, W.-K. K., Spille, J.-H. H., Hecht, M., Lee, C., Li, C., Grube, V., et al. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 4199:eaar4199. doi: 10.1126/science.aar4199

Dash, S., and Trainor, P. A. (2020). The development, patterning and evolution of neural crest cell differentiation into cartilage and bone. *Bone* 137:115409. doi: 10.1016/j.bone.2020.115409

de Laat, W., and Dekker, J. (2012). 3C-based technologies to study the shape of the genome. *Methods* 58, 189–191. doi: 10.1016/j.ymeth.2012.11.005

de Wit, E., Vos, E. S. M., Holwerda, S. J. B., Valdes-Quezada, C., Verstegen, M. J. A. M., Teunissen, H., et al. (2015). CTCF binding polarity determines chromatin looping. *Mol. Cell* 60, 676–684. doi: 10.1016/j.molcel.2015.09.023

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. doi: 10.1016/j.cell.2014.09.030

Ferri, A. L. M., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., et al. (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* 131, 3805–3819. doi: 10.1242/dev.01204

Fisher, J. B., McNulty, M., Burke, M. J., Crispino, J. D., and Rao, S. (2017). Cohesin mutations in myeloid malignancies. *Trends Cancer* 3, 282–293. doi: 10.1016/j.trecan.2017.02.006

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer control of transcriptional bursting. *Cell* 166, 358–368. doi: 10.1016/j.cell.2016.05.025

Ghimire, S., Van Der Jeught, M., Neupane, J., Roost, M. S., Anckaert, J., Popovic, M., et al. (2018). Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Sci. Rep.* 8:5884. doi: 10.1038/s41598-018-24051-5

Gomez-Velazquez, M., Badia-Careaga, C., Lechuga-Vieco, A. V., Nieto-Arellano, R., Tena, J. J., Rollan, I., et al. (2017). CTCF counter-regulates cardiomyocyte development and maturation programs in the embryonic heart. *PLoS Genet.* 13:e1006985. doi: 10.1371/journal.pgen.1006985

Gordon, C. T., Attanasio, C., Bhatia, S., Benko, S., Ansari, M., Tan, T. Y., et al. (2014). Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in pierre robin sequence. *Hum. Mutat.* 35, 1011–1020. doi: 10.1002/humu.22606

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function *in vivo*. *Nat. Cell Biol.* 19, 952–961. doi: 10.1038/ncb3573

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P., et al. (2016). Genetic dissection of the α-globin super-enhancer *in vivo*. *Nat. Genet.* 48, 895–903. doi: 10.1038/ng.3605

Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., et al. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* 32, 26–41. doi: 10.1101/gad.309351.117.GENES

Hnisz, D., Day, D. S., and Young, R. A. (2016). Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* 167, 1188–1200. doi: 10.1016/J.CELL.2016.10.024

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., et al. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* 58, 362–370. doi: 10.1016/j.molcel.2015.02.014

Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S. H., et al. (2017). Dissecting super-enhancer hierarchy based on chromatin interactions. *bioRxiv [Preprint]* doi: 10.1101/149583

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., et al. (2016). Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev. Cell* 36, 9–23. doi: 10.1016/j.devcel.2015.12.014

Ing-Simmons, E., Seitan, V. C., Faure, A. J., Flicek, P., Carroll, T., Dekker, J., et al. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res.* 25, 504–513. doi: 10.1101/gr.184986.114

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D. J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* 133, 761–772. doi: 10.1242/dev.02239

Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* 40, 1348–1353. doi: 10.1038/ng.230

Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.* 28, 152–161. doi: 10.1038/s41594-020-00539-5

Lee, Y. H., and Saint-Jeannet, J. P. (2011). Sox9 function in craniofacial development and disease. *Genesis* 49, 200–208. doi: 10.1002/dvg.20717

Lettice, L. A., Devenney, P., De Angelis, C., and Hill, R. E. (2017). The conserved sonic hedgehog limb enhancer consists of discrete functional elements that regulate precise spatial expression. *Cell Rep.* 20, 1396–1408. doi: 10.1016/j.celrep.2017.07.037

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735. doi: 10.1093/hmg/ddg180

Li, W., Notani, D., and Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* 17, 207–223. doi: 10.1038/nrg.2016.4

Liu, C. F., and Lefebvre, V. (2015). The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res.* 43, 8183–8203. doi: 10.1093/nar/gkv688

Long, H. K., Osterwalder, M., Welsh, I. C., Hansen, K., Davies, J. O. J., Liu, Y. E., et al. (2020). Loss of extreme long-range enhancers in human neural crest drives a craniofacial disorder. *Cell Stem Cell* 27, 765–783.e14. doi: 10.1016/j.stem.2020.09.001

Messerschmidt, D. M., and Kemler, R. (2010). Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev. Biol.* 344, 129–137. doi: 10.1016/j.ydbio.2010.04.020

Moore, J. M., Rabaia, N. A., Smith, L. E., Fagerlie, S., Gurley, K., Loukinov, D., et al. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. *PLoS One* 7:e34915. doi: 10.1371/journal.pone.0034915

Nagakura, R., Yamamoto, M., Jeong, J., Hinata, N., Katori, Y., Chang, W. J., et al. (2020). Switching of Sox9 expression during musculoskeletal system development. *Sci. Rep.* 10:8425. doi: 10.1038/s41598-020-65339-9

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., et al. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017–1021. doi: 10.1126/science.1262088

Noe Gonzalez, M., Blears, D., and Svejstrup, J. Q. (2020). Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat. Rev. Mol. Cell Biol.* 22, 3–21. doi: 10.1038/s41580-020-00308-8

Nora, E., Caccianini, L., Fudenberg, G., Kameswaran, V., Nagle, A., Uebersohn, A., et al. (2019). Molecular basis of CTCF binding polarity in genome folding. *bioRxiv [Preprint]* doi: 10.1101/2019.12.13.876177

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., et al. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930–944.e22. doi: 10.1016/j.cell.2017.05.004

Ong, C.-T., and Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246. doi: 10.1038/nrg3663

Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17921–17926. doi: 10.1073/pnas.1317023110

Phillips, J. E., and Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211. doi: 10.1016/j.cell.2009.06.001

Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295. doi: 10.1016/j.cell.2013.04.053

Pulakanti, K., Pinello, L., Stelloh, C., Blinka, S., Allred, J., Milanovich, S., et al. (2013). Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics* 8, 1303–1320. doi: 10.4161/epi.26597

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–285. doi: 10.1038/nature09692

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., et al. (2017). Cohesin loss eliminates all loop domains. *Cell* 171, 305–320.e24. doi: 10.1016/j.cell.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Rayon, T., Menchero, S., Rollán, I., Ors, I., Helness, A., Crespo, M., et al. (2016). Distinct mechanisms regulate Cdx2 expression in the blastocyst and in trophoblast stem cells. *Sci. Rep.* 6:27139. doi: 10.1038/srep27139

Rowley, M. J., and Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800. doi: 10.1038/s41576-018-0060-8

Sabari, B. R., Agnese, A. D., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361:eaar3958. doi: 10.1126/science.aar3958

Schrode, N., Saiz, N., Di Talia, S., and Hadjantonakis, A. K. (2014). GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Dev. Cell* 29, 454–467. doi: 10.1016/j.devcel.2014.04.011

Schuijers, J., Manteiga, J. C., Weintraub, A. S., Day, D. S., Zamudio, A. V., Hnisz, D., et al. (2018). Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. *Cell Rep.* 23, 349–360. doi: 10.1016/j.celrep.2018.03.056

Sheridan, R. M., Fong, N., D'Alessandro, A., and Bentley, D. L. (2019). Widespread backtracking by RNA Pol II Is a major effector of gene activation, 5′ pause release, termination, and transcription elongation rate. *Mol. Cell* 73, 107–118.e4. doi: 10.1016/j.molcel.2018.10.031

Shin, H. Y. (2019). The structural and functional roles of CTCF in the regulation of cell type-specific and human disease-associated super-enhancers. *Genes Genomics* 41, 257–265. doi: 10.1007/s13258-018-0768-z

Shin, H. Y., Willi, M., Yoo, K. H., Zeng, X., Wang, C., Metser, G., et al. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* 48, 904–911. doi: 10.1038/ng.3606

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286. doi: 10.1038/nrg3682

Strumpf, D., Mao, C. A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., et al. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* 132, 2093–2102. doi: 10.1242/dev.01801

Thiecke, M. J., Wutz, G., Muhar, M., Tang, W., Bevan, S., Malysheva, V., et al. (2020). Cohesin-dependent and -independent mechanisms mediate chromosomal contacts between promoters and enhancers. *Cell Rep.* 32:107929. doi: 10.1016/j.celrep.2020.107929

Thomas, H. F., Kotova, E., Jayaram, S., Halbritter, F., Wysocka, J., Correspondence, C. B., et al. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol. Cell* 81, 969–982.e13. doi: 10.1016/j.molcel.2020.12.047

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858. doi: 10.1038/nature07730

Wang, J., Wang, J., Yang, L., Zhao, C., Wu, L. N., Xu, L., et al. (2020). CTCF-mediated chromatin looping in EGR2 regulation and SUZ12 recruitment critical for peripheral myelination and repair. *Nat. Commun.* 11:4133. doi: 10.1038/s41467-020-17955-2

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319. doi: 10.1016/j.cell.2013.03.035

Willi, M., Yoo, K. H., Reinisch, F., Kuhns, T. M., Lee, H. K., Wang, C., et al. (2017). Facultative CTCF sites moderate mammary superenhancer activity and regulate juxtaposed gene in non-mammary cells. *Nat. Commun.* 8:16069. doi: 10.1038/ncomms16069

Williamson, I., Lettic, L. A., Hill, R. E., and Bickmore, W. A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* 143, 2994–3001. doi: 10.1242/dev.139188

Xenopoulos, P., Kang, M., Puliafito, A., DiTalia, S., and Hadjantonakis, A. K. (2015). Heterogeneities in nanog expression drive stable commitment to pluripotency in the mouse blastocyst. *Cell Rep.* 10, 1508–1520. doi: 10.1016/j.celrep.2015.02.010

Yao, B., Wang, Q., Liu, C. F., Bhattaram, P., Li, W., Mead, T. J., et al. (2015). The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. *Nucleic Acids Res.* 43, 5394–5408. doi: 10.1093/nar/gkv426

Yao, Y., Minor, P. J., Zhao, Y. T., Jeong, Y., Pani, A. M., King, A. N., et al. (2016). Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nat. Genet.* 48, 575–580. doi: 10.1038/ng.3542

Zamudio, A. V., Dall'Agnese, A., Henninger, J. E., Manteiga, J. C., Afeyan, L. K., Hannett, N. M., et al. (2019). Mediator condensates localize signaling factors to key cell identity genes. *Mol. Cell* 76, 753–766.e6. doi: 10.1016/j.molcel.2019.08.016

Zatreanu, D., Han, Z., Mitter, R., Tumini, E., Williams, H., Gregersen, L., et al. (2019). Elongation factor TFIIS prevents transcription stress and R-Loop accumulation to maintain genome stability. *Mol. Cell* 76, 57–69.e9. doi: 10.1016/j.molcel.2019.07.037

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., et al. (2014). A *Sox2* distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* 28, 2699–2711. doi: 10.1101/gad.248526.114

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 31 March 2021
doi: 10.3389/fcell.2021.654344

Functional and Pathological Roles of AHCY

Pedro Vizán[1], Luciano Di Croce[1,2,3] and Sergi Aranda[1]**

[1] Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain, [2] Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

Adenosylhomocysteinase (AHCY) is a unique enzyme and one of the most conserved proteins in living organisms. AHCY catalyzes the reversible break of *S*-adenosylhomocysteine (SAH), the by-product and a potent inhibitor of methyltransferases activity. In mammals, AHCY is the only enzyme capable of performing this reaction. Controlled subcellular localization of AHCY is believed to facilitate local transmethylation reactions, by removing excess of SAH. Accordingly, AHCY is recruited to chromatin during replication and active transcription, correlating with increasing demands for DNA, RNA, and histone methylation. AHCY deletion is embryonic lethal in many organisms (from plants to mammals). In humans, AHCY deficiency is associated with an incurable rare recessive disorder in methionine metabolism. In this review, we focus on the AHCY protein from an evolutionary, biochemical, and functional point of view, and we discuss the most recent, relevant, and controversial contributions to the study of this enzyme.

Keywords: adenosylhomocysteinase, *S*-adenosylhomocysteine hydrolase, *S*-adenosylmethionine, gene regulation, epigenetics, chromatin, embryo development

OPEN ACCESS

Edited by:
José Luis Sardina, Josep Carreras Leukaemia Research Institute (IJC), Spain

Reviewed by:
Carolina M. Greco, University of California, Irvine, United States
Janine Santos, National Institute of Environmental Health Sciences (NIEHS), United States

****Correspondence:***
Luciano Di Croce
Luciano.DiCroce@crg.eu
Sergi Aranda
sergi.aranda@crg.eu

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *16 January 2021*
Accepted: *24 February 2021*
Published: *31 March 2021*

Citation:
Vizán P, Di Croce L and Aranda S (2021) Functional and Pathological Roles of AHCY. Front. Cell Dev. Biol. 9:654344. doi: 10.3389/fcell.2021.654344

INTRODUCTION

The metabolic enzyme adenosylhomocysteinase (AHCY; also alternatively called SAHH for *S*-adenosyl-L-homocysteine hydrolase, according to HUGO Gene Nomenclature Committee) is one of the most conserved enzymes in living organisms, including bacteria, nematodes, yeast, plants, insects, and vertebrates (Kusakabe et al., 2015). In mammals, AHCY is the only enzyme that mediates the reversible catalysis of *S*-adenosylhomocysteine (SAH) to adenosine and L-homocysteine (Turner et al., 1998) (**Figure 1**). AHCY acts within the one-carbon metabolic cycle, a universal metabolic process that enables the transfer of one-carbon units for biosynthetic processes (purines and thymidine), amino acid homeostasis (cysteine, serine, and methionine), redox cellular control, and epigenetic regulation (Ducker and Rabinowitz, 2017). SAH is the product of *S*-adenosyl-L-methionine (SAM)-dependent methyltransferases (MTases), which transfers the methyl group from SAM to a variety of cellular substrates, including nucleic acids and proteins.

S-adenosyl-L-methionine is the major methyl donor in the cell and the second most widely used cofactor after ATP (Clarke, 2001; Ye and Tu, 2018). SAH is structurally similar to SAM and binds with a similar range of affinity to the SAM-binding pocket of MTases, thereby inhibiting their activity in a negative feedback loop (Clarke, 2001; Richon et al., 2011; Zhang and Zheng, 2016) (**Figure 1**). As SAM and SAH compete for the SAM-binding pocket of MTases, the SAM:SAH ratio is considered an indicator of the methylation capacity of cells, and its decrease can correlate with a reduction of the methylation potential (Clarke, 2001; Petrossian and Clarke, 2011).

Unlike acetylation, which only occurs on proteins, methylation can be found at DNA, RNA, and proteins (Clarke, 2001). Chromatin is a macromolecular complex formed by DNA, together with physically associated proteins and non-coding RNAs (Espejo et al., 2020). Nucleosomes are the functional unit of the chromatin and comprise two of each of the core histones H2A, H2B, H3, and H4, which wrap up ∼147 base pairs of DNA. Chemical modifications on DNA and histones modulate gene expression and, in some instances, are responsible for conveying the epigenetic information to control cell-type-specific gene expression programs across cell division (Bonasio et al., 2010; Aranda et al., 2015). Indeed, the potential impact of AHCY in controlling methylation of DNA, RNA, and histones has been studied in many model organisms and cell types.

Here, we review the evolutionary conservation of AHCY across living organisms, its enzymatic structure, and its catalysis. We also detail the functional and molecular roles of AHCY in cellular homeostasis in different model organisms. Finally, we discuss the pathogenic influence of AHCY in human disease.

EVOLUTION, STRUCTURE, AND REGULATION OF ENZYMATIC ACTIVITY OF AHCY

Evolution

Adenosylhomocysteinase is a highly conserved enzyme present in all three major domains: Archaea, Bacteria, and Eukarya (Tehlivets et al., 2013; Kusakabe et al., 2015) (**Figure 1**). Comprehensive phylogenetic analyses suggest that AHCY evolved from a common ancestor in Archaea and Eukarya, and that orthologs transferred horizontally from these clades to bacteria lineage multiple times during evolution (Bujnicki et al., 2003; Stepkowski et al., 2005). In the absence of AHCY, many bacterial species (e.g., *Escherichia coli*) rely on the 5′-methylthioadenosine/*S*-adenosylhomocysteine (MTA/SAH) nucleosidase (which is lacking in mammals) for SAH withdrawal (Stepkowski et al., 2005).

The evolutionary reconstruction of the AHCY protein highlights two major divergences in its sequence. First, an insertion of about 40 amino acids is present within the catalytic domain in many bacteria and some eukaryotes, including plants and some pathogens (such as *Plasmodium falciparum*). In contrast, this insertion is absent Archaea and several eukaryotes, including model organisms such as *Dictyostelium discoideum*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*, as well as vertebrates (Bujnicki et al., 2003; Stepkowski et al., 2005). The insertion does not interfere with the folding of the catalytic domain (Tanaka et al., 2004; Reddy et al., 2008), and rodent insert-less AHCY can functionally restore loss-of-function mutants of *Ahcy* (with insertion) in the purple bacteria *Rhodobacter capsulatus* (Aksamit et al., 1995). Of note, in *Arabidopsis thaliana*, specific deletion of this insertion reveals its role in mediating interactions with proteins, such as adenosine kinase and the mRNA cap MTase, as well as the nuclear localization of AHCY, suggesting a regulatory role for this specific insert in plants (Lee et al., 2012).

The second divergent region in AHCY during evolution is a stretch of about eight amino acids in the *C*-terminal domain present in bacteria and eukaryotes (but not in Archaea), which stabilizes the interaction of AHCY with the NAD^+ cofactor (Hu et al., 1999; Komoto et al., 2000; Kusakabe et al., 2015). In Archaea, the observed strong affinity for the NAD^+ cofactor suggests that the lack of a *C*-terminal tail in AHCY must be replaced by other specific residues (Porcelli et al., 1993).

In eukaryotes, AHCY is among the top 100 most-conserved proteins between yeast and mammals, with about 70% of identity (Mushegian et al., 1998). From zebrafish to humans, *AHCY* encompasses 10 exons with nearly identical size, indicating that the high conservation in vertebrates is also maintained at a structural genomic level (Zhao et al., 2009). The human genome harbors one *AHCY* gene in chromosome 20, and 7 pseudogenes in different chromosomes. *AHCY* is ubiquitously expressed in adult tissues (Chen et al., 2010) and encodes three annotated splicing isoforms (AHCY1 (NP_000678), −2 (NP_001155238), and −3 (NP_001309015), with 432, 404, and 434 amino acids, respectively) that differ only in their first 30 amino acids. These isoforms have not been formally compared side-by-side, although ectopically expressed hAHCY1 mutant lacking the first 14 amino acids accumulates in the nucleus and form catalytically dead tetramers (Grbesa et al., 2017). In-depth molecular and functional analyses are now required to verify and then elucidate a functional regulatory role of this *N*-terminal region.

Structure

Several structures of AHCY have been resolved at an atomic level from bacteria (i.e., *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa*), pathogenic eukaryotes (i.e., *Trypanosoma brucei* and *Plasmodium falciparum*), plants, rodents, and human, all of which share a high structure conservation (Brzezinski, 2020). While some evidence indicates that AHCY can exist in monomeric, homodimeric, and homotetrameric forms within cells (Seo and Ettinger, 1993a,b), all AHCY structures (except that of the plant enzyme *Lupinus luteus*) have been resolved as active homotetramers, with one NAD^+ cofactor bound in each subunit (Turner et al., 1998; Hu et al., 1999; Brzezinski et al., 2008; Kusakabe et al., 2015). Each monomer comprises three conserved domains (**Figure 1**): (*i*) a substrate-binding/catalytic domain (SBD; amino acids [aa] 1–181 and 355–385 in human AHCY); (*ii*) a NAD cofactor-binding domain (CBD; aa 197–351); and (*iii*) a *C*-terminal tail (aa 386–432). Both the SBD and the CBD are connected by two hinge regions (aa 182–196 *N*-terminal hinge, and 352–354 *C*-terminal hinge) (Turner et al., 1998; Kusakabe et al., 2015). The central core of the homotetramer is occupied by the cofactor-binding domains from different subunits, and the substrate-binding domains reside at the surface of the tetramer. Notably, the cofactor-binding domain is spatially separated from the substrate-binding domain by a deep cleft in each monomer. The homotetramer is a dimer of dimers, as the *C*-tail domains from two pairs extend and contact, reciprocally, the NAD-binding domain of the other pair (Turner et al., 1998; Yang et al., 2003; Kusakabe et al.,

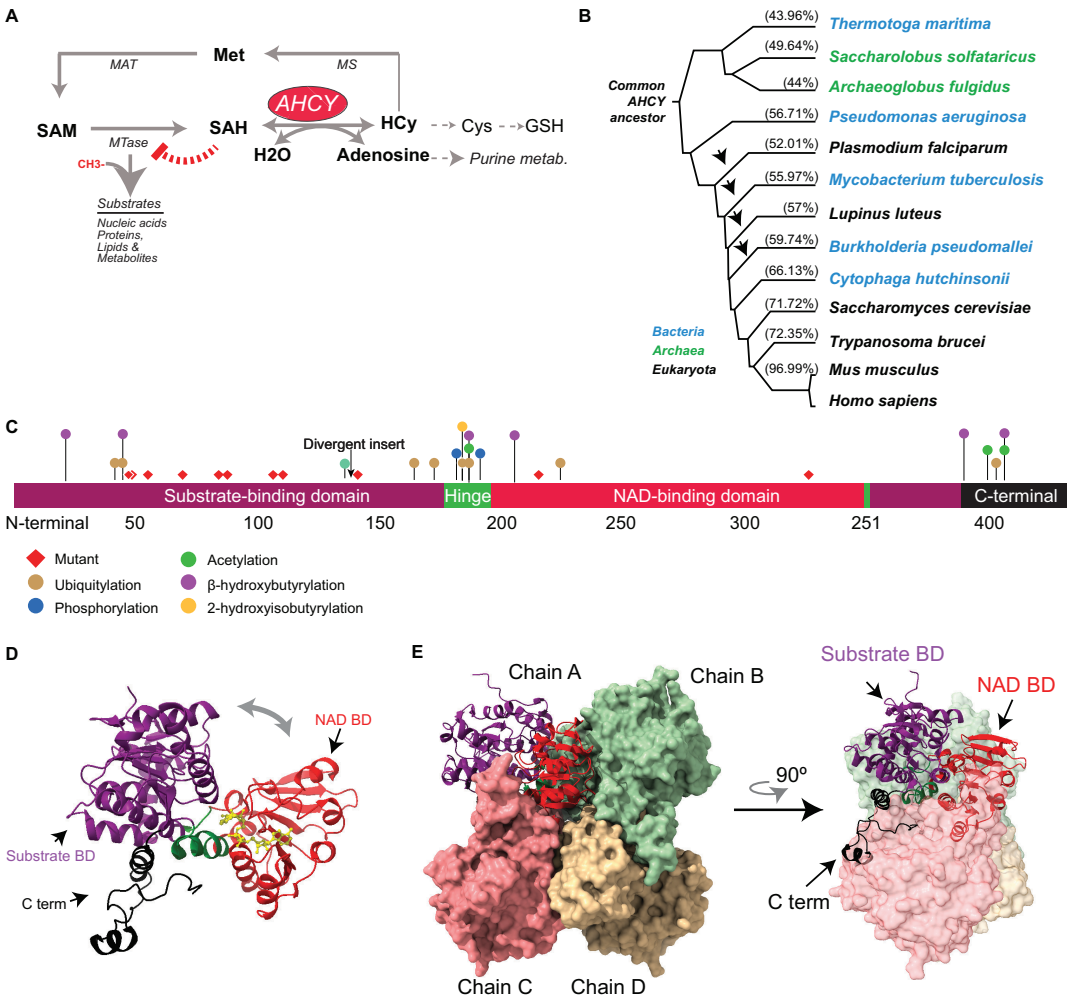


FIGURE 1 | Evolution and structure of adenosylhomocysteinase (AHCY). **(A)** Scheme of the methionine (Met) metabolic pathway. Enzymes of the pathway are indicated AHCY, methionine adenosyltransferase (MAT), methionine synthase (MS), and SAM-dependent methyltransferase (MTase). MTase transfer a methyl group from SAM to substrates, thereby generating SAH. AHCY breaks SAH into adenosine and homocysteine (Hcy). Hcy can be recycled to methionine (Met) coupled to folate metabolism or to produce glutathione. **(B)** Selected AHCY amino acid sequences were obtained from Uniprot (https://www.uniprot.org/) using the following accession numbers *Thermotoga maritima* (O51933), *Saccharolobus solfataricus* (P50252), *Archaeoglobus fulgidus* (O28279), *Pseudomonas aeruginosa* (Q9I685), *Plasmodium falciparum* (P50250), *Mycobacterium tuberculosis* (P9WGV3), *Lupinus luteus* (Q9SP37), *Burkholderia pseudomallei* (Q3JY79), *Cytophaga hutchinsonii* (A0A6N4SNR7), *Saccharomyces cerevisiae* (A0A6A5PY71) *Trypanosoma brucei* (Q383X0), *Mus musculus* (P50247), and *Homo sapiens* (P23526). The tree was generated by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and Phylodendron (http://iubioarchive.bio.net/treeapp/treeprint-form.html). The % indicates the similarity with human protein (P23526) calculated using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The arrow indicates the species where the ∼40-amino acid insert is present. **(C)** The different functional modules of human AHCY are represented. The numbers indicate the position of amino acids. The arrow indicates the position of the divergent insert found in some bacteria and eukaryotes, including plants. Red diamonds indicate the position of the twelve mutations founds in patients with AHCY deficiency (R49C, R49H, A50T, T57I, G71S, D86G, A89V, E108K, T112stop, Y143C, V217M, and Y328D). The colored dots indicate the position of posttranslational modifications found in PhosphoSitePlus, with at least five references, and in the literature. **(D)** Structural model for the human AHCY monomer in the open conformation, the co-factor is depicted in yellow in association with the NAD-binding domain (Protein Data Bank-PDB: 4yvf). The different modules are colored as in **(C)**. **(E)** Structural model of human AHCY tetrameric enzyme in close conformation (protein data bank-PDB: 3nj4). The different modules are colored as in **(C)**.

2015). The tetrameric structure is sustained by intersubunit interactions between two alpha helices (aa 184–187 and aa 190–207) from the adjacent subunits (Turner et al., 1998; Yang et al., 2003).

Structural analyses and computational modeling using different resolved structures indicate that each monomer can exist in an "open" (bound to NAD^+) or "closed" (bound to NAD^+ and substrate) conformation. Upon substrate binding, an ∼18° rotation of the hinge brings together the cofactor- and substrate-binding domains, followed by a rotation of the dimers by ∼14°C (Yin et al., 2000; Yang et al., 2003). Once the reaction has occurred, the enzyme reverts to an open conformation and the product is released (Wang et al., 2005). These transitions from open-to-closed conformation would facilitate the steps of

the enzymatic reaction as well as the diffusion of substrates (Wang et al., 2005).

Enzymatic Activity Regulation

The enzymatic reaction for the reversible catalysis of SAH, first described in 1979 (Palmer and Abeles, 1979), was later supported by analysis of the mammalian structure of AHCY (Kusakabe et al., 2015). At the mechanistic level, AHCY mediates a nucleophilic enzymatic cascade enabled by redox steps (Walsh and Moore, 2019). The reaction is initiated (both *via* the forward breakup or reverse synthesis of SAH) by an oxidation of SAH or adenosine substrates by the enzyme-bound NAD^+. The oxidated intermediate is then cleaved to release homocysteine (Hcy) or water (depending on the substrate) and then is subsequently reduced by NADH to form the final product (adenosine or SAH). In thermodynamic equilibrium, the reaction is largely favored toward the synthesis of SAH *in vitro*, but efficient removal of adenosine and homocysteine enables the net breakup of SAH *in vivo* (Palmer and Abeles, 1979; Kusakabe et al., 2015; Walsh and Moore, 2019).

Proteomic analysis has revealed that mammalian AHCYs are acetylated at lysines 401 and 408 within their *C*-terminal tail (Choudhary et al., 2009). *In vitro*, bi-acetylated-K401/408 human AHCY displays a threefold reduction of the catalytic constant and two-fold increase of SAH K*m* (Wang et al., 2014). Comparative analyses between unmodified and acetylated structures of AHCY indicate a local hydrogen-bound alteration in the vicinity of the modified lysine residues, indicating that slight structural changes in AHCY might have a significant functional impact on its catalytic activity. Two additional modifications, based on the conjugation with short fatty acid derivatives on lysine residues, have recently been reported: the 2-hydroxyisobutyrylation (hib) of lysine (K) 186, and the β-hydroxybutyrylation (bhb) of several lysines (20, 43, 188, 204, 389, and 405) (Huang et al., 2018; Koronowski et al., 2021). In particular, forced K-bhb inhibits AHCY activity in mouse embryonic fibroblast (MEFs) and mouse liver (Koronowski et al., 2021). Enzymatic assays from cells ectopically expressing mutants show that K188R, K389R, and K405R substitutions compromise AHCY activity (Koronowski et al., 2021). In addition to lysine modification, mouse AHCY is posttranslationally modified with an *O*-linked β-*N*-acetylglucosamine sugar (*O*-GlcNAcylation) at threonine 136 (Zhu et al., 2020). The oligomerization capacity of AHCY (and therefore, its enzymatic activity) is reduced by mutation of threonine 136 to alanine (T136A), as well as by pharmacological inhibition of glycosylation (Zhu et al., 2020). Importantly, mouse embryonic stem cells (mESCs) expressing AHCY-T136A mutant display a reduced proliferation and low alkaline phosphatase staining (Zhu et al., 2020), suggesting that AHCY glycosylation is important to balance self-renewal and pluripotency in mESCs. Thus, posttranslational modifications in AHCY can impact the structure of the enzyme, thereby modulating its enzymatic activity and its biological role. In line with this, mass spectrometry-based proteomic discovery has revealed an accumulation of other posttranslational modifications (i.e., phosphorylation, acetylation, and ubiquitylation) in the vicinity of the non-catalytic, movable hinge domain and the *C*-terminal domain (source www.Phosphosite.com) (**Figure 1**), pointing toward their potential roles as AHCY regulatory domains.

Several resolved AHCY structures contain monovalent cations, which would facilitate the catalytic activity. For instance, the resolved structure of the mouse AHCY has sodium cation allocated in the *C*-terminal hinge region (Kusakabe et al., 2015), contributing to the recognition of the substrate, similarly as the plant enzyme (Brzezinski et al., 2008). In addition, the AHCY from *Pseudomonas aeruginosa* binds potassium cations, and kinetic studies indicate that potassium stimulates AHCY enzymatic activity and ligand binding (Czyrko et al., 2018). Conversely, divalent cations, such as zinc and copper, have been shown to inhibit AHCY activity. Structural studies of AHCY from *P. aeruginosa* indicate that zinc ions are coordinated with AHCY homotetramers near the gate of the active sites, thus preventing its accessibility (Czyrko et al., 2018). Even though a copper ion is not required for its protein stabilization or catalytic activity, AHCY is a strong copper binder both *in vitro* and *in vivo*, with a K*d* of about 1×10^{-12} for free-copper, and K*d* of about 1×10^{-17} in the presence of EDTA (Bethin et al., 1995; Li et al., 2004). *In vitro*, copper is a non-competitive inhibitor for the substrates, facilitating the dissociation of NAD^+ in a concentration-dependent manner (Li et al., 2007). Although Cu-AHCY structures have not been resolved, computational predictions as well as the mechanism of inhibition suggest that Cu^{2+} binds the central core of the AHCY tetramer, preventing subunit interaction and/or decreasing NAD^+ affinity (Li et al., 2007). Both AHCY's capacity to bind copper, which is comparable with albumin (Masuoka et al., 1993), and its high abundance in the liver, where it comprises 0.5% of total hepatic cytosolic protein (Bethin et al., 1995), indicate that AHCY could contribute significantly to the hepatic copper metabolism and copper-associated human syndromes, such as Wilson disease. All these data suggest that cation-AHCY interaction may have significant relevance and that the biological and pathological cross-talk between AHCY and cations needs further exploration.

In vitro binding experiments have shown that bovine AHCY endows two adenosine binding sites, and their usage depends on the enzyme-bound NAD^+/NADH ratio (Kloor et al., 2003). With a low affinity, adenosine binds the AHCY-NAD^+ at the catalytic domain while, with high affinity, adenosine binds the enzymatically inactive AHCY-NADH at the cofactor domain (Kloor et al., 2003). Despite the importance of NAD^+ as a cofactor, whether an intracellular fluctuation in NAD^+/NADH concentrations influences AHCY activity or its adenosine binding *in vivo* remains unknown.

BIOLOGICAL AND FUNCTIONAL ROLES OF AHCY

AHCY Function in Embryonic Development

Considering its strong conservation during evolution and its ability to affect the methylation potential in cells, the functionality of AHCY has been explored in multiple organisms.

The *Arabidopsis thaliana* genome harbors two AHCY paralogs (*AtSAHH1* and *AtSAHH2*) with non-redundant functions. While AtSAHH2 seems functionally dispensable in plants (Rocha et al., 2005), a loss-of-function mutation in AtSAHH1 is embryonic lethal and impairs the DNA methylation-dependent gene silencing (Furner et al., 1998). Moreover, genetic and pharmacological reduction of AtSAHH1 activity results in developmental abnormalities (i.e., slow growth and reduced fertility) (Furner et al., 1998; Rocha et al., 2005; Wu et al., 2009). At the molecular level, AtSAHH1 mutant plants display a general DNA hypomethylation and alterations in the expression of key developmental genes, a defect attributed to direct inhibition of plant DNA methyltransferases or histone methyltransferases required to sustain genome methylation (Rocha et al., 2005; Mull et al., 2006; Jordan et al., 2007; Ouyang et al., 2012). Similarly, tobacco plants with reduced expression or pharmacological blockade of AHCY show defects in growth and flowering, reduced DNA methylation on repetitive elements, and ectopic expression of genes critical for flower development in different organs (Tanaka et al., 1997; Fulnecek et al., 2011). Finally, the tomato genome contains three related genes encoding for highly conserved SAHH that display functional redundancy, in contrast to those in *A. thaliana* (Li et al., 2015). Simultaneous silencing of all three genes inhibits vegetative growth (Li et al., 2015). Overall, these studies indicate the essentiality of AHCY for proper plant development and growth.

The mouse genome harbors the *Ahcy* gene in chromosome 2, and at least five intron-less genomic sequences similar to the complementary DNA (cDNA) in chromosomes X, 1, and 16 (from 86 to 99.7% identity). These latter likely correspond to pseudogenes (based on the alignment of mouse *Ahcy* cDNA-CCDS16942.1, using BLAT; Kent, 2002), and at least one (in chr16) encodes for a potentially complete wild-type form of AHCY protein. Phenotypic analysis of a large knockout mouse line collection indicates that homozygous deletion of *Ahcy* is embryonic lethal before E9.5 (Dickinson et al., 2016), and chromosomal microdeletions (of about 100 kb) encompassing the *Ahcy* gene in mice cause recessive lethality at the blastula stage (Miller et al., 1994). Several pieces of evidence point to the essential role of *Ahcy* during early mouse embryo development at the pre-implantation stage. First, during the first 4.5 days post-coitum, *Ahcy* expression increases gradually from the zygote to blastula stages, following the acquisition of the pluripotency (Aranda et al., 2019). Second, the pharmacological treatment of mouse embryos with AHCY inhibitors severely compromises the blastula progression (Miller et al., 1994; Aranda et al., 2019). Third, specific depletion of *Ahcy* drastically reduces the proliferation of mESCs derived from blastocyst (Aranda et al., 2019), and a prolonged period of *Ahcy* depletion promotes spontaneous mESC differentiation (Zhu et al., 2020). Further generation of inducible or tissue-specific Ahcy knockout mice models would be of great interest to dissect its role during development or adult tissue homeostasis.

AHCY Function in Cellular Stress

Metabolomics analysis of young and old *Drosophila melanogaster* indicates that aging in flies correlates with a remarkable accumulation of SAH. Conversely, fly strains naturally selected for their longevity display reduced levels of SAH (Parkhitko et al., 2016). Reduction of *dAhcy* by RNA interference in the whole body or in a neural-specific manner drastically increases SAH levels and reduces lifespan. Conversely, the suppression of two enzymatically inactive paralogs of *dAhcy* (*dAhcyL1* and *dAhcyL2*) increase lifespan, presumably because they act as dominant-negative *dAhcy* interactors (Parkhitko et al., 2016). In zebrafish, loss-of-function mutations in *Ahcy* are lethal, causing defects in exocrine pancreas development (Yee et al., 2005). In addition, mutants display hepatic steatosis (accumulation of lipids in hepatocytes) and liver degeneration early at larvae stage (Yee et al., 2005; Matthews et al., 2009), two common features observed in human patients with AHCY deficiency (see below). These studies suggest that AHCY plays a major role in controlling cellular homeostasis and tissue damage in animals. Although the mechanisms underlying increased aging and tissue degeneration in AHCY-deficient animals have not been experimentally determined, they are likely due to: (*i*) a reduced potential of the transmethylation reaction for specific substrates; (*ii*) increased oxidative stress, as a consequence of a reduction in the flux homocysteine into the transsulfuration pathway for glutathione production; and/or (*iii*) replication stress, caused by decreased availability of adenosine for nucleotide production.

In agreement with its potential role in cellular toxicity, a knockdown of AHCY in the hepatocellular carcinoma cell line HepG2 results in attenuation of cell cycle progression and increased DNA damage (Beluzic et al., 2018). The cytotoxic effects can be reversed by adenosine supplementation, thus suggesting that mild inactivation of AHCY may cause cellular stress in liver cells, due to adenosine depletion (Beluzic et al., 2018).

Molecular Effects on DNA, RNA, and Histone Methylation

DNA methylation is a well-studied and paradigmatic epigenetic mark (Bonasio et al., 2010). Once deposited, the 5-methylcytosine mark can self-perpetuate by a well-known molecular mechanism, influencing genome functionality. DNA methyltransferase 1 (DNMT1), the enzyme responsible for the maintenance of DNA methylation, recognizes hemimethylated DNA during replication and establishes the methyl group in the newly synthesized DNA (Bonasio et al., 2010). This process is assisted by several accessory proteins, including the multidomain protein UHRF1 (Blanchart et al., 2018). In a proteomic survey during the cell cycle of HeLa cells, AHCY was found as a partner of DNMT1 (Ponnaluri et al., 2018) (**Figure 2**). *In vitro*, methyltransferase assays indicated that AHCY enhances the DNMT1 activity, and *in vivo*, AHCY overexpression induces a pervasive increase in DNA methylation in HEK293 cells (Ponnaluri et al., 2018). As this suggests that AHCY is a rate-limiting factor for DNA methylation maintenance, its role during epigenomic reprogramming throughout embryo development and/or in disease progression, such as cancer, must now be investigated.

In addition to maintaining DNA methylation, AHCY also exerts a role in regulating *de novo* DNA methylation (Zhou

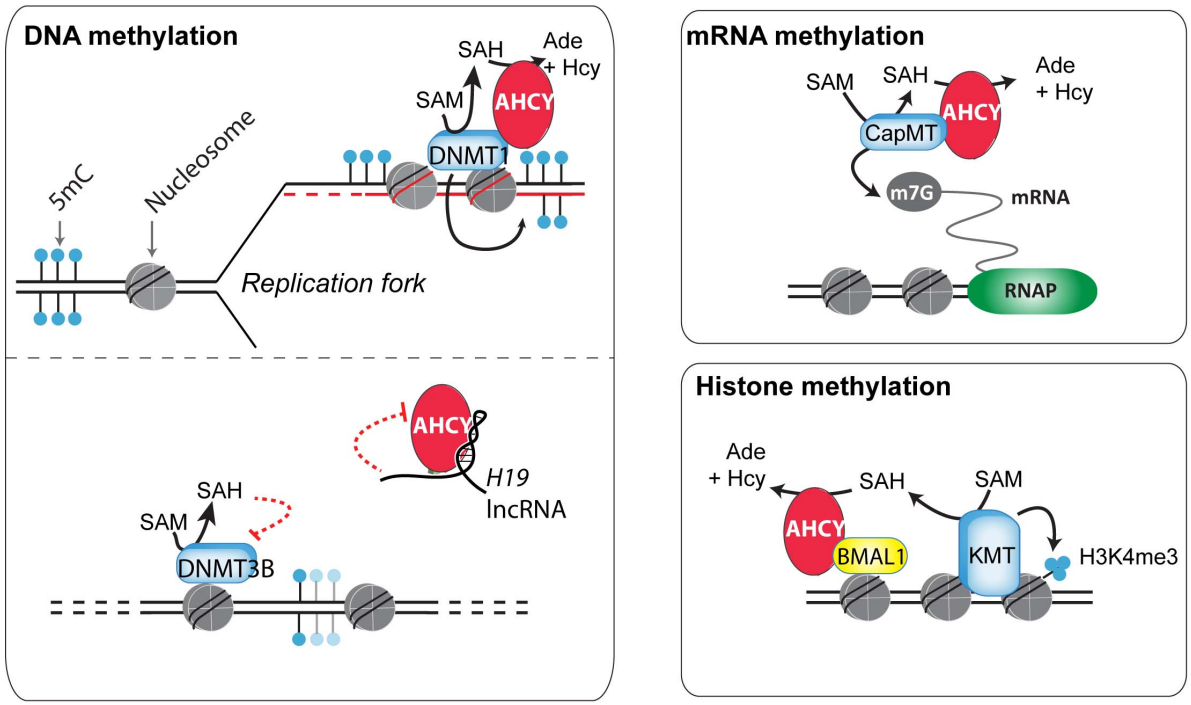


FIGURE 2 | Potential mechanisms of action of AHCY in mammals. Different mechanisms of action have been proposed for the control of DNA, RNA, and histone methylation (discussed in the main text).

et al., 2015). This role is proposed to be mediated throughout the AHCY inhibition by the developmentally regulated long non-coding RNA *H19*. *In vitro*, U-rich nucleotide elements present in *H19* reduce the ability of recombinant AHCY to breakdown SAH *in vitro* (Zhou et al., 2015). *In vivo*, knockdown of *H19* increases the cellular activity of AHCY in myotubes differentiated from mouse myoblasts. As a consequence, *H19* knockdown induces extensive changes in the pattern of genome methylation in myotubes, with specific DNMT3B-target genes gaining DNA methylation in an AHCY-dependent manner. Overall, this indicates that, consistent with its function, AHCY deficiency impacts both in DNA methylation maintenance as well as in *de novo* methylation, thus profoundly influencing the epigenome of cells (**Figure 2**).

RNA methyltransferases are highly sensitive to the reduction of the SAM:SAH ratio (Clarke, 2001). Indeed, pharmacological treatment with 3-deazaadenosine, a specific AHCY inhibitor, induces reduction of methyl RNA levels in mouse macrophage and osteosarcoma cells, including loss of the N6-methyladenosine (m6A) (Backlund et al., 1986). In mouse embryonic fibroblasts, this m6A reduction upon AHCY inhibition is functionally linked to a delay in processing RNAs from circadian rhythm genes, thus impacting their oscillatory behavior response (Fustin et al., 2013). Extending this observation, Fustin et al. (2020) have recently found that targeting AHCY activity with 3-deazaneplanocin A (DZnep) produces a loss in the oscillatory transcriptional response of circadian-associated genes in different organisms (from unicellular algae to humans). The loss of the oscillatory transcriptional activation of the *Hes7* gene during somitogenesis on DZnep-treated mouse embryos suggests that methylation flux, directed by AHCY, is a universal regulator of biological circadian rhythms in metazoans (Fustin et al., 2020).

Global steady-state levels of cap methylation on mRNA (7-methylguanosine, m7G) are seemingly unaffected upon AHCY inhibition in mouse embryonic fibroblasts (measured as the ratio m7G methylated/unmethylated mRNA) (Fustin et al., 2013). However, considering the tight association of m7G with mRNA stability, the uncapping of pre-mRNA could be linked with an increase in pre-mRNA degradation, and, therefore, the impact of AHCY inhibition in m7G regulation might be overlooked. Indeed, several experimental evidences show that increased m7G demands are dependent on AHCY functionality in some instances. In developing Xenopus, xAHCY, which remains nearly cytosolic until the blastula stage, starts to accumulate in the nucleus of differentiating cells, which experience a boost in transcriptional activity during gastrulation (Radomski et al., 1999). In Xenopus oocytes, xAHCY displays a nuclear accumulation on active transcribing loops on lampbrush chromosomes (Radomski et al., 1999) and interacts with mRNA (guanine-7-) methyltransferase through its proximal *N*-terminal region (Radomski et al., 2002). Furthermore, its pharmacological inhibition compromises both methylation and synthesis of nuclear RNA (Radomski et al., 1999, 2002). Similarly, in human and murine cells, an increased demand for m7G induced upon MYC activation depends on AHCY activity (Fernandez-Sanchez

et al., 2009). Upon AHCY knockdown, the MYC-induced m7G is compromised and correlates with reduction in Myc-induced core effects in protein synthesis, cell proliferation, and transformation. Interestingly, *AHCY* is a direct MYC target and it is transcriptionally upregulated in an MYC-dependent manner in numerous cell types (Fernandez-Sanchez et al., 2009). Thus, AHCY appears to be a key response molecule in controlling the molecular and functional effects of Myc.

Based on its role on controlling RNA methylation, AHCY has been long considered a target for antiviral strategies (Bader et al., 1978; De Fazio et al., 1990; Snoeck et al., 1993; Masuta et al., 1995; Daelemans et al., 1997; Chiang, 1998; Daelemans et al., 1998; Huggins et al., 1999; Bray et al., 2000; De Clercq, 2005, 2015; Yoon et al., 2019; Shin et al., 2020). The antiviral properties of AHCY inhibition are based on the metabolic accumulation of SAH, which directly inhibits viral or host RNA MTases required for successful viral spreading (Daelemans et al., 1998; De Clercq, 2005; Liu and Kiledjian, 2006; Leyssen et al., 2008). AHCY inhibitory compounds show a broad-spectrum antiviral activity against RNA viruses, including rhabdo-, filo-, arena-, paramyxo-, reo-, and retroviruses. *In vivo* efficacy of AHCY inhibitors is remarkable, as a single dose of either of the three distinct AHCY inhibitors 3-deazaneplanocin A, 3-deazaaristeromycin, or 3-deazaadenosine protects mice against a lethal infection of the Ebola virus (subtype Zaire) (Huggins et al., 1999; Bray et al., 2000, 2002). Recently, the list of AHCY-sensitive viruses has increased, as the novel potent AHCY inhibitors of 6′-fluorinatedaristeromycin analogs have notable activity against emerging positive-stranded RNA (+RNA) viruses, such as the severe acute respiratory syndrome coronavirus (SARS-CoV1), Middle East respiratory syndrome coronavirus (MERS-CoV), chikungunya virus (CHIKV), and Zika virus (ZIKV) (Yoon et al., 2019; Shin et al., 2020). These results have placed 6′-fluorinatedaristeromycin as a drug candidate against COVID-19.

The methylation of core nucleosome histones, which are the most abundant eukaryotic nuclear proteins, influences protein–protein interactions, chromatin structure, and gene expression (Espejo et al., 2020). The potential role of the nucleosome as a SAM consumer is high, as methylation of ∼0.1% of all histone residues can consume the complete cellular SAM pool in mammalian cells (Ye and Tu, 2018). Arginines and lysines can be extensively methylated on histone proteins, with the histone 3 lysines 4, 9, 27, 36, and 79 (H3K4, H3K9, H3K27, H3K36, and H3K79, respectively) being the most well-characterized (Zhao and Garcia, 2015). Among the different potential effects of AHCY blockade over histone methylation, the global and local di- and tri-methylation of lysine 4 on histone H3 (H3K4me2/3) appears the most sensitive (Fustin et al., 2013; Greco et al., 2020; Zhu et al., 2020). Recently, Greco et al. (2020) found that the oscillatory increase of H3K4me3 in circadian rhythm-related genes is compromised after pharmacological inhibition of AHCY in mouse embryonic fibroblasts. At the molecular level, AHCY interacts with BMAL1, a key circadian transcription factor, and regulates the oscillatory recruitment of BMAL1 to chromatin, thus impacting the amplitude of the transcriptional oscillations of circadian rhythm-related genes (Greco et al., 2020) (**Figure 2**). The authors suggest that the dimer AHCY-BMAL1 cooperates with histone methyltransferases for gene activation and that AHCY removes excess of SAH to facilitate local histone methylation demands during the transcription peaks.

IMPACT OF AHCY IN HUMAN DISEASES

AHCY Involvement on Rare Metabolic Disorders

Wilson disease (WD) is a rare, inborn disorder with severe hepatic and neurological dysfunctions due to the pathological accumulation of copper (Czlonkowska et al., 2018). The clinical symptomatology in WD is highly variable for age of onset (ranging from 5 to 35 years) and severity, varying from nearly asymptomatic with mild changes in personality to acute liver failure or parkinsonian-like effects. Although over 300 disease-causing mutations have been reported in the copper transporter *ATP7B*, the genotype–phenotype correlation in WD is very poor (Czlonkowska et al., 2018). This suggests the existence of other genetic risk factors influencing WD. As AHCY is a strong copper binder, and copper inhibits AHCY activity in a dose-dependent and non-competitive manner (see above), a pathological link between copper accumulation and the inhibition of AHCY has been proposed. Indeed, the pathological accumulation of copper in the Jackson toxic milk mutant mice (tx-J), which carry a missense mutation in the *Atp7b* gene, reduces AHCY expression (Bethin et al., 1995; Li et al., 2004, 2007). In this WD murine model, deficit in AHCY associates with an increase of SAH levels, reduction in the SAM:SAH ratio, and DNA hypomethylation (Medici et al., 2013, 2014, 2016). Interestingly, feeding tx-J mice with diets supplemented with betaine or choline normalizes the SAM:SAH ratio, global DNA hypomethylation, and the expression of hepatic genes related to methionine homeostasis (Medici et al., 2013, 2014, 2016). These data provide a link between hepatic copper accumulation and the function on AHCY in the liver.

In humans, at least 204 missense *AHCY* variants are annotated in public databases[1] or reported in the literature. Twelve of these variants are linked with a rare autosomal recessive disorder in the methionine metabolism of hypermethioninemia (R49C, R49H, A50T, T57I, G71S, D86G, A89V, E108K, T112stop, Y143C, V217M, and Y328D), characterized by a deficiency of AHCY (OMIM: 180960) (Baric et al., 2004, 2005; Buist et al., 2006; Vugrek et al., 2009; Honzik et al., 2012; Stender et al., 2015; Strauss et al., 2015; Vivante et al., 2017; Bas et al., 2020; Grudzinska Pechhacker et al., 2020). Patients typically display severe hepatic, muscle, and cognitive dysfunction, including multiorgan failure followed by death soon after birth (Tehlivets et al., 2013; Stender et al., 2015; Judkins et al., 2018). In addition to this most frequent and severe version, a mild version of the disease has been reported in one family, who remain nearly asymptomatic during childhood but presented severe hepatic disorders during adulthood (Stender et al., 2015). Dietary interventions that reduce methionine uptake have shown variable benefits for patients, questioning its therapeutic advantage in

[1]https://gnomad.broadinstitute.org/

AHCY deficiency syndrome. For one 40-month-old child for whom dietary therapy was ineffective, liver transplantation restored metabolic parameters and reversed psychomotor and cognitive deficits after 6 months (Strauss et al., 2015). However, the lack of further assessment on the course of the disease, and the difficulties of finding a rapidly compatible donor soon after diagnosis, would argue against liver transplantation as a general indication for AHCY deficiency. Thus, finding an effective therapy is urgently needed for treating AHCY deficiency.

Role of AHCY in Cancer

Adenosylhomocysteinase was initially defined as a tumor suppressor (Leal et al., 2008). In agreement, AHCY-deficient mouse embryonic fibroblasts have a higher proliferative capacity and the ability to escape from replicative senescence (Leal et al., 2008). At a mechanistic level, an AHCY knockdown overcomes the p53-induced cell cycle arrest and abrogates p53-dependent transcriptional activity (Leal et al., 2008). In agreement with a tumor suppressor role, *AHCY* mRNA expression was found to be reduced in a large number of different solid tumor samples as compared with non-tumor paired tissue (Leal et al., 2008).

However, AHCY's function as a tumor suppressor seems to be cell type specific, as AHCY inhibition is also linked to anti-migratory and anti-invasive activity in breast cancer cells (Kim et al., 2013; Park et al., 2015) and to enhanced apoptosis in high aggressive neuroblastoma (Westermann et al., 2008; Chayka et al., 2015). In neuroblastoma, AHCY expression is elevated in MYCN-amplified tumor samples and neuroblastoma cell lines (Westermann et al., 2008; Chayka et al., 2015). Interestingly, AHCY knockdown or drug-mediated inhibition induces an increase in apoptosis specifically on MYCN-amplified neuroblast cells (Chayka et al., 2015), thus showing a specific synthetic lethality and making AHCY inhibition a promising therapeutic possibility for personalized treatment of high-risk neuroblastoma.

CONCLUSION

Methylation is one of the most ubiquitous chemical reactions in living organisms (Clarke, 2001). While some methyltransferases retain most of their activity at different SAM:SAH ratios, such as the glycine *N*-methyltransferase, others are more susceptible to SAH elevation, including tRNA or arginine methyltransferases (Clarke, 2001; Richon et al., 2011; Zhang and Zheng, 2016). Thus, considering the large number of SAM-dependent MTases, which account for nearly 1% of the human protein-coding genes (∼200), their different functionalities, and their distinct sensitivity to SAH accumulation (Clarke, 2001; Petrossian and Clarke, 2011; Richon et al., 2011; Zhang and Zheng, 2016), it is difficult to foresee which is the most relevant methyltransferase affected upon AHCY inhibition. Thus, after AHCY blockade, the influence of one molecular outcome (e.g., loss of DNA methylation, RNA methylation, or protein methylation) in a particular molecular process will depend very much on the following: (*i*) the biological context (e.g., cell type, stimulus); (*ii*) the molecular event involved in this biological process (e.g., transcription); (*iii*) the most relevant methyltransferase involved; and (*iv*) their relative sensitivity to SAH accumulation.

Also, as the result of the AHCY activity, the methionine cycle can supply the production of adenosine for nucleotide biosynthesis and for homocysteine, the precursor for the biosynthesis of glutathione in higher eukaryotes (Ye and Tu, 2018). Therefore, in addition to increasing the methylation potential, SAH turnover could be essential to cope with replicative-associated oxidative stress, thus expanding the repertoire of the biological process where AHCY plays a role.

Methylations occur in all subcellular compartments. The intracellular concentrations of SAM (∼10 μM) is at the range of the K*m* of the SAM-dependent MTases, but it can fluctuate 10–100-fold in normal physiological conditions (Clarke, 2001; Ye and Tu, 2018). The nuclear heterogeneity at the transcriptional level and during replication, as well as the existence of phase-separated membrane-less subnuclear compartments (Erdel and Rippe, 2018), likely produce large differences in local fluxes of SAH due to localized transmethylation reactions, such as DNA methylation maintenance at replication sites, or mRNA cap methylation at transcriptionally active chromatin regions (Espejo et al., 2020). Hence, the local recruitment of AHCY seems required to sustain effective transmethylation reactions in time and space. We and others have found that AHCY is recruited to chromatin at specific sites (Aranda et al., 2019; Greco et al., 2020). Our recent discoveries unveil a novel chromatin function of AHCY in controlling gene expression, which is in line with recent data identifying the association of other metabolic enzymes with chromatin (Li et al., 2018). However, our discoveries also present new questions at the mechanistic level: How is AHCY recruited to specific chromatin regions in mammals? How does AHCY control local gene expression? Further studies on protein–protein interactions within chromatin will clarify the molecular mechanisms underlying AHCY's role in gene regulation. This research direction might also have an impact in AHCY-related pathologies as changes in the nucleocytoplasmic distribution have been reported in several pathological AHCY variants (Grbesa et al., 2017). Therefore, the potential pathogenicity caused by mutants that cause aberrant nuclear localization of AHCY might be linked with its chromatin-related functions.

Hundreds of missense polymorphisms are present in the human *AHCY* gene across the population (see text footnote 1), some of which have a similar allelic frequency as those identified as pathological variants. This, together with the asymptomatic evolution of some patients with defects in AHCY activity until fertile adulthood (Stender et al., 2015), suggest that AHCY deficiency could be underdiagnosed, and that the impact of AHCY in human pathologies could be underestimated in the general population.

In 2004, AHCY deficiency was first identified in a 5-month-old Croatian infant with delayed psychomotor development, hypotonia, attention defects, and poor head control since birth, who also presented with elevated levels of aminotransferase, methionine, SAH, and SAM in serum (Baric et al., 2004). Since then, additional patients harboring *AHCY* mutations have been described with similar clinical traits. However, only few associated mutations have been characterized at the molecular

level. In addition to the pathological variants that strongly affect AHCY activity, non-pathologic polymorphic isoforms of AHCY found in different populations display reduced thermal stability of the protein and enzymatic activity (Fumic et al., 2007). More recently, genetic variations of *AHCY* have been linked to different prognosis of children with neuroblastoma (Novak et al., 2016), thus providing evidence of the potential use of AHCY variants as a molecular biomarker. Considering the pleiotropic effects of AHCY, timely identification of AHCY deficiency may allow for targeted intervention to hinder the progression of related diseases. A systematic analysis of all potentially deficient AHCY variants would provide molecular tools for genetic counseling, as well as for improving the prediction, diagnosis, and clinical outcome of AHCY deficiencies.

The aggressiveness of neuroblastoma cancer subtypes seems to rely on AHCY functionality (Westermann et al., 2008; Chayka et al., 2015). This exciting result places AHCY inhibitors as potential agents to target metabolic dependencies in some cancer types. Supporting this notion, tumor-initiating cells from non-small-cell lung cancer (NSCLC) adenocarcinoma display metabolic reprograming and rely on methionine to fuel their tumorigenicity (Wang et al., 2019). Interestingly, pharmacological alterations of the methionine cycle, by blocking AHCY with a specific inhibitor, reduce the colony- and tumor-forming abilities of tumor-initiating cells (Wang et al., 2019). Understanding the basis of AHCY-directed metabolic alterations is therefore essential to gain insights into the sensitivity of specific cancer types to therapeutic interventions based on AHCY inhibition and is a prerequisite for developing effective therapies.

AUTHOR CONTRIBUTIONS

SA wrote the manuscript and created the figures, together with PV and LD. All authors contributed to the article and approved the submitted version.

FUNDING

LD was supported by grants from the Spanish of Economy, Industry and Competitiveness (MEIC) (PID2019-108322GB-100), "Fundación Vencer El Cancer" (VEC), the European Regional Development Fund (FEDER), and from AGAUR. SA was funded by Ramón y Cajal Program MICIU-FSE (RYC2018-025002-I) and ISCIII-FEDER (PI19/01814). We acknowledge support of the Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya.

ACKNOWLEDGMENTS

We thank Isabel Espejo for helpful discussions and V. A. Raker for scientific editing.

REFERENCES

Aksamit, R. R., Buggy, J. J., and Bauer, C. E. (1995). Expression of rat liver AdoHcy hydrolase in a *Rhodobacter capsulatus* ahcY mutant restores pigment formation and photosynthetic growth. *Biochem. Biophys. Res. Commun.* 207, 265–272. doi: 10.1006/bbrc.1995.1182

Aranda, S., Alcaine-Colet, A., Blanco, E., Borras, E., Caillot, C., Sabido, E., et al. (2019). Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci. Adv.* 5:eaav2448. doi: 10.1126/sciadv.aav2448

Aranda, S., Mas, G., and Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* 1:e1500737. doi: 10.1126/sciadv.1500737

Backlund, P. S. Jr., Carotti, D., and Cantoni, G. L. (1986). Effects of the S-adenosylhomocysteine hydrolase inhibitors 3-deazaadenosine and 3-deazaaristeromycin on RNA methylation and synthesis. *Eur. J. Biochem.* 160, 245–251. doi: 10.1111/j.1432-1033.1986.tb09963.x

Bader, J. P., Brown, N. R., Chiang, P. K., and Cantoni, G. L. (1978). 3-Deazaadenosine, an inhibitor of adenosylhomocysteine hydrolase, inhibits reproduction of rous sarcoma virus and transformation of chick embryo cells. *Virology* 89, 494–505. doi: 10.1016/0042-6822(78)90191-5

Baric, I., Cuk, M., Fumic, K., Vugrek, O., Allen, R. H., and Glenn, B. (2005). S-Adenosylhomocysteine hydrolase deficiency: a second patient, the younger brother of the index patient, and outcomes during therapy. *J. Inherit. Metab. Dis.* 28, 885–902. doi: 10.1007/s10545-005-0192-9

Baric, I., Fumic, K., Glenn, B., Cuk, M., Schulze, A., Finkelstein, J. D., et al. (2004). S-adenosylhomocysteine hydrolase deficiency in a human: a genetic disorder of methionine metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4234–4239. doi: 10.1073/pnas.0400658101

Bas, H., Cilingir, O., Tekin, N., Saylisoy, S., Durak Aras, B., Uzay, E., et al. (2020). A Turkish patient with novel AHCY variants and presumed diagnosis of S-adenosylhomocysteine hydrolase deficiency. *Am. J. Med. Genet. A* 182, 740–745. doi: 10.1002/ajmg.a.61489

Beluzic, L., Grbesa, I., Beluzic, R., Park, J. H., Kong, H. K., Kopjar, N., et al. (2018). Knock-down of AHCY and depletion of adenosine induces DNA damage and cell cycle arrest. *Sci. Rep.* 8:14012.

Bethin, K. E., Petrovic, N., and Ettinger, M. J. (1995). Identification of a major hepatic copper binding protein as S-adenosylhomocysteine hydrolase. *J. Biol. Chem.* 270, 20698–20702. doi: 10.1074/jbc.270.35.20698

Blanchart, A., Navis, A. C., Assaife-Lopes, N., Usoskin, D., Aranda, S., Sontheimer, J., et al. (2018). UHRF1 licensed self-renewal of active adult neural stem cells. *Stem Cells* 36, 1736–1751. doi: 10.1002/stem.2889

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. *Science* 330, 612–616. doi: 10.1126/science.1191078

Bray, M., Driscoll, J., and Huggins, J. W. (2000). Treatment of lethal Ebola virus infection in mice with a single dose of an S-adenosyl-L-homocysteine hydrolase inhibitor. *Antivir. Res.* 45, 135–147. doi: 10.1016/s0166-3542(00)00066-8

Bray, M., Raymond, J. L., Geisbert, T., and Baker, R. O. (2002). 3-deazaneplanocin a induces massively increased interferon-alpha production in Ebola virus-infected mice. *Antivir. Res.* 55, 151–159. doi: 10.1016/s0166-3542(02)00018-9

Brzezinski, K. (2020). S-adenosyl-l-homocysteine Hydrolase: a structural perspective on the enzyme with two rossmann-fold domains. *Biomolecules* 10:1682. doi: 10.3390/biom10121682

Brzezinski, K., Bujacz, G., and Jaskolski, M. (2008). Purification, crystallization and preliminary crystallographic studies of plant S-adenosyl-L-homocysteine hydrolase (*Lupinus luteus*). *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 64, 671–673. doi: 10.1107/s1744309108017703

Buist, N. R., Glenn, B., Vugrek, O., Wagner, C., Stabler, S., Allen, R. H., et al. (2006). S-adenosylhomocysteine hydrolase deficiency in a 26-year-old man. *J. Inherit. Metab. Dis.* 29, 538–545. doi: 10.1007/s10545-006-0240-0

Bujnicki, J. M., Prigge, S. T., Caridha, D., and Chiang, P. K. (2003). Structure, evolution, and inhibitor interaction of S-adenosyl-L-homocysteine hydrolase from *Plasmodium falciparum*. *Proteins* 52, 624–632. doi: 10.1002/prot.10446

Chayka, O., D'acunto, C. W., Middleton, O., Arab, M., and Sala, A. (2015). Identification and pharmacological inactivation of the MYCN gene network as a therapeutic strategy for neuroblastic tumor cells. *J. Biol. Chem.* 290, 2198–2212. doi: 10.1074/jbc.m114.624056

Chen, N. C., Yang, F., Capecci, L. M., Gu, Z., Schafer, A. I., Durante, W., et al. (2010). Regulation of homocysteine metabolism and methylation in human and mouse tissues. *FASEB J.* 24, 2804–2817. doi: 10.1096/fj.09-143651

Chiang, P. K. (1998). Biological effects of inhibitors of S-adenosylhomocysteine hydrolase. *Pharmacol. Ther.* 77, 115–134. doi: 10.1016/s0163-7258(97)00089-2

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., et al. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840. doi: 10.1126/science.1175371

Clarke, S. (2001). "S-adenosylmethionine-dependent methyltransferases," in *Homocysteine in Health and Disease*, eds R. Carmel, and D. W. Jackobsen, (Cambridge: Cambridge University Press).

Czlonkowska, A., Litwin, T., Dusek, P., Ferenci, P., Lutsenko, S., Medici, V., et al. (2018). Wilson disease. *Nat. Rev. Dis. Primers* 4:21.

Czyrko, J., Sliwiak, J., Imiolczyk, B., Gdaniec, Z., Jaskolski, M., and Brzezinski, K. (2018). Metal-cation regulation of enzyme dynamics is a key factor influencing the activity of S-adenosyl-L-homocysteine hydrolase from *Pseudomonas aeruginosa*. *Sci. Rep.* 8:11334.

Daelemans, D., Este, J. A., Witvrouw, M., Pannecouque, C., Jonckheere, H., and Aquaro, S. (1997). S-adenosylhomocysteine hydrolase inhibitors interfere with the replication of human immunodeficiency virus type 1 through inhibition of the LTR transactivation. *Mol. Pharmacol.* 52, 1157–1163. doi: 10.1124/mol.52.6.1157

Daelemans, D., Vandamme, A. M., Shuto, S., Matsuda, A., and De Clercq, E. (1998). Stereospecificity of 6'-C-neplanocin a analogues as inhibitors of S- adenosylhomocysteine hydrolase activity and human immunodeficiency virus replication. *Nucleosides Nucleotides* 17, 479–486. doi: 10.1080/07328319808005192

De Clercq, E. (2005). John montgomery's legacy: carbocyclic adenosine analogues as SAH hydrolase inhibitors with broad-spectrum antiviral activity. *Nucleosides Nucleotides Nucleic Acids* 24, 1395–1415. doi: 10.1080/15257770500265638

De Clercq, E. (2015). Curious discoveries in antiviral drug development: the role of serendipity. *Med. Res. Rev.* 35, 698–719. doi: 10.1002/med.21340

De Fazio, G., Alba, A. P., Vicente, M., and De Clercq, E. (1990). Antiviral activity of S-adenosylhomocysteine hydrolase inhibitors against plant viruses. *Antivir. Res.* 13, 219–226. doi: 10.1016/0166-3542(90)90067-h

Dickinson, M. E., Flenniken, A. M., Ji, X., Teboul, L., Wong, M. D., and White, J. K. (2016). High-throughput discovery of novel developmental phenotypes. *Nature* 537, 508–514.

Ducker, G. S., and Rabinowitz, J. D. (2017). One-carbon metabolism in health and disease. *Cell Metab.* 25, 27–42. doi: 10.1016/j.cmet.2016.08.009

Erdel, F., and Rippe, K. (2018). Formation of chromatin subcompartments by phase separation. *Biophys. J.* 114, 2262–2270. doi: 10.1016/j.bpj.2018.03.011

Espejo, I., Di Croce, L., and Aranda, S. (2020). The changing chromatome as a driver of disease: a panoramic view from different methodologies. *Bioessays* 42:e2000203.

Fernandez-Sanchez, M. E., Gonatopoulos-Pournatzis, T., Preston, G., Lawlor, M. A., and Cowling, V. H. (2009). S-adenosyl homocysteine hydrolase is required for Myc-induced mRNA cap methylation, protein synthesis, and cell proliferation. *Mol. Cell. Biol.* 29, 6182–6191. doi: 10.1128/mcb.00973-09

Fulnecek, J., Matyasek, R., Votruba, I., Holy, A., Krizova, K., and Kovarik, A. (2011). Inhibition of SAH-hydrolase activity during seed germination leads to deregulation of flowering genes and altered flower morphology in tobacco. *Mol. Genet. Genomics* 285, 225–236. doi: 10.1007/s00438-011-0601-8

Fumic, K., Beluzic, R., Cuk, M., Pavkov, T., Kloor, D., Baric, I., et al. (2007). Functional analysis of human S-adenosylhomocysteine hydrolase isoforms SAHH-2 and SAHH-3. *Eur. J. Hum. Genet.* 15, 347–351. doi: 10.1038/sj.ejhg.5201757

Furner, I. J., Sheikh, M. A., and Collett, C. E. (1998). Gene silencing and homology-dependent gene silencing in Arabidopsis: genetic modifiers and DNA methylation. *Genetics* 149, 651–662.

Fustin, J. M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., et al. (2013). RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* 155, 793–806. doi: 10.1016/j.cell.2013.10.026

Fustin, J. M., Ye, S., Rakers, C., Kaneko, K., Fukumoto, K., Yamano, M., et al. (2020). Methylation deficiency disrupts biological rhythms from bacteria to humans. *Commun. Biol.* 3:211.

Grbesa, I., Kalo, A., Beluzic, R., Kovacevic, L., Lepur, A., Rokic, F., et al. (2017). Mutations in S-adenosylhomocysteine hydrolase (AHCY) affect its nucleocytoplasmic distribution and capability to interact with S-adenosylhomocysteine hydrolase-like 1 protein. *Eur. J. Cell Biol.* 96, 579–590. doi: 10.1016/j.ejcb.2017.05.002

Greco, C. M., Cervantes, M., Fustin, J. M., Ito, K., Ceglia, N., Samad, M., et al. (2020). S-adenosyl-l-homocysteine hydrolase links methionine metabolism to the circadian clock and chromatin remodeling. *Sci. Adv.* 6:eabc5629. doi: 10.1126/sciadv.abc5629

Grudzinska Pechhacker, M. K., Di Scipio, M., Vig, A., Tumber, A., Roslin, N., Tavares, E., et al. (2020). CRB1-related retinopathy overlapping the ocular phenotype of S-adenosylhomocysteine hydrolase deficiency. *Ophthalmic Genet.* 41, 457–464. doi: 10.1080/13816810.2020.1790013

Honzik, T., Magner, M., Krijt, J., Sokolova, J., Vugrek, O., Beluzic, R., et al. (2012). Clinical picture of S-adenosylhomocysteine hydrolase deficiency resembles phosphomannomutase 2 deficiency. *Mol. Genet. Metab.* 107, 611–613. doi: 10.1016/j.ymgme.2012.08.014

Hu, Y., Komoto, J., Huang, Y., Gomi, T., Ogawa, H., Takata, Y., et al. (1999). Crystal structure of S-adenosylhomocysteine hydrolase from rat liver. *Biochemistry* 38, 8323–8333. doi: 10.1021/bi990332k

Huang, H., Luo, Z., Qi, S., Huang, J., Xu, P., Wang, X., et al. (2018). Landscape of the regulatory elements for lysine 2-hydroxyisobutyrylation pathway. *Cell Res.* 28, 111–125. doi: 10.1038/cr.2017.149

Huggins, J., Zhang, Z. X., and Bray, M. (1999). Antiviral drug therapy of filovirus infections: s-adenosylhomocysteine hydrolase inhibitors inhibit Ebola virus in vitro and in a lethal mouse model. *J. Infect. Dis.* 179(Suppl. 1), S240–S247.

Jordan, N. D., West, J. P., Bottley, A., Sheikh, M., and Furner, I. (2007). Transcript profiling of the hypomethylated hog1 mutant of Arabidopsis. *Plant Mol. Biol.* 65, 571–586. doi: 10.1007/s11103-007-9221-4

Judkins, A. J., Macqueen, B. C., Christensen, R. D., Comstock, J., Mao, R., and Flores-Daboub, J. (2018). Non-immune hydrops, hypotonia, encephalopathy, and liver failure with novel compound heterozygous AHCY mutations. *Neonatology* 114, 337–340. doi: 10.1159/000489292

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202.

Kim, J. H., Kim, J. H., Kim, S. C., Yi, Y. S., Yang, W. S., Yang, Y., et al. (2013). Adenosine dialdehyde suppresses MMP-9-mediated invasion of cancer cells by blocking the Ras/Raf-1/ERK/AP-1 signaling pathway. *Biochem. Pharmacol.* 86, 1285–1300. doi: 10.1016/j.bcp.2013.08.022

Kloor, D., Ludtke, A., Stoeva, S., and Osswald, H. (2003). Adenosine binding sites at S-adenosylhomocysteine hydrolase are controlled by the NAD+/NADH ratio of the enzyme. *Biochem. Pharmacol.* 66, 2117–2123. doi: 10.1016/s0006-2952(03)00581-1

Komoto, J., Huang, Y., Gomi, T., Ogawa, H., Takata, Y., Fujioka, M., et al. (2000). Effects of site-directed mutagenesis on structure and function of recombinant rat liver S-adenosylhomocysteine hydrolase. Crystal structure of D244E mutant enzyme. *J. Biol. Chem.* 275, 32147–32156. doi: 10.1074/jbc.m003725200

Koronowski, K. B., Greco, C. M., Huang, H., Kim, J.-K., Fribourgh, J. L., Crosby, P., et al. (2021). Ketogenesis impact on liver metabolism revealed by proteomics of lysine β-hydroxybutyrylation. *bioRxiv [Preprint]* doi: 10.1101/2021.01.21.427645

Kusakabe, Y., Ishihara, M., Umeda, T., Kuroda, D., Nakanishi, M., Kitade, Y., et al. (2015). Structural insights into the reaction mechanism of S-adenosyl-L-homocysteine hydrolase. *Sci. Rep.* 5:16641.

Leal, J. F., Ferrer, I., Blanco-Aparicio, C., Hernandez-Losa, J., Ramon, Y. C. S., Carnero, A., et al. (2008). S-adenosylhomocysteine hydrolase downregulation contributes to tumorigenesis. *Carcinogenesis* 29, 2089–2095. doi: 10.1093/carcin/bgn198

Lee, S., Doxey, A. C., Mcconkey, B. J., and Moffatt, B. A. (2012). Nuclear targeting of methyl-recycling enzymes in *Arabidopsis thaliana* is mediated by specific protein interactions. *Mol. Plant* 5, 231–248. doi: 10.1093/mp/ssr083

Leyssen, P., De Clercq, E., and Neyts, J. (2008). Molecular strategies to inhibit the replication of RNA viruses. *Antivir. Res.* 78, 9–25. doi: 10.1016/j.antiviral.2008.01.004

Li, M., Li, Y., Chen, J., Wei, W., Pan, X., Liu, J., et al. (2007). Copper ions inhibit S-adenosylhomocysteine hydrolase by causing dissociation of NAD+ cofactor. *Biochemistry* 46, 11451–11458. doi: 10.1021/bi700395d

Li, X., Egervari, G., Wang, Y., Berger, S. L., and Lu, Z. (2018). Regulation of chromatin and gene expression by metabolic enzymes and metabolites. *Nat. Rev. Mol. Cell Biol.* 19, 563–578. doi: 10.1038/s41580-018-0029-7

Li, X., Huang, L., Hong, Y., Zhang, Y., Liu, S., Li, D., et al. (2015). Co-silencing of tomato S-adenosylhomocysteine hydrolase genes confers increased immunity against *Pseudomonas syringae* pv. tomato DC3000 and enhanced tolerance to drought stress. *Front. Plant Sci.* 6:717.

Li, Y., Chen, J., Liu, J., Yang, X., and Wang, K. (2004). Binding of Cu2+ to S-adenosyl-L-homocysteine hydrolase. *J. Inorg. Biochem.* 98, 977–983. doi: 10.1016/j.jinorgbio.2004.02.013

Liu, H., and Kiledjian, M. (2006). Decapping the message: a beginning or an end. *Biochem. Soc. Trans.* 34, 35–38. doi: 10.1042/bst0340035

Masuoka, J., Hegenauer, J., Van Dyke, B. R., and Saltman, P. (1993). Intrinsic stoichiometric equilibrium constants for the binding of zinc(II) and copper(II) to the high affinity site of serum albumin. *J. Biol. Chem.* 268, 21533–21537. doi: 10.1016/s0021-9258(20)80574-2

Masuta, C., Tanaka, H., Uehara, K., Kuwata, S., Koiwai, A., and Noma, M. (1995). Broad resistance to plant viruses in transgenic plants conferred by antisense inhibition of a host gene essential in S-adenosylmethionine-dependent transmethylation reactions. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6117–6121. doi: 10.1073/pnas.92.13.6117

Matthews, R. P., Lorent, K., Manoral-Mobias, R., Huang, Y., Gong, W., Murray, I. V., et al. (2009). TNFalpha-dependent hepatic steatosis and liver degeneration caused by mutation of zebrafish S-adenosylhomocysteine hydrolase. *Development* 136, 865–875. doi: 10.1242/dev.027565

Medici, V., Kieffer, D. A., Shibata, N. M., Chima, H., Kim, K., Canovas, A., et al. (2016). Wilson disease: epigenetic effects of choline supplementation on phenotype and clinical course in a mouse model. *Epigenetics* 11, 804–818. doi: 10.1080/15592294.2016.1231289

Medici, V., Shibata, N. M., Kharbanda, K. K., Islam, M. S., Keen, C. L., Kim, K., et al. (2014). Maternal choline modifies fetal liver copper, gene expression, DNA methylation, and neonatal growth in the tx-j mouse model of Wilson disease. *Epigenetics* 9, 286–296. doi: 10.4161/epi.27110

Medici, V., Shibata, N. M., Kharbanda, K. K., Lasalle, J. M., Woods, R., Liu, S., et al. (2013). Wilson's disease: changes in methionine metabolism and inflammation affect global DNA methylation in early liver disease. *Hepatology* 57, 555–565. doi: 10.1002/hep.26047

Miller, M. W., Duhl, D. M., Winkes, B. M., Arredondo-Vega, F., Saxon, P. J., Wolff, G. L., et al. (1994). The mouse lethal nonagouti (a(x)) mutation deletes the S-adenosylhomocysteine hydrolase (Ahcy) gene. *EMBO J.* 13, 1806–1816. doi: 10.1002/j.1460-2075.1994.tb06449.x

Mull, L., Ebbs, M. L., and Bender, J. (2006). A histone methylation-dependent DNA methylation pathway is uniquely impaired by deficiency in Arabidopsis S-adenosylhomocysteine hydrolase. *Genetics* 174, 1161–1171. doi: 10.1534/genetics.106.063974

Mushegian, A. R., Garey, J. R., Martin, J., and Liu, L. X. (1998). Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res.* 8, 590–598. doi: 10.1101/gr.8.6.590

Novak, E. M., Halley, N. S., Gimenez, T. M., Rangel-Santos, A., Azambuja, A. M., Brumatti, M., et al. (2016). BLM germline and somatic PKMYT1 and AHCY mutations: genetic variations beyond MYCN and prognosis in neuroblastoma. *Med. Hypotheses* 97, 22–25. doi: 10.1016/j.mehy.2016.10.008

Ouyang, B., Fei, Z., Joung, J. G., Kolenovsky, A., Koh, C., Nowak, J., et al. (2012). Transcriptome profiling and methyl homeostasis of an Arabidopsis mutant deficient in S-adenosylhomocysteine hydrolase1 (SAHH1). *Plant Mol. Biol.* 79, 315–331. doi: 10.1007/s11103-012-9914-1

Palmer, J. L., and Abeles, R. H. (1979). The mechanism of action of S-adenosylhomocysteinase. *J. Biol. Chem.* 254, 1217–1226. doi: 10.1016/s0021-9258(17)34190-x

Park, S. J., Kong, H. K., Kim, Y. S., Lee, Y. S., and Park, J. H. (2015). Inhibition of S-adenosylhomocysteine hydrolase decreases cell mobility and cell proliferation through cell cycle arrest. *Am. J. Cancer Res.* 5, 2127–2138.

Parkhitko, A. A., Binari, R., Zhang, N., Asara, J. M., Demontis, F., and Perrimon, N. (2016). Tissue-specific down-regulation of S-adenosyl-homocysteine via suppression of dAhcyL1/dAhcyL2 extends health span and life span in Drosophila. *Genes Dev.* 30, 1409–1422. doi: 10.1101/gad.282277.116

Petrossian, T. C., and Clarke, S. G. (2011). Uncovering the human methyltransferasome. *Mol. Cell. Proteomics* 10:M110000976.

Ponnaluri, V. K. C., Esteve, P. O., Ruse, C. I., and Pradhan, S. (2018). S-adenosylhomocysteine hydrolase participates in DNA methylation inheritance. *J. Mol. Biol.* 430, 2051–2065. doi: 10.1016/j.jmb.2018.05.014

Porcelli, M., Cacciapuoti, G., Fusco, S., Iacomino, G., Gambacorta, A., De Rosa, M., et al. (1993). S-adenosylhomocysteine hydrolase from the thermophilic archaeon Sulfolobus solfataricus: purification, physico-chemical and immunological properties. *Biochim. Biophys. Acta* 1164, 179–188. doi: 10.1016/0167-4838(93)90246-n

Radomski, N., Barreto, G., Kaufmann, C., Yokoska, J., Mizumoto, K., and Dreyer, C. (2002). Interaction of S-adenosylhomocysteine hydrolase of *Xenopus laevis* with mRNA(guanine-7-)methyltransferase: implication on its nuclear compartmentalisation and on cap methylation of hnRNA. *Biochim. Biophys. Acta* 1590, 93–102. doi: 10.1016/s0167-4889(02)00205-7

Radomski, N., Kaufmann, C., and Dreyer, C. (1999). Nuclear accumulation of S-adenosylhomocysteine hydrolase in transcriptionally active cells during development of *Xenopus laevis*. *Mol. Biol. Cell* 10, 4283–4298. doi: 10.1091/mbc.10.12.4283

Reddy, M. C., Kuppan, G., Shetty, N. D., Owen, J. L., Ioerger, T. R., and Sacchettini, J. C. (2008). Crystal structures of *Mycobacterium tuberculosis* S-adenosyl-L-homocysteine hydrolase in ternary complex with substrate and inhibitors. *Protein Sci.* 17, 2134–2144. doi: 10.1110/ps.038125.108

Richon, V. M., Johnston, D., Sneeringer, C. J., Jin, L., Majer, C. R., Elliston, K., et al. (2011). Chemogenetic analysis of human protein methyltransferases. *Chem. Biol. Drug Des.* 78, 199–210. doi: 10.1111/j.1747-0285.2011.01135.x

Rocha, P. S., Sheikh, M., Melchiorre, R., Fagard, M., Boutet, S., Loach, R., et al. (2005). The Arabidopsis homology-dependent gene silencing1 gene codes for an S-adenosyl-L-homocysteine hydrolase required for DNA methylation-dependent gene silencing. *Plant Cell* 17, 404–417. doi: 10.1105/tpc.104.028332

Seo, H. C., and Ettinger, M. J. (1993a). Identification and purification of a self-associating copper-binding protein from mouse hepatic cytosols. *J. Biol. Chem.* 268, 1151–1159. doi: 10.1016/s0021-9258(18)54053-9

Seo, H. C., and Ettinger, M. J. (1993b). Purification and properties of a self-associating, 50-kDa copper-binding protein from brindled mouse livers. *J. Biol. Chem.* 268, 1160–1165. doi: 10.1016/s0021-9258(18)54054-0

Shin, Y. S., Jarhad, D. B., Jang, M. H., Kovacikova, K., Kim, G., Yoon, J. S., et al. (2020). Identification of 6'-beta-fluoro-homoaristeromycin as a potent inhibitor of chikungunya virus replication. *Eur. J. Med. Chem.* 187:111956. doi: 10.1016/j.ejmech.2019.111956

Snoeck, R., Andrei, G., Neyts, J., Schols, D., Cools, M., Balzarini, J., et al. (1993). Inhibitory activity of S-adenosylhomocysteine hydrolase inhibitors against human cytomegalovirus replication. *Antivir. Res.* 21, 197–216. doi: 10.1016/0166-3542(93)90028-h

Stender, S., Chakrabarti, R. S., Xing, C., Gotway, G., Cohen, J. C., and Hobbs, H. H. (2015). Adult-onset liver disease and hepatocellular carcinoma in S-adenosylhomocysteine hydrolase deficiency. *Mol. Genet. Metab.* 116, 269–274. doi: 10.1016/j.ymgme.2015.10.009

Stepkowski, T., Brzezinski, K., Legocki, A. B., Jaskolski, M., and Bena, G. (2005). Bayesian phylogenetic analysis reveals two-domain topology of S-adenosylhomocysteine hydrolase protein sequences. *Mol. Phylogenet. Evol.* 34, 15–28. doi: 10.1016/j.ympev.2004.09.008

Strauss, K. A., Ferreira, C., Bottiglieri, T., Zhao, X., Arning, E., Zhang, S., et al. (2015). Liver transplantation for treatment of severe S-adenosylhomocysteine hydrolase deficiency. *Mol. Genet. Metab.* 116, 44–52. doi: 10.1016/j.ymgme.2015.06.005

Tanaka, H., Masuta, C., Uehara, K., Kataoka, J., Koiwai, A., and Noma, M. (1997). Morphological changes and hypomethylation of DNA in transgenic tobacco expressing antisense RNA of the S-adenosyl-L-homocysteine hydrolase gene. *Plant Mol. Biol.* 35, 981–986.

Tanaka, N., Nakanishi, M., Kusakabe, Y., Shiraiwa, K., Yabe, S., Ito, Y., et al. (2004). Crystal structure of S-adenosyl-L-homocysteine hydrolase from the human malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 343, 1007–1017. doi: 10.1016/j.jmb.2004.08.104

Tehlivets, O., Malanovic, N., Visram, M., Pavkov-Keller, T., and Keller, W. (2013). S-adenosyl-L-homocysteine hydrolase and methylation disorders: yeast as a model system. *Biochim. Biophys. Acta* 1832, 204–215. doi: 10.1016/j.bbadis.2012.09.007

Turner, M. A., Yuan, C. S., Borchardt, R. T., Hershfield, M. S., Smith, G. D., and Howell, P. L. (1998). Structure determination of selenomethionyl S-adenosylhomocysteine hydrolase using data at a single wavelength. *Nat. Struct. Biol.* 5, 369–376. doi: 10.1038/nsb0598-369

Vivante, A., Ityel, H., Pode-Shakked, B., Chen, J., Shril, S., Van Der Ven, A. T., et al. (2017). Exome sequencing in Jewish and Arab patients with rhabdomyolysis reveals single-gene etiology in 43% of cases. *Pediatr. Nephrol.* 32, 2273–2282. doi: 10.1007/s00467-017-3755-8

Vugrek, O., Beluzic, R., Nakic, N., and Mudd, S. H. (2009). S-adenosylhomocysteine hydrolase (AHCY) deficiency: two novel mutations with lethal outcome. *Hum. Mutat.* 30, E555–E565.

Walsh, C. T., and Moore, B. S. (2019). Enzymatic cascade reactions in biosynthesis. *Angew. Chem. Int. Ed. Engl.* 58, 6846–6879. doi: 10.1002/anie.201807844

Wang, M., Borchardt, R. T., Schowen, R. L., and Kuczera, K. (2005). Domain motions and the open-to-closed conformational transition of an enzyme: a normal mode analysis of S-adenosyl-L-homocysteine hydrolase. *Biochemistry* 44, 7228–7239. doi: 10.1021/bi047524m

Wang, Y., Kavran, J. M., Chen, Z., Karukurichi, K. R., Leahy, D. J., and Cole, P. A. (2014). Regulation of S-adenosylhomocysteine hydrolase by lysine acetylation. *J. Biol. Chem.* 289, 31361–31372. doi: 10.1074/jbc.m114.597153

Wang, Z., Yip, L. Y., Lee, J. H. J., Wu, Z., Chew, H. Y., and Chong, P. K. W. (2019). Methionine is a metabolic dependency of tumor-initiating cells. *Nat. Med.* 25, 825–837.

Westermann, F., Muth, D., Benner, A., Bauer, T., Henrich, K. O., and Oberthuer, A. (2008). Distinct transcriptional MYCN/c-MYC activities are associated with spontaneous regression or malignant progression in neuroblastomas. *Genome Biol.* 9:R150.

Wu, X., Li, F., Kolenovsky, A., Caplan, A., Cui, Y., and Cutler, A. (2009). A mutant deficient in S-adenosylhomocysteine hydrolase in Arabidopsis shows defects in root-hair development this paper is one of a selection of papers published in a special issue from the national research council of canada – plant biotechnology institute. *Botany* 87, 571–584. doi: 10.1139/b08-124

Yang, X., Hu, Y., Yin, D. H., Turner, M. A., Wang, M., Borchardt, R. T., et al. (2003). Catalytic strategy of S-adenosyl-L-homocysteine hydrolase: transition-state stabilization and the avoidance of abortive reactions. *Biochemistry* 42, 1900–1909. doi: 10.1021/bi0262350

Ye, C., and Tu, B. P. (2018). Sink into the epigenome: histones as repositories that influence cellular metabolism. *Trends Endocrinol. Metab.* 29, 626–637. doi: 10.1016/j.tem.2018.06.002

Yee, N. S., Lorent, K., and Pack, M. (2005). Exocrine pancreas development in zebrafish. *Dev. Biol.* 284, 84–101. doi: 10.1016/j.ydbio.2005.04.035

Yin, D., Yang, X., Hu, Y., Kuczera, K., Schowen, R. L., Borchardt, R. T., et al. (2000). Substrate binding stabilizes S-adenosylhomocysteine hydrolase in a closed conformation. *Biochemistry* 39, 9811–9818. doi: 10.1021/bi000595a

Yoon, J. S., Kim, G., Jarhad, D. B., Kim, H. R., Shin, Y. S., and Qu, S. (2019). Design, synthesis, and Anti-RNA virus activity of 6'-fluorinated-aristeromycin analogues. *J. Med. Chem.* 62, 6346–6362. doi: 10.1021/acs.jmedchem.9b00781

Zhang, J., and Zheng, Y. G. (2016). SAM/SAH Analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chem. Biol.* 11, 583–597. doi: 10.1021/acschembio.5b00812

Zhao, J. N., Wang, Y., Zhao, B. S., and Chen, L. L. (2009). Analyzing S-adenosylhomocysteine hydrolase gene sequences in deuterostome genomes. *J. Biomol. Struct. Dyn.* 27, 371–380. doi: 10.1080/07391102.2009.10507323

Zhao, Y., and Garcia, B. A. (2015). Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.* 7:a025064. doi: 10.1101/cshperspect.a025064

Zhou, J., Yang, L., Zhong, T., Mueller, M., Men, Y., and Zhang, N. (2015). H19 lncRNA alters DNA methylation genome wide by regulating S-adenosylhomocysteine hydrolase. *Nat. Commun.* 6:10221.

Zhu, Q., Cheng, X., Cheng, Y., Chen, J., Xu, H., and Gao, Y. (2020). O-GlcNAcylation regulates the methionine cycle to promote pluripotency of stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 117, 7755–7763. doi: 10.1073/pnas.1915582117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

REVIEW
published: 20 April 2021
doi: 10.3389/fcell.2021.654915

Cell Fate Decisions in the Wake of Histone H3 Deposition

Reuben Franklin, Jernej Murn and Sihem Cheloufi**

Department of Biochemistry, Stem Cell Center, University of California, Riverside, Riverside, CA, United States

An expanding repertoire of histone variants and specialized histone chaperone partners showcases the versatility of nucleosome assembly during different cellular processes. Recent research has suggested an integral role of nucleosome assembly pathways in both maintaining cell identity and influencing cell fate decisions during development and normal homeostasis. Mutations and altered expression profiles of histones and corresponding histone chaperone partners are associated with developmental defects and cancer. Here, we discuss the spatiotemporal deposition mechanisms of the Histone H3 variants and their influence on mammalian cell fate during development. We focus on H3 given its profound effect on nucleosome stability and its recently characterized deposition pathways. We propose that differences in deposition of H3 variants are largely dependent on the phase of the cell cycle and cellular potency but are also affected by cellular stress and changes in cell fate. We also discuss the utility of modern technologies in dissecting the spatiotemporal control of H3 variant deposition, and how this could shed light on the mechanisms of cell identity maintenance and lineage commitment. The current knowledge and future studies will help us better understand how organisms employ nucleosome dynamics in health, disease, and aging. Ultimately, these pathways can be manipulated to induce cell fate change in a therapeutic setting depending on the cellular context.

Keywords: chromatin, histone H3, nucleosome diversity, histone chaperone, cell cycle, reprogramming, development, cellular plasticity

OPEN ACCESS

Edited by:
Justin Brumbaugh, University of Colorado Boulder, United States

Reviewed by:
Andrew P. Hutchins, Southern University of Science and Technology, China
Genevieve Almouzni, UMR 3664 Dynamique du noyau, France

****Correspondence:***
Jernej Murn
jernej.murn@ucr.edu
Sihem Cheloufi
sihem.cheloufi@ucr.edu

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *17 January 2021*
Accepted: *18 March 2021*
Published: *20 April 2021*

Citation:
Franklin R, Murn J and Cheloufi S (2021) Cell Fate Decisions in the Wake of Histone H3 Deposition. Front. Cell Dev. Biol. 9:654915. doi: 10.3389/fcell.2021.654915

INTRODUCTION

The Nucleosome and the Histone H3 Family

Cell fate decisions are central to development, normal homeostasis, and responding to infections, injury, and aging. During these processes, stem cells sustain the ability to self-renew and differentiate. These stem cell properties are tightly controlled by signaling pathways that orchestrate complex transcriptional and posttranscriptional layers of gene regulation. The structural foundation of these cell type-specific transcriptional programs is determined by DNA-protein-RNA complexes within the nuclear space. In 1879, Walther Flemming first described this complex structure in mitotic salamander cells, terming it "chromatin" from the Greek word chroma, referring to the color affinity of the intensely stained nuclear content. Almost a century later, X-ray diffraction patterns of chromatin by Maurice Wilkins, Vittorio Luzzati, and Aaron Klug suggested a repeating building unit and that histones are involved in packaging DNA (Luzzati and Nicolaieff, 1959; Wilkins et al., 1959). Indeed, subsequent enzymatic digestion of chromatin isolated from rat liver cells using DNA nuclease revealed multiples of 200 base pair DNA fragments (Hewish and Burgoyne, 1973). Electron micrographs of chromatin fibers also revealed

that these repeating units, known as nucleosomes, were composed of DNA wrapped around histone molecules (Kornberg, 1974; Olins and Olins, 1974; Oudet et al., 1975). Two decades later, Karoline Luger's structural studies determined that the core of the nucleosome consists of 147 base pairs of DNA wrapped around an octamer of histones assembled from a tetramer of histone H3:H4 dimers that is flanked by two Histone H2A:H2B dimers (Luger et al., 1997). Since then, nucleosomes have become known as highly dynamic hubs of DNA-protein-RNA interactions, that not only allow for cell-type specific gene regulation, but for higher order chromatin organization important to many cellular processes.

The histones within the core nucleosome are interchangeable with different isoforms, identified as histone variants by Franklin and Zweidler (1977). While the repertoire of histones continues to expand, the Histone H3 family in particular has been in the spotlight of chromatin and cellular plasticity research. H3 carries the majority of well characterized heritable posttranslational modifications (PTMs) known to date, evolved a centromere specific histone variant, has a pronounced effect on nucleosome stability compared to other histones and can act as an oncogene due to mutations within critical residues subject to PTMs (Filipescu et al., 2014). Moreover, the current research on H3 shows how profound this integral nucleosome component is to the regulation of chromatin states and cell identity (Filipescu et al., 2014; Loppin and Berger, 2020; Martire and Banaszynski, 2020). The histone H3 family is composed of 8 members, H3.1, H3.2, H3.3, CENPA, H3.4, H3.5, H3.X and H3.Y. While the latter 4 members are poorly characterized, the replicative H3.1/H3.2 variants and the non-replicative H3.3 and CENPA variants have received much attention.

H3 histone forms differ markedly in their gene structure, expression profiles, deposition mode and post translational modifications (Mendiratta et al., 2018; Martire and Banaszynski, 2020). H3.1 and H3.2 are found in multiple copies in the genome. In dividing cells, they are defined as replicative histones due to their S-phase specific expression and replication-dependent deposition, which allows for chromatin assembly in the wake of DNA synthesis when parental histones are diluted (**Figure 1A**; Mendiratta et al., 2018; Grover et al., 2018). The H3.3 variant differs from the H3.1 and H3.2 by only 5 and 4 amino acids, respectively. On the other hand, there are two H3.3 genes in mammals, H3f3a and H3f3b, that encode identical amino acid sequence but are different in their primary DNA sequence and are tightly regulated transcriptionally and post-transcriptionally in different cell types (Muhire et al., 2019). H3.3 genes are expressed throughout the cell cycle in dividing cells (**Figure 1**) and are highly abundant, if not the predominant H3, in non-dividing cells. Finally, CENPA, the centromere specific H3 variant, shares less than 51% sequence identity with the replicative histones and forms a highly compacted nucleosome core that is wrapped by only 121 base pairs of DNA. It is encoded by one gene expressed during G2 and mitosis in preparation for new CENPA incorporation in centromeres (**Figure 1B**; Jansen et al., 2007; Martire and Banaszynski, 2020). For recent evolutionary analysis of H3 variants and their role in development and disease, readers are referred to (Buschbeck and Hake, 2017; Loppin and Berger, 2020).

Overall, the positioning, modifications, and histone composition of nucleosomes can have profound effects on chromatin accessibility to transcription factors at actively transcribed (euchromatic) and repressed (heterochromatic) compartments, whose activity determines cell identity. Indeed, recent integration of different epigenomic maps, including higher order chromatin structures, nucleosome positioning, histone distribution and modifications during early development, and different cell fate change paradigms, demonstrates the complexity of spatiotemporal chromatin rearrangements (Eckersley-Maslin et al., 2018; Pérez-Palacios and Bourc'his, 2018; Fang et al., 2018; Ishiuchi et al., 2021). However, how nucleosome components and assembly pathways contribute to this regulation is still being dissected.

Histone Chaperone Roles in Nucleosome Dynamics and Beyond

Nucleosomes are diverse and dynamic. They can be shifted, assembled, or disassembled, and organized into different chromatin compartments through cooperation of histone chaperones, chromatin remodelers, and chromatin modifying factors (Dixon et al., 2012; Struhl and Segal, 2013). In particular, histone chaperones are the life partners of the nucleosome's core histones. They are involved in escorting histones from their synthesis, storage, and transport, to histone modification, deposition, eviction, and recycling in the nucleosome (Grover et al., 2018).

Originally, the term "molecular chaperone" was used by Ron Laskey who isolated and characterized the function of nucleoplasmin as the first histone chaperone using Xenopus egg extracts (Laskey et al., 1977, 1978). This was the proof of principle that histone chaperones are involved in nucleosome assembly by binding directly to histones, neutralizing their positive charges and preventing non-specific interactions and aggregates formed with DNA *in vitro* under physiological salt concentrations (Dilworth et al., 1987). This seminal discovery laid the groundwork for our knowledge today that nucleosome assembly involves a step wise transfer of H3:H4 and H2A:H2B dimers and a complex network of histone chaperone partners (Hammond et al., 2017). As a whole, histone chaperones have no consensus sequences or structural motifs, making the discovery of novel chaperones more challenging. They exhibit considerable differences perhaps due to specialized functions in (1) recognition of distinct histone variants, (2) dedicated activities in different DNA transactions: replication, transcription, repair, and recombination, (3) diverse complex formation in histone dependent or independent manners and (4) spatiotemporal requirements in different cell-types.

Histone H3 and its variants especially exhibit a highly complex and specialized histone chaperone network in addition to more general chaperone interactions as seen with other histones (**Figure 1**). However, studying the interplay between the different histone chaperone pathways has been challenging to disentangle in the context of cell fate transitions. Recent technological advances combining histone labeling, genetic

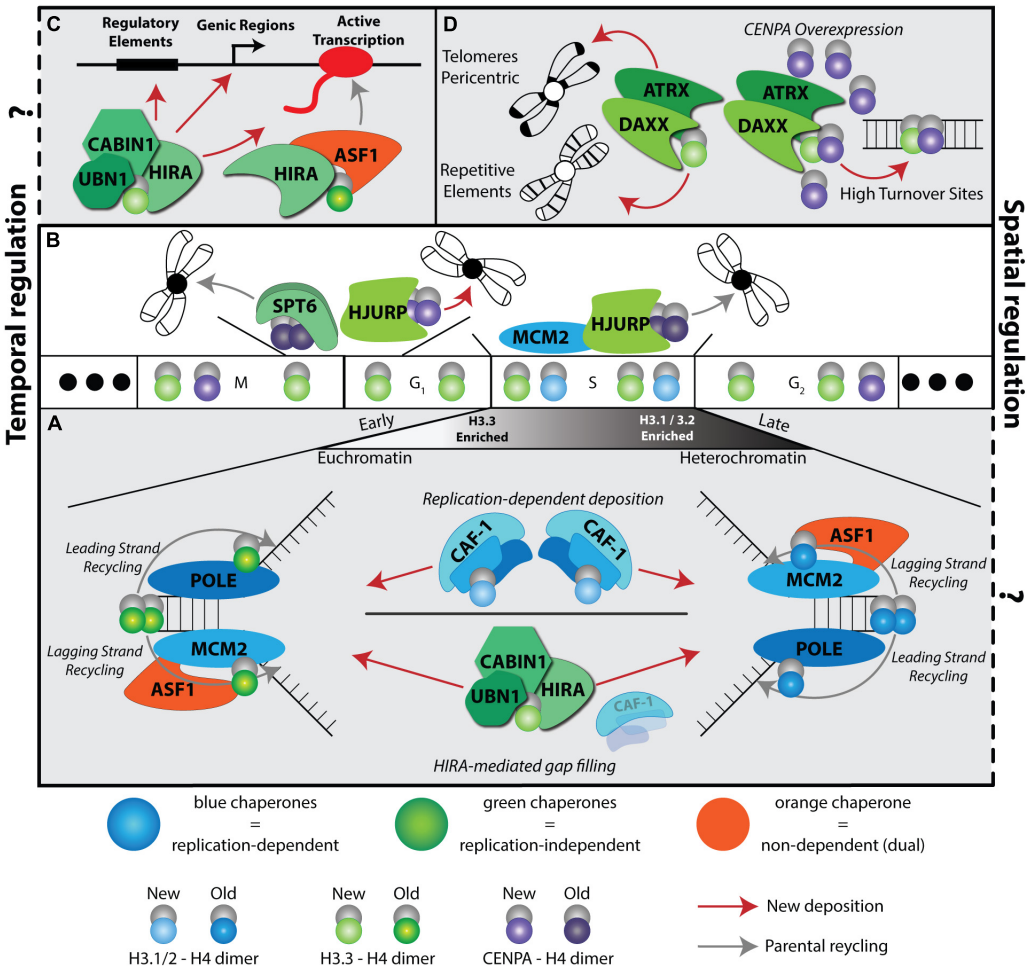


FIGURE 1 | A complex histone chaperone network regulates spatiotemporal H3 deposition. The temporal H3 expression and deposition pathways are illustrated using the cell cycle phases (G1, S, G2, M) as a centerpiece. Three solid dots on the ends indicate continuous cycling. Histones H3.1/2 (blue) are expressed during S phase, H3.3 (green) is expressed throughout the cell cycle and CENPA is expressed during G2 and mitosis. Histone deposition molecules are color coded: RD (blue), RI (green), Dual (orange). Recycling of old histones (dark shaded H3 molecules) and deposition of new histones (light shaded H3 molecules) are indicated by gray and red arrows, respectively. The spatial deposition is illustrated either in the context of heterochromatin or euchromatin compartments, transcription sites or repetitive elements. **(A)** CAF-1 trimer deposits new H3.1/2 during S phase. ASF1 and MCM2 cooperate to promote histone recycling to the lagging strand, while subunits POLE3-POLE4 recycle histones to the leading strand. Early replicating regions occur in euchromatin and are enriched with H3.3 histones, while late replicating regions occur in heterochromatic regions and are enriched with replicative histone H3.1/2. In the absence of CAF-1, HIRA deposits H3.3 at replication sites in a gap-filling mechanism. **(B)** CENPA is deposited in a spatiotemporal manner. HJURP deposits new CENPA at centromeres during late M and early G1 phases. SUPT6 recycles parental CENPA during transcription in late M and early G1 to prevent CENPA eviction at centromeres. HJURP cooperates with MCM2 during S phase to recycle parental CENPA at the centromere. **(C)** HIRA-mediated H3.3 deposition at a transcriptional unit. Two different HIRA complexes deposit new and parental H3.3. **(D)** DAXX-ATRX deposits H3.3 at telomeres, pericentric regions, and repetitive elements. When CENPA is overexpressed, DAXX-ATRX deposits heterotypic tetramers containing both H3.3 and CENPA at sites of high histone turnover.

engineering, epigenomics, high resolution microscopy, and structural and biochemical approaches in different contexts have started to shed light on understanding the role of nucleosome dynamics in cell fate decisions. In this review, we focus on H3 deposition pathways in the context of the cell cycle and how they relate to cell fate transitions during early development and several culture systems (**Figures 2**, **3** and **Table 1**). The roles of other histone variants, accompanying chaperones, chromatin remodelers and modifiers in cell fate transitions are reviewed in recent publications, including this special issue.

HISTONE H3 DEPOSITION PATHWAYS

Spatiotemporal Regulation of CAF-1, HIRA and DAXX Pathways

Early work on nucleosome assembly pathways demonstrated a specificity of histone H3 chaperones to assemble nucleosomes in a DNA replication dependent (RD) or independent (RI) manner (Almouzni and Méchali, 1988; Smith and Stillman, 1989; Ray-Gallet et al., 2002). This cell cycle determinant of histone chaperone pathways is further complicated by the specific

partnerships of histone chaperones with different variants, parental "old" histones versus newly synthesized histones and the deposition coordinates in the genome (**Figure 1**). Assembly of newly synthesized histones has been extensively studied and recent work is beginning to uncover the recycling mechanisms of old histones (Venkatesh and Workman, 2015; Serra-Cardona and Zhang, 2018). Moreover, some plasticity and redundancy across these pathways has been observed, especially in terms of handing over new histones, recycling old histones, or when some pathways are absent or compromised (placeholding/gap filling) (**Figures 1A,B**) (Dunleavy et al., 2011; Ray-Gallet et al., 2011; Schneiderman et al., 2012).

The first discovery of a RD histone chaperone was reported more than three decades ago from human cell extracts by the purification and characterization of the Chromatin Assembly Factor complex CAF-1, a trimeric subunit composed of p150, p60 and Rbbp4 (Smith and Stillman, 1989). To date, CAF-1 is the only known histone chaperone that loads newly synthesized H3.1/2:H4 dimers onto DNA in a RD manner (**Figure 1A**).

It took another decade to discover the Histone H3.3-specific chaperone HIRA (histone cell cycle regulator). Subsequently, HIRA was found to also function as a trimeric complex with UBN1 and CABIN subunits (Lorain et al., 1998; Magnaghi et al., 1998). The HIRA complex incorporates newly synthesized H3.3 in a RI manner and was initially associated with histone deposition at active sites of transcription (Ahmad and Henikoff, 2002; Ray-Gallet et al., 2002; **Figure 1C**).

Genome wide distribution of H3.3 deposition in mouse embryonic stem cells (mESCs) led to the discovery of an additional H3.3-specific histone chaperone death domain associated protein (DAXX) (Tang et al., 2004; Goldberg et al., 2010; Lewis et al., 2010). DAXX in complex with the chromatin remodeler ATRX specializes in the deposition of H3.3 at repetitive elements, including telomeres, pericentromeric DNA, and a subset of endogenous retroviral elements (Goldberg et al., 2010; Lewis et al., 2010; Elsässer et al., 2015; Hoelper et al., 2017; **Figure 1D**). Although, the cell cycle timing of DAXX mediated H3.3 deposition is unclear, considering that the deposition of newly synthesized CENPA on centromeres via the Holliday junction recognition protein (HJURP) histone chaperone occurs in late mitosis/early G1 phase (Dunleavy et al., 2009; **Figure 1B**), it is tempting to speculate that the H3.3 deposition at repetitive elements coincides with this timing.

During S phase, HJURP mediates parental CENPA recycling with MCM2, a subunit of the helicase complex, and H3 histones act as place holders until new CENPA is deposited (Dunleavy et al., 2011; Zasadzińska et al., 2018; **Figure 1B**). Notably, CENPA overexpression leads to heterotypic deposition with H3.3 by the histone chaperone DAXX, reinforcing the plasticity of the histone chaperone network (**Figure 1D**; Arimura et al., 2014; Lacoste et al., 2014).

The ASF1 Hub

The ASF1 histone chaperone, originally discovered in yeast (Tyler et al., 1999), serves dual RD and RI nucleosome assembly pathways as it functions in handing over newly synthesized H3.1:H4 and H3.3:H4 dimers to CAF-1 and HIRA complexes, respectively (Grover et al., 2018). In mammals, the ASF1 network is diversified by the emergence of two paralogs, ASF1A and ASF1B, with preferences for HIRA and CAF-1 p60, respectively (Tang et al., 2006; Abascal et al., 2013). Interestingly, recent work by Almouzni's group demonstrated that ASF1 coordinates with HIRA to distinguish between old and new histone incorporation during transcription. In this model, ASF1 participates in the HIRA complex to recycle old H3.3, while new H3.3 is deposited via a UBN1-containing complex (**Figure 1C**; Torné et al., 2020).

Depletion of both ASF1 paralogs showed ASF1 is important for histone recycling during replication. This recycling is accomplished in partnership with MCM2, a subunit of the helicase complex. Together, they promote the recycling of old H3:H4 dimers in a RD manner (Groth et al., 2007; Huang et al., 2015; **Figure 1A**). Furthermore, recent evidence indicates MCM2 promotes symmetric loading of parental H3:H4 during DNA replication by preventing biased parental histone loading on the leading strand (Petryk et al., 2018). Conversely, newly identified histone chaperones POLE3-POLE4, subunits of the leading strand polymerase POLE, are proposed to load parental H3:H4 on the leading strand (Bellelli et al., 2018; **Figure 1A**). Considering these intriguing findings, it will be interesting to probe the interplay between POLE subunits, MCM2, and ASF1 and how leading and lagging strand deposition is balanced during self-renewal or differentiation.

Genome wide distribution and high-resolution microscopy mapping of parental histones in human cells demonstrate that H3.1 and H3.3 associate with late and early replicating regions, respectively (Clément et al., 2018; Mendiratta et al., 2018; **Figure 1A**). Therefore, it is tempting to hypothesize that due to its preference for HIRA interaction, ASF1A could act by recycling H3.3 while ASF1B could participate in recycling H3.1/2:H4 at the replication fork. It would therefore be interesting to investigate how ASF1 paralogs could participate in loading parental H3.1/2:H4 and H3.3:H4 dimers depending on their associated partners, replication sites/timing and leading versus lagging strand preference.

Another histone chaperone discovered in yeast, Suppressor of Ty 6 (SPT6) (Kaplan et al., 2003) plays a role in recycling parental H3:H4 during transcription and ASF1 can fill in the nucleosome gaps in spt6 yeast mutants (Jeronimo et al., 2019). Interestingly, yeast has only one form of H3 that is closely related to the mammalian H3.3 variant (Talbert and Henikoff, 2010). Given that SPT6 has an important role in transcription elongation (Vos et al., 2018, 2020), the interplay of ASF1 and SPT6 in histone deposition and recycling during transcription may yield further insights into nucleosome dynamics during this process. While commonly cited as a H3 histone chaperone, as shown in yeast, it remains to be concretely determined if SPT6 functions in H3:H4 deposition in mammalian cells. However, in support of this hypothesis, a recent study proposes a role for the histone chaperone SPT6 as a recycling factor for CENPA, with evidence in Drosophila and HeLa cells (Bobkov et al., 2020; **Figure 1B**).

Altogether the recent advances in labeling and mapping spatiotemporal distribution of old and new histones, structural mechanisms of histone-chaperone recognition, how they interact directly with DNA templates, and mapping nucleosome positions during different DNA processes have deepened our understanding on how the cell uses nucleosome assembly to maintain or reprogram chromatin organization. While mechanistic studies are largely performed *in vitro* or in immortal human or mouse cell lines, this knowledge could provide the fundamental mechanisms at play for stem cell maintenance and lineage commitment during development and tissue homeostasis (see sections below).

H3 DEPOSITION IN DIFFERENT PHYSIOLOGICAL SETTINGS

How the nucleosome assembly pathways discussed above (**Figure 1**) modulate cell fate decisions and cell identity maintenance puzzled scientists for decades. The recent breakthroughs in the field coincide with a burst of technological advances and their relevant applications in studying culture models and developmental processes (**Figure 2**). A wide spectrum of phenotypes has been observed upon manipulation of RD and RI nucleosome assembly pathways that depend on spatiotemporal histone requirements, with RD pathways arguably more challenging to study due to their requirement in cellular proliferation and subsequent lethality. Here, we will describe some examples and discuss possible mechanisms along with future implications.

CULTURE MODELS

Culture models to study cellular differentiation and reprogramming are powerful platforms to explore the molecular mechanisms orchestrated by the histone variants-histone chaperone network because they provide an opportunity to study cell autonomous effects within specific lineages, with some systems being more homogenous compared to others and are compatible with biochemical approaches (**Figure 2A**). Here we describe stem cell-based systems that mimic normal development and that have proved useful in understanding H3 deposition pathways.

Embryonic Stem Cells

ESCs are the earliest embryonic cells that can be captured *in vitro* from the blastocyst and propagated without compromising their pluripotent potential (**Figure 2A**; Evans and Kaufman, 1981; Martin, 1981). A change in culture conditions and/or intrinsic factors can coax ESCs to interchange their potency levels and/or commit to different lineages (**Figure 2A**). For example, ESCs can be maintained in culture in various states of pluripotency reflecting naïve (ground) or primed developmental states. Human ESCs (hESCs) derived from blastocysts reflect an even later pluripotency state during mouse development akin to the epiblast stem cells (epiSCs) that can be derived from the post implantation mouse embryo (**Figure 2A**).

In contrast to the developmental arrests that have been observed in CAF-1 depleted embryos (Houlard et al., 2006), it has been more amenable to probe the function of H3 deposition pathways particularly in mESCs without compromising cellular viability. The loss of the Chaf1a and Chaf1b subunits of the CAF-1 complex in ESCs results in their reprogramming to an earlier embryonic cell state mimicking the two-cell stage of embryonic development (2C-like cells) (Ishiuchi et al., 2015, 2021). The conversion of ESCs to 2C-like cells is dependent on cell progression through S-phase and on the chromatin assembly activity of CAF-1, displaying similar molecular features to spontaneously derived 2C-like cells and 2-cell stage embryos. Although, the recent finding that H3.3 deposition resumes non-canonical distribution upon CAF-1 ablation in ESCs suggests that this 2C-like cell fate induction is in fact reflecting a transient or even earlier embryonic cell state (Ishiuchi et al., 2021; **Table 1**). It will be interesting to compare additional epigenetic features and transcriptomes during this developmental window in the CAF-1 mutants.

This induction of a permissive chromatin state upon CAF-1 loss in ESCs is consistent with the initial observations where CAF-1 loss in ESCs affects heterochromatin features specific to stem cells (Houlard et al., 2006). Indeed, recent work demonstrated that CAF-1 loss impairs ESC differentiation in an embryoid body assay (Cheng et al., 2019; **Figure 2A**). Interestingly, this defect was linked to failure of establishing H3K27me3 marks at pluripotency promoters upon differentiation through CAF-1-PCNA and CAF-1-Polycomb (PRC2) recruitment to the replication fork. In this context, Cheng et al. (2019) detected a reduced association of H3.1 and H3K27me3 with replicating chromatin. It would be interesting to test how H3.1/H3.3 ratio affects H3K27me3 establishment and how parental histone inheritance is influenced during this process, potentially conferring a resistance of CAF-1 ESCs to differentiation.

Parental and new histone distribution was examined at a single cell level using a Wnt3a-induced asymmetric ESC division model, demonstrating there is differential distribution of old and new canonical histones in the daughter cells (Ma et al., 2020). This suggests a specialized action of histone chaperones during asymmetric division. Additionally, in light of the new implication that MCM2 promotes symmetric cell division through RD histone recycling to the lagging strand (Bellelli et al., 2018; Petryk et al., 2018), possibly with ASF1 as seen in HeLa cells (Bellelli et al., 2018; Petryk et al., 2018; **Figure 1A**), it will be interesting to examine how perturbation of different histone chaperones in ESC self-renewal and differentiation affects histone distribution during replication. Of note, the involvement of CAF-1 and histone mutations has been previously proposed to play a role in asymmetric histone deposition during *C. elegans* development (Nakano et al., 2011). Whether similar mechanisms are conserved in mammals remains to be tested.

Contrary to the loss of CAF-1 in ESCs, perturbation of RI nucleosome assembly in ESCs does not alter gene expression

profiles nor compromise ESC identity under self-renewal conditions. However, H3.3 together with its partners DAXX and ATRX are involved in silencing repetitive elements in ESCs, including a subset of retroelements and telomeres (Goldberg et al., 2010; Lewis et al., 2010; Elsässer et al., 2015; Hoelper et al., 2017; **Figure 1D**). Remarkably, this effect is more pronounced in more naïve or hypomethylated ESC cultures reflecting an important role in early preimplantation development (He et al., 2015). Whether this effect is purely a function in safeguarding genome stability or fine tuning transcriptional programs co-opted by repetitive elements remains to be explored (Macfarlan et al., 2012). The physiological effect of RI nucleosome assembly pathway depletion is exacerbated upon differentiation of ESCs where lineage specific gene expression programs are perturbed.

Interestingly the loss of ASF1A, HIRA and H3.3 affect Histone H3 K27 methylation (H3K27me3) specifically at developmentally regulated genes (Banaszynski et al., 2013; Gehre et al., 2020; Gao et al., 2018; **Figure 2A** and **Table 1**). HIRA dependent deposition of H3.3 is proposed to establish bivalent marks in ESCs at developmentally regulated genes while ASF1A dependent disassembly of nucleosomes facilitates resolution of bivalent domains upon ESC differentiation. H3.3 loss in ESCs also reduces enhancer H3 acetylation marks including H3K27ac, H3K18ac, H3K64ac, and H3K122ac (Martire et al., 2019). H3K27ac, in particular, a mark known to coincide with active enhancers is stimulated by the phosphorylation of the serine 31 residue on the H3.3 tail in mESCs. H3.3 serine 31 (H3.3S31) is one of the amino acids unique to H3.3. Supplementing H3.3 KO mESCs with replicative histone H3.2, bearing an alanine 31 residue, cannot rescue the enhancer acetylation defect despite being deposited at these sites. Moreover, the loss of H3.3 in ESCs does not affect chromatin accessibility or the recruitment of p300 histone acetyltransferase at enhancer elements suggesting that the H3.3S31 residue is uniquely required downstream of HIRA mediated deposition for subsequent chromatin signaling pathways. Consistent with the loss of H3K27me3 or DNA methylation in ESCs, the reduced acetylation of the H3.3 KO is tolerated by ESCs under self-renewing conditions with no dramatic effect on gene expression (Martire et al., 2019). However, their differentiation triggers defects in chromatin accessibility and establishing active enhancer elements and subsequent activation of differentiation genes.

A recent systematic characterization of all four H3.3 specific residues in a Xenopus gastrulation model reinforces the essential role of H3.3S31 specific phosphorylation during this developmental process (Sitbon et al., 2020). Strikingly, the replacement of all three H3.3 residues that are required for specific RI chaperone interactions with their H3.2 replicative counterparts was compatible with normal gastrulation. It will be interesting to perform similar genetic analyses in the context of ESC differentiation.

Recent work interrogated the function of H3.3 lysine residues (K4 and K36) in ESCs (Gehre et al., 2020). Alanine substitutions of H3.3K4 and H3.3K36 did not compromise ESC self-renewal but perturbed lineage specific transcriptional programs and differentiation, albeit with varying degrees. H3.3K4, but not H3.3K36, mutant ESCs exhibited severe defects and resulted in reduced H3.3 deposition at regulatory elements, especially promoters, independently of the lysine charge. While wild type replicative histones share these same residues with H3.3 and are able to compensate and maintain normal nucleosome density around transcription start sites (TSS), this is not sufficient to maintain the correct chromatin state. This observation reinforces the importance of H3.3 specific residues. Interestingly, H3.3K4 mutation did not perturb H3.3 histone chaperone expression or binding but diminished the interactions with chromatin remodelers and increased RNA polymerase activity. The authors thus propose a role for K4 in maintaining H3.3 at regulatory elements through proper recruitment of remodelers and accurate transcriptional activity. This study highlights how histone chaperones act in concert with remodelers and accompanying PTM signals to regulate nucleosome dynamics.

Taken together, these H3.3 studies in ESCs and model organisms justify some of the needs to incorporate H3.3 at regulatory elements and highlight the relevance of unique and common H3 residues in regulating nucleosome dynamics and setting specialized chromatin environments post nucleosome assembly (**Figure 2** and **Table 1**).

Considering these findings, it is tempting to speculate that during mESC differentiation, RD assembly pathways play a passive role in diluting ESC identity and RI pathways play an active role in establishing new identity. However, discrepancies in the effect of manipulating these pathways between hESCs compared to mESCs still need to be resolved. For example, the loss of both HIRA and ASF1 compromise hESCs self-renewal (Gonzalez-Muñoz et al., 2014; Zhu et al., 2017). HIRA loss in hESCs results in downregulation of pluripotency factors, activation of various lineage markers and differentiation. Moreover, in hESCs, the HIRA complex is proposed to associate with a stem cell specific subunit PROHIBITIN that stabilizes distinct complexes and cooperates with HIRA to regulate the metabolic circuitry in hESCs through H3.3 deposition. Considering that hESCs resemble mouse epiSCs (**Figure 2A**), which reflect a more primed pluripotent cell state, it is possible that phenotypes similar to mESCs could arise when examined in more naïve hESCs (Brumbaugh et al., 2019). It would be exciting to probe histone exchange dynamics and histone chaperone networks during interconversion of these pluripotency states to build on the current study documenting the changes in histone modifications to shape chromatin environments (De Clerck et al., 2019).

Reprogramming and Transdifferentiation

Reprogramming and transdifferentiation platforms have proved valuable in revealing unprecedented physiological roles of nucleosome assembly pathways in somatic cells (**Figure 2A**). For example, probing the function of CAF-1 in the context of transcription factor mediated reprogramming of mouse embryonic fibroblasts to induced pluripotent stem cells (iPSCs) implicated its role in maintaining somatic cell identity. In this system, CAF-1 is proposed to act in part through its nucleosome assembly function by restricting access to pluripotency transcription factors (Cheloufi et al., 2015;

Cheloufi and Hochedlinger, 2017). Supporting the role of CAF-1 in reprogramming, CAF-1 depletion in mESCs facilitates the generation of cloned blastocysts using somatic cell nuclear transfer technology and transdifferentiation between different lineages (Cheloufi et al., 2015; Ishiuchi et al., 2015). Contrary to CAF-1, ASF1A loss inhibits reprogramming of human somatic cells to iPSCs (Gonzalez-Muñoz et al., 2014). In this system, ASF1A co-expression with pluripotent transcription factor OCT4 is sufficient to reprogram human adult dermal fibroblasts when exposed to the oocyte-specific paracrine growth factor GDF9. In this context, ASF1A is proposed to work by promoting acetylation of histone H3K56 and cooperating with OCT4 to activate the pluripotency transcriptional network. ASF1A acts upstream of CAF-1 as a donor of newly synthesized histones but its functions also overlap with other nucleosome assembly pathways (**Figures 1A,C**). Thus, this discrepancy in reprogramming phenotypes between CAF-1 and ASF1A can be purely dependent on the spatiotemporal requirement of histone deposition and/or histone chaperone-independent functions. The implication of ASF1A in cellular reprogramming stemmed from it being a maternally deposited factor in the oocyte cytoplasm. Similarly, the H3.3 histone variant proved to be an essential maternal factor for reprogramming and the development of fertilized, parthenogenetically derived and SCNT embryos (Wen D. et al., 2014) (**Figure 2A**). In this context, H3.3 plays an important role in nucleosome remodeling in either the parental pronuclei or the donor nucleus (**Figure 2A**). Consistent with a spatiotemporal requirement of histone deposition pathways in shaping cellular identity, a recent study demonstrates a dual role of HIRA mediated H3.3 deposition in maintaining somatic cell identity and establishing pluripotency during reprogramming (Fang et al., 2018). Thus, this global rearrangement of H3.3 deposition akin to the one observed during oogenesis and the early cleavage embryo represents an important mechanism in preparation for cell fate conversions (see preimplantation development & Ishiuchi et al., 2021). However, the interplay with other histone chaperone pathways remains to be determined especially in a setting where the cell cycle is required for cell fate switches.

In light of these observations, we propose that nucleosome pathways at different potency states during development can dictate cell identity maintenance versus cell fate commitment or reprogramming toward different lineages. This could be purely dependent on specific remodeling of histone variants distribution and cell cycle properties (**Figure 3**).

FROM GAMETOGENESIS TO EARLY EMBRYONIC DEVELOPMENT

Gametogenesis

The sperm and oocyte are highly specialized cell types that transmit both genetic and epigenetic information through generations (**Figure 2B**). During spermatogenesis, the genome undergoes a stepwise replacement of histones with transition proteins and ultimately protamines to form the highly condensed nucleus of the sperm (Raja and Renkawitz-Pohl, 2005; Torres-Flores and Hernández-Hernández, 2020). This process is thought to prevent DNA damage, confer better sperm quality, and reprogram the paternal nucleosomes in preparation for fertilization as protamine knockouts result in defective sperm and developmental arrest (Cho et al., 2001, 2003). The nuclear condensation within the sperm head is accompanied with complex PTMs of the disassembled histones and the newly deposited protamines which could potentially involve the action of different histone chaperones whose identity remains to be determined. However, despite the removal of nearly 90% of all histones in the sperm, CENPA is retained. Also, select nucleosomes at regulatory DNA elements retain H3.1/2 and H3.3 (Hammoud et al., 2009; Erkek et al., 2013; Das et al., 2017). The retention of nucleosomes containing specific histone variants and corresponding PTMs on the paternal genome is thought to be a mechanism for transmitting epigenetic information to the embryo (Champroux et al., 2018). Of note, profiling the accurate histone distribution in the sperm nucleus has proved to be technically challenging depending on the method used to purify mature sperm that have undergone proper histone replacement and chromatin digestion for histone pull downs (Yoshida et al., 2018).

Consistent with the histone retention in the mature sperm, genetic studies support these observations. To date, several mouse knockout and conditional alleles of the two H3.3 genes have been generated albeit with variable phenotypic consequences on the germline and embryonic development (see post-implantation development) possibly due to the different targeting strategies, genetic heterogeneity of the mouse strains as well as possible redundancy with testis specific H3 variants. For example, in a mixed C57BL/6 and 129 mouse background, H3f3a+/−; H3f3b−/− compound mutant with one remaining copy of the H3f3a gene are male sterile (Jang et al., 2015) while other studies reported that the surviving single H3f3a and H3f3b knockouts have variable levels of sterility (Couldrey et al., 1999; Bush et al., 2013; Tang et al., 2013, 2015; Yuen et al., 2014). Regardless of these differences, accumulating evidence supports a unique role of H3.3 in chromatin remodeling in the male germline.

The effect of H3.3 loss in the female germline is more debatable. In contrast to previous studies reporting female sterility of single H3.3 knockouts, H3f3a+/−; H3f3b−/− compound mutant females are viable and fertile (Bush et al., 2013; Jang et al., 2015; Tang et al., 2013, 2015; **Figure 2B**). This is surprising given that mature oocytes are devoid of replicative histones in their genome and that there is H3.3 redistribution during oogenesis in preparation for embryogenesis (Ishiuchi et al., 2021) (see preimplantation development). Furthermore, the requirement of H3.3 histone partners during gametogenesis warrant further investigations. Interestingly, Asf1b knockout mice are viable but have reduced reproductive capacity showing a more severe defect in females versus males (Messiaen et al., 2016). This study showed that ASF1B is specifically expressed in the female gonads during development and propose its role in regulating meiotic entry. In light of these findings and the proposed molecular function of ASF1 (see The ASF1 Hub), it is

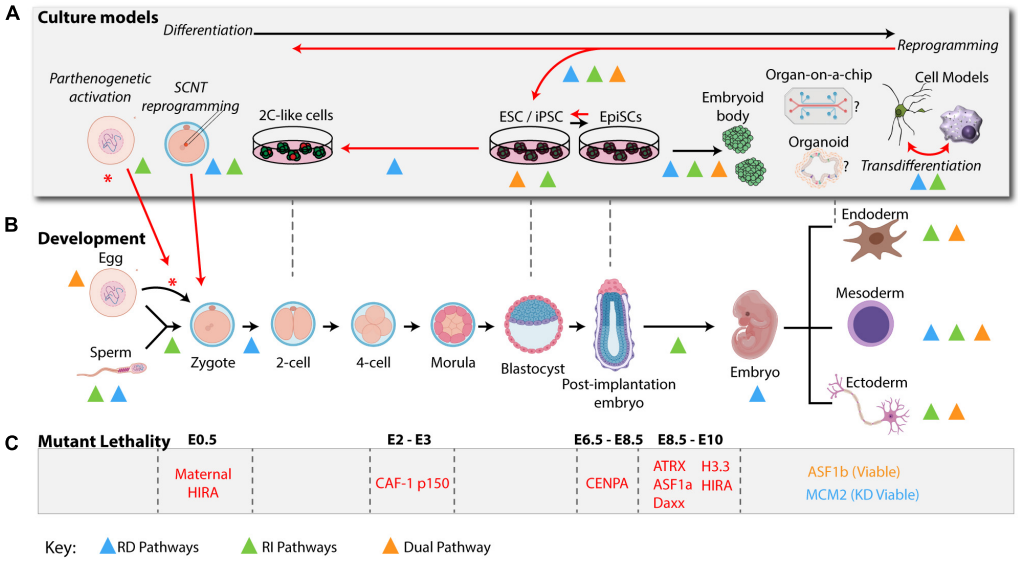


FIGURE 2 | H3 deposition pathways in a physiological setting: development and culture models. H3 Deposition pathways characterized in cell fate transitions during development or in cellular systems are indicated by colored triangles using the same color scheme used in **Figure 1** (RD: blue, RI: green, and dual: orange). **(A)** *Ex vivo* models are represented above their most relevant developmental stages as indicated in the developmental timeline (B) Black arrows indicate *differentiation* of the zygote from a higher potency to a lower potency. Red arrows indicate key cell plasticity pathways, including, zygote reprogramming following oocyte activation via *parthenogenesis* or *SCNT*, *reprogramming* of somatic cells to pluripotency and *transdifferentiation* of cells directly from one lineage to another. **(B)** A mouse developmental timeline, depicting the sperm and oocyte generating the zygote, early cleavage embryos, blastocyst, and early post-implantation embryo followed by specialized lineages discussed in the text (mouse development icons were created using BioRender software). **(C)** Summary of histone variants or histone chaperone mutant lethality in early embryo development. See **Table 1** for a summary of phenotypes and corresponding references.

tempting to speculate that the Asf1b paralog plays a specialized role in retaining H3.3 containing nucleosomes in a RD manner in the oocytes. On the other hand, given that it does not have an effect on the male germline, it might function independent of its histone chaperone role.

Notably, due to the early lethality of most histone chaperone mutants (see Preimplantation Development), the use of conditional knockouts and the development of *ex vivo* gametogenesis culture systems (Hamazaki et al., 2021) will be instrumental in resolving these limitations. Furthermore, this will shed light on the mechanisms and differences in spatiotemporal H3 re-distribution during spermatogenesis and oogenesis and how germline reprogramming may impact epigenetic inheritance.

Preimplantation Development

The fertilization between the sperm and oocyte gives rise to the zygote. In the zygote, both the paternal and maternal genomes undergo dramatic reprogramming events to give rise to the most plastic embryonic cell state known as "totipotency" (**Figures 2A,B**). If successful, the zygote will ultimately give rise to all cell types necessary for the development of an organism including the extraembryonic tissues. During this process, both the paternal and maternal pronuclei undergo major chromatin remodeling using maternally deposited factors in preparation for the first mitotic divisions and the transition to zygotic transcription (**Figure 2B**; Probst and Almouzni, 2011; Eckersley-Maslin et al., 2018). Accumulating evidence supports the idea that maternally deposited histones and histone chaperones are essential for reprogramming the zygote following fertilization. Indeed, the paternal genome is decondensed when incorporation of maternally deposited H3.3 replaces protamines, allowing for genome reprogramming (Loppin et al., 2005; Torres-Padilla et al., 2006). This is now known to be triggered by site specific phosphorylation of protamines by the RNA splicing factor SRPK1 which permits recruitment of nucleoplasmin (NPM2) and HIRA for protamine unloading and H3.3 deposition, respectively (Gou et al., 2020).

The manipulation of maternally deposited factors in oocytes followed by natural fertilization, parthenogenetic activation, or somatic cell nuclear transfer has been instrumental in understanding the mechanisms of RI incorporation of H3.3 onto parental chromatin (**Figure 2B** and **Table 1**; Lin et al., 2014; Wen D. et al., 2014). For example, deletion of HIRA in mouse oocytes results in inhibition of nucleosome assembly in the male genome and oocytes are unable to develop parthenogenetically. This study links HIRA-dependent H3.3 deposition to active transcription of ribosomal RNA in the zygote (Lin et al., 2014).

It will be interesting to probe the function of other H3.3 mediated site-specific histone chaperone pathways in the oocytes. Using ultra-low input native CHIP-seq a recent study generated a spatiotemporal map of H3.3 distribution during oogenesis, the zygote, and the early cleavage stage embryos (Ishiuchi et al., 2021). Interestingly, H3.3 undergoes a gradual global rearrangement during oogenesis forming a unique non-canonical pattern in the mature oocyte and zygote. At this developmental

window, H3.3 is more broadly distributed across the genome and exhibits some enrichment at heterochromatic regions. Remarkably, the non-canonical H3.3 distribution is similar between the maternal and paternal pronuclei in the zygote but is different from other post-mitotic cells, such as neurons. Interestingly, this unique chromatin incorporation of H3.3 in the oocyte and zygote coincides with previously reported distinct epigenetic features, including chromatin accessibility, histone marks and DNA methylation during preimplantation development (Eckersley-Maslin et al., 2018; Burton et al., 2020). For example, temporal regulation of histone methyltransferases (SUV39H1&H2) involved in the deposition of the H3K9me3 repressive mark post fertilization results in establishing an accessible and non-repressive constitutive heterochromatin in the zygote that ultimately matures in later stages to a compacted and repressive state. While Ishiuchi et al., 2021 did not report a correlation between H3.3 deposition and H3K9me3 profiles in the zygote, it is tempting to speculate that this early marking of heterochromatin is established as a consequence of the slight preferential loading of H3.3 on heterochromatin compartments in the zygote. H3.3 deposition within these domains could create a chromatin environment to recruit SUV39H1 similar to the mechanism proposed for PRC2 recruitment in ESCs at developmentally regulated genes (**Figure 2** and **Table 1**; Banaszynski et al., 2013). However, comparison of existing ChIP-seq data over constitutive heterochromatin domains may be challenging due to variable chromatin fragmentations, timing of the embryos, and considering multi-mapping reads at repetitive elements. Notably, as the zygote transitions to the 2-cell stage, the broad H3.3 distribution is reprogrammed to a more localized pattern reminiscent of the known canonical pattern initially described in ESCs. The reorganization and/or retention of H3.3 in the zygote occurs with the loading of replicative H3.1&H3.2 in a RD and transcription independent manner. Furthermore, it is regulated by CAF-1 as injecting a dominant negative form of the CAF-1 p150 subunit in the zygote reduces the canonical H3.3 rearrangement. Importantly this unique H3.3 rearrangement and deposition of replicative histones is essential for development as inhibition of maternally deposited CAF-1 results in developmental arrest at the 4-cell stage, consistent with previous reports ablating CAF-1 in the embryo (**Figure 2C** and **Table 1**; Houlard et al., 2006; Akiyama et al., 2011).

Considering the oocyte is devoid of replicative histones and that the expression of H3.1/2 peaks only after the embryo has undergone one cell division, it will be interesting to investigate how the non-canonical H3.3 distribution is propagated during the first round of replication, and how its recycling is regulated by histone chaperones in the context of replication timing. Therefore, it will be important to further characterize the histone chaperones required for H3.3 nucleosome exchange during these early cell divisions and how they are involved in preparing for zygotic genome activation and the establishment of heterochromatin and euchromatin domains. We think that the H3.3 broad distribution in the oocytes and zygote is possibly pre-programmed because of the need for fast RI eviction of protamines genome wide in the paternal pronucleus to ensure near-equal reprogramming of parental pronuclei in the zygote.

Post-implantation Development

In contrast to the severity of maternal or zygotic CAF-1 loss, mice lacking ASF1A survive to mid-gestation (**Figure 2C**; Hartford et al., 2011). Given that Asf1b knockout mice are viable, it would be interesting to probe phenotypic consequences of Asf1a/b double knockouts. Similarly, in mice lacking either H3.3 gene and their histone chaperone partners, HIRA or DAAX, embryos also progress to mid-gestation (**Figure 2C**; Michaelson et al., 1999; Roberts et al., 2002; Jang et al., 2015). Interestingly, H3f3a/b double knockout mice progress through early patterning of the embryo but are lethal 2 days after implantation (**Figure 2C**). Single knockouts of H3f3a and H3f3b, however, are reported with a spectrum of phenotypes depending on the study resulting in compromised viability and sterility (see gametogenesis). Similarly, CENP-A knockout mice are also lethal shortly after implantation (Howman et al., 2000). Of note, chromatin defects in these different mutants suggest a convergent mechanism between RD and RI pathways where heterochromatin structures are primarily affected leading to mitotic defects and developmental arrest. The relatively late phenotypic manifestations of RI nucleosome assembly pathways could be due to maternal deposition of mRNA and proteins of histones and histone chaperones, or redundancy of histone chaperone pathways. It would therefore be interesting to test the effects of their maternal contribution and the consequences of individual histone chaperone perturbation on histone deposition and chromatin accessibility in the early embryo.

While examining histone exchange and chromatin dynamics in the early embryo remain challenging, current developments in CRISPR CAS9 gene editing (Adli, 2018; Anzalone et al., 2020), chromatin profiling technologies such as ATAC-seq, CUT&RUN and CUT&Tag that rely on a small number of cells (Buenrostro et al., 2013; Corces et al., 2017; Kaya-Okur et al., 2019; Meers et al., 2019), single cell multi-omics (Pérez-Palacios and Bourc'his, 2018), and culture systems are instrumental in understanding the mechanisms of nucleosome assembly during these most plastic cell states during development.

LINEAGE SPECIFIC DIFFERENTIATION, A MIX OF IN VIVO AND CULTURE SYSTEMS

Given the early lethal phenotypes of the H3 deposition pathways during development (**Figures 2B,C**), it has been challenging to probe their function in normal homeostasis. However, recent work using lineage specific differentiation systems as well as cancer and injury models (Evano et al., 2020) has not only shed light on how some mechanisms described above (**Figure 1**) are at play, but also associates histone chaperones with histone deposition independent functions. We postulate that these differences could be intimately linked to the spatiotemporal expression of the H3 deposition machinery and the specific cell cycle properties (e.g.: short versus long) within different lineages (**Figure 3**). Evidence so far suggests an important role of both RD and RI pathways in lineage restriction and maintenance. For

TABLE 1 | Histone and histone chaperone roles in cell fate decisions.

Pathway	Protein	Function/ Phenotype	Mechanism	System	References
RD	H3.1/2	Male fertility	–	Human/mouse sperm	Hammoud et al., 2009; Bush et al., 2013; Erkek et al., 2013; Tang et al., 2013, 2015; Jang et al., 2015; Das et al., 2017
		Zygote/embryo development	Replication-associated deposition (?)	Mouse embryo	Ishiuchi et al., 2021
RD	CAF-1	Hematopoiesis	–	Mouse bone marrow	Volk et al., 2018
		CD8 + T-cell identity	HDAC and LSD1 CD4 silencing in conjunction with DNMT3a	Mouse T-cell	Ng et al., 2019
		Restricts plasticity	Chromatin accessibility and heterochromatin maintenance	Mouse MEF/HSPC reprogramming, ESCs, B-cell to Mac transition, MEF to neuron	Cheloufi et al., 2015; rev. in Cheloufi and Hochedlinger, 2017
		Differentiation	H3K27me3 mediated silencing	mESCs	Cheng et al., 2019
		Heterochromatin organization	LSD1?	mESCs/embryo	Houlard et al., 2006
		Zygote/embryo development	Replication-associated deposition	Mouse embryo	Ishiuchi et al., 2015, 2021
RD	MCM2	Decidualization	Cell cycle arrest	Mouse endometrial stromal cells	Kong et al., 2016
		Symmetric cell division	Symmetric histone recycling	mESC/HeLa	Petryk et al., 2018; Ma et al., 2020
		Adult SC deficiency	DNA damage/replication deficiency (?)	Mouse *in vivo*	Pruitt et al., 2007
		CD8 + T-cell identity	–	T-Cells	Ng et al., 2019
RI	H3.3	Male fertility	–	Human/mouse sperm	Hammoud et al., 2009; Erkek et al., 2013; Das et al., 2017
		Fertilization	Sperm decondensation		
		Muscle Differentiation	MyoD/MEF2 expression through H3.3 deposition	C2C12 to myotube	Yang et al., 2011, 2016
		Osteoblast conversion	H3.3 deposition	C2C12 to osteoblast	Song et al., 2012
		Pluripotency	PRC2 Recruitment	mESCs	Banaszynski et al., 2013; Schlesinger et al., 2017
		MEF reprogramming	HIRA-mediated H3.3 deposition at promoters	MEFs/iPSC	Fang et al., 2018
		Transdifferentiation	–	MEFs/iHPs	Fang et al., 2018
		Differentiation	–	mESC to neuron	Fang et al., 2018
		Macrophage activation	H3.3S31ph SETD2 recruitment	mouse macrophage	Armache et al., 2020
		Differentiation	H3.3S31ph-mediated p300 activity and enhancer acetylation	mESC	Martire et al., 2019
RI	DAXX	Restricts plasticity	ERV Accessibility	Mouse pancreas	Wasylishen et al., 2020; Wu et al., 2020
		Neuron activation	Daxx-phosphorylation H3.3 deposition.	Mouse CNS	Michod et al., 2012
RI	HIRA	Fertility	rRNA Transcription; H3.3-deposition	Mouse zygote	Lin et al., 2014
		Muscle Differentiation	MyoD/MEF2 expression through H3.3 deposition	C2C12 to myotube	Yang et al., 2011, 2016
		Osteoblast conversion	H3.3 deposition	C2C12 to osteoblast	Song et al., 2012
		Adult hematopoiesis	Chromatin accessibility	Mouse bone marrow	Chen et al., 2020
		Cardiac differentiation	H3.3 deposition	mESC differentiation to cardiac	Dilg et al., 2016; Saleh et al., 2018
		Myofiber maintenance	–	Muscle mouse myocytes *in vivo*	Valenzuela et al., 2017
		Pluripotency/self-renewal	PHB/H3.3 deposition chromatin promoters	hESC	Zhu et al., 2017
		Mesoderm development	–	Mouse embryo	Roberts et al., 2002
		Neurogenesis	SETD1A-mediated beta-catenin regulation	Mouse CNS	Li and Jiao, 2017; Jeanne et al., 2021
RI	CENPA	Gametogenesis	CENPA retention	Sperm and oocyte	Rev in. Das et al., 2017

(Continued)

TABLE 1 | Continued

Pathway	Protein	Function/ Phenotype	Mechanism	System	References
RI	HJURP	Cellular senescence	P53-dependent (?)	HDFs and HUVEC	Heo et al., 2013
		Cellular quiescence	centromere identity	human RPE1 and starfish oocyte	Swartz et al., 2019
Dual	ASF1A	Lineage differentiation	Histone eviction at promoters	Mouse EBs, neural differentiation	Gao et al., 2018
	ASF1A	Pluripotency/reprogramming	H3K56 acetylation	H9 ESCs, hADFs	Gonzalez-Muñoz et al., 2014
	ASF1A	Muscle differentiation	MyoD/MEF2 expression through H3.3 deposition	C2C12 to myotube	Yang et al., 2011
	ASF1A	Osteoblast conversion	H3.3 deposition	C2C12 to osteoblast	Song et al., 2012
	ASF1B	Cell proliferation	H3.3 recruitment	Human beta-cells	Paul et al., 2016
	ASF1B	Female fertility	–	Mouse oocyte	Messiaen et al., 2016

Summary of histone or histone chaperone roles in cell fate decisions in context of the systems studied and illuminated mechanisms.

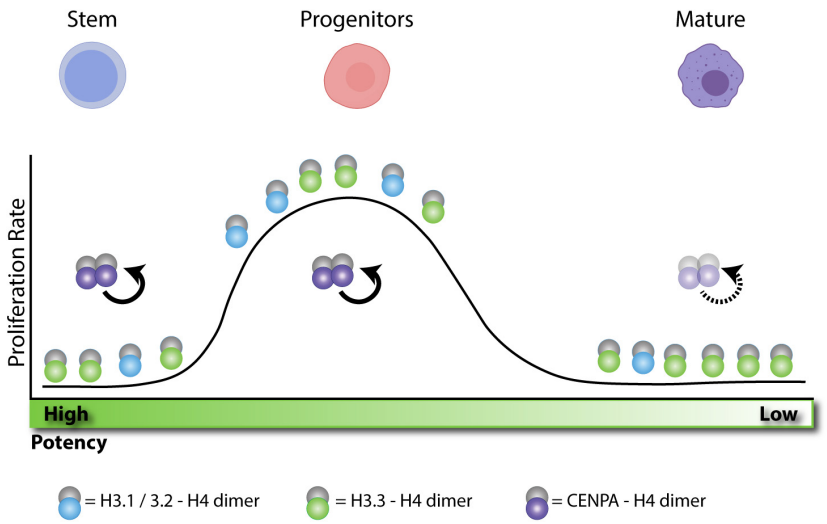


FIGURE 3 | H3 enrichment depends on the proliferative state and cellular potency. Adult stem cells remain in a quiescent state. Without replication these cells become enriched in H3.3. After activation, stem cells differentiate into progenitors and rapidly expand, becoming more enriched in H3.1/2 histone. After expansion, progenitors terminally differentiate into post-mitotic cells. Without replication, these cells also become enriched in H3.3. In quiescent cells, CENPA is actively maintained at the centromere during transcription to preserve proliferative potential. During expansion, centromeres are maintained by HJURP and SUPT6 as indicated in **Figure 1**. After terminal differentiation, post-mitotic cell centromeres can lose stability over time.

example, CENPA in quiescent cells is specifically maintained to preserve proliferative potential (**Figure 3**; Swartz et al., 2019). Additionally, many cellular differentiation paradigms implicate CAF-1 and HIRA as a transcriptional repressor or activator, respectively (**Figure 2B** and **Table 1**).

Volk et al. demonstrated that while complete loss of CAF-1 in the mouse inhibits normal hematopoiesis, its reduced levels is tolerated (Volk et al., 2018). Low levels of CAF-1 protect the mice from cancer progression by triggering differentiation of MLL/AF9 leukemic cells into mature myeloid cells. In this setting, CAF-1 is proposed to maintain leukemic cell identity via its RD nucleosome assembly activity as well as its competitive binding to sites of myeloid specifying transcription factors.

In a screen for chromatin regulators, silencing the CD4 gene in CD8 + cytotoxic T cells, CAF-1 was also identified as a transcriptional repressor among other fork components, including MCM2 (Ng et al., 2019). In this setting, CAF-1, in addition to its nucleosome assembly function, is proposed to cooperate with DNA and histone modifying enzymes by binding directly to histone deacetylase and histone demethylases to ensure heritable silencing of the CD4 gene. More recently, single cell profiling demonstrates that CAF-1 loss in myeloid progenitor cells triggers their partial differentiation leading to a mixed cellular state (Preprint, Guo et al., 2020). Interestingly, in comparison to normal myeloid differentiation, CAF-1 loss triggers a unique chromatin accessibility environment and activation of multi-lineage specific transcription factors. How the transcriptionally repressive role of CAF-1 in these systems is linked to its H3 deposition, alternative deposition of histone variants and/or recruitment of chromatin regulators remain to be determined.

In contrast to CAF-1, HIRA is widely studied in various cellular models including several mesoderm-derived tissues as well as during neurogenesis (**Figure 2B** and **Table 1**). For example, during normal hematopoiesis, Chen et al. (2020) demonstrated that upon conditional deletion of HIRA in

the mouse, long term hematopoietic stem cell (LT-HSC) function is impaired, leading to lethality. Interestingly, LT-HSC is thought to be in a more quiescent state. Accordingly, HIRA deletion had no effect in fetal hematopoiesis where hematopoietic stem cells are actively cycling. As seen with the mature oocyte and the early zygote, it is tempting to speculate that H3.3 deposition by HIRA could be involved in maintaining a unique chromatin environment in LT-HSCs that is subsequently remodeled during their self-renewal and differentiation (see model in **Figure 3**). HIRA is also required to maintain leukemic cells. Majumder et al. (2019) show that down-regulation of HIRA in chronic myeloid leukemia leads to a differentiation phenotype and implanting HIRA KO progenitors results in increased megakaryocyte differentiation. In this context, depletion of HIRA causes enrichment of H3.3 at promoters of key megakaryocyte differentiation factors GATA2 and MKL1, and a loss of H3.3 at erythroid differentiation promoters. It will be intriguing to dissect the mechanisms of differential H3.3 deposition upon HIRA loss. In addition to hematopoiesis, HIRA and ASF1 have been implicated in C2C12 myoblast cellular plasticity. Both HIRA and ASF1 are important for myoblast differentiation into myotubes and for their osteogenic conversion (Yang et al., 2011; Song et al., 2012). In myoblast differentiation HIRA and ASF1 drive MyoD expression and allow for H3.3 accumulation at critical enhancer regions. Additional studies implicate the role of HIRA in neurogenesis showing that it can interact with B-catenin to promote neurogenesis (Li and Jiao, 2017). In addition, *in vivo* conditional deletion of HIRA causes widespread defects in neurogenesis (Jeanne et al., 2021). Taken together in these systems, the HIRA mediated H3 deposition mechanisms are poorly characterized.

Interestingly, a recent study highlights HIRA's gap filling mechanisms in the context of metastatic transformation of breast and colon cancer tissues (Gomes et al., 2019) (**Figure 1A**). In this setting, the activation of epithelial-to-mesenchymal transition genes to promote metastasis is dependent on downregulation of canonical H3 deposition of CAF-1 and HIRA-mediated H3.3 deposition at regulatory sites. It will be interesting to determine the mechanism underlying the specific deposition of H3.3 at these sites. Given HIRA's selective deposition of H3.3 to regulatory elements and the body of active genes, HIRA's function as a histone chaperone in addition to H3.3S31 specific phosphorylation could create a chromatin environment to facilitate the binding of transcription factors and chromatin regulators to maintain cell identity or instruct cell fate change depending on the cellular context and environment. Interestingly, H3.3 is required for neuronal stem cell proliferation and differentiation via promoting H3K16 acetylation. Whether H3.3S31 is required in this context remains to be explored (Xia and Jiao, 2017).

Another example of H3.3 guided recruitment of chromatin factors was recently highlighted while dissecting the transcriptional response to pathogens. In this context, selective phosphorylation of H3.3S31 at rapidly induced genes triggers a chromatin signaling cascade via recruiting a histone methyltransferase that promotes transcriptional elongation and repulsing a chromatin reader that inhibits transcription (Thorne et al., 2012; Guo et al., 2014; Wen H. et al., 2014; Armache et al., 2020).

Considering other H3.3 deposition pathways, conditional deletion of DAXX in pancreatic tissues supports its role in ERV silencing (**Figure 2B** and **Table 1**; Wasylishen et al., 2020). While no phenotypes were observed, the more permissible transcriptional state is proposed to increase responses to stressors and to impair recovery. Outside of DAXX function at repetitive elements, another study in neurons reported a non-canonical DAXX mediated H3.3 deposition at regulatory elements that is linked to neuronal activation (Michod et al., 2012). Furthermore, DAXX is responsible for the ectopic deposition of overexpressed CENPA which is a hallmark of many cancers (**Figure 1D**; Sharma et al., 2019). It will be interesting to probe how DAXX responds to the loss of HIRA in these systems and investigate the function of other histone chaperone mediated deposition of H3, including ASF1A/B, HJURP, SPT6 in these cellular settings.

DISCUSSION

Histone chaperones are in place to modulate the deposition of histones at the right place and right time and coordinate the action of accompanying chromatin factors, including lineage-specific transcription factors during quiescence, stem cell self-renewal, differentiation, or reprogramming. While the expression of histone variants during the cell cycle and development is well documented, the activity, complex diversity, and interplay of histone chaperones during these processes is poorly understood, especially in the context of cell fate transitions. This is clearly complicated by the multifunctional characteristics of histone chaperones as they play additional roles independent of nucleosome assembly that are in turn linked to chromatin regulation.

The lessons that we learned from studying H3 deposition pathways in the context of normal development, culture model and disease state suggest that the RD and RI H3 deposition pathways act in a balanced manner to maintain lineage identity and instruct cell fate change in response to signals. Aside from the traditional culture models, it will be interesting to exploit newly developed organoid culture models and gametogenesis platforms to characterize the mechanisms of histone exchange and apply the lessons we learned from model organisms (**Figure 2A**). These emerging culture models provide unique systems to perform biochemical studies and create high resolution spatiotemporal maps of histone deposition in the context of cell fate determination. Finally, future therapeutic avenues include (1) the identification of unique histone deposition machinery in disease states and investigating the epigenetic addictions as a consequence of histone mutations or compromised histone chaperone activity and (2) manipulate histone chaperone pathways to generate specific cell types for regenerative purposes.

AUTHOR CONTRIBUTIONS

SC proposed the topic of the review and outlined the structure. SC and JM mentored RF throughout the analysis and discussions.

SC, JM, and RF reviewed the literature, wrote the review, designed the figures, and compiled the table. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the UC Cancer Research Coordinating Committee (CRCC – C21CR2175) and the City of Hope – UC Riverside Biomedical Research Initiative seed grants. RF is supported by UC Riverside's Eugene Cota-Robles Fellowship.

ACKNOWLEDGMENTS

The authors thank Dr. Qi Chen, Dr. Andrew Volk, and members of the SC and JM lab for critical reading of the manuscript. The authors apologize to colleagues whose work could not be cited owing to space limitations.

REFERENCES

Abascal, F., Corpet, A., Gurard-Levin, Z. A., Juan, D., Ochsenbein, F., Rico, D., et al. (2013). Subfunctionalization via Adaptive Evolution Influenced by Genomic Context: The Case of Histone Chaperones ASF1a and ASF1b. *Mole. Biol. Evol.* 30, 1853–1866. doi: 10.1093/molbev/mst086

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9:1911. doi: 10.1038/s41467-018-04252-2

Ahmad, K., and Henikoff, S. (2002). Histone H3 variants specify modes of chromatin assembly. *PNAS* 99, 16477–16484. doi: 10.1073/pnas.172403699

Akiyama, T., Suzuki, O., Matsuda, J., and Aoki, F. (2011). Dynamic Replacement of Histone H3 Variants Reprograms Epigenetic Marks in Early Mouse Embryos. *PLoS Genetics* 7:e1002279. doi: 10.1371/journal.pgen.1002279

Almouzni, G., and Méchali, M. (1988). Assembly of spaced chromatin promoted by DNA synthesis in extracts from Xenopus eggs. *EMBO J.* 7, 665–672. doi: 10.1002/j.1460-2075.1988.tb02861.x

Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38, 824–844. doi: 10.1038/s41587-020-0561-9

Arimura, Y., Shirayama, K., Horikoshi, N., Fujita, R., Taguchi, H., Kagawa, W., et al. (2014). Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci. Rep.* 4:7115. doi: 10.1038/srep07115

Armache, A., Yang, S., Martínez, de Paz, A., Robbins, L. E., Durmaz, C., et al. (2020). Histone H3.3 phosphorylation amplifies stimulation-induced transcription. *Nature* 583, 852–857. doi: 10.1038/s41586-020-2533-0

Banaszynski, L. A., Wen, D., Dewell, S., Whitcomb, S. J., Lin, M., Diaz, N., et al. (2013). Hira-Dependent Histone H3.3 Deposition Facilitates PRC2 Recruitment at Developmental Loci in ES Cells. *Cell* 155, 107–120. doi: 10.1016/j.cell.2013.08.061

Bellelli, R., Belan, O., Pye, V. E., Clement, C., Maslen, S. L., Skehel, J. M., et al. (2018). POLE3-POLE4 Is a Histone H3-H4 Chaperone that Maintains Chromatin Integrity during DNA Replication. *Mol. Cell* 72, 112.e–126.e. doi: 10.1016/j.molcel.2018.08.043

Bobkov, G. O. M., Huang, A., van den Berg, S. J. W., Mitra, S., Anselm, E., Lazou, V., et al. (2020). Spt6 is a maintenance factor for centromeric CENP-A. *Nat. Commun.* 11:2919. doi: 10.1038/s41467-020-16695-7

Brumbaugh, J., Stefano, B. D., and Hochedlinger, K. (2019). Reprogramming: identifying the mechanisms that safeguard cell identity. *Development* 146:dev.182170. doi: 10.1242/dev.182170

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218. doi: 10.1038/nmeth.2688

Burton, A., Brochard, V., Galan, C., Ruiz-Morales, E. R., Rovira, Q., Rodriguez-Terrones, D., et al. (2020). Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat. Cell Biol.* 22, 767–778. doi: 10.1038/s41556-020-0536-6

Buschbeck, M., and Hake, S. B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. *Nat. Rev. Mole. Cell Biol.* 18, 299–314. doi: 10.1038/nrm.2016.166

Bush, K. M., Yuen, B. T., Barrilleaux, B. L., Riggs, J. W., O'Geen, H., Cotterman, R. F., et al. (2013). Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. *Epigenet. Chromatin* 6:7. doi: 10.1186/1756-8935-6-7

Champroux, A., Cocquet, J., Henry-Berger, J., Drevet, J. R., and Kocer, A. (2018). A Decade of Exploring the Mammalian Sperm Epigenome: Paternal Epigenetic and Transgenerational Inheritance. *Front. Cell Dev. Biol.* 6:00050. doi: 10.3389/fcell.2018.00050

Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y. L., Murn, J., Ninova, M., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* 528, 218–224. doi: 10.1038/nature15749

Cheloufi, S., and Hochedlinger, K. (2017). Emerging roles of the histone chaperone CAF-1 in cellular plasticity. *Curr. Opin. Genet. Dev.* 46, 83–94. doi: 10.1016/j.gde.2017.06.004

Chen, C., Sun, M., Warzecha, C., Bachu, M., Dey, A., Wu, T., et al. (2020). HIRA, a DiGeorge Syndrome Candidate Gene, Confers Proper Chromatin Accessibility on HSCs and Supports All Stages of Hematopoiesis. *Cell Reports* 30, 2136.e–2149.e. doi: 10.1016/j.celrep.2020.01.062

Cheng, L., Zhang, X., Wang, Y., Gan, H., Xu, X., Lv, X., et al. (2019). Chromatin Assembly Factor 1 (CAF-1) facilitates the establishment of facultative heterochromatin during pluripotency exit. *Nucleic Acids Res.* 47, 11114–11131. doi: 10.1093/nar/gkz858

Cho, C., Jung-Ha, H., Willis, W. D., Goulding, E. H., Stein, P., Xu, Z., et al. (2003). Protamine 2 Deficiency Leads to Sperm DNA Damage and Embryo Death in Mice1. *Biol. Reproduct.* 69, 211–217. doi: 10.1095/biolreprod.102.015115

Cho, C., Willis, W. D., Goulding, E. H., Jung-Ha, H., Choi, Y.-C., Hecht, N. B., et al. (2001). Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat. Genet.* 28, 82–86. doi: 10.1038/ng0501-82

Clément, C., Orsi, G. A., Gatto, A., Boyarchuk, E., Forest, A., Hajj, B., et al. (2018). High-resolution visualization of H3 variants during replication reveals their controlled recycling. *Nat. Commun.* 9:3181. doi: 10.1038/s41467-018-05697-1

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14, 959–962. doi: 10.1038/nmeth.4396

Couldrey, C., Carlton, M. B. L., Nolan, P. M., Colledge, W. H., and Evans, M. J. (1999). A retroviral Gene Trap Insertion into the Histone 3.3A Gene Causes Partial Neonatal Lethality, Stunted Growth, Neuromuscular Deficits and Male Sub-fertility in Transgenic Mice. *Hum. Mole. Genetics* 8, 2489–2495. doi: 10.1093/hmg/8.13.2489

Das, A., Smoak, E. M., Linares-Saldana, R., Lampson, M. A., and Black, B. E. (2017). Centromere inheritance through the germline. *Chromosoma* 126, 595–604. doi: 10.1007/s00412-017-0640-y

De Clerck, L., Taelman, J., Popovic, M., Willems, S., Van der Jeught, M., Heindryckx, B., et al. (2019). Untargeted histone profiling during naive conversion uncovers conserved modification markers between mouse and human. *Sci. Rep.* 9:17240. doi: 10.1038/s41598-019-53681-6

Dilg, D., Saleh, R. N. M., Phelps, S. E. L., Rose, Y., Dupays, L., Murphy, C., et al. (2016). HIRA is required for heart development and directly Regulates Tnni2 and Tnnt3. *PLoS One* 11:e0161096. doi: 10.1371/journal.pone.0161096

Dilworth, S. M., Black, S. J., and Laskey, R. A. (1987). Two complexes that contain histones are required for nucleosome assembly in vitro: Role of nucleoplasmin and N1 in Xenopus egg extracts. *Cell* 51, 1009–1018. doi: 10.1016/0092-8674(87)90587-3

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082

Dunleavy, E. M., Almouzni, G., and Karpen, G. H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G1 phase. *Nucleus* 2, 146–157. doi: 10.4161/nucl.2.2.15211

Dunleavy, E. M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., et al. (2009). HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell* 137, 485–497. doi: 10.1016/j.cell.2009.02.040

Eckersley-Maslin, M. A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mole. Cell Biol.* 19, 436–450. doi: 10.1038/s41580-018-0008-z

Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D., and Banaszynski, L. A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* 522, 240–244. doi: 10.1038/nature14345

Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., et al. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mole. Biol.* 20, 868–875. doi: 10.1038/nsmb.2599

Evano, B., Khalilian, S., Le Carrou, G., Almouzni, G., and Tajbakhsh, S. (2020). Dynamics of Asymmetric and Symmetric Divisions of Muscle Stem Cells In Vivo and on Artificial Niches. *Cell Rep.* 30, 3195.e–3206.e. doi: 10.1016/j.celrep.2020.01.097

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156. doi: 10.1038/292154a0

Fang, H.-T., El Farran, C. A., Xing, Q. R., Zhang, L.-F., Li, H., Lim, B., et al. (2018). Global H3.3 dynamic deposition defines its bimodal role in cell fate transition. *Nat. Commun.* 9:1537. doi: 10.1038/s41467-018-03904-7

Filipescu, D., Müller, S., and Almouzni, G. (2014). Histone H3 Variants and Their Chaperones During Development and Disease: Contributing to Epigenetic Control. *Ann. Rev. Cell Dev. Biol.* 30, 615–646. doi: 10.1146/annurev-cellbio-100913-013311

Franklin, S. G., and Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature* 266, 273–275. doi: 10.1038/266273a0

Gao, Y., Gan, H., Lou, Z., and Zhang, Z. (2018). Asf1a resolves bivalent chromatin domains for the induction of lineage-specific genes during mouse embryonic stem cell differentiation. *PNAS* 115, E6162–E6171. doi: 10.1073/pnas.1801909115

Gehre, M., Bunina, D., Sidoli, S., Lübke, M. J., Diaz, N., Trovato, M., et al. (2020). Lysine 4 of histone H3.3 is required for embryonic stem cell differentiation, histone enrichment at regulatory regions and transcription accuracy. *Nat. Genet.* 52, 273–282. doi: 10.1038/s41588-020-0586-5

Goldberg, A. D., Banaszynski, L. A., Noh, K.-M., Lewis, P. W., Elsaesser, S. J., Stadler, S., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* 140, 678–691. doi: 10.1016/j.cell.2010.01.003

Gomes, A. P., Ilter, D., Low, V., Rosenzweig, A., Shen, Z.-J., Schild, T., et al. (2019). Dynamic Incorporation of Histone H3 Variants into Chromatin Is Essential for Acquisition of Aggressive Traits and Metastatic Colonization. *Cancer Cell* 36, 402.e–417.e. doi: 10.1016/j.ccell.2019.08.006

Gonzalez-Muñoz, E., Arboleda-Estudillo, Y., Out, H. H., and Cibelli, J. B. (2014). Histone chaperone ASF1A is required for maintenance of pluripotency and cellular reprogramming. *Science* 2014:1254745. doi: 10.1126/science.1254745

Gou, L.-T., Lim, D.-H., Ma, W., Aubol, B. E., Hao, Y., Wang, X., et al. (2020). Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation. *Cell* 180, 1212.e–1227.e. doi: 10.1016/j.cell.2020.02.020

Groth, A., Corpet, A., Cook, A. J. L., Roche, D., Bartek, J., Lukas, J., et al. (2007). Regulation of Replication Fork Progression Through Histone Supply and Demand. *Science* 318, 1928–1931. doi: 10.1126/science.1148992

Grover, P., Asa, J. S., and Campos, E. I. (2018). H3–H4 Histone Chaperone Pathways. *Ann. Rev. Genet.* 52, 109–130. doi: 10.1146/annurev-genet-120417-031547

Guo, R., Zheng, L., Park, J. W., Lv, R., Chen, H., Jiao, F., et al. (2014). BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation-Decorated Chromatin to Regulated Pre-mRNA Processing. *Mole. Cell* 56, 298–310. doi: 10.1016/j.molcel.2014.08.022

Guo, Y., Ji, F., Murn, J., Frankhouser, D., Blanco, M. A., Chiem, C., et al. (2020). The Histone Chaperone CAF-1 Sustains Myeloid Lineage Identity. *bioRxiv* 2020:350447. doi: 10.1101/2020.10.22.350447

Hamazaki, N., Kyogoku, H., Araki, H., Miura, F., Horikawa, C., Hamada, N., et al. (2021). Reconstitution of the oocyte transcriptional network with transcription factors. *Nature* 589, 264–269. doi: 10.1038/s41586-020-3027-9

Hammond, C. M., Strømme, C. B., Huang, H., Patel, D. J., and Groth, A. (2017). Histone chaperone networks shaping chromatin function. *Nat. Rev. Mole. Cell Biol.* 18, 141–158. doi: 10.1038/nrm.2016.159

Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T., and Cairns, B. R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 460, 473–478. doi: 10.1038/nature08162

Hartford, S. A., Luo, Y., Southard, T. L., Min, I. M., Lis, J. T., and Schimenti, J. C. (2011). Minichromosome maintenance helicase paralog MCM9 is dispensible for DNA replication but functions in germ-line stem cells and tumor suppression. *PNAS* 108, 17702–17707. doi: 10.1073/pnas.1113524108

He, Q., Kim, H., Huang, R., Lu, W., Tang, M., Shi, F., et al. (2015). The Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by Promoting H3K9 Trimethylation. *Cell Stem Cell* 17, 273–286. doi: 10.1016/j.stem.2015.07.022

Heo, J.-I., Cho, J. H., and Kim, J.-R. (2013). HJURP regulates cellular senescence in human fibroblasts and endothelial cells via a p53-dependent pathway. *J. Gerontol. A Biol. Sci. Med. Sci.* 68, 914–925. doi: 10.1093/gerona/gls257

Hewish, D. R., and Burgoyne, L. A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* 52, 504–510. doi: 10.1016/0006-291X(73)90740-7

Hoelper, D., Huang, H., Jain, A. Y., Patel, D. J., and Lewis, P. W. (2017). Structural and mechanistic insights into ATRX-dependent and -independent functions of the histone chaperone DAXX. *Nat. Commun.* 8:1193. doi: 10.1038/s41467-017-01206-y

Houlard, M., Berlivet, S., Probst, A. V., Quivy, J.-P., Héry, P., Almouzni, G., et al. (2006). CAF-1 Is Essential for Heterochromatin Organization in Pluripotent Embryonic Cells. *PLoS Genetics* 2:e181. doi: 10.1371/journal.pgen.0020181

Howman, E. V., Fowler, K. J., Newson, A. J., Redward, S., MacDonald, A. C., Kalitsis, P., et al. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *PNAS* 97, 1148–1153. doi: 10.1073/pnas.97.3.1148

Huang, H., Strømme, C. B., Saredi, G., Hödl, M., Strandsby, A., González-Aguilera, C., et al. (2015). A unique binding mode enables MCM2 to chaperone histones H3–H4 at replication forks. *Nat. Struct. Mole. Biol.* 22, 618–626. doi: 10.1038/nsmb.3055

Ishiuchi, T., Abe, S., Inoue, K., Yeung, W. K. A., Miki, Y., Ogura, A., et al. (2021). Reprogramming of the histone H3.3 landscape in the early mouse embryo. *Nat. Struct. Mole. Biol.* 28, 38–49. doi: 10.1038/s41594-020-00521-1

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., et al. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671. doi: 10.1038/nsmb.3066

Jang, C.-W., Shibata, Y., Starmer, J., Yee, D., and Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. *Genes Dev.* 29, 1377–1392. doi: 10.1101/gad.264150.115

Jansen, L. E. T., Black, B. E., Foltz, D. R., and Cleveland, D. W. (2007). Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* 176, 795–805. doi: 10.1083/jcb.200701066

Jeanne, M., Vuillaume, M.-L., Ung, D. C., Vancollie, V. E., Wagner, C., Collins, S. C., et al. (2021). Haploinsufficiency of the HIRA gene located in the 22q11 deletion syndrome region is associated with abnormal neurodevelopment and impaired dendritic outgrowth. *Hum. Genet.* 2021, 2252–2251. doi: 10.1007/s00439-020-02252-1

Jeronimo, C., Poitras, C., and Robert, F. (2019). Histone Recycling by FACT and Spt6 during Transcription Prevents the Scrambling of Histone Modifications. *Cell Rep.* 28, 1206.e–1218.e. doi: 10.1016/j.celrep.2019.06.097

Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. *Science* 301, 1096–1099. doi: 10.1126/science.1087374

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., et al. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 10:1930. doi: 10.1038/s41467-019-09982-5

Kong, S., Han, X., Cui, T., Zhou, C., Jiang, Y., Zhang, H., et al. (2016). MCM2 mediates progesterone-induced endometrial stromal cell proliferation and differentiation in mice. *Endocrine* 53, 595–606. doi: 10.1007/s12020-016-0894-9

Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* 184, 868–871. doi: 10.1126/science.184.4139.868

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A. V., Barth, T., Cantaloube, S., et al. (2014). Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. *Mol. Cell* 53, 631–644. doi: 10.1016/j.molcel.2014.01.018

Laskey, R. A., Honda, B. M., Mills, A. D., and Finch, J. T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275, 416–420. doi: 10.1038/275416a0

Laskey, R. A., Mills, A. D., and Morris, N. R. (1977). Assembly of SV40 chromatin in a cell-free system from Xenopus eggs. *Cell* 10, 237–243. doi: 10.1016/0092-8674(77)90217-3

Lewis, P. W., Elsaesser, S. J., Noh, K.-M., Stadler, S. C., and Allis, C. D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *PNAS* 107, 14075–14080. doi: 10.1073/pnas.1008850107

Li, Y., and Jiao, J. (2017). Histone chaperone HIRA regulates neural progenitor cell proliferation and neurogenesis via β-catenin. *J. Cell Biol.* 216, 1975–1992. doi: 10.1083/jcb.201610014

Lin, C.-J., Koh, F. M., Wong, P., Conti, M., and Ramalho-Santos, M. (2014). Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. *Dev. Cell* 30, 268–279. doi: 10.1016/j.devcel.2014.06.022

Loppin, B., and Berger, F. (2020). Histone Variants: The Nexus of Developmental Decisions and Epigenetic Memory. *Annu. Rev. Genet.* 54, 121–149. doi: 10.1146/annurev-genet-022620-100039

Loppin, B., Bonnefoy, E., Anselme, C., Laurençon, A., Karr, T. L., and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature* 437, 1386–1390. doi: 10.1038/nature04059

Lorain, S., Quivy, J.-P., Monier-Gavelle, F., Scamps, C., Lécluse, Y., Almouzni, G., et al. (1998). Core Histones and HIRIP3, a Novel Histone-Binding Protein, Directly Interact with WD Repeat Protein HIRA. *Mole. Cell. Biol.* 18, 5546–5556. doi: 10.1128/MCB.18.9.5546

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. doi: 10.1038/38444

Luzzati, V., and Nicolaieff, A. (1959). Etude par diffusion des rayons X aux petits angles des gels d'acide désoxyribonucléique et de nucléoprotéines: (note préliminaire). *J. Mole. Biol.* 1, 127–IN5. doi: 10.1016/S0022-2836(59)80041-3

Ma, B., Trieu, T.-J., Cheng, J., Zhou, S., Tang, Q., Xie, J., et al. (2020). Differential Histone Distribution Patterns in Induced Asymmetrically Dividing Mouse Embryonic Stem Cells. *Cell Rep.* 32:108003. doi: 10.1016/j.celrep.2020.108003

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi: 10.1038/nature11244

Magnaghi, P., Roberts, C., Lorain, S., Lipinski, M., and Scambler, P. J. (1998). HIRA, a mammalian homologue of Saccharomyces cerevisiae transcriptional co-repressors, interacts with Pax3. *Nat. Genet.* 20, 74–77. doi: 10.1038/1739

Majumder, A., Dharan, A. T., Baral, I., Varghese, P. C., Mukherjee, A., Subhadradevi, L., et al. (2019). Histone chaperone HIRA dictate proliferation vs differentiation of chronic myeloid leukemia cells. *FASEB BioAdv.* 1, 525–537. doi: 10.1096/fba.2019-00014

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci.* 78, 7634–7638. doi: 10.1073/pnas.78.12.7634

Martire, S., and Banaszynski, L. A. (2020). The roles of histone variants in fine-tuning chromatin organization and function. *Nat. Rev. Mole. Cell Biol.* 21, 522–541. doi: 10.1038/s41580-020-0262-8

Martire, S., Gogate, A. A., Whitmill, A., Tafessu, A., Nguyen, J., Teng, Y.-C., et al. (2019). Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and enhancer acetylation. *Nat. Genet.* 51, 941–946. doi: 10.1038/s41588-019-0428-5

Meers, M. P., Bryson, T. D., Henikoff, J. G., and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *eLife* 8:e46314. doi: 10.7554/eLife.46314

Mendiratta, S., Gatto, A., and Almouzni, G. (2018). Histone supply: Multitiered regulation ensures chromatin dynamics throughout the cell cycle. *J. Cell Biol.* 218, 39–54. doi: 10.1083/jcb.201807179

Messiaen, S., Guiard, J., Aigueperse, C., Fliniaux, I., Tourpin, S., Barroca, V., et al. (2016). Loss of the histone chaperone ASF1B reduces female reproductive capacity in mice. *Reproduction* 151, 477–489. doi: 10.1530/REP-15-0327

Michaelson, J. S., Bader, D., Kuo, F., Kozak, C., and Leder, P. (1999). Loss of Daxx, a promiscuously interacting protein, results in extensive apoptosis in early mouse development. *Genes Dev.* 13, 1918–1923. doi: 10.1101/gad.13.15.1918

Michod, D., Bartesaghi, S., Khelifi, A., Bellodi, C., Berliocchi, L., Nicotera, P., et al. (2012). Calcium-Dependent Dephosphorylation of the Histone Chaperone DAXX Regulates H3.3 Loading and Transcription upon Neuronal Activation. *Neuron* 74, 122–135. doi: 10.1016/j.neuron.2012.02.021

Muhire, B. M., Booker, M. A., and Tolstorukov, M. Y. (2019). Non-neutral evolution of H3.3-encoding genes occurs without alterations in protein sequence. *Sci. Rep.* 9:8472. doi: 10.1038/s41598-019-44800-4

Nakano, S., Stillman, B., and Horvitz, H. R. (2011). CAF-1-Mediated Chromatin Assembly Generates a Bilateral Asymmetry in C. elegans Neuroanatomy. *Cell* 147, 1525–1536. doi: 10.1016/j.cell.2011.11.053

Ng, C., Aichinger, M., Nguyen, T., Au, C., Najar, T., Wu, L., et al. (2019). The histone chaperone CAF-1 cooperates with the DNA methyltransferases to maintain Cd4 silencing in cytotoxic T cells. *Genes Dev.* 33, 669–683. doi: 10.1101/gad.322024.118

Olins, A. L., and Olins, D. E. (1974). Spheroid Chromatin Units (ν Bodies). *Science* 183, 330–332. doi: 10.1126/science.183.4122.330

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4, 281–300. doi: 10.1016/0092-8674(75)90149-X

Paul, P. K., Rabaglia, M. E., Wang, C.-Y., Stapleton, D. S., Leng, N., Kendziorski, C., et al. (2016). Histone chaperone ASF1B promotes human β-cell proliferation via recruitment of histone H3.3. *Cell Cycle* 15, 3191–3202. doi: 10.1080/15384101.2016.1241914

Pérez-Palacios, R., and Bourc'his, D. (2018). A single-cell chromatin map of human embryos. *Nat. Cell Biol.* 20, 742–744. doi: 10.1038/s41556-018-0134-z

Petryk, N., Dalby, M., Wenger, A., Stromme, C. B., Strandsby, A., Andersson, R., et al. (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science* 361, 1389–1392. doi: 10.1126/science.aau0294

Probst, A. V., and Almouzni, G. (2011). Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends Genet.* 27, 177–185. doi: 10.1016/j.tig.2011.02.002

Pruitt, S. C., Bailey, K. J., and Freeland, A. (2007). Reduced Mcm2 Expression Results in Severe Stem/Progenitor Cell Deficiency and Cancer. *Stem Cells* 25, 3121–3132. doi: 10.1634/stemcells.2007-0483

Raja, S. J., and Renkawitz-Pohl, R. (2005). Replacement by Drosophila melanogaster Protamines and Mst77F of Histones during Chromatin Condensation in Late Spermatids and Role of Sesame in the Removal of These Proteins from the Male Pronucleus. *Mole. Cell. Biol.* 25, 6165–6177. doi: 10.1128/MCB.25.14.6165-6177.2005

Ray-Gallet, D., Quivy, J.-P., Scamps, C., Martini, E. M.-D., Lipinski, M., and Almouzni, G. (2002). HIRA Is Critical for a Nucleosome Assembly Pathway Independent of DNA Synthesis. *Mole. Cell* 9, 1091–1100. doi: 10.1016/S1097-2765(02)00526-9

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., et al. (2011). Dynamics of Histone H3 Deposition In Vivo Reveal a Nucleosome Gap-Filling Mechanism for H3.3 to Maintain Chromatin Integrity. *Mole. Cell* 44, 928–941. doi: 10.1016/j.molcel.2011.12.006

Roberts, C., Sutherland, H. F., Farmer, H., Kimber, W., Halford, S., Carey, A., et al. (2002). Targeted Mutagenesis of the Hira Gene Results in Gastrulation Defects

and Patterning Abnormalities of Mesoendodermal Derivatives Prior to Early Embryonic Lethality. *Mole. Cell. Biol.* 22, 2318–2328. doi: 10.1128/MCB.22.7.2318-2328.2002

Saleh, R. N. M., Dilg, D., Abou Zeid, A. A., Hashad, D. I., Scambler, P. J., and Chapgier, A. L. A. (2018). HIRA directly targets the enhancers of selected cardiac transcription factors during in vitro differentiation of mouse embryonic stem cells. *Mol. Biol. Rep.* 45, 1001–1011. doi: 10.1007/s11033-018-4247-z

Schlesinger, S., Kaffe, B., Melcer, S., Aguilera, J. D., Sivaraman, D. M., Kaplan, T., et al. (2017). A hyperdynamic H3.3 nucleosome marks promoter regions in pluripotent embryonic stem cells. *Nucleic Acids Res.* 45, 12181–12194. doi: 10.1093/nar/gkx817

Schneiderman, J. I., Orsi, G. A., Hughes, K. T., Loppin, B., and Ahmad, K. (2012). Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. *PNAS* 109, 19721–19726. doi: 10.1073/pnas.1206629109

Serra-Cardona, A., and Zhang, Z. (2018). Replication-Coupled Nucleosome Assembly in the Passage of Epigenetic Information and Cell Identity. *Trends Biochem. Sci.* 43, 136–148. doi: 10.1016/j.tibs.2017.12.003

Sharma, A. B., Dimitrov, S., Hamiche, A., and Van Dyck, E. (2019). Centromeric and ectopic assembly of CENP-A chromatin in health and cancer: old marks and new tracks. *Nucleic Acids Res.* 47, 1051–1069. doi: 10.1093/nar/gky1298

Sitbon, D., Boyarchuk, E., Dingli, F., Loew, D., and Almouzni, G. (2020). Histone variant H3.3 residue S31 is essential for Xenopus gastrulation regardless of the deposition pathway. *Nat. Commun.* 11:1256. doi: 10.1038/s41467-020-15084-4

Smith, S., and Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell* 58, 15–25. doi: 10.1016/0092-8674(89)90398-X

Song, T.-Y., Yang, J.-H., Park, J. Y., Song, Y., Han, J.-W., Youn, H.-D., et al. (2012). The role of histone chaperones in osteoblastic differentiation of C2C12 myoblasts. *Biochem. Biophys. Res. Commun.* 423, 726–732. doi: 10.1016/j.bbrc.2012.06.026

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mole. Biol.* 20, 267–273. doi: 10.1038/nsmb.2506

Swartz, S. Z., McKay, L. S., Su, K.-C., Bury, L., Padeganeh, A., Maddox, P. S., et al. (2019). Quiescent Cells Actively Replenish CENP-A Nucleosomes to Maintain Centromere Identity and Proliferative Potential. *Dev. Cell* 51, 35.e–48.e. doi: 10.1016/j.devcel.2019.07.016

Talbert, P. B., and Henikoff, S. (2010). Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mole. Cell Biol.* 11, 264–275. doi: 10.1038/nrm2861

Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O. G., Shiekhattar, R., et al. (2004). A Novel Transcription Regulatory Complex Containing Death Domain-associated Protein and the ATR-X Syndrome Protein. *J. Biol. Chem.* 279, 20369–20377. doi: 10.1074/jbc.M401321200

Tang, M. C. W., Jacobs, S. A., Mattiske, D. M., Soh, Y. M., Graham, A. N., Tran, A., et al. (2015). Contribution of the Two Genes Encoding Histone Variant H3.3 to Viability and Fertility in Mice. *PLoS Genet.* 11:e1004964. doi: 10.1371/journal.pgen.1004964

Tang, M. C. W., Jacobs, S. A., Wong, L. H., and Mann, J. R. (2013). Conditional allelic replacement applied to genes encoding the histone variant H3.3 in the mouse. *Genesis* 51, 142–146. doi: 10.1002/dvg.22366

Tang, Y., Poustovoitov, M. V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., et al. (2006). Structure of a human ASF1a–HIRA complex and insights into specificity of histone chaperone complex assembly. *Nat. Struct. Mole. Biol.* 13, 921–929. doi: 10.1038/nsmb1147

Thorne, J. L., Ouboussad, L., and Lefevre, P. F. (2012). Heterochromatin protein 1 gamma and IκB kinase alpha interdependence during tumour necrosis factor gene transcription elongation in activated macrophages. *Nucleic Acids Res.* 40, 7676–7689. doi: 10.1093/nar/gks509

Torné, J., Ray-Gallet, D., Boyarchuk, E., Garnier, M., Le Baccon, P., Coulon, A., et al. (2020). Two HIRA-dependent pathways mediate H3.3 de novo deposition and recycling during transcription. *Nat. Struct. Mole. Biol.* 27, 1057–1068. doi: 10.1038/s41594-020-0492-7

Torres-Flores, U., and Hernández-Hernández, A. (2020). The Interplay Between Replacement and Retention of Histones in the Sperm Genome. *Front. Genet.* 11:00780. doi: 10.3389/fgene.2020.00780

Torres-Padilla, M.-E., Bannister, A. J., Hurd, P. J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int. J. Dev. Biol.* 50, 455–461. doi: 10.1387/ijdb.052073mt

Tyler, J. K., Adams, C. R., Chen, S.-R., Kobayashi, R., Kamakaka, R. T., and Kadonaga, J. T. (1999). The RCAF complex mediates chromatin assembly during DNA replication and repair. *Nature* 402, 555–560. doi: 10.1038/990147

Valenzuela, N., Soibam, B., Li, L., Wang, J., Byers, L. A., Liu, Y., et al. (2017). HIRA deficiency in muscle fibers causes hypertrophy and susceptibility to oxidative stress. *J. Cell Sci.* 130, 2551–2563. doi: 10.1242/jcs.200642

Venkatesh, S., and Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mole. Cell Biol.* 16, 178–189. doi: 10.1038/nrm3941

Volk, A., Liang, K., Suraneni, P., Li, X., Zhao, J., Bulic, M., et al. (2018). A CHAF1B-Dependent Molecular Switch in Hematopoiesis and Leukemia Pathogenesis. *Cancer Cell* 34, 707.e–723.e. doi: 10.1016/j.ccell.2018.10.004

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., et al. (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* 560, 607–612. doi: 10.1038/s41586-018-0440-4

Vos, S. M., Farnung, L., Linden, A., Urlaub, H., and Cramer, P. (2020). Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mole. Biol.* 27, 668–677. doi: 10.1038/s41594-020-0437-1

Wasylishen, A. R., Sun, C., Moyer, S. M., Qi, Y., Chau, G. P., Aryal, N. K., et al. (2020). Daxx maintains endogenous retroviral silencing and restricts cellular plasticity in vivo. *Sci. Adv.* 6:eaba8415. doi: 10.1126/sciadv.aba8415

Wen, D., Banaszynski, L. A., Liu, Y., Geng, F., Noh, K.-M., Xiang, J., et al. (2014). Histone variant H3.3 is an essential maternal factor for oocyte reprogramming. *Proc. Natl. Acad. Sci.* 111, 7325–7330. doi: 10.1073/pnas.1406389111

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., et al. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* 508, 263–268. doi: 10.1038/nature13045

Wu, C., Ding, H., Wang, S., Li, Y., Liu, S.-B., Wang, X., et al. (2020). DAXX inhibits cancer stemness and epithelial–mesenchymal transition in gastric cancer. *Br. J. Cancer* 122, 1477–1485. doi: 10.1038/s41416-020-0800-3

Wilkins, M. H. F., Zubay, G., and Wilson, H. R. (1959). X-ray diffraction studies of the molecular structure of nucleohistone and chromosomes. *J. Mole. Biol.* 1, 179–IN10. doi: 10.1016/S0022-2836(59)80046-2

Xia, W., and Jiao, J. (2017). Histone variant H3.3 orchestrates neural stem cell differentiation in the developing brain. *Cell Death Diff.* 24, 1548–1563. doi: 10.1038/cdd.2017.77

Yang, J.-H., Song, Y., Seol, J.-H., Park, J. Y., Yang, Y.-J., Han, J.-W., et al. (2011). Myogenic transcriptional activation of MyoD mediated by replication-independent histone deposition. *PNAS* 108, 85–90. doi: 10.1073/pnas.1009830108

Yang, J.-H., Song, T.-Y., Jo, C., Park, J., Lee, H.-Y., Song, I., et al. (2016). Differential regulation of the histone chaperone HIRA during muscle cell differentiation by a phosphorylation switch. *Exp. Mol. Med.* 48, e252–e252. doi: 10.1038/emm.2016.68

Yoshida, K., Muratani, M., Araki, H., Miura, F., Suzuki, T., Dohmae, N., et al. (2018). Mapping of histone-binding sites in histone replacement-completed spermatozoa. *Nat. Commun.* 9:3885. doi: 10.1038/s41467-018-06243-9

Yuen, B. T. K., Bush, K. M., Barrilleaux, B. L., Cotterman, R., and Knoepfler, P. S. (2014). Histone H3.3 regulates dynamic chromatin states during spermatogenesis. *Development* 141, 3483–3494. doi: 10.1242/dev.106450

Zasadzińska, E., Huang, J., Bailey, A. O., Guo, L. Y., Lee, N. S., Srivastava, S., et al. (2018). Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP. *Dev. Cell* 47, 348.e–362.e. doi: 10.1016/j.devcel.2018.09.003

Zhu, Z., Li, C., Zeng, Y., Ding, J., Qu, Z., Gu, J., et al. (2017). PHB Associates with the HIRA Complex to Control an Epigenetic-Metabolic Circuit in Human ESCs. *Cell Stem Cell* 20, 274.e–289.e. doi: 10.1016/j.stem.2016.11.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/fcell.2021.655201

Polycomb Factor PHF19 Controls Cell Growth and Differentiation Toward Erythroid Pathway in Chronic Myeloid Leukemia Cells

Marc García-Montolio[1†], Cecilia Ballaré[1†], Enrique Blanco[1], Arantxa Gutiérrez[1], Sergi Aranda[1], Antonio Gómez[2], Chung H. Kok[3,4], David T. Yeung[3,4], Timothy P. Hughes[3,4], Pedro Vizán[1‡] and Luciano Di Croce[1,5,6*‡]*

[1] Epigenetics Events in Cancer Laboratory, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain, [2] Rheumatology Department, Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain, [3] Precision Medicine Theme, South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA, Australia, [4] Adelaide Medical School, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, SA, Australia, [5] Universitat Pompeu Fabra (UPF), Barcelona, Spain, [6] ICREA, Barcelona, Spain

OPEN ACCESS

Edited by:
José Luis Sardina, Josep Carreras Leukaemia Research Institute (IJC), Spain

Reviewed by:
Wendy Beguelin, Cornell University, United States
Angel Hernández-Hernández, University of Salamanca, Spain

***Correspondence:**
Luciano Di Croce
Luciano.DiCroce@crg.eu
Pedro Vizán
Pedro.Vizan@crg.eu

[†]These authors share co-authorship

[‡]These authors share senior authorship

Specialty section:
This article was submitted to Stem Cell Research, a section of the journal Frontiers in Cell and Developmental Biology

Received: *18 January 2021*
Accepted: *22 March 2021*
Published: *29 April 2021*

Citation:
García-Montolio M, Ballaré C, Blanco E, Gutiérrez A, Aranda S, Gómez A, Kok CH, Yeung DT, Hughes TP, Vizán P and Di Croce L (2021) Polycomb Factor PHF19 Controls Cell Growth and Differentiation Toward Erythroid Pathway in Chronic Myeloid Leukemia Cells. Front. Cell Dev. Biol. 9:655201. doi: 10.3389/fcell.2021.655201

Polycomb group (PcG) of proteins are a group of highly conserved epigenetic regulators involved in many biological functions, such as embryonic development, cell proliferation, and adult stem cell determination. PHD finger protein 19 (PHF19) is an associated factor of Polycomb repressor complex 2 (PRC2), often upregulated in human cancers. In particular, myeloid leukemia cell lines show increased levels of PHF19, yet little is known about its function. Here, we have characterized the role of PHF19 in myeloid leukemia cells. We demonstrated that PHF19 depletion decreases cell proliferation and promotes chronic myeloid leukemia (CML) differentiation. Mechanistically, we have shown how PHF19 regulates the proliferation of CML through a direct regulation of the cell cycle inhibitor p21. Furthermore, we observed that MTF2, a PHF19 homolog, partially compensates for PHF19 depletion in a subset of target genes, instructing specific erythroid differentiation. Taken together, our results show that PHF19 is a key transcriptional regulator for cell fate determination and could be a potential therapeutic target for myeloid leukemia treatment.

Keywords: chronic myeloid leukemia, polycomb, PHF19, epigenetics, erythroid differentiation

INTRODUCTION

Cell fate decisions rely on the precise control of specific transcription programs, which are governed by multiple layers of regulation. Among them, the epigenetic status of genes and their regulatory regions are paramount, not only because of their direct impact on expression but also for its reversible nature that allows a progressive fine-tuning control of expression along differentiation. The Polycomb group of proteins is one of the most important players in epigenetic regulation. They form multimeric complexes in the nucleus that associate with and modify the chromatin landscape. The Polycomb repressor complex 2 (PRC2) catalyzes the trimethylation of lysine 27 on the histone H3 N-terminal tail (H3K27me3), which is associated with chromatin compaction and gene repression (Schuettengruber et al., 2017). Apart from the PRC2 core components (EZH2,

SUZ12, and EED), several sub-stoichiometric accessory factors have been described to regulate PRC2 genomic localization and function (Vizan et al., 2015; Laugesen et al., 2019). We and others have previously characterized the role of PHF19 (a mammalian homolog of the *Drosophila melanogaster* Polycomb-like, PCL) in embryonic stem cells (Ballare et al., 2012; Brien et al., 2012; Hunkapiller et al., 2012) and, more recently, in the hematopoietic system (Vizan et al., 2020). In both cases, in normal conditions, *PHF19* expression is reduced during differentiation, indicating its potential role in maintaining stem/progenitor characteristics.

Cancer could be considered as a failure to achieve or maintain unambiguous differentiated cell fates. Leukemia is the generic name given to several types of cancers, characterized for the accumulation of undifferentiated cells called blasts, either in the blood system, in the bone marrow, or in the lymph nodes (Islam, 1992). Myeloid leukemias are a group of leukemias where the cancerous cells derive from the myeloid precursors, which in normal conditions give rise to erythrocytes, platelets, monocytes, or granulocytes. Chronic myeloid leukemia (CML) is a clonal disorder of the hematopoietic system that accounts for 15% of the newly diagnosed leukemias in adults (Jabbour and Kantarjian, 2020) and is characterized by the unregulated growth of non-functional erythroid cells and platelets in the peripheral blood, as well as marked myeloid hyperplasia in the bone marrow (Sawyers, 1999). CML was the first type of leukemia associated with a chromosomal translocation: the Philadelphia chromosome [t(9;22) (q34;q11.2)]. This translocation results in a fusion protein called BCR-ABL, which generates an anomalous tyrosine kinase activity that not only leads to the speeding of cell division but also promotes genomic instability, making the cell more susceptible to accumulate extra genetic mutations (Hehlmann et al., 2007). The specific tyrosine kinase inhibitor imatinib is nowadays the most common treatment in CML. However, almost 30% of patients develop resistance to imatinib (Kok et al., 2019). In this context, new treatments are required, and epigenetic factors have lately gained attention (Machova Polakova et al., 2013).

In the last few years, several reports have addressed the role of PHF19 in different types of cancers, including melanoma (Ghislin et al., 2012), prostate (Jain et al., 2020), glioblastoma (Deng et al., 2018), and also hematopoietic malignancies of the lymphoid branch, such as multiple myeloma (Ren et al., 2019; Mason et al., 2020), as well as its specific role in preventing T cell exhaustion (Ji et al., 2019). However, its role in myeloid leukemias has not been elucidated. Interestingly, from those reports, it is becoming clear that, although PHF19 could be considered as an enhancer of cell proliferation, the consequences of its depletion could be hazardous since it would lead to the generation of slow-growing undifferentiated cells that ultimately may increase the malignant properties of tumoral cells (Ghislin et al., 2012; Jain et al., 2020). Therefore, the role of PHF19 in myeloid leukemias, which are a clear example of the accumulation of undifferentiated progenitors, needs to be further investigated. In this study, we demonstrated not only the antiproliferative effects of targeting PHF19 in myeloid leukemias but also how its reduction leads to specific differentiation of CML cells toward the erythroid pathway.

RESULTS

PHF19 Expression Regulates Chronic Myeloid Leukemia Performance

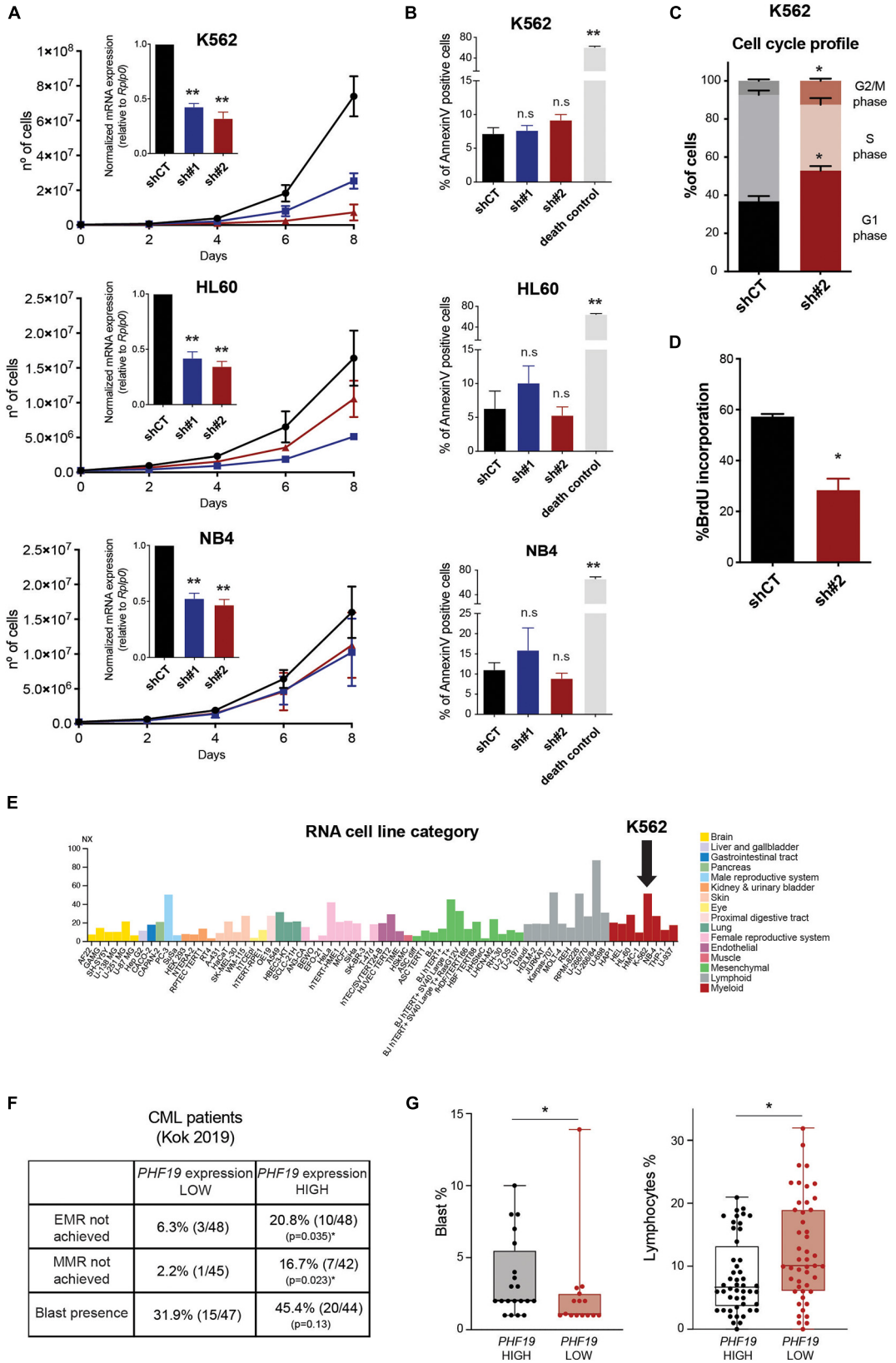
Three human myeloid-related leukemia cell lines (one of chronic myeloid leukemia origin: K562; two from acute promyelocytic leukemia: HL60 and NB4) were depleted for the long isoform of *PHF19* using two different short hairpin RNAs (shRNA) (**Figure 1A**, inner panels). In all cellular models, we observed a decrease in cell growth (**Figure 1A**), which was not due to an increase in apoptosis induction (**Figure 1B**). K562 was the most affected cell line, particularly when using sh*PHF19*#2. Of note is that K562 cell growth reduction upon EZH2 depletion using shRNAs has been reported, but accompanied by an increased apoptotic rate (Xie et al., 2016; Liu et al., 2017). Hence, we decided to profile the cell cycle by propidium iodide staining. Interestingly, a shortening in the S phase could be detected, together with an enrichment in both the G1 and G2/M phases upon *PHF19* depletion (**Figure 1C** and **Supplementary Figure 1A**): G1 increase from 36.8 $\pm$ 5.6% to 52.9 $\pm$ 4.9% and G2/M from 7.3 $\pm$ 1.5% to 12.5 $\pm$ 2.4%. This slow cell growth was confirmed by a reduction in the incorporation of the synthetic nucleoside BrdU (**Figure 1D** and **Supplementary Figure 1B**).

Among the human cancer cell lines reported in the Protein Atlas Database (**Figure 1E**; Uhlen et al., 2017)[1], myeloid malignancies showed a high level of expression for *PHF19*, with K562 falling into the top three among all cell lines analyzed. We then wondered whether PHF19 plays a significant role in CML. Thus, we analyzed the expression data from blood samples in a cohort of almost 100 CML patients who were followed after imatinib treatment (Kok et al., 2019). When segregated by *PHF19* expression (50% highest vs. 50% lowest expressing patients), a differential response to treatment was observed: a high expression of *PHF19* correlated with poorer clinical outcomes (**Figure 1F**), i.e., a higher proportion of the patients with high *PHF19* expression failed to achieve early molecular response (EMR) or multiple molecular response (MMR) after the treatment. Although not statistically significant, the presence of blasts was also proportionately increased in the *PHF19* high group of patients. Moreover, when the percentage of blasts for positive samples was examined, it was significantly greater in the *PHF19* high fraction, which was also accompanied with a decrease in the percentage of total lymphocytes in the blood taking into account the whole cohort (**Figure 1G**). In this regard, the combination of PHF19 depletion in K562 cells with a lower concentration of imatinib causes a similar effect on cell growth (**Supplementary Figure 1C**). Taken together, these results pointed to a specific role of PHF19 in CML.

PHF19 Expression Controls the Differentiation Transcription Program Toward Erythroid

To further characterize the role of PHF19 in K562 cells, we performed transcriptomic analysis after RNA isolation and

[1]https://www.proteinatlas.org/ENSG00000119403-PHF19/cell



	PHF19 expression LOW	*PHF19* expression HIGH
EMR not achieved	6.3% (3/48)	20.8% (10/48) (p=0.035)*
MMR not achieved	2.2% (1/45)	16.7% (7/42) (p=0.023)*
Blast presence	31.9% (15/47)	45.4% (20/44) (p=0.13)

FIGURE 1 | Continued

FIGURE 1 | (A) Accumulative growth curve of cells infected with an empty short hairpin RNA (shRNA, *shCT*) and two shRNAs against *PHF19* (*sh#1* and *sh#2*). The *X*-axis shows the days after puromycin selection (see section "Materials and Methods"). *Inner panels*: levels of *PHF19* expression measured by qPCR relative to the housekeeping gene *Rplp0* and normalized by shCT expression 2 days after selection. Mean ($n > 3$ for all cases) + SEM. **(B)** Apoptosis assessed by the percentage of annexin V-positive cells in shCT-, sh# 1-, and sh#2-infected cells. Cell death control: non-infected cells under puromycin selection for 3 days. Mean ($n = 2$) + SEM. **(C)** Percentage of cells in the G1, S, and G2/M phases after propidium iodide (PI) staining 4 days after selection. Mean ($n = 4$) + SEM. **(D)** Percentage of BrdU-positive cells in shCT- and sh#2-infected K562 cells 4 days after selection. Mean ($n = 3$) + SEM. **(E)** Expression levels of *PHF19* in a panel of cell lines obtained from the Protein Atlas Database (cited in the main text). **(F)** Percentage and absolute number of patients that did not achieve EMR (early molecular response) or MMR (multiple molecular response) and presented blasts at peripheral blood samples, grouped by *PHF19* levels. **(G)** Boxplots and individual values of the percentage of blasts in patients where they could be detected (*left panel*) and percentage of lymphocytes (*right panel*) in peripheral blood samples. $*p < 0.05$; $**p < 0.01$.

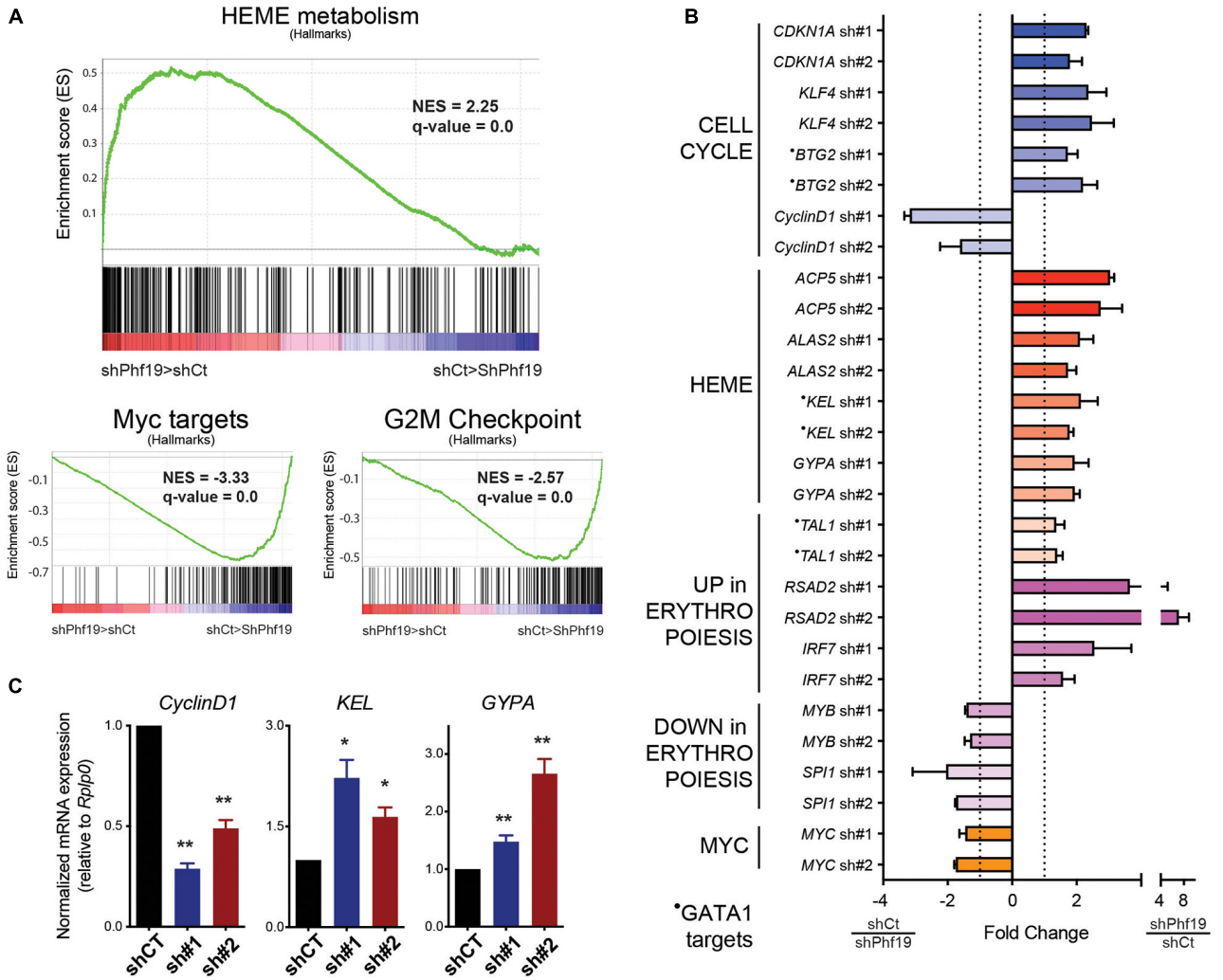


FIGURE 2 | (A) Gene set enrichment analysis (GSEA) showing positive enrichment in the *PHF19*-depleted transcriptome for *Heme metabolism* and negative enrichment for *Myc targets* and *G2/M checkpoint* (hallmarks). **(B)** Normalized fold change expression (using reads per kilobase of transcript per million mapped reads, RPKMs) for the two short hairpin RNAs (shRNAs) against PHF19 (sh#1 and sh#2) with respect to cells infected with shCT (empty shRNA) in relevant genes from two independent RNA-seq experiments. Mean + SEM. **(C)** Expression of cyclin D1 (*CCND1*), *KEL*, and *GYPA* in sh#1 and sh#2 K562 cells, measured by qPCR relative to the housekeeping gene *RPLP0* and normalized by shCT expression. Mean ($n > 3$ in all cases) + SEM. $*p < 0.05$; $**p < 0.01$.

massive sequencing (RNA-seq) in control cells and upon *PHF19* depletion. Both the shRNAs applied in **Figure 1A** were used for two independent replicates, and the expression differences were ranked using DESeq2 (Love et al., 2014) for gene set enrichment analysis (GSEA) (Subramanian et al., 2005). Unbiased GSEA pre-ranked analysis using hallmark signatures (Liberzon et al., 2015) rendered *Heme metabolism* as the top enriched category in depleted cells (**Figure 2A**). Interestingly, the K562 cell line could be potentially differentiated toward erythroid cell fate. Among the enriched categories in the control cells, we found *Myc targets*, whose decrease has also been clearly documented for erythroid differentiation (Jayapal et al., 2010), and *G2/M*

checkpoint (**Figure 2A**), which correlated with the cell cycle arrest observed in **Figures 1A–D**.

We further analyzed our data by selecting the top 300 upregulated and downregulated genes according to DESeq2 ranking and performed Gene Ontology (GO) using the Enrichr web tool (Kuleshov et al., 2016). Among the categories we found in the upregulated group were *regulation of G1/S transition of mitotic cell cycle* (*GO:2000045*) (GO_Biological_Process, $p = 0.031$), *negative regulation of mitotic cell cycle* (*GO:0045930*) (GO_Biological_Process, $p = 0.035$), *Heme Biosynthesis WP561* (WikiPathways, $p = 0.0075$), *GATA1 CHEA* (ENCODE_and_ChEA Consensus_TFs_from_ChIP-X, $p = 0.019$), and *CD71 + EarlyErythroid* (Human_Gene_Atlas, $p = 0.02$). Conversely, *MYC ENCODE* (ENCODE_and_ChEA Consensus_TFs_from_ChIP-X, $p = 0.00000025$), and *CD34*$^+$ (Human_Gene_Atlas, $p = 0.00012$) were found in the downregulated group. Examples of the upregulated and downregulated genes are depicted in **Figure 2B**, and some were validated in independent experiments (**Figure 2C**). Cell cycle categories were found in the upregulated group of genes, in agreement with the data reported in **Figures 1A–D** and with the GSEA analysis (**Figure 2A**). Of note is that GATA1 transcriptional activity is required for erythroid differentiation (Tanimura et al., 2016). Contrary to *Myc*, which is directly downregulated upon *PHF19* depletion, *GATA1* expression was not directly affected upon *PHF19* knockdown, but many of its targets were found upregulated, mimicking its effects. The complementary upregulation on genes assigned to CD71$^+$ cells and the downregulation of CD34$^+$, which is a marker of undifferentiated cells (Orfao et al., 2019), are consistent with the role of PHF19 maintenance of undifferentiated cell status (Ballare et al., 2012). Among the upregulated genes, we also noted *Acp5*, whose expression is required for fetal erythroid differentiation (Ying et al., 2014); *Gypa*, which translates in the surface marker CD235a widely used for functional characterization of erythroid commitment (Fajtova et al., 2013; Li et al., 2014); as well as other markers [both upregulated (*Rsad2* and *Irf7*) and downregulated (*Myb* and *SpiI*)] used for erythroid cell determination (Liang et al., 2015). In sum, transcriptomic profiling indicated that PHF19 plays an important role in K562 cell fate determination.

PHF19 Regulates p21 Expression in CML

Since PHF19 is an epigenetic factor associated with PRC2, we hypothesize that its chromatin profile could offer valuable clues to understanding its role in regulating cell cycle and differentiation. To determine PHF19 localization on chromatin, we performed chromatin immunoprecipitation followed by massive sequencing (ChIP-seq) in control and *PHF19*-depleted K562 cells, together with IgG as the ChIP control (**Figure 3A**). We identified 2,328 PHF19 peaks genome-wide, significantly enriched in the promoter regions (**Supplementary Figure 2A**), corresponding to 1,297 target genes (**Figure 3A**). Several target genes (depicted in **Supplementary Figure 2B**) were further validated by ChIP-quantitative PCR (qPCR) in independent experiments (**Figure 3B**). Then, we defined bona fide PRC2 targets by intersecting EZH2, SUZ12, and H3K27me3 target genes from the K562 ChIP-seq data released by the ENCODE project (Consortium, 2012; **Figure 3C**, top panel). More than 90% of the PHF19 target genes were shared with PRC2 (**Figure 3C**, bottom panel), indicating a very low impact of PHF19 outside its canonical function as associated sub-stoichiometric Polycomb factor. We then reasoned that, as part of PRC2, PHF19 may have a general role as a transcriptional repressor. Indeed, the GSEA of the PHF19 gene targets showed significant enrichment in the upregulated genes upon depletion (**Figure 3D**). Among those, *CDKN1A* (p21) called our attention and was further validated by ChIP-qPCR around its transcription start site (**Figure 3E**). Moreover, we observed an increased expression at the RNA and protein levels in the *PHF19* knockdown conditions (**Figures 3F,G**). We thus wondered whether the PHF19-dependent control of p21 was a consistent mechanism for cell cycle modulation at the bone marrow of CML patients, where most cycling cells are found. Remarkably, using published data (Abraham et al., 2016), we corroborated the anticorrelation between *Cdkn1a* and *PHF19* expressions (**Figure 3H**).

Specific MTF2 Compensation Upon PHF19 Reduction Defines PRC2 Role in K562

Overexpression of p21 may not be sufficient to induce differentiation toward erythroid fate in K562. Recently, the depletion of the PHF19 homolog MTF2 has been linked to erythropoiesis in a knockout mouse model (Rothberg et al., 2018). These two factors may be competing for interaction with the PRC2 core components. Although the *MTF2* levels did not change upon *PHF19* depletion, we wondered whether MTF2 occupancy would be affected by changes in PHF19 expression. Thus, we performed MTF2 ChIP-seq in *PHF19*-depleted cells and found that: (i) most of the PHF19 targets overlapped with the MTF2 targets (**Figure 4A**, and **Supplementary Figure 2C**), indicating redundancy not only for the complex formation but also for the same genomic regions; (ii) MTF2 binding was more spread across the genome; and (iii) there was an increase in the number of target genes upon *PHF19* depletion (**Figure 4A**). In fact, this increase was also noticeable in the ChIP-seq signal strength upon *PHF19* knockdown, which we validated for several genomic targets (**Figure 4B** and **Supplementary Figure 2D**). However, the observed general derepression of PHF19 target genes upon depletion of *PHF19* seemed incompatible with MTF2 compensation. Certainly, the expression changes are not homogeneous (**Figure 3D**), and we hypothesized that, although MTF2 might be able to compensate the lack of PHF19 in many target genes, there is a subset of genes where the reduction of PHF19 directly leads to the upregulation of expression. To investigate this possibility, we selected the top 200 ranked upregulated PHF19 target genes and analyzed changes in the MTF2 levels upon *PHF19* knockdown. We corroborated that the fold change expression of these selected genes was consistent with the DESeq2 ranking (**Figure 4C**, top left panel). Interestingly, the MTF2 levels did not significantly change for those targets that were highly derepressed. In contrast, a significant increase of the MTF2 signal was observed in the rest of the targets (**Figure 4C**, bottom

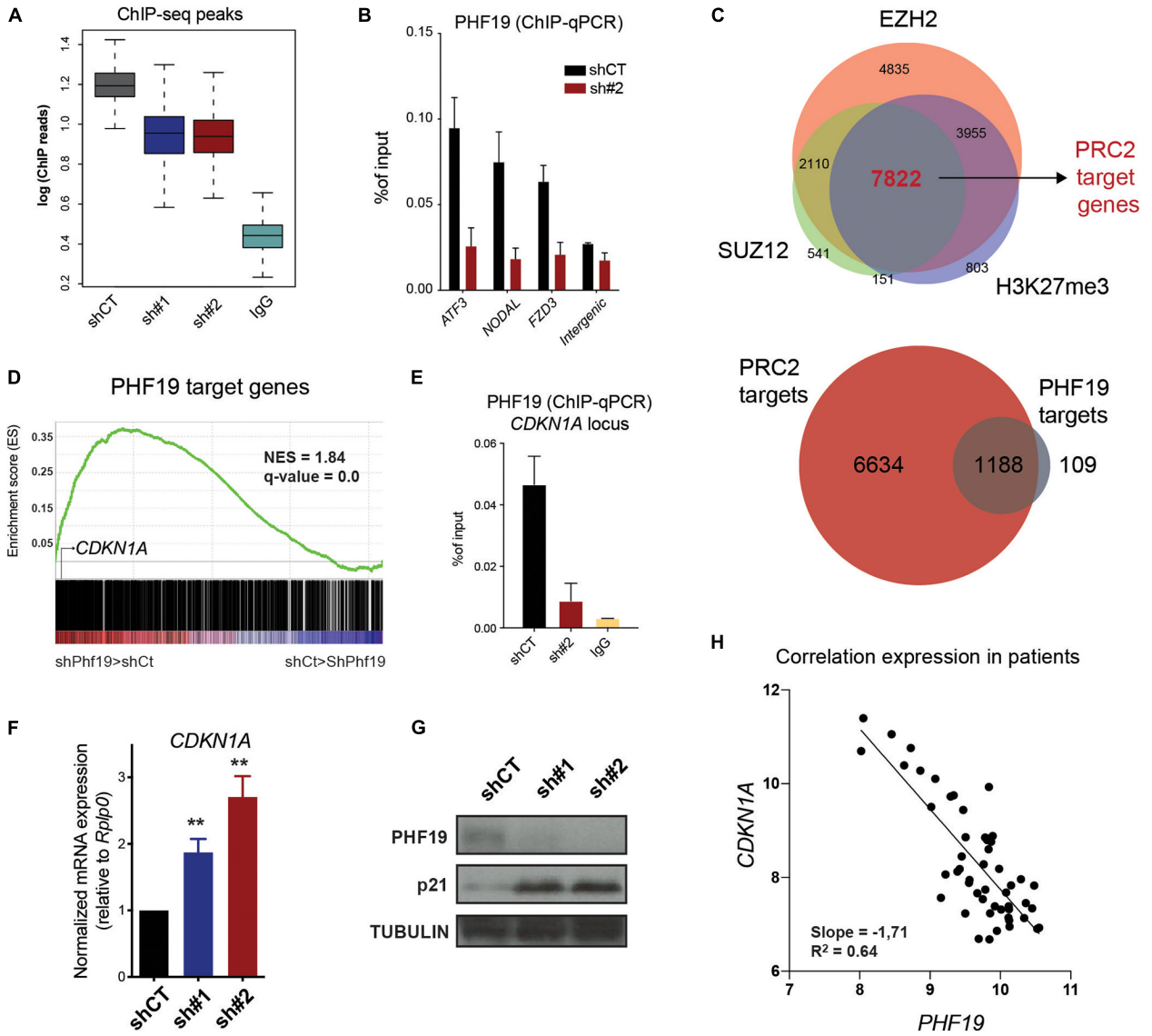


FIGURE 3 | (A) Boxplot of the total number of reads in PHF19 chromatin immunoprecipitation sequencing (ChIP-seq) peaks. **(B)** PHF19 ChIP enrichment (with respect to the input) measured by qPCR of several genes and intergenic region as a negative control for the cells infected with an empty short hairpin RNA (shRNA, *shCT*) and a shRNA against PHF19 (sh#2). Mean ($n > 2$) + SEM. **(C)** Venn diagrams showing the intersection between the EZH2, SUZ12, and H3K27me3 target genes (*top*) and between these common targets and PHF19 target genes (*bottom*). **(D)** Gene set enrichment analysis (GSEA) showing positive enrichment in the *PHF19*-depleted transcriptome for the PHF19 target genes. **(E)** PHF19 ChIP enrichment (with respect to the input) measured by qPCR of the *CDKN1A* gene for shCT- and sh#2-infected K562 cells, as well as ChIP using IgG in shCT cells. **(F)** Expression of *CDKN1A* measured in sh#1- and sh#2-infected K562 cells, measured by qPCR relative to the housekeeping gene *RPLP0* and normalized by the expression of shCT-infected cells. Mean ($n = 8$) + SEM. **(G)** Western blot for PHF19 and p21 of shCT-, sh# 1-, and sh#2-infected K562 cells 5 days after selection; TUBULIN is used as a loading control. **(H)** Scatter plot showing the correlation of *PHF19* and *CDKN1A* microarray expression values measured from all bone marrow precursors of chronic myeloid leukemia (CML) patients (cited in the main text).
$^{**}p < 0.01$.

left panel). Moreover, we reasoned that the absence of PHF19 would allow an increased interaction of MTF2 with the PRC2 core components, which would affect its function beyond PHF19 targets. Therefore, we performed the same analysis for the top 200 MTF2 target genes upregulated upon *PHF19* depletion. Similarly, the derepressed genes did not show an increase in MTF2 levels, as observed and expected for the rest of the target genes (**Figure 4C**, right panels).

These results suggested a distinct behavior of MTF2 compensation after *PHF19* reduction. To explore what caused this difference irrespective of the expression, we selected the top 200 MTF2 targets that showed the stronger increase of MTF2 occupancy upon *PHF19* depletion (**Figure 4E**, left panel). Firstly, we studied the GO of these selected targets, and strikingly, the top two categories obtained by GO were the *Wnt signaling pathway* and *regulation of canonical Wnt signaling*. Wnt-related categories

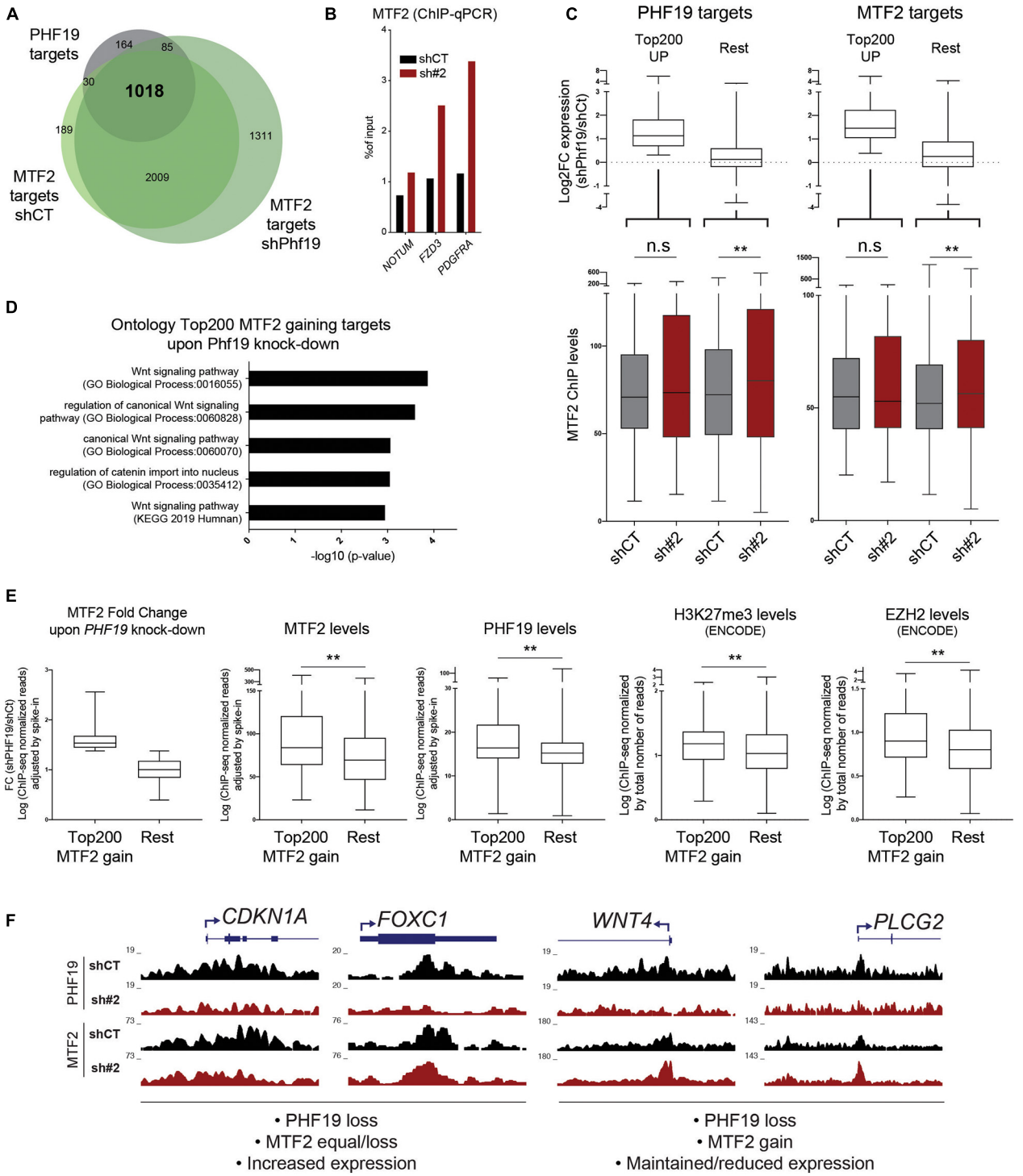


FIGURE 4 | (A) Venn diagram showing the intersection between the PHF19 gene targets and the MTF2 gene targets for cells infected with an empty short hairpin RNA (shRNA, *shCT*) and a shRNA against PHF19 (*sh#2*). **(B)** MTF2 chromatin immunoprecipitation (ChIP) enrichment (with respect to the input) measured by qPCR of several genes and intergenic region as a negative control in shCT- and sh#2-infected K562 cells. **(C)** *Top*: Fold change of sh*PHF19* vs. shCT (Log2) of the top 200 upregulated vs. the rest of the PHF19 (*left*) and MTF2 (*right*) gene targets. *Bottom*: Levels (maximum peak high) of the MTF2 ChIP signal in shCT- and sh#2-infected K562 cells of the top 200 upregulated vs. the rest of the PHF19 (*left*) and MTF2 (*right*) gene targets. **(D)** Gene ontology of the top 200 MTF2 targets that increase their levels upon *PHF19* depletion. **(E)** Characterization of the top 200 MTF2 targets that increase their levels upon *PHF19* depletion versus the rest of the MTF2 levels. From *left* to *right*: Fold change of MTF2 gain and levels (maximum peak high) of MTF2, PHF19, H3K27me3, and EZH2. **(F)** Representative screenshots modified from the UCSC Genome Browser. $^{**}p < 0.01$.

were also found significantly enriched when pathways (Kyoto Encyclopedia of Genes and Genomes, KEGG) were queried (**Figure 4D**). Of note is that, in the GO of the 200 lower PHF19 target genes ranked by expression upon *PHF19* depletion, we could also detect Wnt-related categories: *Regulation of Wnt signaling pathway* (GO_Biological_Process, $p = 0.00079$) and *Wnt signaling pathway* (KEGG, $p = 0.0052$). This is concordant with the reported MTF2 role in hematopoiesis since repression of the Wnt signaling pathway by MTF2 instructs erythroid differentiation (Rothberg et al., 2018). To gain insights into why MTF2 is increasingly deposited in a subset of targets, we studied their epigenetic status in control conditions: as depicted in **Figure 4E**, the ChIP-seq levels of MTF2, PHF19, H3K27me3, and EZH2 were higher in the top 200 MTF2 target genes with respect to the rest, indicating they were significantly occupied by PRC2 prior to *PHF19* depletion. Finally, **Figure 4F** depicts representative examples of the chromatin profiles for PHF19 and MTF2 upon *PHF19* depletion. As could be observed, *CDKN1A* or *FOXC1* (transcriptionally upregulated upon *PHF19* depletion) did not display an MTF2 increase, contrary to what could be observed in the Wnt-related genes *WNT4* (transcriptomically unaffected) and *PLCG2* (transcriptionally downregulated).

PHF19 Depletion Enhances Erythroid Differentiation While Impeding Megakaryocyte Cell Fate Induction

Throughout the experiments with shRNAs, we noticed that cell pellets from the *PHF19*-depleted cells acquired a pale reddish color (**Figure 5A**), possibly indicating a hasty production of hemoglobin. Therefore, we reasoned that, beyond epigenetic and transcription indications toward erythroid differentiation, K562 cells were already acquiring phenotypic erythroid characteristics. To assess this, we analyzed by flow cytometry the presence of the well-known erythroid precursor marker CD235a (Fajtova et al., 2013; Li et al., 2014) using cells treated with Ara-C (cytarabine) as a positive control (Zhang et al., 2007; Cai et al., 2014). Both shRNAs caused an increase of CD235a (**Figure 5B**). To corroborate that the effect on differentiation upon *PHF19* depletion is specific, we forced the differentiation of K562 cells into the megakaryocyte lineage by phorbol myristate acetate (PMA) treatment (Cai et al., 2014) after *PHF19* depletion and measured the megakaryocyte surface marker CD61 (Ogino et al., 2014). Indeed, the cells with reduced levels of *PHF19* were resistant to acquire megakaryocyte characteristics (**Figure 5C**). Finally, Ara-C is used as a treatment in CML and other leukemias[2], and according to our results, reduction of the PHF19 levels could be proposed for cooperative CML treatments. To test this, we reduced a hundred times the concentration of Ara-C used as a positive control in **Figure 5B** to produce a more modest increase in the CD235a marker, and we measured differentiation and cell growth upon PHF19 depletion. Interestingly, Ara-C cell growth inhibition was enhanced by PHF19 depletion (**Supplementary Figure 2E**); more interestingly, it was able to further enhance the differentiation of K562 cells in the presence of low Ara-C-treated cells (**Figure 5D**).

[2]https://www.drugs.com/ppa/cytarabine-conventional.html

DISCUSSION

Here, we have shown that reduced PHF19 levels in CML cells arrest the cell cycle and promote differentiation toward erythroid fate. The role of PRC2 in leukemia has been studied (Carlo et al., 2019), and in particular, the catalytic PRC2 core component EZH2 has been reported to be overexpressed in CML (Xie et al., 2016). Moreover, EZH2 inhibition reduces cell growth and sensitizes CML cells to tyrosine kinase inhibitors (Scott et al., 2016; Xie et al., 2016). Furthermore, in line with our results, it has been recently reported that the oncogene MYCN regulates EZH2 expression in CML cells, which leads to p21 repression, increasing proliferation and blocking differentiation (Liu et al., 2017). However, due to its pleiotropic functions, targeting the PRC2 core components may disrupt many cellular functions, and thus focusing on the sub-stoichiometric accessory factors, such as PHF19, may be an advantage. A fine example of this was our recent study on the role of PHF19 in normal mouse hematopoiesis: previous loss-of-function studies of PRC2 core components demonstrated their essential role in hematopoiesis (Xie et al., 2014; Lee et al., 2015; Yu et al., 2017), but the associated lethality had hampered an in-deep characterization of the transcriptional pathways or H3K27me3 (re)distribution in hematopoietic stem cells (HSCs). We generated a *Phf19* knockout mice, which resulted viable and then allowed us to unveil its role in controlling adult HSC dormancy (Vizan et al., 2020). Similarly, this study has allowed us to determine a very specific epigenetic mechanism by which high levels of PHF19 impede erythroid differentiation, which could have been impossible to evaluate by blocking the entire PRC2 activity.

To achieve differentiated characteristics, the cell cycle has to be arrested. The CML cell line K562 harbors the BCR-ABL fused tyrosine kinase, which activates the signaling cascades in charge of cell cycle regulation. At the same time, p53 is truncated and is not functional in these cells (Law et al., 1993), leaving p21 as one of the few factors able to avoid uncontrolled cell proliferation. We demonstrated that PHF19 occupies the p21 promoter, and its depletion coincides with p21 derepression, pointing to a plausible mechanism of how PHF19 overexpression impacts on cell growth. Surprisingly, a recent study has reported no effects on cell proliferation in K562 upon *PHF19* depletion (Ren et al., 2019). Nonetheless, we hypothesize that this may be due to different targeting and/or reduction efficiency. On the other hand, p21 overexpression may account for cell cycle arrest, but it might not be enough to induce the specific erythroid differentiation reported. The K562 cell line has been previously described as erythroleukemia (Chylicki et al., 2000), but it also has been largely known as a model for megakaryoblast differentiation (Alitalo, 1990). In fact, it has been previously shown that p21-induced cell cycle arrest favors megakaryocyte differentiation (Munoz-Alonso et al., 2005). Therefore, additional cellular and molecular mechanisms might be triggered upon *PHF19* depletion.

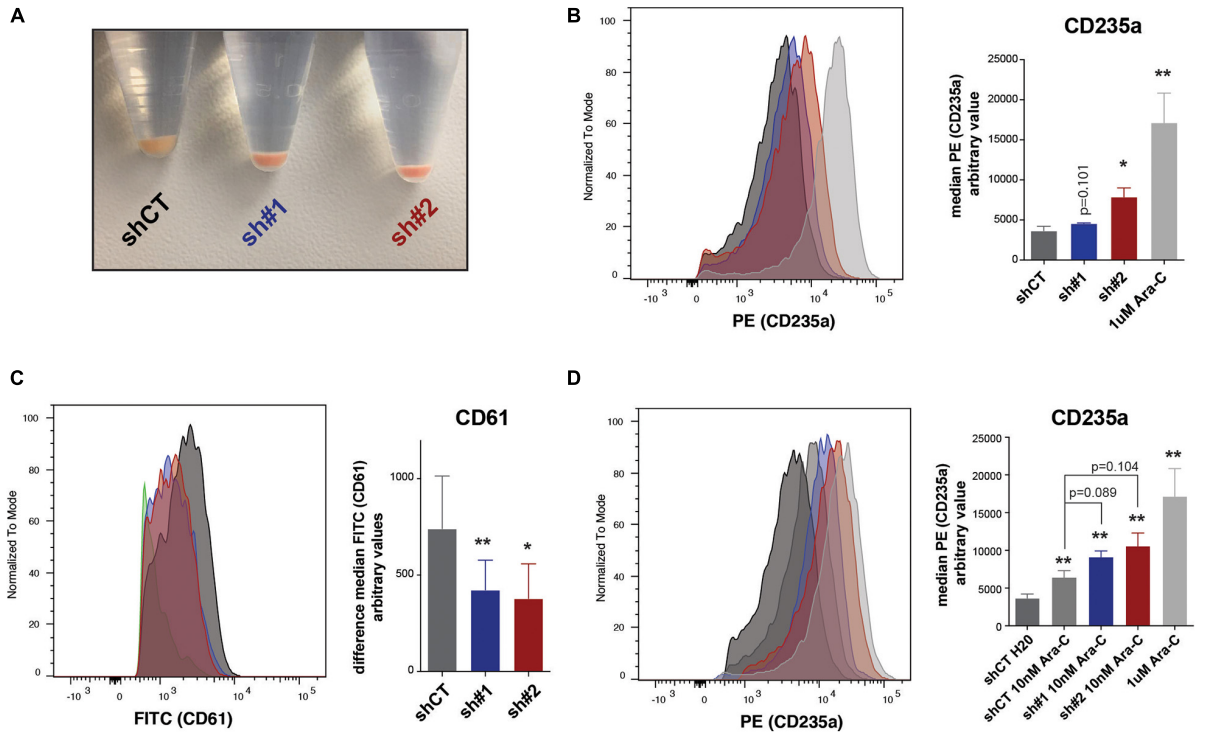


FIGURE 5 | (A) Picture of pelleted K562 infected with an empty short hairpin RNA (shRNA, *shCT*) and two shRNAs against PHF19 (*sh*#1 and *sh#2*) 5 days after puromycin selection. **(B)** CD235a marker levels measured by flow cytometry in shCT-, sh# 1-, sh#2-infected cells 4 days after puromycin selection and 1 μM of Ara-C treatment in uninfected cells as a positive control. *Left*, representative plot of a single experiment. *Right*, bar plot of the mean of median cytometry values (shCt and sh#1, $n = 4$; sh#1 and 1 μM of Ara-C, $n = 3$) + SEM. **(C)** CD61 marker levels measured by flow cytometry in shCT-, sh# 1-, and sh#2-infected cells 4 days after puromycin selection and treated with 1 nM phorbol myristate acetate (PMA) for 24 h. *Left*, representative plot of a single experiment. *Right*, bar plot of the mean of median cytometry values ($n = 3$) + SEM. **(D)** CD235a marker levels measured by flow cytometry at 4 days in shCT-, sh# 1-, and sh#2-infected cells after puromycin selection and treated with 10 nM of Ara-C for 72 h, with 1 μM of Ara-C treatment in uninfected cells as a positive control. *Left*, representative plot of a single experiment. *Right*, bar plot of the mean of median cytometry values (shCt, 10 nM Ara-C; sh#1, 10 nM Ara-C, $n = 24$; sh#1, 10 nM Ara-C, and 1 μM of Ara-C, $n = 3$) + SEM. $^*p < 0.05$; $^{**}p < 0.01$.

We noticed that, although the main function of PHF19 is transcriptional gene repression, a subset of the PHF19 target genes remain unaffected or are even more repressed upon its depletion. In this sense, a recent report has demonstrated that another PCL homolog, MTF2, is required for normal erythropoiesis (Rothberg et al., 2018). We have studied the occupancy of MTF2 in response to *PHF19* depletion, and we have observed that it differentially compensates the PHF19 loss in a subset of targets. Interestingly, the gain of MTF2 is more pronounced among those genes whose expressions are non-derepressed, including the Wnt pathway, which remarkably needs to remain repressed in order to achieve erythroid characteristics (Rothberg et al., 2018). Some of these genes reduce their expressions, but also a part of them remains unaffected, probably because their expression levels were already low to ensure further differentiation steps.

Which other chromatin or regulatory features determine where in the genome MTF2 is compensating for PHF19 reduction remains to be comprehensively elucidated. Worthy of note is our observation that those genes in which MTF2 did not compensate for PHF19 loss displayed reduced levels of the PRC2 components, which led us to hypothesize that repressive chromatin status is warranted specifically in a subset of genes to ensure cell differentiation. Therefore, we foresee a model where high levels of PHF19 in CML cancer cells would lead to small albeit enough accumulation in other genomic targets. This would induce a degree of unexpected gene repression, heightening the already disturbed normal differentiation. In other words, a gain in *PHF19* expression would intensify the resemblance to myeloid precursors, where cell fate has not been yet decided. In conclusion, this study confirms the necessity of maintaining tight control of epigenetic regulation to sustain proper adult stem cell differentiation as well as reinforces the possibility of using specific acquired epigenetic vulnerabilities for tumor differentiating therapies.

MATERIALS AND METHODS

Cell Culture and Growth Curves

NB4, HL60, and K562 cells were cultured at 37°C and 5% of CO_2 in RPMI medium supplemented with 10% fetal bovine serum. HEK293T cells were cultured at 37°CC and 5% CO_2

in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum.

To monitor cell growth, leukemic cell lines were seeded at 2.5×10^5 cells/ml for the NB4 and HL60 cell lines and 3×10^5 cells/ml for the K562 cell line, then the cells were counted and seeded again every other day. Imatinib (Thermo Fisher Scientific) and Ara-C (Sigma) were added after puromycin selection and cell growth monitored for 6 days.

Lentiviral Production and Infection

The PLKO.1 lentiviral system was used for the production of shRNAs against *PHF19*. The target sequences were AAGCTTCCATCCACATGTGTT for shRNA#1 and GCCACACATTTGAGAGCATCA for shRNA#2. Empty vector was used as the control (shCT). Viral particles were produced in HEK293T, which were plated at a density of 2×10^6 cells in a p10 plate and transfected the following day by adding dropwise while vortexing a $CaCl_2$–DNA solution (10 μg of pLKO.1 of plasmid, 6 μg of pCML-dR8.91, 5 μg of pCMV-VSGV, and 62 μl of 2 M $CaCl_2$ in a final volume of 0.5 ml) previously incubated at room temperature (RT) for 15 min to an equal volume of HBS 2 × (HEPES-buffered saline solution, pH 7.05, 0.28 NaCl, 0.05 M HEPES, and 1.5 mM Na_2HPO_4). After cell incubation for 14–16 h, the transfection medium was replaced by fresh medium and the cells were incubated for 24 h. The medium with the lentiviral particles was harvested and filtered through a 45-μm filter. Then, fresh medium was added to the HEK293T and the media with lentiviral particles were harvested again the following day.

Two rounds of leukemic cell infection were performed on six-well plates according to the days the medium was harvested. Of the cells, 5×10^5 were plated in 1 ml of medium, and then 1 ml of the medium with lentivirus was added. Cells were spinoculated (1,000 × *g*, 90 min, 32°C) in the presence of protamine sulfate (1 μg/ml). After the two rounds of infection, the cells were selected with 1 μg/ml of puromycin.

Apoptosis, Cell Cycle, and BrdU Incorporation

The cell apoptosis assay was performed using violet annexin V/Dead Cell Apoptosis Kit (Invitrogen) according to the manufacturer's protocol. Post-staining, the cells were analyzed by flow cytometry. For cell cycle, 1×10^6 cells were washed with cold phosphate-buffered saline (PBS) and resuspended in 0.9 ml of EDTA-PBS (5 mM EDTA). Then, the cells were permeabilized by adding 2.1 ml of 100% cold ethanol dropwise while the mixture was softly shaken. The cells were kept at 4°C until the following day, when they were resuspended in propidium iodide (PI) staining buffer that contains 955 μl PBS + 30 μl of solution A (38 μl of 0.5 M sodium citrate + 562 μl of 500 μg/ml propidium iodide) + 2 μl of RNAse A (Thermo Fisher). The cells then were incubated at 37°C for 1 h and analyzed by flow cytometry. Cell cycle profiles were analyzed with the ModFit LT™ software. For BrdU assay, the cells were treated with 10 μM of BrdU solution for 30 min and processed using the BrdU Flow Kit (BD Pharmingen) according to the manufacturer's protocol. The percentage of BrdU-positive cells was analyzed by flow cytometry.

Cell Differentiation by Surface Markers

The K562 cells either in normal conditions or after Ara-C (Sigma) or PMA (Sigma) treatment were rinsed twice with PBS and incubated for 45 min with conjugated antibodies against CD235a-PE (Invitrogen #12-9987-82) or CD61-FITC (eBioscience #11-0619-42) at 4°C, protected from light. After washing twice with cold PBS, the cells were suspended in PBS with DAPI to discard the non-viable cells and then analyzed by flow cytometry. Analysis was performed using FlowJo software.

Western Blot

Four days after selection, infected (shCT, sh#1, and sh#2) cell pellets were washed twice with PBS and resuspended in hypotonic buffer (10 mM Tris–HCl, pH 7.4, 10 mM KCl, and 15 mM $MgCl_2$) in the presence of protease and phosphatase inhibitors. After 10 min of incubation on ice, the resuspension was centrifuged (700 × *g*) for 5 min at 4°C. The pellet was resuspended in nuclear lysis buffer (300 mM NaCl, 50 mM HEPES, pH 7.5, 0.5% NP40, and 2.5 mM $MgCl_2$) in the presence of protease and phosphatase inhibitors and benzonase (50 U/500 μl of buffer). The resuspension was centrifuged (16,000 × *g*) for 30 min at 4°C and the supernatant was considered the protein extract of the nuclear protein. Protein concentration was quantified by the Bradford assay (Bio-Rad). Seventy micrograms of protein was diluted in 5 × Laemmli buffer and heated for 5 min at 100°C. Protein samples were loaded in a NuPAGE™ 4–12% Bis-Tris precast gel (Invitrogen). The proteins were transferred onto nitrocellulose membranes at 300 mA for 70 min at 4°C, blocked with 5% milk in TBS-Tween (10 mM Tris–HCl, pH 7.5, 100 mM NaCl, and 0.1% Tween-20) for 30 min at room temperature, and incubated at 4°C in TBS-Tween 5% milk overnight with the following primary antibodies: PHF19 (Cell Signaling #77271), p21 (Cell Signaling #2947), and TUBULIN (Abcam #7291). The following day, the membranes were washed with TBS-Tween followed by incubation of the secondary antibody conjugated to horseradish peroxidase (1:5,000) TBS-Tween for 1 h at room temperature. Then, the membranes were washed twice with TBS-Tween at room temperature. The proteins were then detected with enhanced chemiluminescence reagent (Pierce ECL Western Blotting Substrate, Thermo Scientific).

Gene Expression

RNA was extracted with the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol.

Expression by qPCR

cDNA was generated from 1 μg of RNA with the First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's instructions. Real-time PCR (qPCR) reactions were performed using the SYBR Green I PCR Master Mix (Roche) and the Roche LightCycler 480. The primers used were the following: *PHF19* Fw (CAGCAGAAAAGGCGAGTTTATAG), *PHF19* Rv (CTCCAGGCTGAGGTGAAGTC); *CCND1* Fw (GCCGA

GAAGCTGTGCATC), *CCND1* Rv (CCACTTGAGCTTG TTCACCA); *KEL* Fw (ACCATGGGGAGACTGTCCT), *KEL* Rv (GGGCTTCCTACACATCACCT); *GYPA* Fw (CAA ACGGGACACATATGCAG), *GYPA* Rv (TCCAATAACACCAG CCATCA); *CDKN1A* Fw (CAGCTGCCGAAGTCAGTTCC), *CDKN1A* Rv (GTTCTGACATGGCGCCTCC).

RNA-Seq

RNA samples were quantified and the quality evaluated using Bioanalyzer. Libraries were prepared at the UPF/CRG Genomics Unit using 1 μg total RNA and sequenced using the Illumina HiSeq2000 sequencer. RNA-seq reads were mapped against the hg19 human genome assembly using TopHat (Trapnell et al., 2009) with the option –g 1 to discard those reads that could not be uniquely mapped in just one region. DESeq2 (Love et al., 2014) was run to quantify the expression of every annotated transcript using the RefSeq catalog of exons and to identify each set of differentially expressed genes. Reads per kilobase of transcript per million mapped reads (RPKMs) were used for boxplots and to calculate the fold change differences between conditions. GSEA of the pre-ranked lists of genes by DESeq2 stat value was performed with the GSEA software (Subramanian et al., 2005).

Chromatin Immunoprecipitation and Analysis

Cells (25×10^6) were harvested and washed twice with PBS and cross-linked in two steps. Firstly, the cells were resuspended in 10 ml of PBS 1 mM $MgCl_2$ and 40 μl of ChIP Cross-link Gold (Diagenode) and incubated, shaking for 30 min at RT. Secondly, the cells were resuspended in 1% formaldehyde for 10 min at RT and fixation stopped by adding glycine to a final concentration of 0.125 M, then incubated for 5 min at RT. Then, the cells were washed twice with cold PBS. Chromatin preparation and ChIP experiments were performed with the ChIP-IT High Sensitivity Kit from Active Motif (#53040) according to the manufacturer's instructions. ChIPs were performed using 5 μg/ChIP of the following antibodies: PHF19 (Cell Signaling #77271), MTF2 (ProteinTech 16208-1-AP), and control IgG (Abcam #ab172730). For spike-in control, an equal amount of *D. melanogaster* S2 cell chromatin was added to each ChIP reaction (0.1% of the K562 cell chromatin), together with 1 μg of an antibody against a *Drosophila*-specific histone variant, H2Av (Active Motif, #61686).

ChIP Quantification by qPCR

Real-time PCR (qPCR) reactions were performed using the SYBR Green I PCR Master Mix (Roche) and the Roche LightCycler 480. The primers used were the following: *ATF3* Fw (GTGGGTGGTCTGAGTGAGGT), *ATF3* Rv (CACAGTT TGGTAATTTGGGGTAG); *NODAL* Fw (GCGACTTCCTTAC TCGACCTC), *NODAL* Rv (CACAGTTTGGTAATTTGGGGT AG); *FZD3* Fw (AAAAGCACGTGCCATGAAT), *FZD3* Rv (CCTCCTTCATGGAGCCAGT); *CDKN1A* Fw (ATGTCATCC TCCTGATCTTTTCA), *CDKN1A* Rv (AGAATGAGTTGGCA CTCTCCAG); *NOTUM* Fw (CCGAGGCTGGGCTTATTT), *NOTUM* Rv (GGGAAGAAAAGGCGATGC); *PDGFRA* Fw (GGGGTGTCAGTTACAGAAGGTCT), *PDGFRA* Rv (CTGCCTGGATTAAAGTGTTAGGG); *INTERGENIC* Fw (ACAGGATAAAGTTGGCATAACCA); *INTERGENIC* Rv (CAACAAAACCGTTTGGAATACAT).

ChIP-Seq

For ChIP-seq experiments, library preparation was performed from 2–10 ng of precipitated chromatin at the UPF/CRG Genomics Unit. The libraries were sequenced using the Illumina HiSeq2000 sequencer. ChIP-seq reads containing spike-in were mapped against a synthetic genome constituted by human and fruit fly chromosomes (hg19 + dm3) using Bowtie with the option -m 1 to discard the reads that did not map uniquely to one region (Langmead et al., 2009). MACS was run with the default parameters, but with the shift size adjusted to 100 bp to perform the peak calling against the corresponding control sample (Zhang et al., 2008). In the PHF19 ChIP-seqs, only peaks with tags >70 were considered as positive. The genome distribution of each set of peaks was calculated by counting the number of peaks fitted on each class of region according to RefSeq annotations. Promoter is the region between 2.5 Kbp upstream and 2.5 Kbp downstream of the transcription start site (TSS). Genic regions correspond to the rest of the gene (the part that is not classified as promoter), and the rest of the genome is considered to be intergenic. Peaks that overlapped with more than one genomic feature were proportionally counted the same number of times. Each set of target genes was retrieved by matching the ChIP-seq peaks in the region 2.5 Kbp upstream of the TSS until the end of the transcripts as annotated in RefSeq. The signal strength or ChIP-seq level was calculated as the maximum high of peaks within the same region normalized by the fly spike-in number of reads of the same experiment. For EZH2, SUZ12, and H3K27me3 ENCODE data, raw reads and peaks were downloaded from GEO series GSE29611 (GSM1003576, GSM1003545, and GSM733658). The UCSC Genome Browser was used to generate the screenshots of each group of experiments along the manuscript (Kent et al., 2002).

Data Availability

The datasets generated and analyzed for this study can be found in the National Center for Biotechnology Information Gene Expression Omnibus (Barrett et al., 2013) repository under the accession number GSE164804.

Statistics

The number of replicates for each experiment is detailed in the corresponding figure legends or main text. For *PHF19* levels, apoptosis, cell cycle, BrdU incorporation, and qPCR expression data, paired *t* test was used. For lymphocyte counts and ChIP-seq levels, unpaired *t* test was used. For the patient data in **Figure 1F**, Fisher's exact test was used. For blast counts in positive patient samples, the Mann–Whitney test was used. For the ratio of CD235a and CD61 levels, paired *t* test was used. Significance was set as $^*p < 0.05$; $^{**}p < 0.01$ throughout the study.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be found in the National Center for Biotechnology Information Gene Expression Omnibus (Barrett et al., 2013) repository under the accession number GSE164804.

AUTHOR CONTRIBUTIONS

MG-M and CB planned and performed the experiments, and analyzed and interpreted the data. EB performed the bioinformatic analysis of genome-wide data, and analyzed and interpreted the results. ArG performed the experiments, and analyzed and interpreted the data. SA analyzed and interpreted the data. AnG performed analysis of patient data, and analyzed and interpreted the results. CK, DY, and TH provided patient data and interpreted the results. PV and LD conceived and planned the project, analyzed and interpreted the data, and wrote the manuscript, with the assistance and final approval of all authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Di Croce Laboratory is supported by grants from the Spanish Ministry of Science and Innovation (BFU2016-75008-P and PID2019-108322GB-100), "Fundación Vencer El Cancer" (VEC), the European Regional Development Fund (FEDER), and from AGAUR (SGR 2017-2019). We acknowledge the support of the Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya. PV was supported by the Fundación Científica de la Asociación Española Contra el Cáncer. SA was funded by the Ramon y Cajal program of the Ministerio de Ciencia, Innovación y Universidades, the European Social Fund under the reference number RYC-2018-025002-I, and the Instituto de Salud Carlos III-FEDER (PI19/01814).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.655201/full#supplementary-material

Supplementary Figure 1 | (A) Representative analysis of cell cycle phases in shCT and sh*PHF19*#2. **(B)** Representative analysis of BrdU incorporation analysis in shCT and sh*PHF19*#2. **(C)** Cell growth of cells infected with an shCT and shPHF19#2 in the absence or presence of 10 nM Ara-C for 6 days.

Supplementary Figure 2 | (A) Genomic distribution of ChIP-seq peaks of PHF19. The spie-chart represents the genomic distribution of ChIP-seq peaks (outer circle) corrected by the whole-genome distribution of each gene genomic feature (indicated in the background circle distribution). **(B)** PHF19 ChIP-seq screenshots modified from UCSC genome browser of genes validated in **Figure 3B**. **(C)** Genomic distribution of ChIP-seq peaks of MTF2 (in shCT and sh*PHF19*#2). The spie-chart represents the distribution of peaks corrected by the genome-wide distribution of each gene genomic feature (indicated in the background circle distribution). **(D)** MTF2 ChIP-seq screenshots modified from UCSC genome browser of genes validated in **Figure 4B**. **(E)** Accumulative growth of cells infected with an shCT and sh*PHF19*#2 in the absence or presence of different doses of Imatinib for 6 days.

REFERENCES

Abraham, S. A., Hopcroft, L. E., Carrick, E., Drotar, M. E., Dunn, K., Williamson, A. J., et al. (2016). Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells. *Nature* 534, 341–346.

Alitalo, R. (1990). Induced differentiation of K562 leukemia cells: a model for studies of gene expression in early megakaryoblasts. *Leuk. Res.* 14, 501–514. doi: 10.1016/0145-2126(90)90002-q

Ballare, C., Lange, M., Lapinaite, A., Martin, G. M., Morey, L., Pascual, G., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* 19, 1257–1265.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 41, D991–D995.

Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S. A., Egan, C. M., et al. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* 19, 1273–1281. doi: 10.1038/nsmb.2449

Cai, J., Gong, R., Yan, F., Yu, C., Liu, L., Wang, W., et al. (2014). ZNF300 knockdown inhibits forced megakaryocytic differentiation by phorbol and erythrocytic differentiation by arabinofuranosyl cytidine in K562 cells. *PLoS One* 9:e114768.

Carlo, V. Di, Mocavini, I., and Di Croce, L. (2019). Polycomb complexes in normal and malignant hematopoiesis. *J. Cell Biol.* 218, 55–69. doi: 10.1083/jcb.201808028

Chylicki, K., Ehinger, M., Svedberg, H., Bergh, G., Olsson, I., and Gullberg, U. (2000). p53-mediated differentiation of the erythroleukemia cell line K562. *Cell Growth Differ.* 11, 315–324.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247

Deng, Q., Hou, J., Feng, L., Lv, A., Ke, X., Liang, H., et al. (2018). PHF19 promotes the proliferation, migration, and chemosensitivity of glioblastoma to doxorubicin through modulation of the SIAH1/beta-catenin axis. *Cell Death Dis.* 9:1049.

Fajtova, M., Kovarikova, A., Svec, P., Kankuri, E., and Sedlak, J. (2013). Immunophenotypic profile of nucleated erythroid progenitors during maturation in regenerating bone marrow. *Leuk. Lymphoma* 54, 2523–2530. doi: 10.3109/10428194.2013.781167

Ghislin, S., Deshayes, F., Middendorp, S., Boggetto, N., and Alcaide-Loridan, C. (2012). PHF19 and Akt control the switch between proliferative and invasive states in melanoma. *Cell Cycle* 11, 1634–1645. doi: 10.4161/cc.20095

Hehlmann, R., Hochhaus, A., Baccarani, M., and European, L. (2007). Chronic myeloid leukaemia. *Lancet* 370, 342–350.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., et al. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genet.* 8:e1002576. doi: 10.1371/journal.pgen.1002576

Islam, A. (1992). The origin and spread of human leukemia. *Med. Hypotheses* 39, 110–118. doi: 10.1016/0306-9877(92)90149-7

Jabbour, E., and Kantarjian, H. (2020). Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *Am. J. Hematol.* 95, 691–709. doi: 10.1002/ajh.25792

Jain, P., Ballare, C., Blanco, E., Vizan, P., and Di Croce, L. (2020). PHF19 mediated regulation of proliferation and invasiveness in prostate cancer cells. *eLife* 9:e51373.

Jayapal, S. R., Lee, K. L., Ji, P., Kaldis, P., Lim, B., and Lodish, H. F. (2010). Down-regulation of Myc is essential for terminal erythroid maturation. *J. Biol. Chem.* 285, 40252–40265. doi: 10.1074/jbc.m110.181073

Ji, Y., Fioravanti, J., Zhu, W., Wang, H., Wu, T., Hu, J., et al. (2019). miR-155 harnesses Phf19 to potentiate cancer immunotherapy through epigenetic reprogramming of CD8(+) T cell fate. *Nat. Commun.* 10:2157.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006. doi: 10.1101/gr.229102.

Kok, C. H., Yeung, D. T., Lu, L., Watkins, D. B., Leclercq, T. M., Dang, P., et al. (2019). Gene expression signature that predicts early molecular response failure in chronic-phase CML patients on frontline imatinib. *Blood Adv.* 3, 1610–1621. doi: 10.1182/bloodadvances.2019000195

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Laugesen, A., Hojfeldt, J. W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell.* 74, 8–18. doi: 10.1016/j.molcel.2019.03.011

Law, J. C., Ritke, M. K., Yalowich, J. C., Leder, G. H., and Ferrell, R. E. (1993). Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk. Res.* 17, 1045–1050. doi: 10.1016/0145-2126(93)90161-d

Lee, S. C., Miller, S., Hyland, C., Kauppi, M., Lebois, M., Di Rago, L., et al. (2015). Polycomb repressive complex 2 component Suz12 is required for hematopoietic stem cell function and lymphopoiesis. *Blood* 126, 167–175. doi: 10.1182/blood-2014-12-615898

Li, J., Hale, J., Bhagia, P., Xue, F., Chen, L., Jaffray, J., et al. (2014). Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. *Blood* 124, 3636–3645. doi: 10.1182/blood-2014-07-588806

Liang, R., Camprecios, G., Kou, Y., McGrath, K., Nowak, R., Catherman, S., et al. (2015). Identifies essential FOXO3 functions at key steps of terminal erythropoiesis. *PLoS Genet.* 11:e1005526. doi: 10.1371/journal.pgen.1005526

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi: 10.1016/j.cels.2015.12.004

Liu, L., Xu, F., Chang, C. K., He, Q., Wu, L. Y., Zhang, Z., et al. (2017). MYCN contributes to the malignant characteristics of erythroleukemia through EZH2-mediated epigenetic repression of p21. *Cell Death Dis.* 8:e3126. doi: 10.1038/cddis.2017.526

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Machova Polakova, K., Koblihova, J., and Stopka, T. (2013). Role of epigenetics in chronic myeloid leukemia. *Curr. Hematol. Malig. Rep.* 8, 28–36. doi: 10.1007/s11899-012-0152-z

Mason, M. J., Schinke, C., Eng, C. L. P., Towfic, F., Gruber, F., Dervan, A., et al. (2020). Multiple Myeloma DREAM Challenge reveals epigenetic regulator PHF19 as marker of aggressive disease. *Leukemia* 34, 1866–1874. doi: 10.1038/s41375-020-0742-z

Munoz-Alonso, M. J., Acosta, J. C., Richard, C., Delgado, M. D., Sedivy, J., and Leon, J. (2005). p21Cip1 and p27Kip1 induce distinct cell cycle effects and differentiation programs in myeloid leukemia cells. *J. Biol. Chem.* 280, 18120–18129. doi: 10.1074/jbc.m500758200

Ogino, T., Kobuchi, H., Fujita, H., Matsukawa, A., and Utsumi, K. (2014). Erythroid and megakaryocytic differentiation of K562 erythroleukemic cells by monochloramine. *Free Radic. Res.* 48, 292–302. doi: 10.3109/10715762.2013.865840

Orfao, A., Matarraz, S., Perez-Andres, M., Almeida, J., Teodosio, C., Berkowska, M. A., et al. (2019). Immunophenotypic dissection of normal hematopoiesis. *J. Immunol. Methods* 475:112684. doi: 10.1016/j.jim.2019.112684

Ren, Z., Ahn, J. H., Liu, H., Tsai, Y. H., Bhanu, N. V., Koss, B., et al. (2019). PHF19 promotes multiple myeloma tumorigenicity through PRC2 activation and broad H3K27me3 domain formation. *Blood* 134, 1176–1189. doi: 10.1182/blood.2019000578

Rothberg, J. L. M., Maganti, H. B., Jrade, H., Porter, C. J., Palidwor, G. A., Cafariello, C., et al. (2018). Mtf2-PRC2 control of canonical Wnt signaling is required for definitive erythropoiesis. *Cell Discov.* 4:21.

Sawyers, C. L. (1999). Chronic myeloid leukemia. *N. Engl. J. Med.* 340, 1330–1340.

Schuettengruber, B., Bourbon, H. M., Di Croce, L., and Cavalli, G. (2017). Genome regulation by polycomb and trithorax: 70 years and counting. *Cell* 171, 34–57. doi: 10.1016/j.cell.2017.08.002

Scott, M. T., Korfi, K., Saffrey, P., Hopcroft, L. E., Kinstrie, R., Pellicano, F., et al. (2016). Epigenetic reprogramming sensitizes CML stem cells to combined EZH2 and tyrosine kinase inhibition. *Cancer Discov.* 6, 1248–1257. doi: 10.1158/2159-8290.cd-16-0263

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Tanimura, N., Miller, E., Igarashi, K., Yang, D., Burstyn, J. N., Dewey, C. N., et al. (2016). Mechanism governing heme synthesis reveals a GATA factor/heme circuit that controls differentiation. *EMBO Rep.* 17, 249–265. doi: 10.15252/embr.201541465

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Uhlen, M., Zhang, C., Lee, S., Sjostedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357:eaan2507.

Vizan, P., Beringer, M., Ballare, C., and Di Croce, L. (2015). Role of PRC2-associated factors in stem cells and disease. *FEBS J.* 282, 1723–1735. doi: 10.1111/febs.13083

Vizan, P., Gutierrez, A., Espejo, I., Garcia-Montolio, M., Lange, M., Carretero, A., et al. (2020). The Polycomb-associated factor PHF19 controls hematopoietic stem cell state and differentiation. *Sci. Adv.* 6:eabb2745. doi: 10.1126/sciadv.abb2745

Xie, H., Peng, C., Huang, J., Li, B. E., Kim, W., Smith, E. C., et al. (2016). Chronic myelogenous leukemia- initiating cells require polycomb group protein EZH2. *Cancer Discov.* 6, 1237–1247. doi: 10.1158/2159-8290.cd-15-1439

Xie, H., Xu, J., Hsu, J. H., Nguyen, M., Fujiwara, Y., Peng, C., et al. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* 14, 68–80. doi: 10.1016/j.stem.2013.10.001

Ying, W., Wang, H., Bazer, F. W., and Zhou, B. (2014). Pregnancy-secreted Acid phosphatase, uteroferrin, enhances fetal erythropoiesis. *Endocrinology* 155, 4521–4530. doi: 10.1210/en.2014-1397

Yu, W., Zhang, F., Wang, S., Fu, Y., Chen, J., Liang, X., et al. (2017). Depletion of polycomb repressive complex 2 core component EED impairs fetal hematopoiesis. *Cell Death Dis.* 8:e2744. doi: 10.1038/cddis.2017.163

Zhang, D., Cho, E., and Wong, J. (2007). A critical role for the co-repressor N-CoR in erythroid differentiation and heme synthesis. *Cell. Res.* 17, 804–814. doi: 10.1038/cr.2007.72

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership