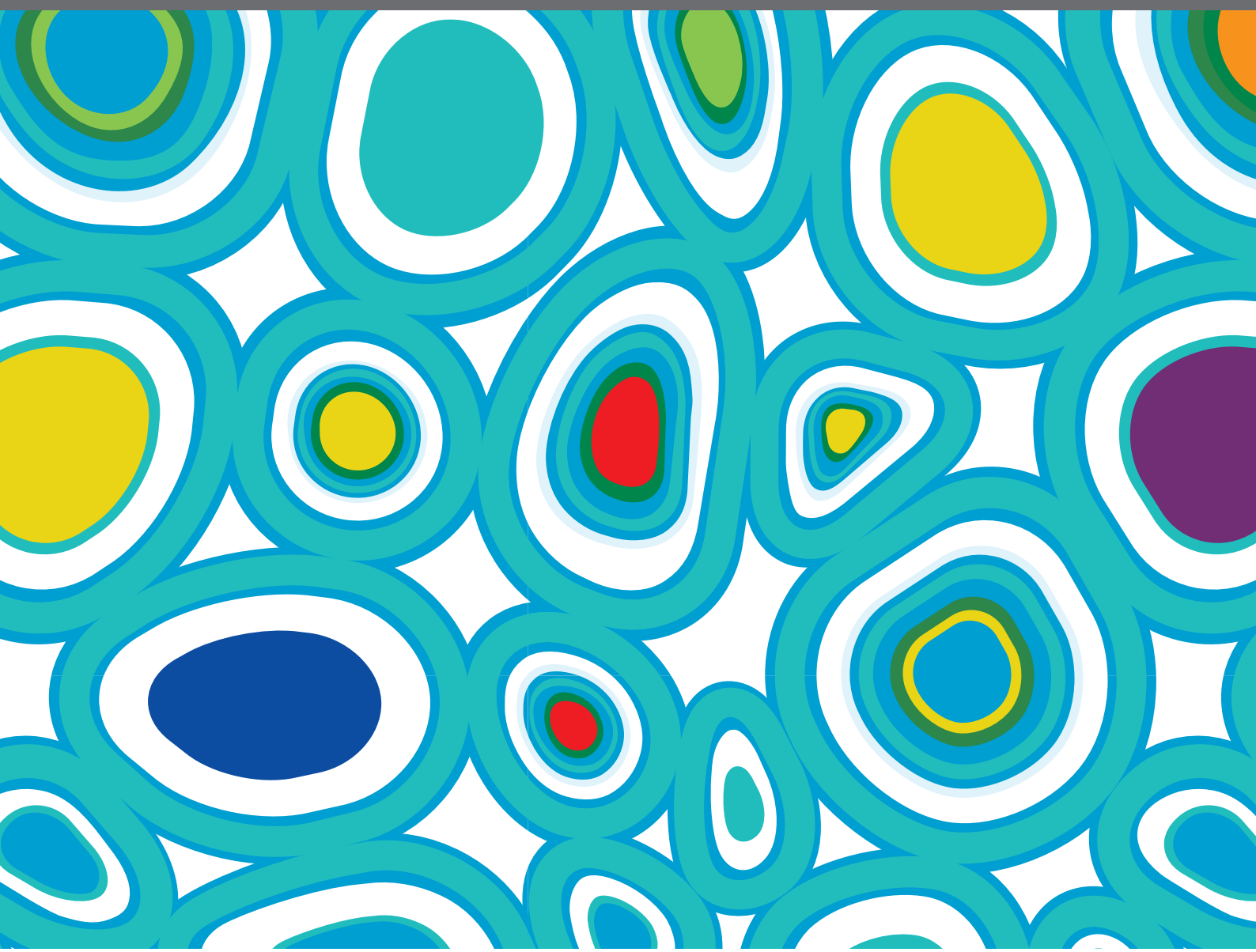
MONOALLELIC EXPRESSION IN DEVELOPMENT AND DISEASES

EDITED BY: Jin Xu, Lan Jiang, Qiaolin Deng, Zhana Duren,
Rosalind M. John, Louis Lefebvre and Anthony Isles
PUBLISHED IN: Frontiers in Cell and Developmental Biology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-653-8
DOI 10.3389/978-2-88974-653-8

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

MONOALLELIC EXPRESSION IN DEVELOPMENT AND DISEASES

Topic Editors:
Jin Xu, Sun Yat-Sen University, China
Lan Jiang, Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences (CAS), China
Qiaolin Deng, Karolinska Institutet (KI), Sweden
Zhana Duren, Clemson University, United States
Rosalind M. John, Cardiff University, United Kingdom
Louis Lefebvre, University of British Columbia, Canada
Anthony Isles, Cardiff University, United Kingdom

Citation: Xu, J., Jiang, L., Deng, Q., Duren, Z., John, R. M., Lefebvre, L., Isles, A., eds. (2022). Monoallelic Expression in Development and Diseases. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-653-8

Table of Contents

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/fcell.2021.642212

Gene Families With Stochastic Exclusive Gene Choice Underlie Cell Adhesion in Mammalian Cells

*Mikhail Iakovlev†, Simone Faravelli† and Attila Becskei**

Biozentrum, University of Basel, Basel, Switzerland

Exclusive stochastic gene choice combines precision with diversity. This regulation enables most T-cells to express exactly one T-cell receptor isoform chosen from a large repertoire, and to react precisely against diverse antigens. Some cells express two receptor isoforms, revealing the stochastic nature of this process. A similar regulation of odorant receptors and protocadherins enable cells to recognize odors and confer individuality to cells in neuronal interaction networks, respectively. We explored whether genes in other families are expressed exclusively by analyzing single-cell RNA-seq data with a simple metric. This metric can detect exclusivity independently of the mean value and the monoallelic nature of gene expression. Chromosomal segments and gene families are more likely to express genes concurrently than exclusively, possibly due to the evolutionary and biophysical aspects of shared regulation. Nonetheless, gene families with exclusive gene choice were detected in multiple cell types, most of them are membrane proteins involved in ion transport and cell adhesion, suggesting the coordination of these two functions. Thus, stochastic exclusive expression extends beyond the prototypical families, permitting precision in gene choice to be combined with the diversity of intercellular interactions.

Keywords: allelic exclusion, carbonic anhydrase, cell identity, Poisson-binomial distribution, single-cell RNA-seq, basigin, olfactory receptor, mouse

OPEN ACCESS

Edited by:
Lan Jiang,
Beijing Institute of Genomics (CAS), China

Reviewed by:
Qiang Wu,
Shanghai Jiao Tong University, China
Ivan Garcia-Bassets,
University of California, San Diego, United States

***Correspondence:**
Attila Becskei
attila.becskei@unibas.ch

† These authors have contributed equally to this work and share first authorship

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

Received: *15 December 2020*
Accepted: *30 March 2021*
Published: *29 April 2021*

Citation:
Iakovlev M, Faravelli S and Becskei A (2021) Gene Families With Stochastic Exclusive Gene Choice Underlie Cell Adhesion in Mammalian Cells. Front. Cell Dev. Biol. 9:642212. doi: 10.3389/fcell.2021.642212

INTRODUCTION

The combinatorial principle plays an important role in the evolution of complex organisms. A large proportion of the mammalian genomes encodes regulators, especially transcription factors (Vaquerizas et al., 2009), which determine what combination of genes will be turned on and off. Each cell type expresses a distinct set of genes, a form of phenotypic diversity that has been studied by single cell expression profiling, such as single-cell RNA-seq, with an unprecedented throughput (Baran-Gale et al., 2018). The study of the combinatorial expression patterns of genes belonging to a gene family or gene array is of particular relevance, among which the exclusive gene choice of the odorant and T-cell receptors has received widespread attention.

Each olfactory neuron expresses a single odorant receptor isoform randomly selected from more than a thousand gene isoforms (Massah et al., 2015; Khamlichi and Feil, 2018) and triggers a signal in response to a particular odor. Thus, precision of expression in a single cell is combined with diversity in a cell population. A similar principle underlies the immune response: each lymphocyte expresses a single antigen receptor randomly chosen from a large repertoire. The receptor isoforms

are diversified, in part, due to the stochastic gene choice of the variable domain. With the in-depth study of these systems, it became apparent that a non-negligible proportion of cells expresses more than one, typically two gene isoforms (Brady et al., 2010). These cells with dual T cell receptors may enhance the antiviral response but can also underlie autoimmune disorders (Ji et al., 2010; Bradley et al., 2017). Thus, stochastic gene choice has clear physiological implications.

A slightly different form of exclusivity was observed in the protocadherin (Pcdh) array, which encodes multi-subunit membrane proteins mediating cell-to-cell interactions between neurons (Yagi, 2012). In this array, most cells express at least two distinct variable α-isoforms from a repertoire of 12 genes, one from the paternal, one from the maternal chromosome (Esumi et al., 2005). These findings indicate that the strict definition of exclusivity—one gene (isoform) per single cell—needs extending to account for the observed distributions and for averages greater than one.

These observations lead to the question about how to define exclusive expression in terms of a probability distribution. Is the expression of T-cell receptor isoforms exclusive if cells with dual T-cell receptors constitute 1, 50, or 90% of the population? What if three different receptor isoforms were to be expressed in some of the cells (Vatakis et al., 2013)? Recently, the degree of exclusivity in the stochastic gene choice of the Pcdh gene array was quantified with a probabilistic approach that defines exclusivity independently of the mean number of expressed genes in an array (Wada et al., 2018). This definition of stochastic exclusivity implies that the distribution of the number of expressed gene isoforms is narrower than expected from the purely random, independent expression of the genes in the array. For example, gene choice is precise when the majority of cells express three gene isoforms and only a few cells express less or more than three isoforms. Thus, stochastic exclusivity reflects simply the precision in gene choice irrespective of the underlying mechanism, let it be chromosomal looping during gene activation, negative feedback or allelic exclusion after DNA recombination.

Here, we examined single-cell RNA-seq data and established the exclusivity in the classic gene arrays and families, the odorant receptors, the T-cell receptors and the Pcdh-α array in some cell types, with a simple metric, regardless of whether gene expression is monoallelic or has a mean value of one. After this validation of our approach, we examined how the genome-wide organization of the genes affects stochastic gene choice and detected gene families (paralogs) with exclusive gene choice.

RESULTS

Single Cell RNA-Seq Datasets

We analyzed RNA-seq datasets consisting of at least 100 single cell measurements of a well-defined cell type isolated from the mouse *Mus musculus*. Neurons from various locations in the nervous system were included, such as somatosensory neurons from dorsal root ganglions (Li et al., 2016), dopaminergic neurons (Hook et al., 2018) and corticostriatal neurons from the visual cortex (Tasic et al., 2016). Non-neuronal cell types encompassed nearly all organs: two types of lymphocytes, $CD8^+$ T-cells (Kakaradov et al., 2017) and type 17 helper cells (Th17) (Gaublomme et al., 2015); dendritic cells from the bone marrow (Schlitzer et al., 2015), cardiomyocytes (Nomura et al., 2018), endothelial cells (Veerman et al., 2019), enterocytes (Haber et al., 2017), fibroblasts (Reinius et al., 2016), kidney duct cells (Chen et al., 2017), thymus epithelial cells (Sansom et al., 2014), prostate stromal cells (Kwon et al., 2019), type I and II alveolar cells from the lung (Guo et al., 2019); hepatoblasts and hepatocytes from the liver (Yang et al., 2017), pancreatic endocrine cells (Yu et al., 2019). Undifferentiated cell types were represented by embryonic stem cells isolated from embryos (Cheng et al., 2019) and embryonic stem cell (ESC) cultures (Klein et al., 2015). The gene expression has UMI units in two studies, while all other studies have FPKM/TPM units (**Supplementary Table 1**). The libraries in most studies were generated by Smart-Seq2 or its variants, which typically capture more genes than other technologies (Baran-Gale et al., 2018).

Dichotomization of RNA-Seq Counts

The distribution of the RNA counts in a single-cell RNA-seq dataset is determined by various factors, in particular, the stochastic processes in gene expression and the methods for amplifying and detecting the RNA molecules. Gene expression is stochastic due to the low copy number of genes and mRNA molecules, and due to the spatiotemporal nature of biochemical processes in the cell (Battich et al., 2015; Baudrimont et al., 2019; Finn and Misteli, 2019; Friedrich et al., 2019; Rodrigo, 2019). When the expression has two states (OFF and ON states), the resulting distribution can be bimodal, often referred to as stochastic gene choice. Many genes display bimodal expression (**Supplementary Figure 1A**; Shalek et al., 2013).

To determine the proportion of OFF and ON cells, the RNA distribution must be dichotomized. For this purpose, we compared two classes of methods. In the moment-based methods, the averages or variances of the total distribution or parts of it are calculated. The second class of methods relies on the fitting of probability mass or density functions (pdf). The moment-based methods are more robust but lack a uniform mathematical framework (**Supplementary Figures 1B–D**). Conversely, the pdfs have mathematically well-defined dichotomization points but their fitting is less robust. In order to combine the advantages of the two approaches, we aimed at selecting the moment based approach that correlates the most with the dichotomization using pdfs.

We tested three types of moment-based methods: the Variance Reduction Score (VRS), Fraction of Maximal values (FM) and Geometric Trimmed Mid-Extreme threshold (GTME) (section "Materials and Methods"). The VRS quantifies the extent to which a given threshold reduces the sum of the variances of the two subpopulations relative to the unsplit population (Hellwig et al., 2010). The threshold minimizing the VRS was selected for the dichotomization. We devised two additional methods based on biological control principles, the FM and GTME. The FM is based on the assumption that a biological function can be performed as long as a variable in the ON state does not deviate

too much from an optimal level. Accordingly, we defined the FM-threshold as the one tenth of the observed maximal values in the distribution. The GTME threshold is the geometric mean of the extreme values of the distribution; thus, it combines information on both the minimal and maximal values of the distribution.

To find the appropriate distribution, the specific pdf was selected in an unbiased way from a large number of known probability mass and density functions according to the Bayesian information criteria, and the parameters were fitted simultaneously. Whenever a mixture distribution, the sum of two or more probability functions, was selected, the antimode, the minimum value between two modes of the pdf, was determined (section "Materials and Methods"). The antimode was then used as the threshold to dichotomize the cell population.

The dichotomization is illustrated using the Pcdhac2 RNA counts from the somatosensory neuron dataset (**Figure 1A**). The values of the four thresholds differed up to about ten times. The corresponding ON cell frequencies differed less since few cells have RNA counts between the two peaks of the distribution where most of the thresholds are positioned (**Figure 1A**). Indeed, when comparing ON cell frequencies, all methods were closely correlated (**Figures 1B,C**); even the lowest correlation had a large value (0.79). For comparison, we also show the dichotomization with a constant threshold at 0.5 TPM. The dichotomization with the antimode correlated most strongly with the GTME-dichotomization (Spearman rank correlation = 0.86), followed by the FM and constant thresholds and last by the VRS (0.79). Therefore, we applied the GTME to all datasets with TPM/FPKM units. It is important to note that GTME thresholds were calculated also for those genes with high bimodality coefficient that yielded unimodal probability density functions, which is often the case, when there are few cells in the OFF or ON expression states (**Supplementary Figure 2**, **Supplementary Table 2**, and **Supplementary Text 1**).

Some datasets had UMI units (**Supplementary Table 1**). For these distributions, the Bayesian selection and fitting typically

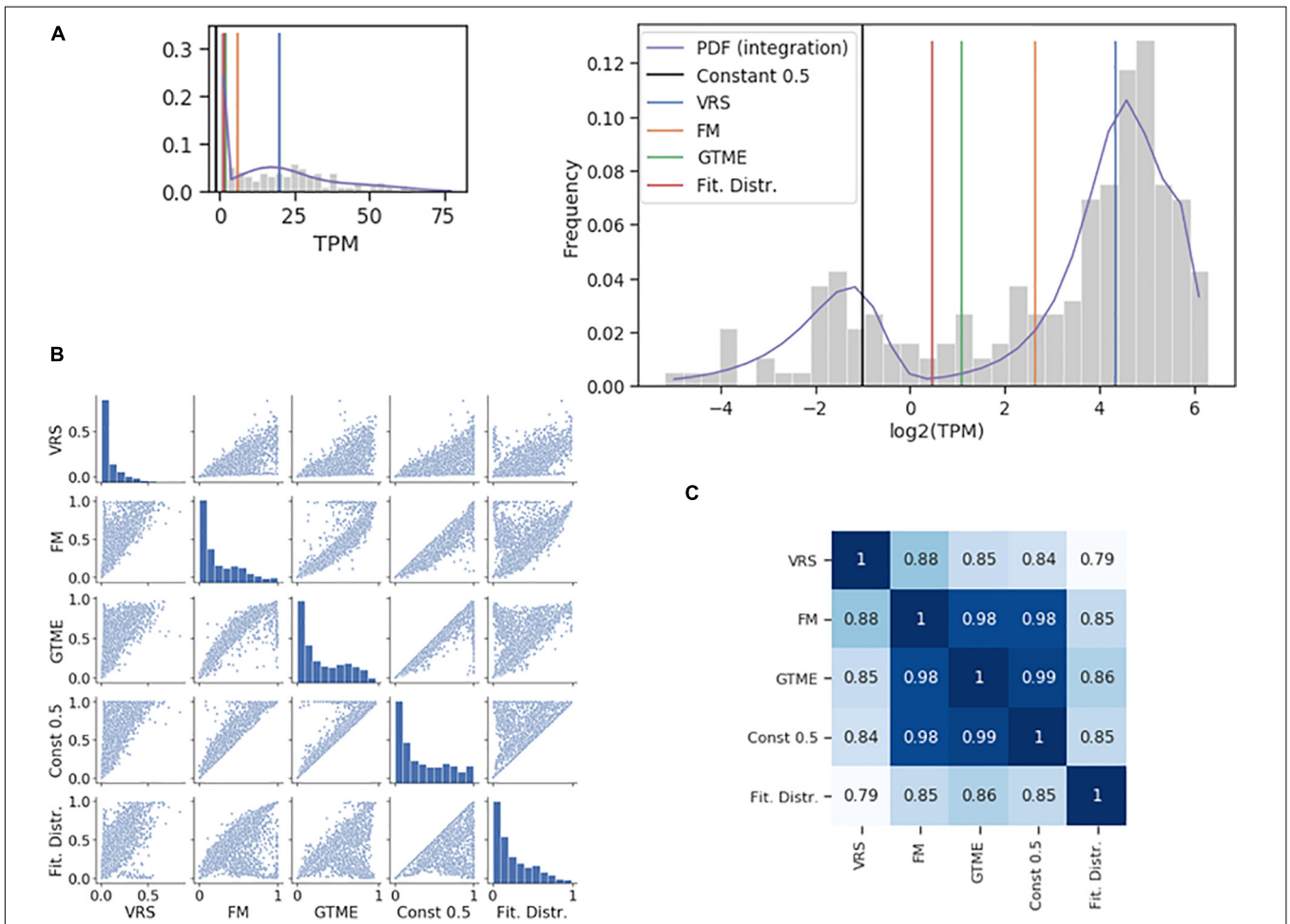


FIGURE 1 | Comparison of dichotomization methods. **(A)** The histogram of the Pcdhac2 transcript numbers in the somatosensory neurons. The dichotomization yielded the following thresholds: 2.11 (GTME), 6.28 (FM), and 20.21 (VRS) TPM. The fitted probability density function (pdf) is a mixture of normal distributions, with an antimode at 1.37. The pdf is integrated piecewise according to the logarithmic bins. The thumbnail plot is a version of the main plot with a linearly scaled *x*-axis. **(B)** Pair-wise scatter plots showing the ON cell frequencies of each gene in the somatosensory neuron dataset with a bimodality coefficient greater than 0.55, after dichotomization with different methods. **(C)** The Spearman rank correlation of the ON cell frequencies shown in **(B)**.

returned Poisson or Yule-Simon distributions, and rarely mixture distributions, which precluded the determination of the antimodes. Therefore, we compared the thresholds according to their ability to dichotomize RNA counts of marker genes of specific cell types (**Supplementary Figure 3**). This led to the selection of the FM-threshold. For most genes, the threshold was positioned between zero and one, simply equating the zero RNA count with the OFF state.

Effect of Proximity on Stochastic Gene Choice

All RNA distributions were converted into ON cell frequencies with the dichotomization described above. We then examined how proximity affects stochastic gene choice, as genes are often located side by side in gene families with exclusivity. Proximity can influence gene expression in many ways, by promoting the interaction of genes with enhancers via looping, by modifying epigenetic signatures, by relocating chromosomes into active or inactive nuclear compartments, such as transcription factories and heterochromatic compartments (Finn and Misteli, 2019; Monahan et al., 2019).

If a chromosomal segment shuttles back and forth between sufficiently large active and inactive nuclear compartments, all or none of the genes in that segment will be expressed, which will result in a large cell-to-cell variation in the number of expressed genes in that segment. The all-or-none response is an example of stochastic co-occurrence (a.k.a. concurrence, **Figure 2A**). In contrast, although each gene is randomly chosen to be expressed, the number of genes expressed in each cell may be the same or similar (**Figure 2A**, exclusivity). While exclusivity is often equated with the expression of a single gene isoform, this is not necessary as long as the overlap among the chosen genes is small. It is the constant number of expressed isoforms that matters, which is particularly important for protein complexes with fixed stoichiometry. Alternative chromosomal configurations in which a fixed number of genes is located in active nuclear compartments while preventing the remaining genes in the segment from being activated can produce stochastic exclusive gene choice.

The all-or-none response and the fixed, constant number of ON genes in each cell are extreme cases of stochastic co-occurrence and exclusivity, respectively. In this work, we use the terms co-occurrence or exclusivity in a probabilistic (stochastic) sense, and in order to quantify the range of their values, we calculated the interdependence coefficient (IC). IC is the ratio of the cell-to-cell variance in the number of genes chosen to be expressed to the variance of the Poisson-binomial distribution expected from the ON state frequencies of each gene under consideration (see "The Interdependence Coefficient (IC)" in section "Materials and Methods") (Wada et al., 2018). An IC less than one indicates exclusivity, while an IC greater than one indicates concurrence in stochastic gene choice. When IC is one, the choice of the genes is unbiased, which can reflect independent expression of these genes. Thus, IC enables the detection of exclusive gene choice even when the mean number of expressed genes is greater than one (as in **Figure 2A**). This illustration shows that the mean number of the ON genes with the exclusive expression can be greater than with concurrent expression (3 versus 2.5), although the variance is significantly lower (0 versus 12).

If a gene affects the probability of the ON and OFF states of the genes in its vicinity, chromosomal segments with exclusive or concurrent expression will be overrepresented. To test this hypothesis, we calculated the IC for segments comprising 14 genes sampled along the chromosomes, which corresponds to the number of genes in the Pcdh-α array. This calculation gives the distribution of the IC values for the original genome. Next, we reshuffled the genes in the genome and calculated the IC for the segments, and by repeating the reshuffling, we obtained a representative distribution of the IC values (**Figures 2B,C**). To characterize the differences in the distributions, we compared the location of 10th or 90th percentiles (i.e., 1st and 9th decile) to assess the enrichment of the exclusive and concurrent segments, respectively.

In the somatosensory neurons and the prostate stromal cells, the 10th percentile shifted to higher values after the reshuffling, which indicates that the closeness of the genes promotes exclusivity (**Figures 2B,D** and **Supplementary Figure 4A**). In some cell types, there is no significant difference in the location of the 10th percentiles (**Figure 2C**). In the majority of the cell types, exclusivity is suppressed (**Figure 2D**, top panel). The 90th percentile shifts to substantially lower values when the genome is reshuffled, namely by more than 0.5 in some cells, revealing that all cell types except the prostate stromal cells were enriched in concurrent segments (**Figures 2B–D** and **Supplementary Figures 4A,B**). In the original genomes of most cell types, the IC values are more broadly distributed than in the reshuffled genome, as reflected by the quantile ratio of the 9th decile to the 1st decile (**Figure 2D**, bottom panel), which is mostly due to the overrepresentation of concurrence.

In summary, the permutation tests have shown that the proximity of the genes shifts stochastic gene choice to co-occurrence and suppresses exclusivity in most cell types.

Stochastic Gene Choice in the Protocadherin Cluster

The effect of gene proximity can be specifically assessed for the Pcdh family by comparing the Pcdh genes in the α-, β-, and γ-arrays to the Pcdh genes scattered throughout the genome. Most of the scattered isoforms belong to the [δ-protocadherins (Pcdh-1, -7, 8, -9, -10, -11, -17, -18, and -19)] (Redies et al., 2005; Harrison et al., 2020). Especially, the α-array is relevant since the expression there is controlled by chromosomal looping mediated by the CTCF (Jia et al., 2020). The expression of the isoforms varies with the cell type. For example, αC2, α11, and α5 are the most frequently expressed isoforms in the somatosensory, dopaminergic, and corticostriatal neurons, respectively (**Figure 3A**). The corticostriatal neurons express relatively few α-isoforms with a pronounced exclusivity (**Figure 3B** and **Supplementary Data 1**). On the other hand, unbiased choice (or independence, IC not significantly different from one) is observed in somatosensory neurons, and weak concurrence in the corticostriatal neurons.

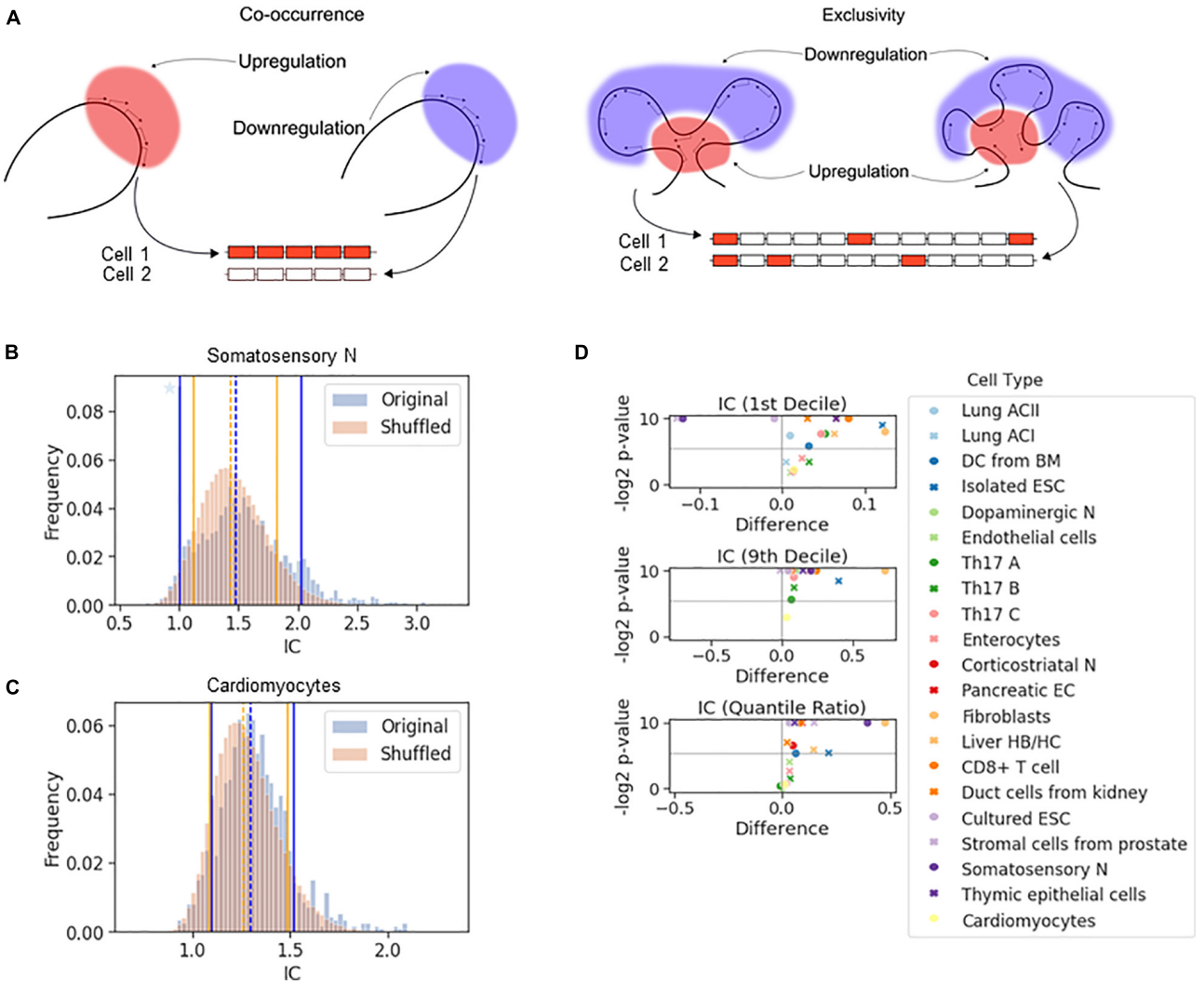


FIGURE 2 | The effect of chromosomal adjacency on stochastic gene choice. **(A)** Schemes showing examples of how the two major forms of stochastic gene choice, concurrence and exclusivity, can arise from alternative chromosomal configurations. **(B,C)** The IC distributions calculated from the original and the shuffled genomes of the somatosensory neurons **(B)** and cardiomyocytes **(C)**. Segmentation size: 14 genes. The blue star denotes a high bar in the histogram hidden by the full line. The location of the 1st (full line), 5th (dashed line) and 9th (full line) deciles is given in the order of original and reshuffled distribution, followed by the *P*-values for the differences: 1.00, 1.48, 2.02; 1.12, 1.44, 1.82; 0.001, 0.016, 0.001. **(B)** 1.10, 1.29, 1.52; 1.09, 1.26, 1.49; 0.217, 0.002, 0.137 **(C). (D)** Volcano plots showing the difference of 1st decile, 9th decile and quantile ratio IC values between the original and the shuffled genomes, along with the corresponding *P*-values (permutation test, segment size: 14 genes). The gray horizontal line at 0.025 corresponds to a two-tailed significance level of 0.05.

The α-array can be conveniently compared with the scattered Pcdhs in the somatosensory neurons, as they have a similar number of isoforms, 14 and 12, and the mean number of expressed isoforms is also similar (3.2 and 3.0 genes, respectively). The IC of the scattered isoforms is more than twice as large as the IC of the α-array (**Figure 3B**). In both the somatosensory and corticostriatal neurons, α-array belongs to the lowest decile of IC distribution. Thus, the α-array in particular gains exclusivity due to the gene adjacency and proximity.

To get a more detailed view of how stochastic gene choice varies along the chromosomal region containing the Pcdh cluster, we moved a 14-gene window along the chromosome to calculate the IC (green horizontal rectangle in **Figure 3C**). In somatosensory and corticostriatal neurons, the resulting IC profiles are similar along the portion of the chromosome comprising the α- and β-arrays and the region upstream of the array, with the corticostriatal cells having lower IC. The lower IC values in the arrays of the corticostriatal cells can be explained by the lower IC values in the genome when compared to the somatosensory cells, as indicated by the range delimited by the 2.5 and 97.5 percentiles of the genomic IC distribution (**Figure 3C**).

The above results suggest that IC profiles can be conserved between different cell types. Interestingly, the conserved exclusivity extends upstream of the Pcdh α-array involving the Zmat2 and Vaultrc5 genes (**Figure 3C**), which suggests that they may be also linked mechanistically and/or functionally to the

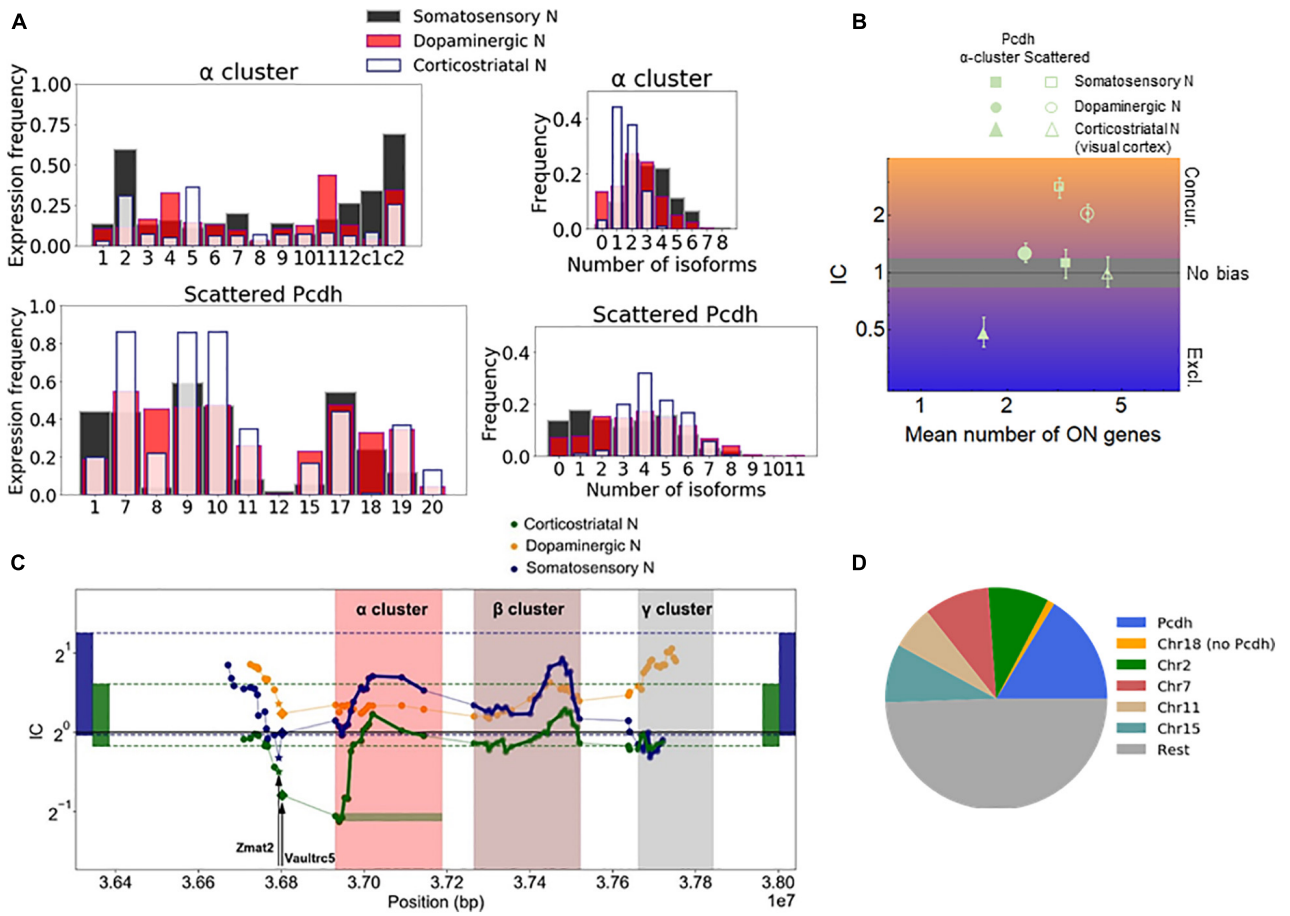


FIGURE 3 | Chromosomal segments with stochastic exclusive gene choice in multiple cell types. **(A)** The expression of the Pcdh α-array and the scattered Pcdhs in different neuronal types. The expression frequency indicates the proportion of cells expressing a particular gene isoform. The distribution of the number of expressed gene isoforms per cell indicates the proportion of cells expressing 0, 1, 2 or more isoforms per cell at the RNA level. **(B)** The IC values calculated from the data shown in **(A)**. The error bars denote the 95% confidence intervals obtained by bootstrapping. **(C)** The IC of segments with 14 genes along the chromosome. The symbols indicate the position of the most upstream gene in each segment. A full segment is denoted by the green horizontal rectangle, at the first gene of the Pcdh-α cluster. The two genes upstream of the Pcdh α cluster (Vaultrc5 and Zmat2) are marked with a star and diamond. The rectangles located at the two extremes of the plot indicate the 2.5 and 97.5 percentiles of the IC distribution calculated for the chromosomal segments in the genome. **(D)** The number of chromosomal segments with exclusive gene choice (as shown in **Supplementary Figure 5**) in each chromosome for all cell types combined.

array. This effect is particularly strong in the somatosensory neurons; in these cells, the segment starting with Zmat2, and comprehending the Vaultrc5 and the 12 variable α isoforms has the lowest IC value altogether in the relevant portion of the chromosome (**Figure 3C**).

Chromosomal Segments With Stochastic Exclusive Choice

The above findings suggest that segments with exclusive gene choice can be longer or shorter than previously assumed. To identify chromosomal segments of various lengths that conserve stochastic exclusive expression in multiple cell types, we have segmented the genome into segments comprising 7, 14, or 21 genes. In order to compare different cell types, it is important to take into account that cells in different studies have IC distributions with different mean values (see e.g., **Figures 2B,C** and **Supplementary Figures 4A,B**). The difference persists even after the reshuffling, suggesting that it originates from a systemic intrinsic or extrinsic variable. For example, the procedure used for the isolation of cells and RNA and for the RNA detection can introduce positive correlations extrinsically, making the average genomic IC appear larger.

To take into account the above differences, we selected all segments that belong to the lowest 2.5 percentile of the IC distribution in at least two different cell types (or cells cultured in different conditions). We then combined all the segments having 7, 14, or 21 genes that belong to the lowest 2.5 percentile. The two criteria above have been expanded to include a third, stating that the IC must be significantly less than one in at least one of the cell types, i.e., the 95% confidence interval must be below one.

Next, we analyzed the location of these segments. Interestingly, the segments overlapping with the Pcdh array represented the largest fraction (**Figure 3D**). Segments from

the Pcdh array were identified in all analyzed types of neurons (corticostriatal, dopaminergic, and somatosensory), and even in non-neuronal cells, such as endothelial and the lung alveolar cells (**Supplementary Figure 5**). The Pcdh beta isoforms play a role in tumor suppression in lung cancer (Ting et al., 2019), implying the possibility that exclusive Pcdh expression may diversify cellular identity in non-neuronal cells, as well.

The chromosome 6 harbors a second prominent gene array, the Trbv, which encodes the variable domains of the T-cell receptor. The low IC values of the overlapping segments indicate a strong exclusivity: it is significantly below one in one of Th17 cell variant and numerically less than one in another Th17 cell variant (**Supplementary Figure 5**). It is important to note that the list of identified arrays with exclusive gene choice is unlikely to be exhaustive because some genes are not detected in a particular cell type. For example, the RNA-seq data cover the expression of Trbv in Th17 cells but not in CD8+ lymphocytes, even though stochastic gene choice and allelic exclusion have been primarily studied in CD8+ lymphocytes. The importance of the exclusivity in T-cell receptor expression in Th17 lymphocytes is underscored by the presence of IL-17 in the cytokine storms, which are thought to contribute to the lethality of the coronavirus disease Covid-19 (Wu and Yang, 2020). Dual reactive lymphocytes that recognize endogenous, neurologically relevant, antigens as well as the coronavirus have also been detected (Boucher et al., 2007).

Gene Families Shift the Stochastic Gene Choice Toward Co-occurrence

The successful detection of Trbv and Pcdh arrays based on their low IC values indicates that exclusive gene choice can be identified solely based on RNA-seq counts without any information on the alleles and sequence similarity. These gene families have two characteristic features: they are encoded by similar sequences and form an array along the chromosome. The gene family aspect may be more important for the odorant receptors since more than a thousand receptor isoforms are encoded by multiple arrays scattered over a large number of chromosomes. Therefore, after having explored the effect of chromosomal proximity, we turned our attention to gene families.

To dichotomize the RNA counts for the gene families, we have not imposed the criterion based on the bimodality coefficient. Instead, we combined the information on the RNA counts of all genes to define the tails of the distribution to calculate a single threshold for all genes in the family. A familywise threshold was used also in a recent study examining how the olfactory receptor expression changes during cell differentiation (Hanchate et al., 2015). We have adapted the GTME to calculate the familywise threshold (fGTME, section "Materials and Methods"). The fGTME threshold resulted in an IC = 0.48 and the mean number of ON genes was 1.0 (**Figure 4A**), evidencing a marked exclusivity in the choice of olfactory receptors. For comparison, a constant threshold at 0.5 resulted in IC = 3.33 and the mean number of ON genes being around 2 (**Supplementary Table 3**). Thus, the constant threshold fails to detect the well-established single isoform expression per cell (**Supplementary Text 1**).

Next, we dichotomized gene expression in each family in various cell types and reshuffled all the genes belonging to a family encompassing at least five genes (**Figure 4B**). In the somatosensory neurons, there were many gene families with an IC larger than the 97.5 percentile of the IC distribution of the

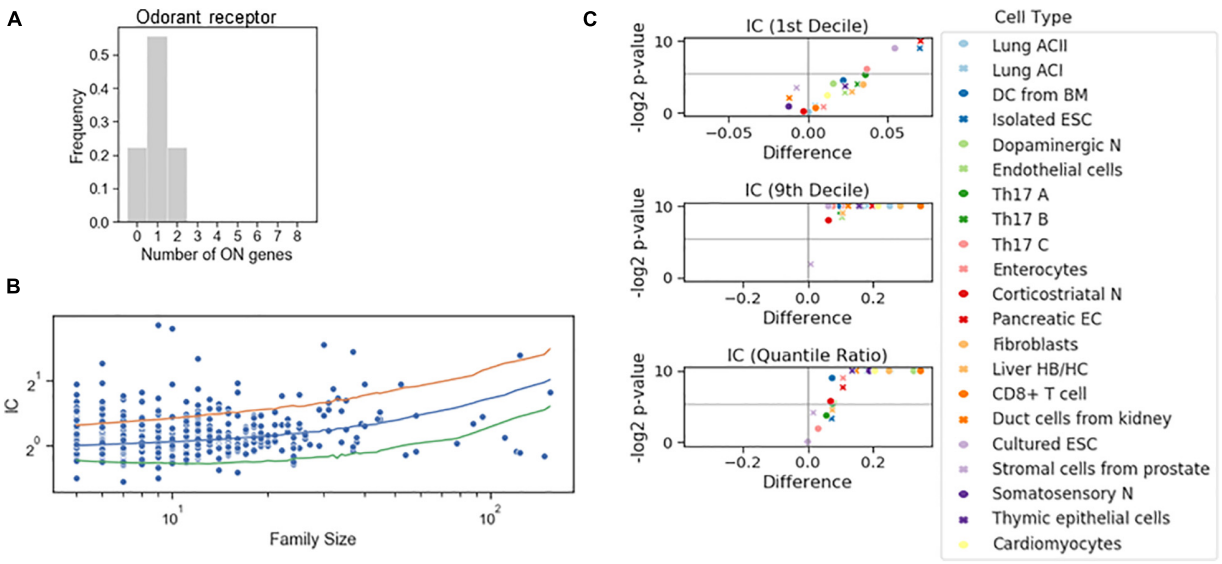


FIGURE 4 | Stochastic interdependence in gene families. **(A)** The number of expressed genes per cell in the family of odorant receptor genes, dichotomized with the familywise threshold (70.7 TPM). Number of cells is $N = 27$. **(B)** IC values of individual families in somatosensory neuron dataset, grouped by the family size. The majority of the families with ICs exceeding either 2.5 or 97.5 IC percentiles of the shuffled genome (orange and green lines, respectively) are concurrent. **(C)** Volcano plots showing the difference of 1st decile, 9th decile and quantile Ratio IC values between the original and the shuffled genomes, along with the corresponding *P*-values (permutation test) calculated for the gene families consisting of 7 genes.

reshuffled genome, but only a few with an IC less than the 2.5 percentile, suggesting that concurrence dominates also in families. Indeed, the systematic examination revealed that the IC at the 10th percentile displayed a significant change in four cell types and the exclusivity was not promoted in any of the cell types. On the other hand, co-occurrence was significantly promoted in all but one cell type (**Figure 4C**), implying that the shared regulation of the genes in a family shifts gene choice toward co-occurrence.

The Relation Between Stochastic Gene Choice and Allelic Exclusion

In addition to the shared regulation of the genes, allelic exclusion may affect stochastic choice in a gene family. The families of the olfactory and T-cell receptors display allelic exclusion, so that only one of the two alleles is expressed, which is also termed monoallellic expression. The molecular mechanisms underlying allelic exclusion can stabilize the gene choice; thus, allelic exclusion may promote stochastic exclusive gene choice. Allelic exclusion takes place after the stochastic choice of the promoter of a T-cell receptor isoform (Ryu et al., 2004). The expression of one allele suppresses the expression of the other allele (Vatakis et al., 2013), a process mediated by various molecular mechanisms. However, allelic exclusion does not necessarily go hand in hand with stochastic gene choice, as the following two examples suggest. Allelic exclusion plays a major compensatory role in the expression of sex chromosomes. In order to compensate for the double dosage of the X chromosomes in females, one of the X chromosomes is inactivated randomly in each cell. Consequently, only one of the gene alleles, the maternal or paternal, is expressed in each cell (Cheng et al., 2019; Zhang et al., 2020); however, this allelic exclusion is not associated with exclusive gene choice because all relevant genes are expressed by one of the chromosomes (**Figure 5A**). In the protocadherin array, the genes can be expressed monoallelically or biallelically (Kaneko et al., 2006).

To assess whether allelic exclusion can contribute to the choice of gene isoforms on a genomic scale, we analyzed RNA-seq data obtained from heterozygous fibroblasts (Larsson et al., 2019), in which the two alleles of most genes can be distinguished. As a measure for allelic exclusion, we calculated the Spearman correlation coefficient between the two alleles for each gene. The overwhelming majority of the genes displayed positive interallelic correlation. Only a small proportion of genes had negative correlation, most of them are located on the X-chromosome, confirming the predominance of this classical form of allelic exclusion (**Figures 5B,C**). The allelic exclusion is evident for genes with mean RNA count above 0.5 UMI (**Figure 5C**). 17 genes from the β- and γ-arrays of the protocadherin cluster are also expressed; all of them have a positive interallelic correlation with a mean value of 0.53 (**Supplementary Figure 6A**). Next, we calculated two variables for each gene family: the mean value of the interallelic correlations and the biallelic IC differential (see section "Materials and Methods"). The biallelic IC differential is negative if the IC is reduced upon combining the alleles from the two haplotypes, implying that allelic exclusion contributes to exclusivity in stochastic choice in the gene family. Nearly all families have positive mean interallelic correlation, the degree of which does not correlate positively with the biallelic IC differential (**Figure 5D**). The only gene family with negative mean interallelic correlation is the melanoma associated antigen family. Interestingly, this family experiences the largest shift toward exclusivity in the stochastic gene choice when the two haplotypes are combined: $IC = 2.47$ and 2.67 for the haplotypes and $IC = 1.52$ for the diplotype. Thus, this shift is substantial but not sufficient to attain exclusivity in stochastic gene choice (**Supplementary Figure 6B**). Most genes of the melanoma associated antigen family are located on the X-chromosome, and the rest of them at the Prader-Willi locus, which is also known to be imprinted (Weon and Potts, 2015; Tacer and Potts, 2017), and explains the marked allelic exclusion in this family. These findings indicate that gene families with allelic exclusion are rare; however, specific gene families can utilize it to enhance exclusivity in stochastic gene choice. Importantly, families with IC less than one have positive mean interallelic correlation (**Supplementary Figure 6B**), suggesting that stochastic exclusive gene choice does not necessarily imply allelic exclusion.

Gene Families With Stochastic Exclusive Gene Choice

After having analyzed the mechanisms that affect stochastic choice in gene families, we examined exclusivity and co-occurrence in all cell types. The T-cell receptor beta-chain family in the Th17 cells was the most exclusive among all families, with an IC between 0.49 and 0.62 (**Figure 6A**), comparable to the odorant receptors (**Figure 4A**). On the other extreme of the scale, the histone 2A family is one of the families with the largest IC values ($IC = 4.80$ in Th17 and 2.38 in liver cells). The histone family nicely illustrates the functional relevance of concurrence: some cells enter the S-phase of the cell-cycle and express the histones to support the ongoing DNA replication, while the cells in the other phases of the cell cycle do not express and/or are degraded (Marzluff and Koreski, 2017), which results in a large coherent cell-to-cell variation in the number of expressed gene isoforms (**Figure 6B**).

Thus, our analysis with appropriate dichotomization and a simple metric confirmed the exclusive choice in all three prototypic families and gene arrays (T-cell receptor, odorant receptor, Pcdh), so they serve as the positive control for the identification of other gene families. To identify families with stochastic exclusive gene choice, we used the robust approach developed for the chromosomal segments, which combined relative and absolute criteria for exclusivity. The relative criterion ensures that families are selected from the lowest 2.5 percentile of the IC distribution of each cell type. The second criterion states that a family is only considered exclusive if it belongs to the lowest 2.5 percentile in at least two cell types. The last, absolute selection criterion states that the IC must be significantly smaller than one in at least one of the cell types.

The clustered Pcdh family is exclusive in corticostriatal neurons and endothelial cells, and also in the somatosensory neurons. However, in the latter cell type it does not belong to the

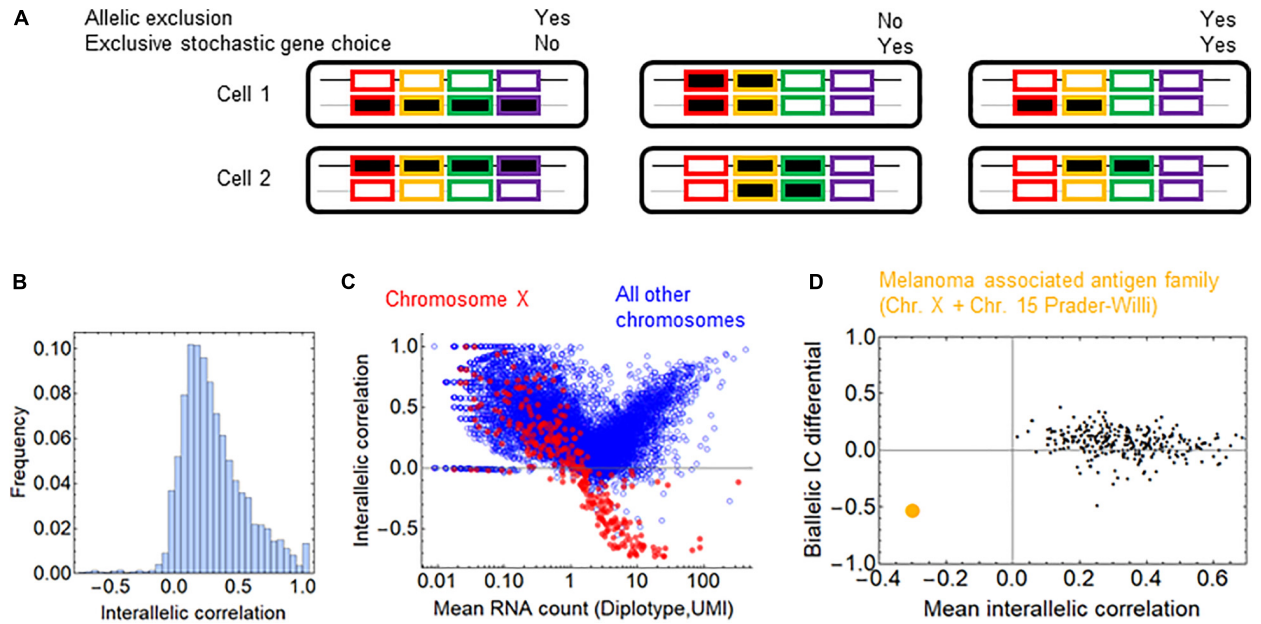


FIGURE 5 | Allelic exclusion and exclusivity in stochastic gene choice. **(A)** Schematic representation of different combinations of exclusivity in allelic and gene choice in an array of four genes. The black and gray lines represent the maternal and paternal chromosomes. The rectangles with no or black filling represent the OFF and ON expression states, respectively **(B)** The interallelic correlation in fibroblasts (Larsson et al., 2019). Negative correlations indicate the allelic exclusion. **(C)** The relation between RNA count and interallelic correlation. The genes on the chromosome X are shown in red. **(D)** The melanoma-associated antigen gene family is highlighted in orange among the gene families. It is the only family with negative mean interallelic correlation.

bottom 2.5 percentile of the IC distribution and consequently, it is not indicated as a hit in **Figure 7**.

The majority of the retrieved families encode membrane proteins (**Figure 7**) like the three prototypic families. Many of them are associated, directly or indirectly, with two processes: transmembrane ion transport and intercellular adhesion (**Figure 8A**), These include well-known families involved in cell adhesion such as the basigin related (Bsg, Ccdc141, Cntn5, Cntn6, Dscam, Dscaml1, Emb, Myot, Mypn, Nexn, Nptn, Nrcam, Prtg, Vstm2l) and the synaptic adhesion-like molecule families (Igsf10, Lrfn1, Lrfn2, Lrfn3, Lrfn4, Lrfn5, Lrit1, Lrit2, Lrit3). There are also families primarily involved in ion transport but many of the genes are also involved in cell adhesion, exemplified by the sodium/potassium transporting ATPase subunit gamma and the carbonic anhydrase and anion exchange proteins (**Figures 7**, **8B**).

The Fxyd1-7 gene isoforms encode the gamma subunit of the Na^+/K^+ ATPase, which is the regulatory subunit of this ion pump. While these ATPases are primarily involved in ion homeostasis, they can also trans-dimerize and thus mediate cell-to-cell interaction (Tokhtaeva et al., 2016). The stochastic exclusivity of the basigin related genes can be observed in somatosensory and corticostriatal neurons (**Figure 7**). The members of this family are named after the immunoglobulin−superfamily molecule basigin and are well known mediators of intercellular adhesion (Muramatsu, 2016), comprising genes such as Contactin 6 (Cntn6), Down syndrome cell adhesion molecule (Dscam) and Neuronal cell adhesion molecule (Nrcam). The basigins often interact with monocarboxylic acid transporters, which catalyze the transport of lactate, pyruvate, etc. (Payen et al., 2020); thus, they indirectly affect the ion transport.

The carbonic anhydrase family displays a similar duality of functions related to ion homeostasis and intercellular adhesion, and have a pronounced exclusivity (IC between 0.76 and 0.87; **Figures 6C**, **7**). The primary role of carbonic anhydrases is the catalysis of the reversible conversion of CO_2 to carbonic acid. However, some isoforms have lost their catalytic activity (Car8, 10, and 11) and they play a role in promoting the diversification in neuronal adhesion and interactions (Sterky et al., 2017).

The analysis of an RNA-seq dataset, which does not meet the inclusion criteria (cells > 100) (Ho et al., 2018), reveals a further gene family involved in cell adhesion, the collagen alpha family, expressed exclusively in corticostriatal and medium spiny neurons (**Supplementary Data 2**).

The Efficiency of RNA Detection by Single Cell RNA-seq and Stochastic Exclusivity

The efficiency of RNA detection by RNA-seq is less than 100% and is not uniform in a cell population (Baran-Gale et al., 2018). One may assume that cells with a low number of captured genes mimic exclusivity since only a few genes or gene isoforms are detected in these cells. To assess how such a cellular heterogeneity affects the quantification of stochastic gene choice, we removed 10 percent of the cells with the lowest number of detected genes and calculated the IC from the truncated population

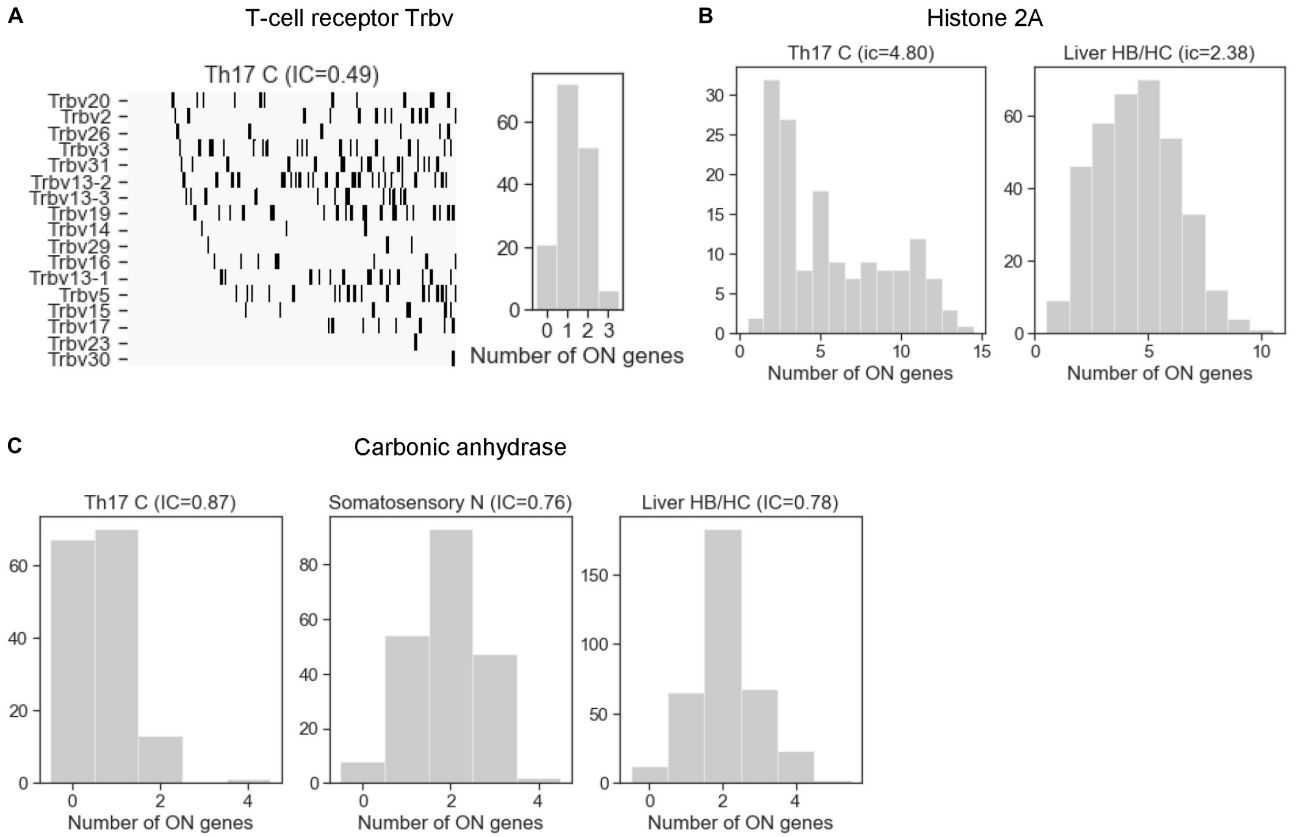


FIGURE 6 | The distribution of the number of expressed gene isoforms (ON genes) per cell in gene families with exclusive and concurrent expression. **(A)** The T-cell receptor beta chain family shows a clear exclusivity in Th17 cells ($IC = 0.49$). The left plot shows the dichotomized expression states. Each column represents a single cell. The gene isoforms are ordered according to expression frequency (highest on the top) and the cells are ordered according to number of expressed isoforms per cell (lowest on the left side). **(B)** The histone 2A family shows co-occurrence in Th17 and liver HB/HC cells, with an IC value of 4.8 and 2.38, respectively. **(C)** The number of expressed carbonic anhydrase genes per cell in Th17 cells, somatosensory neurons, and liver HB/HC (IC = 0.87, 0.76 and 0.78, respectively).

(**Figure 9A**). If the removed cells were accountable for exclusivity, the truncation would have increased the IC. However, the mean IC did not increase; in fact, it decreased slightly in the truncated population of the somatosensory neuron dataset and also in all other datasets (**Figure 9B** and **Supplementary Table 4**). **Figure 9C** shows the exclusive gene families with the lowest IC in the prostate stromal cells and the somatosensory neuron datasets, which have the lowest and highest numbers of detected genes per cell, respectively. The amiloride-sensitive sodium channel family (PTHR11690) has the lowest IC in the somatosensory neurons, whereas the PTHR33589 in the prostate stromal cells, which includes Jacalin-like lectin domain-containing proteins. The exclusive families detected in two cell types are also displayed. After truncation, the mean number of ON genes increases in most of these families, as expected, since cells with a low number of genes are removed. Importantly, the IC remained less than one in all of the families, and in several cases the IC even decreased after the truncation. Similarly, the IC remained less than one in all but two exclusive families shown in **Figure 7**.

Six datasets with TPM units having the largest gene coverage (above 8,000, see **Supplementary Table 4**) yield 34 hits while the remaining 11 TPM datasets yield only 27 hits. Thus, lower gene coverage in these datasets does not seem to lead to spurious hits, but rather reduces the success rate of the detection of exclusive families. Accordingly, the development of newer single cell RNA-seq technologies with higher capture efficiency may enable the detection of more families with exclusive gene choice.

DISCUSSION

Determinants of Exclusivity in Gene Families and Chromosomal Segments

Our results show that stochastic exclusivity is rare in both gene families and segments and concurrence is overrepresented. Multiple mechanisms are likely to underlie this phenomenon. Evolving from a single gene, paralogs have common regulatory sequences. Consequently, a shift from concurrence toward exclusivity is expected only after a sufficient evolutionary divergence in the family. Chromosomal proximity can also promote concurrence when a transcription factor affects multiple genes in a chromosomal segment (Wada et al., 2019). For example, two copies of the same gene at the same chromosomal position experience more correlated fluctuations if

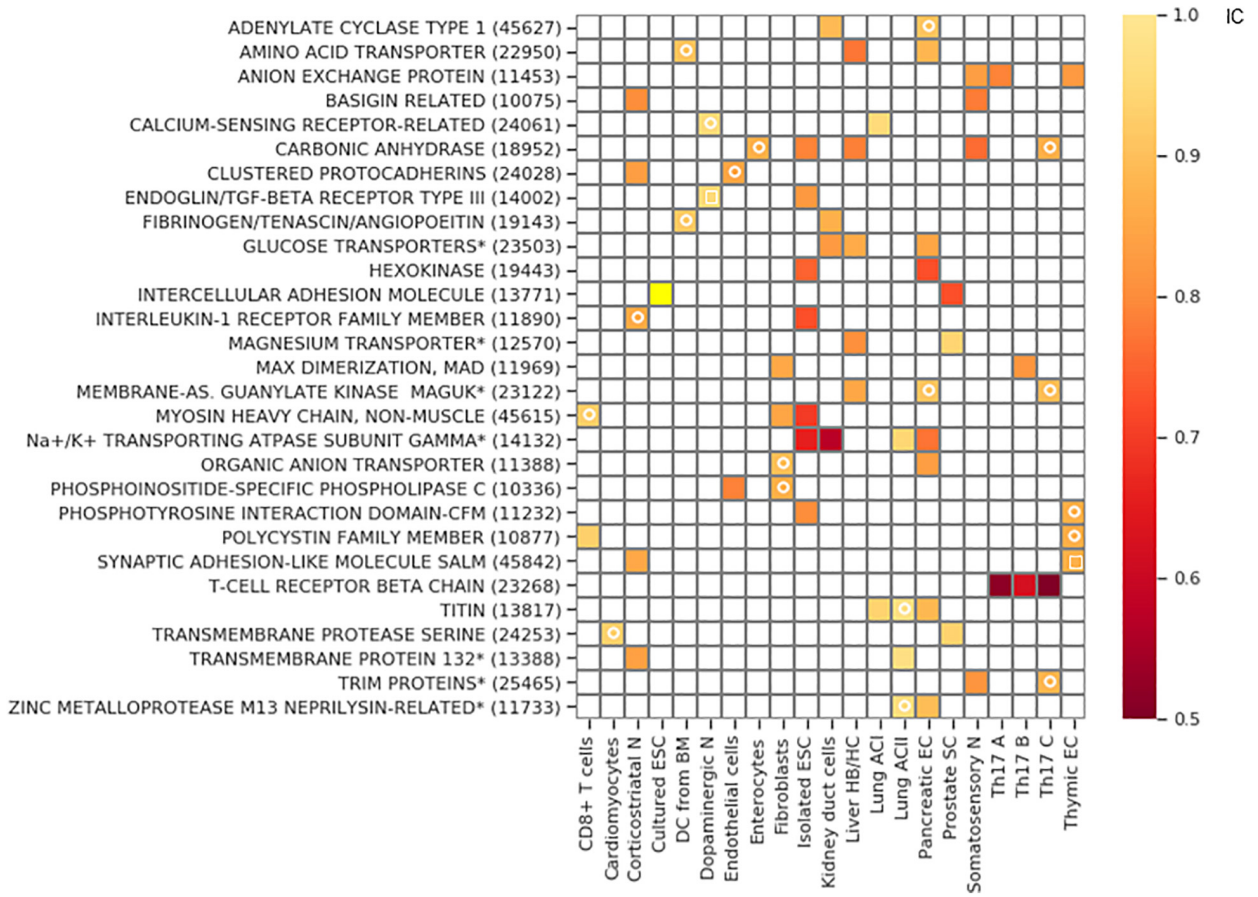


FIGURE 7 | Gene families with exclusive gene choice. Gene families with stochastic exclusive gene choice in two or more cell types; further details of selection as in **Supplementary Figure 5** (see also **Supplementary Data 1**). For the families labeled with star, descriptive names were given instead of the Panther names. The Panther numbers of the families are indicated in parenthesis. The white circle denotes segments with an IC numerically less than 1 without reaching significance. The white empty squares indicates the families that lose exclusivity after truncation of the cell population at the 10th percentile of the total number of detected genes per cell.

they are positioned on linked chromosomes than on physically separated, but homologous, chromosomes (Becskei et al., 2005). Furthermore, the positive correlation in stochastic gene expression has gradient-like features along the mammalian chromosomes (Sun and Zhang, 2019). Thus, the predominance of concurrence in the genome can be viewed as a direct consequence of evolutionary-genetic and biophysical-chemical processes.

Despite the dominance of concurrence in chromosomal segments, chromosomal proximity may promote exclusivity in the appropriate context. A single gene in the Pcdh α-genes can be chosen to be expressed upon the formation of a CTCF-mediated chromosomal loop between the chosen gene and a downstream enhancer (Wu Q. et al., 2020; Wu Y. et al., 2020). This looping mediated gene proximity may promote exclusivity and may explain the much higher exclusivity of the Pcdh α-array in comparison to the scattered Pcdhs. Recent findings indicate that the arrangement of CTCF binding sites as tandems play an important role since they insulate gene expression and thus effect stochastic promoter choice (Jia et al., 2020).

Exclusivity has no general molecular marker for all three classical exclusive gene families. Variations even exist among the Pcdh arrays. CTCF controls the expression of the Pcdh β-isoforms, as well (Hirayama et al., 2012; Sams et al., 2016) but the β-array has a larger IC than the α-array (**Supplementary Data 1**). Furthermore, cell-specific mechanisms are likely to explain why the expression in the Pcdh-α array is exclusive in some neuronal types but unbiased in others (**Figure 3B**). It is also possible that the interactions of the neurons during development determine whether or not stochastic gene choice is exclusive, which means that gene expression and cell adhesion are under mutual control.

The calculation of and analysis with IC has multiple advantages. It can help to define the range of chromosomal segments subject to exclusive gene choice, especially when the genes do not belong to a family. For example, the exclusivity in the α-Pcdh array extends beyond the array and affects two upstream genes, Zmat2 and Vaultrc5. Zmat2 has been shown to regulate the splicing of genes involved in cell adhesion (Tanis et al., 2018). Thus, Zmat2 may directly affect the Pcdh-mediated

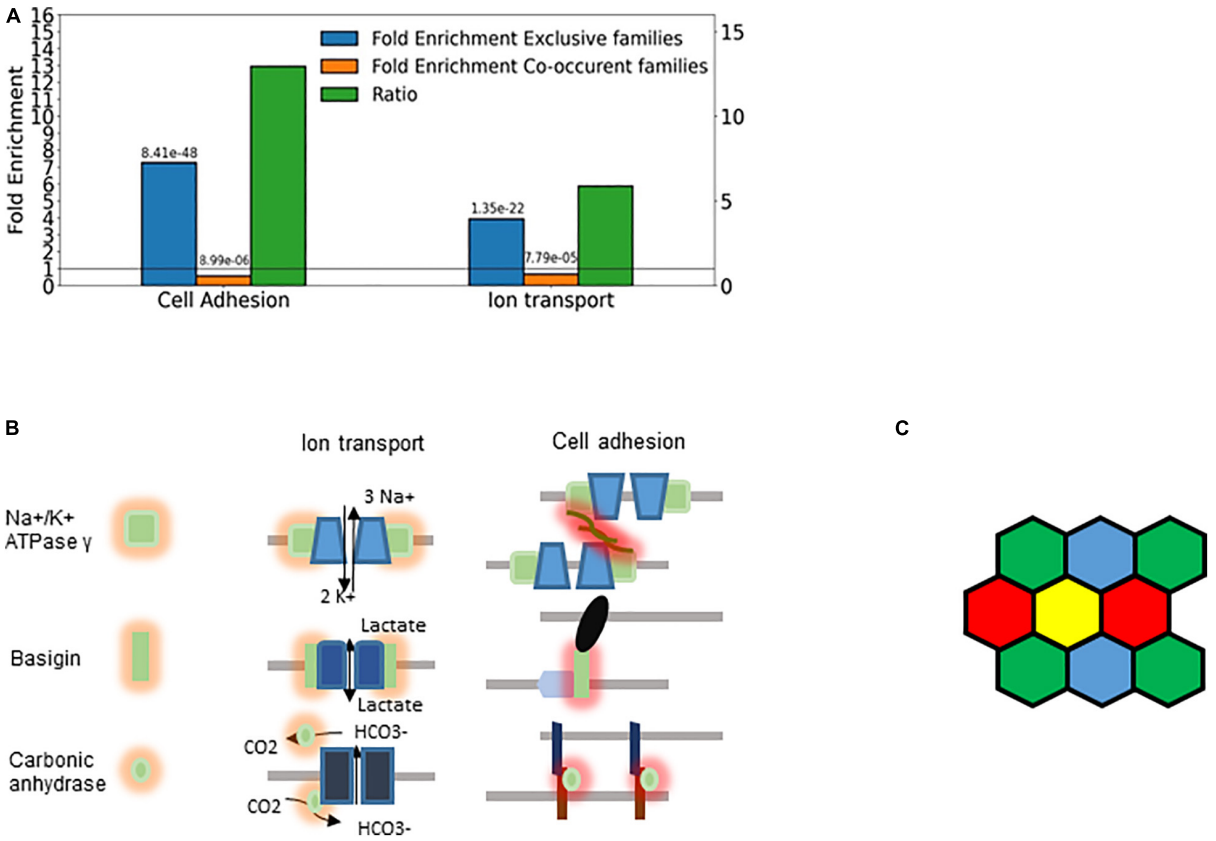


FIGURE 8 | Cellular individuality and cell adhesion. **(A)** Enrichment analysis of the genes belonging to exclusive and concurrent gene families. The *P*-values are indicated on the top of the bars. The exclusive families were selected with the criteria described in **Figures 4B**, **7**. The concurrent families (IC belonging to top 2.5 percentile) were constrained with the following criteria: mean number of expressed genes per cell higher than 0.03, IC significantly higher than 1 and at least 5 non-zero genes per family. We considered all the genes expressed at least in one cell type belonging to the selected families. The enrichment analysis was performed though an enrichment analysis tool (http://geneontology.org/). The figure shows two selected functions: Cell adhesion (GO: 0007155) and Ion transport (GO: 0006811). The ratio of the fold-enrichment in the exclusive to that in concurrent families is shown. **(B)** Schematic representation highlighting the dual role of three gene families (Fxyd, basigin, and carbonic anhydrase genes). On the left side, the cis interaction of the corresponding proteins with channels and pumps is denoted by orange shades. These functions are related to metabolic and ion homeostasis. On the right side, the trans-interaction with ligands on the adjacent cells is labeled with red shades. The glycosylation of the Fxyd protein affects the transdimerization of the Na^+/K^+ ATPase. The carbonic anhydrase interacts with the anion exchange protein, which transports HCO_3^-. **(C)** Schematic representation of cells showing that the exclusive expression of four gene isoforms (colors) is sufficient to confer cellular individuality in a two dimensional tissue.

cell adhesion. Furthermore, Vaultrc5 is a vault RNA, which controls autophagy, and several Pcdh proteins are known to associate with autophagy related proteins (Buscher et al., 2020).

Similarly, the IC formalism does not require predefined sets of genes for the assessment of exclusivity. For example, the αC1 and αC2 isoforms are usually excluded from the analysis when the number of expressed gene isoforms is quantified in the α-array due to their constitutive expression in Purkinje cells (Esumi et al., 2005). However, their expression is not constitutive in other cell types: the αC1 and αC2 isoforms are expressed at a lower frequency than some of the variable isoforms (α1-12) in corticostriatal neurons (**Figure 3A**). Since the IC formalism does not assume a single gene to be expressed in order to be exclusive, it permits the detection of exclusivity in all these cell types with different mean number of expressed genes.

IC has another important aspect, the absolute value. The T-cell receptor family with IC values as low as 0.5 has an unmatched degree of exclusivity in comparison to the other detected exclusive families. This may reflect the fact that multiple different molecular mechanisms cooperate to stabilize exclusive stochastic gene expression: the promoter choice through chromosomal looping is followed by DNA recombination and allelic exclusion (Massah et al., 2015). DNA recombination is unlikely to contribute to the exclusivity in the families involved in cell adhesion. The exact mechanism underlying exclusivity, looping or covalent epigenetic modifications or other processes, remains to be determined (Magklara and Lomvardas, 2013; Almenar-Queralt et al., 2019).

Functional Relevance of Stochastic Exclusive Gene Choice

We have used relatively stringent criteria to identify families with exclusive choice since they had to be detected in at

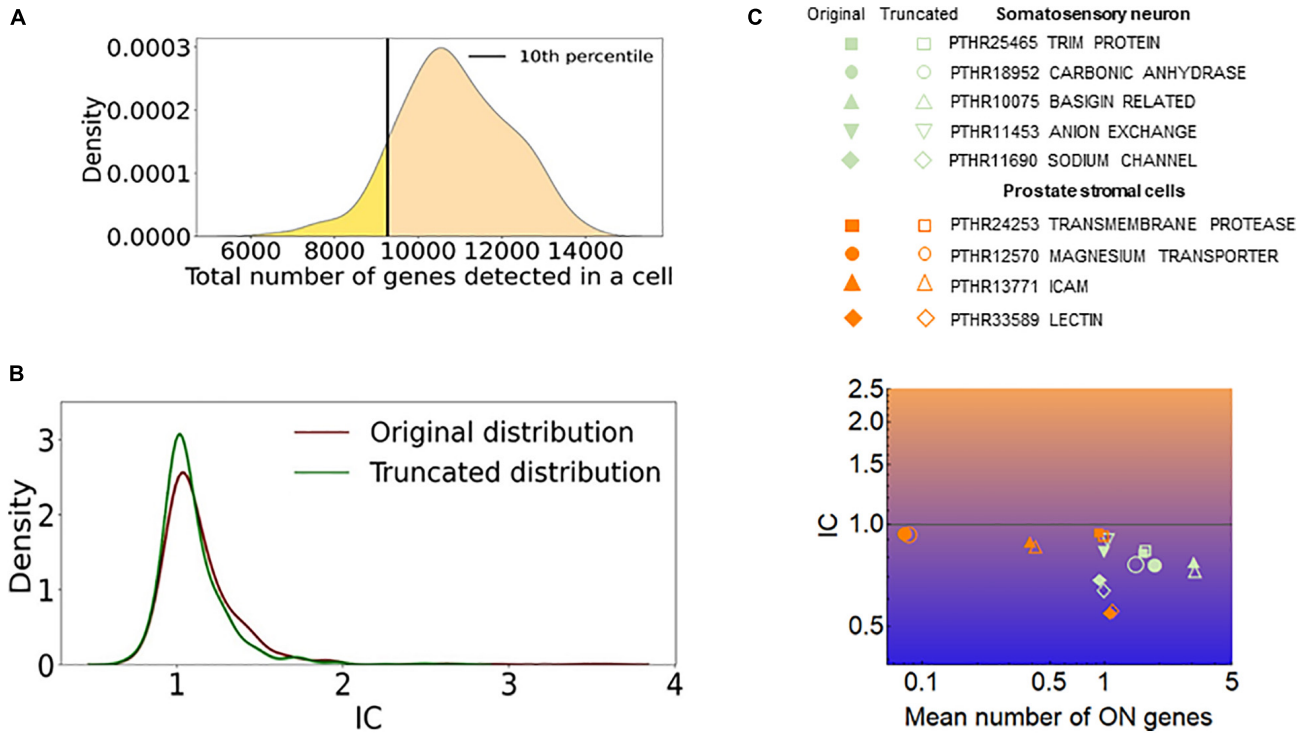


FIGURE 9 | The effect of cells with low number of detected genes on the IC. **(A)** The distribution of the total number of detected genes per cell (dgpc) in the somatosensory neuron dataset. The black line indicates the dgpc below which the cells were removed to obtain the truncated distribution. **(B)** The distribution of IC values of gene families calculated from the original and truncated cell populations shown in **(A)**. **(C)** The IC and the mean number of ON genes calculated with the original (full) and the truncated (empty) datasets. The prostate stromal cell and the somatosensory neuron datasets were used.

least two different cell types. Despite the overrepresentation of concurrence in most genomes, the exclusive gene choice is not restricted to the T-cell receptor, odorant receptor and Pcdh families. Ten other families were identified with pronounced exclusivity, with IC less than 0.8: the anion-exchange and basigin related proteins, the carbonic anhydrases, intercellular adhesion molecule, interleukin-1 receptor family, phospholipase C, the sodium/potassium transporting ATPase gamma subunit, the hexokinases and the non-muscle myosin heavy-chain. Most of them directly affect cell adhesion (**Figure 7**), but even hexokinases can affect motor or cytoskeletal proteins, and thus regulate cellular adhesion (Hsu et al., 2010; Ghosh et al., 2016). Ion transport is the second most overrepresented function in the detected families. Ions have been long known to modulate cell adhesion (Arcangeli and Becchetti, 2006). In addition to calcium, magnesium and pH are of major physiological relevance in cell adhesion (Takeichi and Okada, 1972).

Ion transport and cell adhesion can be regulated by the same protein (**Figure 8B**). For example, the ratio of the Fxyd5 isoform to the α1–β1 heterodimer determines whether the Na^+/K^+ ATPase acts as a positive or negative regulator of intercellular adhesion (Tokhtaeva et al., 2016). This is highly reminiscent of the Pcdh proteins, in which the ratio of the expressed isoforms determined intercellular adhesion (Yagi, 2012; Thu et al., 2014). Interestingly, basigin can also bind the β2−subunit of Na^+/K^+ ATPase (Heller et al., 2003).

The carbonic anhydrase isoforms Car10 and Car11 are secreted glycoproteins that are predominantly expressed in the brain. Car10 was shown to be a conserved pan-neurexin ligand (Sterky et al., 2017). Neurexins, like protocadherins, mediate interneuronal interactions, but the isoform diversity is generated primarily through alternative splicing (Mauger and Scheiffele, 2017) and not by stochastic gene choice. Overexpression of Car10 in neurons creates a shift in neurexin isoforms in mouse and human neurons, which may explain how the stochastic choice of Car isoforms generates diversity. Even catalytic Cars affect intercellular adhesion. For example, Car9, a cancer associated transmembrane isoform of carbonic anhydrase, reduces E-cadherin mediated adhesion (Svastova et al., 2003). The Cars can interact with the anion exchange proteins, Slc4a, which transport bicarbonate (Morgan et al., 2007), which is thought to accelerate CO_2 transport. Thus, two families with exclusive expression can interact physically. It remains to be determined how gene families involved in glucose transport and metabolism profit from exclusive expression. Recent advances in the description of the spatial variations in metabolism across a cell-population (Ben-Moshe and Itzkovitz, 2019; Polyzos et al., 2019) do suggest that not only cell adhesion but also ion homeostasis may profit from stochastic exclusive gene choice. The transmembrane serine proteases (Tmprss) may also affect cell adhesion by regulated proteolysis, which can help cancer cells to spread (Qiu et al., 2007; Tanabe and List, 2017).

Cells interact through homophilic or heterophilic interactions (Ahrens et al., 2002; Thu et al., 2014; Brasch et al., 2018). The affinity of the interaction can depend on the particular combination of the respective protein isoforms (Yagi, 2012). Thus, diversity through gene choice can have functional consequences. For example, choosing two isoforms from a repertoire of five genes permits 10 combinations, and thus 10 cellular identities. It is important to note that cells in a plane can become fully distinguishable with the exclusive expression of four different gene isoforms, according to the four color theorem (**Figure 8C**; Wu et al., 2015). Somewhat higher numbers are needed for cells arranged in 3-dimensional interaction networks. Thus, the detected families with 5 or more members are in principle capable of supporting sufficient diversity to enable each cell to distinguish itself from its neighbors.

The combinatorial diversity due to the random choice of multiple gene isoforms is translated into a diversity of cell-to-cell interactions, while the exclusivity guarantees the precise stoichiometry within the membrane protein complexes. This principle is a conserved property of many gene families involved in cell adhesion and ion transport beyond the protocadherins, suggesting that stochastic exclusive gene choice is an ideal mechanism to link diversity with precision in cell adhesion.

MATERIALS AND METHODS

Data Sources

To define the chromosomal segments, the Genome Reference Consortium Mouse Build 38 patch release 6 (GRCm38.p6) was used[1]. The genes marked as predicted were excluded, and only the genes sourced from Best-placed RefSeq (BestRefSeq) and Curated Genomic were considered.

PANTHER15.0 was used to map genes to their corresponding gene families[2] (Mi et al., 2019).

The single cell RNA-seq datasets are described in **Supplementary Table 1**.

Interconversion of RNA-Seq Quantification Units

TPM (Transcripts Per Million) units were analyzed without conversion. The RPKM (Reads Per Kilobase Million) and FPKM (Fragments Per Kilobase Million) can differ between samples, causing biases for the statistical interpretation of the data (Wagner et al., 2012). Therefore, they were converted into TPM units (Kim et al., 2018):

$$TPM_g = \frac{FPKM_g}{\sum_j FPKM_g} 10^6$$

$FPKM_g$ represents the FPKM values of a given gene. The gene counts are summed over the population of j cells.

[1]https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.26

[2]ftp://ftp.pantherdb.org/sequence_classifications/current_release/PANTHER_Sequence_Classification_files/PTHR15.0_mouse

Datasets with Unique Molecular Identifier (UMI) counts were used without further normalization.

Dichotomization of Expression Into ON and OFF States for the Genes in the Chromosomal Segments

To exclude the genes with unimodal expression, the bimodality coefficient was calculated for each gene:

$$b = \frac{g^2}{k + \frac{3(n-1)^2}{(n-2)(n-3)}}$$

where k is the sample excess kurtosis, g is the sample skewness, n is number of samples (i.e. cells) (Knapp, 2007). Only the genes with $b > 0.55$ were kept since a value of 5/9 or less corresponds to a unimodal distribution. This filtering was applied to data in TPM units for the analyses of chromosomal segments.

Three methods were compared to dichotomize the expression of individual genes: VRS, FM and GTME. The minimum threshold was set to be 0.5 TPM, which is widely used as threshold for a gene considered to be expressed. Thus, when a procedure resulted in a threshold with a value less than 0.5 TPM, it was replaced by 0.5 TPM. Upon determining the threshold, the genes are dichotomized. If the expression value is greater than or equal to a threshold, the gene is marked as expressed in this cell (i.e., with 1), otherwise it is marked as not expressed (i.e., with 0).

Variance Reduction Score (VRS)

VRS is a measure of bimodality, in that it reflects how much the variance of the original distribution is reduced in comparison to the sum of the variances of the two distributions obtained by the splitting of the original distribution with a threshold (Hellwig et al., 2010).

$$VRS = \frac{\sum_{x \in X_{below}} (x - \bar{x}_{below})^2 + \sum_{x \in X_{above}} (x - \bar{x}_{above})^2}{\sum_{x \in X} (x - \bar{x})^2}$$

where X is a total set of expression values of a gene, X_{below} and X_{above} are sets of expression values lower than and greater than or equal to a threshold, respectively. $\bar{x}$, $\bar{x}_{below}$ and $\bar{x}_{above}$ are the mean expression values for the three sets, respectively.

In order to find the threshold with the minimal VRS, a range of threshold values were tested for each gene. This range is a list of geometrically progressing series with the step of 1.2 starting at 0.025 quantile of non-zero expression values up to the 0.975 quantile to get a more granular view of VRS at lower thresholds. The threshold that yields the minimum VRS is chosen as a dichotomization threshold.

Fraction of Maximal Values (FM)

The FM is a biochemically motivated threshold and assumes that the expression of a gene does not vary too much around its activity specific to the ON state. For this purpose, the 1/10th of the TPM value at the 97.5 percentile was chosen. If the number of cells with non-zero expression values (N) is less than 120, then the

1/10th value of the average (arithmetic mean) of the three largest values was calculated.

$$FM = \begin{cases} \frac{x_{0.975}}{10}, & N \geq 120 \\ \frac{\sum_{i=N-2}^{N} x_i}{10}, & N < 120 \end{cases}$$

where x_p is the pth quantile of non-zero expression values, x_i is the ith element of the sorted non-zero expression value list, N is the number of non-zero expression values,

Geometric Trimmed Mid-Extreme (GTME)

The GTME is motivated by the predictions of transition rates in bistable systems: the threshold between the two states is defined as the geometric mean of the low and high states (Hsu et al., 2016). Bistable systems can underlie bimodal distribution but there is no simple relation between them because of the transiency (Pajaro et al., 2019). In order to define the threshold without knowing the exact values of ON and OFF states, the geometric mean of the non-zero TPM values at the bottom and top 2.5 percentiles (40-quantiles) of the distribution were taken. If the number of non-zero TPM values is less than 120, the average (arithmetic mean) of the three least and largest values were used to calculate the geometric mean.

$$GTME = \begin{cases} \sqrt{x_{0.025} \cdot x_{0.975}}, & N \geq 120 \\ \sqrt{\sum_{i=1}^{3} x_i \cdot \sum_{i=N-2}^{N} x_i}, & N < 120 \end{cases}$$

Analogous thresholds allow for the precise calculation of the transition rates in a bistable cell population (Hsu et al., 2016).

Familywise Thresholds

Assuming that the expression values of genes within a family are similar, a common threshold can be defined for all genes within a family. The familywise FM (fFM) and GTME (fGTME) were calculated as follows. The RNA counts larger than 0.5 were considered instead of the x > 0 condition. When the respective cell number N was larger than 120, the $x_{g,\ 0.025}$ and $x_{g,\ 0.975}$ were calculated for each gene. The fFM was calculated from the maximum of the set of $x_{g,\ 0.975}$, $g \in GF$, representing each gene in a gene family (GF). Thus, a single gene in the family determines the threshold for all the genes in the family. Similarly, the two genes corresponding to the minimum of the $x_{g,\ 0.025}$ and the maximum of the $x_{g,\ 0.975}$ $g \in GF$, set determine the fGTME. Analogous calculation were performed for $N < 120$, with mean averages of the three largest and smallest expression values, instead of the values at the percentiles.

Fitting of Distributions

Probability density (or mass) functions, $\varphi(x)$, were fitted with the FindDistribution of Wolfram Mathematica, which combines the Bayesian information criterion with priors over distributions to select both the best distribution and the best parameters for it. Commonly fitted distributions were the Binomial, Cauchy, Exponential, Gamma, Geometric, Normal, Laplace, Logistic, Lognormal, Poisson, Negative Binomial, Yule-Simmons distribution and their mixtures. Whenever a mixture distribution was obtained by the FindDistribution, the antimodes were calculated. The antimodes were determined analytically based on the first and second derivatives of $\varphi(x)$. The smallest antimode in the range $x > 0.5$ was used as thresholds for dichotomization for each gene. As opposed to other methods, the $\varphi(x)$ based thresholds were not used for calculation of IC across the genome, since they were obtained for a smaller number of genes in comparison to the other methods. This is because the fitting of $\varphi(x)$ is less robust, especially when there are few cells in the OFF or ON expression states or when the measurement error is larger.

The Interdependence Coefficient (IC)

The IC is the ratio of the observed variance in the number of expressed genes in a cell population to the variance of the Poisson binomial distribution expected from the expression frequencies (Wada et al., 2018). The variance of the generalized binomial (Poisson-binomial) distribution is a function of the probability of each isoforms i to be expressed (p_i):

$$IC = \frac{\sigma^2_{OBS}}{\sigma^2_{PB}}, \quad \text{where } \sigma^2_{PB} = \sum_{i=1}^{N_\alpha} (1 - p_i) p_i$$

p_i is equal to the ON cell frequency. IC = 1 indicates an unbiased (independent) stochastic gene choice according to the Poisson-binomial distribution, akin to a relation Fano-factor = 1, which indicates a Poisson distribution for a single gene (Ozbudak et al., 2002).

The 95% confidence interval (CI) of the IC was calculated by bootstrapping. After resampling the cell population, the observed variance and the expected Poisson-binomial variance were calculated for each resampling, and IC was calculated. When the 95% CI was below one, exclusivity was considered significant.

Permutation Tests

Permutation tests were used to assess the effect of chromosomal adjacency and family membership on stochastic interdependence. The expression values of the genes are shuffled among all genes but for those that were not measured in a particular dataset or were not bimodal. The shuffling was performed 1000 times. Similarly, the assignment of genes (i.e., their respective expression values) to gene families is shuffled. Only the genes that are present in both the families and the RNA-seq datasets are reassigned in a way that the sizes and number of families are preserved. The distribution of IC values were obtained for each re-shuffling.

The 10th and 90th percentiles and their ratio were calculated as representative quantiles for the exclusivity and concurrence. Therefore, the P-values for the changes in the quantiles were calculated based on the permutation tests (Ernst, 2004). The P-value was calculated as follows

$$P\ value = \frac{1 + \sum_{i=1}^{N} I(|\hat{x} - \bar{x}| \geq |x_i - \bar{x}|)}{1 + N}$$

where $\hat{x}$ is the original statistic, $\bar{x}$ is the mean of the shuffled statistic, x_i is the statistic of the ith permutation, and N is the number of permutations. The pseudocount is added to avoid

P-values of 0. Since 1,000 permutations were performed, the smallest *P*-value is 0.001.

A two-tailed *P*-value of 0.05 was selected for a statistic to be considered significantly higher or lower than the statistic of the shuffled distributions. Exclusivity is promoted when the 10th percentile of the original distribution is significantly smaller. Similarly, co-occurrence is promoted when the 90th percentile is significantly greater. These tests were applied for each chromosomal segment size separately. Families were grouped according to their size, and the same tests were performed as for the chromosomal adjacency. Only family sizes that have 30 or more gene families were taken for the permutation tests.

Identification of Genes Subject to Concurrent or Exclusive Gene Choice in Multiple Cell Types

To assess which sets of genes conserve their mode of interdependence across multiple cell types, the pairwise overlap of gene segments or gene families that are within the bottom or top 2.5 percentiles of their respective IC distributions was determined. In other words, a segment or a family is considered a hit, if it appears in two datasets in the respective tails of IC distributions. The chromosomal segments were overlapped separately for each segment size, whereas all families were considered together (**Supplementary Data 1**). Further conditions to filter the selected genes are described in the relevant context.

Examination of the Relations Between Allelic Exclusion and Stochastic Gene Choice

The mean interallelic correlation was calculated by averaging the Fisher transform of the Spearman correlation coefficient calculated for the two alleles, followed by a back transformation (Alexander, 1990):

$$\overline{\rho_S} = Tanh[\frac{1}{N}\sum_{i=1}^{n} Arctanh[\rho_{Si}]]$$

To calculate the biallelic IC differential between the diplotypes and haplotypes, the following formula was used:

$$\text{Biallelic IC differential} = \text{Log}\left[\frac{IC_{Diplotype}}{\sqrt{IC_{Haplotype\_1}IC_{Haplotype\_2}}}\right]$$

DATA AVAILABILITY STATEMENT

The algorithms used in this work are available in the GitHub repository: https://github.com/d-lowl/stochastic-gene-choice.

AUTHOR CONTRIBUTIONS

AB designed the project and wrote the manuscript. MI wrote the programs. MI, SF, and AB analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded in part by the Swiss National Foundation (SNF 310030_185001).

ACKNOWLEDGMENTS

We thank Vincent Jaquet, Michael Stadler, and Peter Scheiffele for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.642212/full#supplementary-material

REFERENCES

Ahrens, T., Pertz, O., Haussinger, D., Fauser, C., Schulthess, T., and Engel, J. (2002). Analysis of heterophilic and homophilic interactions of cadherins using the c-Jun/c-Fos dimerization domains. *J. Biol. Chem.* 277, 19455–19460. doi: 10.1074/jbc.m200606200

Alexander, R. A. (1990). A note on averaging correlations. *Bull. Psychon. Soc.* 28, 335–336. doi: 10.3758/bf03334037

Almenar-Queralt, A., Merkurjev, D., Kim, H. S., Navarro, M., Ma, Q., Chaves, R. S., et al. (2019). Chromatin establishes an immature version of neuronal protocadherin selection during the naive-to-primed conversion of pluripotent stem cells. *Nat. Genet.* 51, 1691–1701. doi: 10.1038/s41588-019-0526-4

Arcangeli, A., and Becchetti, A. (2006). Complex functional interaction between integrin receptors and ion channels. *Trends Cell Biol.* 16, 631–639. doi: 10.1016/j.tcb.2006.10.003

Baran-Gale, J., Chandra, T., and Kirschner, K. (2018). Experimental design for single-cell RNA sequencing. *Brief. Funct. Genom.* 17, 233–239. doi: 10.1093/bfgp/elx035

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. *Cell* 163, 1596–1610. doi: 10.1016/j.cell.2015.11.018

Baudrimont, A., Jaquet, V., Wallerich, S., Voegeli, S., and Becskei, A. (2019). Contribution of RNA degradation to intrinsic and extrinsic noise in gene expression. *Cell Rep.* 26, 3752–3761 e3755.

Becskei, A., Kaufmann, B. B., and Van Oudenaarden, A. (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* 37, 937–944. doi: 10.1038/ng1616

Ben-Moshe, S., and Itzkovitz, S. (2019). Spatial heterogeneity in the mammalian liver. *Nat. Rev. Gastroenterol. Hepatol.* 16, 395–410. doi: 10.1038/s41575-019-0134-x

Boucher, A., Desforges, M., Duquette, P., and Talbot, P. J. (2007). Long-term human coronavirus-myelin cross-reactive T-cell clones derived from multiple sclerosis patients. *Clin. Immunol.* 123, 258–267. doi: 10.1016/j.clim.2007.02.002

Bradley, C. P., Teng, F., Felix, K. M., Sano, T., Naskar, D., Block, K. E., et al. (2017). Segmented filamentous bacteria provoke lung autoimmunity by inducing gut-lung axis Th17 cells expressing dual TCRs. *Cell Host Microbe* 22, 697–704 e694.

Brady, B. L., Steinel, N. C., and Bassing, C. H. (2010). Antigen receptor allelic exclusion: an update and reappraisal. *J. Immunol.* 185, 3801–3808. doi: 10.4049/jimmunol.1001158

Brasch, J., Katsamba, P. S., Harrison, O. J., Ahlsen, G., Troyanovsky, R. B., Indra, I., et al. (2018). Homophilic and heterophilic interactions of Type II cadherins identify specificity groups underlying cell-adhesive behavior. *Cell Rep.* 23, 1840–1852. doi: 10.1016/j.celrep.2018.04.012

Buscher, M., Horos, R., and Hentze, M. W. (2020). 'High vault-age': non-coding RNA control of autophagy. *Open Biol.* 10:190307. doi: 10.1098/rsob.190307

Chen, L., Lee, J. W., Chou, C. L., Nair, A. V., Battistone, M. A., Paunescu, T. G., et al. (2017). Transcriptomes of major renal collecting duct cell types in mouse identified by single-cell RNA-seq. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9989–E9998.

Cheng, S., Pei, Y., He, L., Peng, G., Reinius, B., Tam, P. P. L., et al. (2019). Single-Cell RNA-Seq reveals cellular heterogeneity of pluripotency transition and X chromosome dynamics during early mouse development. *Cell Rep.* 26, 2593–2607 e2593.

Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Stat. Sci.* 19, 676–685.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., et al. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* 37, 171–176. doi: 10.1038/ng1500

Finn, E. H., and Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science* 365:eaaw9498. doi: 10.1126/science.aaw9498

Friedrich, D., Friedel, L., Finzel, A., Herrmann, A., Preibisch, S., and Loewer, A. (2019). Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency. *Mol. Syst. Biol.* 15:e9068.

Gaublomme, J. T., Yosef, N., Lee, Y., Gertner, R. S., Yang, L. V., Wu, C., et al. (2015). Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* 163, 1400–1412. doi: 10.1016/j.cell.2015.11.009

Ghosh, S., Gupta, P., and Sen, E. (2016). TNFalpha driven HIF-1alpha-hexokinase II axis regulates MHC-I cluster stability through actin cytoskeleton. *Exp. Cell Res.* 340, 116–124. doi: 10.1016/j.yexcr.2015.11.016

Guo, M., Du, Y., Gokey, J. J., Ray, S., Bell, S. M., Adam, M., et al. (2019). Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat. Commun.* 10:37.

Haber, A. L., Biton, M., Rogel, N., Herbst, R. H., Shekhar, K., Smillie, C., et al. (2017). A single-cell survey of the small intestinal epithelium. *Nature* 551, 333–339.

Hanchate, N. K., Kondoh, K., Lu, Z., Kuang, D., Ye, X., Qiu, X., et al. (2015). Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science* 350, 1251–1255. doi: 10.1126/science.aad2456

Harrison, O. J., Brasch, J., Katsamba, P. S., Ahlsen, G., Noble, A. J., Dan, H., et al. (2020). Family-wide structural and biophysical analysis of binding interactions among non-clustered delta-protocadherins. *Cell Rep.* 30, 2655–2671 e2657.

Heller, M., Von Der, Ohe, M., Kleene, R., Mohajeri, M. H., and Schachner, M. (2003). The immunoglobulin-superfamily molecule basigin is a binding protein for oligomannosidic carbohydrates: an anti-idiotypic approach. *J. Neurochem.* 84, 557–565. doi: 10.1046/j.1471-4159.2003.01537.x

Hellwig, B., Hengstler, J. G., Schmidt, M., Gehrmann, M. C., Schormann, W., and Rahnenfuhrer, J. (2010). Comparison of scores for bimodality of gene expression distributions and genome-wide evaluation of the prognostic relevance of high-scoring genes. *BMC Bioinform.* 11:276.

Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered pcdh genes in neurons. *Cell Rep.* 2, 345–357. doi: 10.1016/j.celrep.2012.06.014

Ho, H., De Both, M., Siniard, A., Sharma, S., Notwell, J. H., Wallace, M., et al. (2018). A guide to single-cell transcriptomics in adult rodent brain: the medium spiny neuron transcriptome revisited. *Front. Cell. Neurosci.* 12:159.

Hook, P. W., Mcclymont, S. A., Cannon, G. H., Law, W. D., Morton, A. J., Goff, L. A., et al. (2018). Single-cell RNA-Seq of mouse dopaminergic neurons informs candidate gene selection for sporadic parkinson disease. *Am. J. Hum. Genet.* 102, 427–446. doi: 10.1016/j.ajhg.2018.02.001

Hsu, C., Jaquet, V., Maleki, F., and Becskei, A. (2016). Contribution of bistability and noise to cell fate transitions determined by feedback opening. *J. Mol. Biol.* 428, 4115–4128. doi: 10.1016/j.jmb.2016.07.024

Hsu, R. M., Tsai, M. H., Hsieh, Y. J., Lyu, P. C., and Yu, J. S. (2010). Identification of MYO18A as a novel interacting partner of the PAK2/betaPIX/GIT1 complex and its potential function in modulating epithelial cell migration. *Mol. Biol. Cell* 21, 287–301. doi: 10.1091/mbc.e09-03-0232

Ji, Q., Perchellet, A., and Goverman, J. M. (2010). Viral infection triggers central nervous system autoimmunity via activation of CD8+ T cells expressing dual TCRs. *Nat. Immunol.* 11, 628–634. doi: 10.1038/ni.1888

Jia, Z., Li, J., Ge, X., Wu, Y., Guo, Y., and Wu, Q. (2020). Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol.* 21:75.

Kakaradov, B., Arsenio, J., Widjaja, C. E., He, Z., Aigner, S., Metz, P. J., et al. (2017). Early transcriptional and epigenetic regulation of CD8(+) T cell differentiation revealed by single-cell RNA sequencing. *Nat. Immunol.* 18, 422–432. doi: 10.1038/ni.3688

Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., et al. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J. Biol. Chem.* 281, 30551–30560. doi: 10.1074/jbc.m605677200

Khamlichi, A. A., and Feil, R. (2018). Parallels between mammalian mechanisms of monoallelic gene expression. *Trends Genet.* 34, 954–971. doi: 10.1016/j.tig.2018.08.005

Kim, S. C., Yu, D., and Cho, S. B. (2018). COEX-Seq: convert a variety of measurements of gene expression in RNA-Seq. *Genomics Inform.* 16:e36. doi: 10.5808/gi.2018.16.4.e36

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201. doi: 10.1016/j.cell.2015.04.044

Knapp, T. R. (2007). Bimodality revisited. *J. Mod. Appl. Stat. Methods* 6:3.

Kwon, O. J., Zhang, Y., Li, Y., Wei, X., Zhang, L., Chen, R., et al. (2019). Functional heterogeneity of mouse prostate stromal cells revealed by single-cell RNA-Seq. *iScience* 13, 328–338. doi: 10.1016/j.isci.2019.02.032

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., et al. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* 565, 251–254. doi: 10.1038/s41586-018-0836-1

Li, C. L., Li, K. C., Wu, D., Chen, Y., Luo, H., Zhao, J. R., et al. (2016). Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. *Cell Res.* 26, 83–102. doi: 10.1038/cr.2015.149

Magklara, A., and Lomvardas, S. (2013). Stochastic gene expression in mammals: lessons from olfaction. *Trends Cell Biol.* 23, 449–456. doi: 10.1016/j.tcb.2013.04.005

Marzluff, W. F., and Koreski, K. P. (2017). Birth and death of Histone mRNAs. *Trends Genet.* 33, 745–759. doi: 10.1016/j.tig.2017.07.014

Massah, S., Beischlag, T. V., and Prefontaine, G. G. (2015). Epigenetic events regulating monoallelic gene expression. *Crit. Rev. Biochem. Mol. Biol.* 50, 337–358. doi: 10.3109/10409238.2015.1064350

Mauger, O., and Scheiffele, P. (2017). Beyond proteome diversity: alternative splicing as a regulator of neuronal transcript dynamics. *Curr. Opin. Neurobiol.* 45, 162–168. doi: 10.1016/j.conb.2017.05.012

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D419–D426.

Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* 565, 448–453. doi: 10.1038/s41586-018-0845-0

Morgan, P. E., Pastorekova, S., Stuart-Tilley, A. K., Alper, S. L., and Casey, J. R. (2007). Interactions of transmembrane carbonic anhydrase, CAIX, with bicarbonate transporters. *Am. J. Physiol. Cell Physiol.* 293, C738–C748.

Muramatsu, T. (2016). Basigin (CD147), a multifunctional transmembrane glycoprotein with various binding partners. *J. Biochem.* 159, 481–490. doi: 10.1093/jb/mvv127

Nomura, S., Satoh, M., Fujita, T., Higo, T., Sumida, T., Ko, T., et al. (2018). Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. *Nat. Commun.* 9:4435.

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and Van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.* 31, 69–73.

Pajaro, M., Otero-Muras, I., Vazquez, C., and Alonso, A. A. (2019). Transient hysteresis and inherent stochasticity in gene regulatory networks. *Nat. Commun.* 10:4581.

Payen, V. L., Mina, E., Van Hee, V. F., Porporato, P. E., and Sonveaux, P. (2020). Monocarboxylate transporters in cancer. *Mol. Metab.* 33, 48–66.

Polyzos, A. A., Lee, D. Y., Datta, R., Hauser, M., Budworth, H., Holt, A., et al. (2019). Metabolic reprogramming in astrocytes distinguishes region-specific neuronal susceptibility in huntington mice. *Cell Metab.* 29:e1211.

Qiu, D., Owen, K., Gray, K., Bass, R., and Ellis, V. (2007). Roles and regulation of membrane-associated serine proteases. *Biochem. Soc. Trans.* 35, 583–587. doi: 10.1042/bst0350583

Redies, C., Vanhalst, K., and Roy, F. (2005). delta-Protocadherins: unique structures and functions. *Cell Mol. Life Sci.* 62, 2840–2852. doi: 10.1007/s00018-005-5320-z

Reinius, B., Mold, J. E., Ramskold, D., Deng, Q., Johnsson, P., Michaelsson, J., et al. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* 48, 1430–1435. doi: 10.1038/ng.3678

Rodrigo, G. (2019). Ab initio scaling laws between noise and mean of gene expression. *Phys. Rev. E* 100:032415.

Ryu, C. J., Haines, B. B., Lee, H. R., Kang, Y. H., Draganov, D. D., Lee, M., et al. (2004). The T-cell receptor beta variable gene promoter is required for efficient V beta rearrangement but not allelic exclusion. *Mol. Cell. Biol.* 24, 7015–7023. doi: 10.1128/mcb.24.16.7015-7023.2004

Sams, D. S., Nardone, S., Getselter, D., Raz, D., Tal, M., Rayi, P. R., et al. (2016). Neuronal CTCF is necessary for basal and experience-dependent gene regulation, memory formation, and genomic structure of BDNF and Arc. *Cell Rep.* 17, 2418–2430. doi: 10.1016/j.celrep.2016.11.004

Sansom, S. N., Shikama-Dorn, N., Zhanybekova, S., Nusspaumer, G., Macaulay, I. C., Deadman, M. E., et al. (2014). Population and single-cell genomics reveal the aire dependency, relief from polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res.* 24, 1918–1931. doi: 10.1101/gr.171645.113

Schlitzer, A., Sivakamasundari, V., Chen, J., Sumatoh, H. R., Schreuder, J., Lum, J., et al. (2015). Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat. Immunol.* 16, 718–728. doi: 10.1038/ni.3200

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240. doi: 10.1038/nature12172

Sterky, F. H., Trotter, J. H., Lee, S. J., Recktenwald, C. V., Du, X., Zhou, B., et al. (2017). Carbonic anhydrase-related protein CA10 is an evolutionarily conserved pan-neurexin ligand. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1253–E1262.

Sun, M., and Zhang, J. (2019). Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. *PLoS Genet.* 15:e1008389. doi: 10.1371/journal.pgen.1008389

Svastova, E., Zilka, N., Zat'ovicova, M., Gibadulinova, A., Ciampor, F., Pastorek, J., et al. (2003). Carbonic anhydrase IX reduces E-cadherin-mediated adhesion of MDCK cells via interaction with beta-catenin. *Exp. Cell Res.* 290, 332–345. doi: 10.1016/s0014-4827(03)00351-3

Tacer, K. F., and Potts, P. R. (2017). Cellular and disease functions of the prader-willi syndrome gene MAGEL2. *Biochem. J.* 474, 2177–2190. doi: 10.1042/bcj20160616

Takeichi, M., and Okada, T. S. (1972). Roles of magnesium and calcium ions in cell-to-substrate adhesion. *Exp. Cell Res.* 74, 51–60. doi: 10.1016/0014-4827(72)90480-6

Tanabe, L. M., and List, K. (2017). The role of type II transmembrane serine protease-mediated signaling in cancer. *FEBS J.* 284, 1421–1436. doi: 10.1111/febs.13971

Tanis, S. E. J., Jansen, P., Zhou, H., Van Heeringen, S. J., Vermeulen, M., Kretz, M., et al. (2018). Splicing and chromatin factors jointly regulate epidermal differentiation. *Cell Rep.* 25, 1292–1303 e1295.

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19, 335–346. doi: 10.1038/nn.4216

Thu, C. A., Chen, W. V., Rubinstein, R., Chevee, M., Wolcott, H. N., Felsovalyi, K. O., et al. (2014). Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. *Cell* 158, 1045–1059. doi: 10.1016/j.cell.2014.07.012

Ting, C. H., Lee, K. Y., Wu, S. M., Feng, P. H., Chan, Y. F., Chen, Y. C., et al. (2019). FOSB(-)PCDHB13 axis disrupts the microtubule network in non-small cell lung cancer. *Cancers* 11:107. doi: 10.3390/cancers11010107

Tokhtaeva, E., Sun, H., Deiss-Yehiely, N., Wen, Y., Soni, P. N., Gabrielli, N. M., et al. (2016). The O-glycosylated ectodomain of FXYD5 impairs adhesion by disrupting cell-cell trans-dimerization of Na,K-ATPase beta1 subunits. *J. Cell Sci.* 129, 2394–2406. doi: 10.1242/jcs.186148

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263. doi: 10.1038/nrg2538

Vatakis, D. N., Arumugam, B., Kim, S. G., Bristol, G., Yang, O., and Zack, J. A. (2013). Introduction of exogenous T-cell receptors into human hematopoietic progenitors results in exclusion of endogenous T-cell receptor expression. *Mol. Ther.* 21, 1055–1063. doi: 10.1038/mt.2013.28

Veerman, K., Tardiveau, C., Martins, F., Coudert, J., and Girard, J. P. (2019). Single-cell analysis reveals heterogeneity of high endothelial venules and different regulation of genes controlling lymphocyte entry to lymph nodes. *Cell Rep.* 26, 3116–3131 e3115.

Wada, T., Wallerich, S., and Becskei, A. (2018). Stochastic gene choice during cellular differentiation. *Cell Rep.* 24, 3503–3512. doi: 10.1016/j.celrep.2018.08.074

Wada, T., Wallerich, S., and Becskei, A. (2019). Synthetic transcription factors switch from local to long-range control during cell differentiation. *ACS Synth. Biol.* 8, 223–231. doi: 10.1021/acssynbio.8b00369

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285. doi: 10.1007/s12064-012-0162-3

Weon, J. L., and Potts, P. R. (2015). The MAGE protein family and cancer. *Curr. Opin. Cell Biol.* 37, 1–8. doi: 10.1016/j.ceb.2015.08.002

Wu, D., and Yang, X. O. (2020). TH17 responses in cytokine storm of COVID-19: An emerging target of JAK2 inhibitor Fedratinib. *J. Microbiol. Immunol. Infect.* 53, 368–370. doi: 10.1016/j.jmii.2020.03.005

Wu, M., Ha, S., Abdullah, T., and Kim, C. (2015). Exclusive channel allocation methods based on four-color theorem in clustering sensor networks. *New Trends Comput. Collective Intelligence* 572, 107–116. doi: 10.1007/978-3-319-10774-5_10

Wu, Q., Liu, P., and Wang, L. (2020). Many facades of CTCF unified by its coding for three-dimensional genome architecture. *J. Genet. Genom.* 47, 402–424.

Wu, Y., Jia, Z., Ge, X., and Wu, Q. (2020). Three-dimensional genome architectural CCCTC-binding factor makes choice in duplicated enhancers at Pcdhalpha locus. *Sci. China Life Sci.* 63, 835–844. doi: 10.1007/s11427-019-1598-4

Yagi, T. (2012). Molecular codes for neuronal individuality and cell assembly in the brain. *Front. Mol. Neurosci.* 5:45.

Yang, L., Wang, W. H., Qiu, W. L., Guo, Z., Bi, E., and Xu, C. R. (2017). A single-cell transcriptomic analysis reveals precise pathways and regulatory mechanisms underlying hepatoblast differentiation. *Hepatology* 66, 1387–1401. doi: 10.1002/hep.29353

Yu, X. X., Qiu, W. L., Yang, L., Zhang, Y., He, M. Y., Li, L. C., et al. (2019). Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. *EMBO J.* 38:e100164.

Zhang, X., Hong, D., Ma, S., Ward, T., Ho, M., Pattni, R., et al. (2020). Integrated functional genomic analyses of klinefelter and turner syndromes reveal global network effects of altered X chromosome dosage. *Proc. Natl. Acad. Sci. U. S. A.* 117, 4864–4873. doi: 10.1073/pnas.1910003117

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: 22 June 2021
doi: 10.3389/fcell.2021.676543

Imprinted Gene Expression and Function of the Dopa Decarboxylase Gene in the Developing Heart

*Adam R. Prickett[1], Bertille Montibus[1], Nikolaos Barkas[1†], Samuele M. Amante[1†], Maurício M. Franco[1†], Michael Cowley[1†], William Puszyk[1], Matthew F. Shannon[1], Melita D. Irving[1,2], Marta Madon-Simon[3], Andrew Ward[3], Reiner Schulz[1], H. Scott Baldwin[4] and Rebecca J. Oakey[1]**

[1] Department of Medical and Molecular Genetics, King's College London, London, United Kingdom, [2] Department of Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom, [3] Department of Biology and Biochemistry and Centre for Regenerative Medicine, University of Bath, Bath, United Kingdom, [4] Department of Pediatrics (Cardiology), Vanderbilt University Medical Center, Nashville, TN, United States

OPEN ACCESS

Edited by:
Anthony Isles,
Cardiff University, United Kingdom

Reviewed by:
Ionel Sandovici,
University of Cambridge, United Kingdom
Rosalind M. John,
Cardiff University, United Kingdom

***Correspondence:**
Rebecca J. Oakey
rebecca.oakey@kcl.ac.uk

†Present address:
Nikolaos Barkas,
Human Cell Atlas Data Coordination Platform, Data Sciences Platform, Broad Institute, Cambridge, MA, United States
Samuele M. Amante,
Centre for Genomics and Child Health, Barts and The London School of Medicine and Dentistry, Blizard Institute, Queen Mary University of London, London, United Kingdom
Maurício M. Franco,
Embrapa Genetic Resources and Biotechnology, Laboratory of Animal Reproduction, Brasília, Brazil
Michael Cowley,
Department of Biological Sciences, Center for Human Health and the Environment, North Carolina State University, Raleigh, NC, United States

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

Received: 05 March 2021
Accepted: 19 May 2021
Published: 22 June 2021

Dopa decarboxylase (DDC) synthesizes serotonin in the developing mouse heart where it is encoded by *Ddc_exon1a*, a tissue-specific paternally expressed imprinted gene. *Ddc_exon1a* shares an imprinting control region (ICR) with the imprinted, maternally expressed (outside of the central nervous system) *Grb10* gene on mouse chromosome 11, but little else is known about the tissue-specific imprinted expression of *Ddc_exon1a*. Fluorescent immunostaining localizes DDC to the developing myocardium in the pre-natal mouse heart, in a region susceptible to abnormal development and implicated in congenital heart defects in human. *Ddc_exon1a* and *Grb10* are not co-expressed in heart nor in brain where *Grb10* is also paternally expressed, despite sharing an ICR, indicating they are mechanistically linked by their shared ICR but not by *Grb10* gene expression. Evidence from a *Ddc_exon1a* gene knockout mouse model suggests that it mediates the growth of the developing myocardium and a thinning of the myocardium is observed in a small number of mutant mice examined, with changes in gene expression detected by microarray analysis. Comparative studies in the human developing heart reveal a paternal expression bias with polymorphic imprinting patterns between individual human hearts at *DDC_EXON1a,* a finding consistent with other imprinted genes in human.

Keywords: dopa decarboxylase, knock-out, imprinting, heart, mouse, human

INTRODUCTION

There are over a hundred imprinted genes in the mouse and human genome (Kelsey and Bartolomei, 2012) many of which contribute to mammalian growth and development (Cleaton et al., 2014). Genomic imprinting is the parent-of-origin-dependent, allele-specific expression of a gene (Ideraabdullah et al., 2008). Imprinted genes are regulated by epigenetic mechanisms including parent-of-origin-dependent, allele-specific DNA methylation of CpG-rich differentially methylated regions (DMRs). There are two classes of DMRs: germline DMRs that are established during gametogenesis are maintained throughout development and act as imprinting control regions (ICRs); and somatic DMRs that arise post-fertilization, are often tissue-specific and can contribute to the regulation of imprinted gene clusters together with the ICR (Edwards and Ferguson-Smith, 2007). Both classes are maintained during cell division (Lewis and Reik, 2006).

The Dopa decarboxylase gene (*Ddc)* has two transcript isoforms, one is expressed from both parental alleles and the other is imprinted. *Ddc*, is expressed from both parental alleles in the urinary system, eye, nervous system, liver, limbs, alimentary system and ear. Deficiency of this canonical form of the gene in humans results in an autosomal recessive inborn error of metabolism (MIM #608643) (Lee et al., 2013; Smith et al., 2014). The second transcript isoform, known as *Ddc_exon1a* was identified through the analysis of a differential gene expression screen designed to detect novel imprinted genes and it is imprinted, being expressed only from the paternally inherited allele in the developing mouse heart (Menheniott et al., 2008).

Imprinted genes are typically found in clusters in the genome (Barlow and Bartolomei, 2014) and *Ddc_exon1a* is no exception lying adjacent to *Grb10,* an imprinted gene that encodes an intracellular signaling adaptor protein. *Grb10* is typically maternally expressed (Charalambous et al., 2003) and acts to restrict fetal growth and promote adipose deposition in adulthood (Smith et al., 2007; Madon-Simon et al., 2014). Unusually, *Grb10* is expressed from the opposite (paternal) allele in the CNS but the mechanism that underlies this switch between maternal and paternal expression is unclear, as is the role for paternally expressed *Grb10* in neurons (Plasschaert and Bartolomei, 2015). The two genes therefore comprise an imprinting cluster where imprinted expression is directed via the shared ICR in the 5′ untranslated region (UTR) of *Grb10* [also known as the *Grb10* CpG island 2 (CGI 2) DMR]. When the ICR is ablated, it results in loss of both *Grb10* and *Ddc_exon1a* imprinting (Shiura et al., 2009). Since the regulation of imprinted gene clusters is typically co-ordinated (Ideraabdullah et al., 2008), we reasoned that *Ddc_exon1a* expression could be co-ordinately regulated with the expression of *Grb10*.

Ddc_exon1a is the only variant of Ddc expressed in heart and is a unique example of a transcript that shows heart-specific genomic imprinting. *Grb10* has a more complex imprinted expression pattern in the developing embryo but exhibits paternal expression in the CNS. There are varying reports regarding *Ddc_exon1a* expression in the brain (Shiura et al., 2009; Madon-Simon et al., 2014; Smith et al., 2014; Plasschaert and Bartolomei, 2015). As *Grb10* is expressed maternally in most tissues but shows paternal expression specifically in the brain and in subsets of cells in the heart, this suggested to us that *Ddc_exon1A* and *Grb10* might be linked and coordinately expressed in brain and heart.

We sought to investigate the allelic expression of *Ddc_exon1a* in the brain and the heart using allele-specific assays in tissues from reciprocal hybrid mouse strains. As allelic expression was predominantly observed in the heart, the spatial pattern of DDC_EXON1A protein was delineated in the developing mouse heart by immunostaining embryonic sections. Spatial distribution was also compared to its ICR partner *Grb10* in a gene trap transgenic mouse line. To investigate function, the phenotype of a knockout mouse model of *Ddc_exon1a* was examined and changes in the developing myocardium were seen along with gene expression changes associated with tissue development and cellular organization. An antisense transcript overlaps *Ddc_exon1a* but no evidence was found for it influencing imprinted *Ddc_exon1a* expression *in cis*. A comparative study examining the expression of *DDC_EXON1A* in 40 human fetal hearts including fetal-maternal pairs, reveals a paternal expression bias and a polymorphic pattern of imprinted gene expression.

MATERIALS AND METHODS

Allele-Specific RT-PCR Assays

RNA was extracted from tissue using the RNAeasy Kit™ (Qiagen), assessed for purity using NanoDrop (requiring a 260/280 ratio of ∼2.0) and integrity using Agilent 2100 Bioanalyzer, and converted to cDNA with a Superscript II™ (Invitrogen) kit, as per manufacturers' instructions. For mouse *Ddc_exon1a* the allele was identified via a G/A single nucleotide polymorphism (SNP) between *Mus musculus domesticus* C57BL/6J (B) and *Mus musculus castaneus* CAST/EiJ (C) in exon 3 (mm9, chr11:11776278). Transcripts with this SNP were amplified by PCR from reciprocal BxC and CxB hybrids (by convention, the maternal genotype is listed first) and sequenced. *Ddc_exon1a* and *Ddc_canonical* transcripts were amplified using exon-specific forward primers EXON1A-F (5′-TGTCACCAAGGAGAGAGAGAGA-3′) and EXON1-F (5′- AGTTGTGTCGCCACCTCCT-3′) and a common reverse primer, EXON4-R (5′-CAGCTCTTCCAGCCAAAAAG-3′). PCR: 94°C for 3 min, 34 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min, with a final extension of 72°C for 5 min and Sanger sequencing using an ABI 3730xl.

For AK0066690, nested primers were used AK006690_ F_outer: CCAGCCTCCATTTCAGAGTT, AK006690_R_outer: TTGACTAGGAATATTTCCTTCCAT, amplicon size: 250bp. Inner primers were AK006690_F_inner: TTCAGCCAAGAG TGCTTAGG, AK006690_R_inner: GCTGCTGCATGCTTAT TTGT, amplicon size: 184 bp.

Immunostaining

E15.5 wildtype embryos were fixed in 4%PFA for 1 h at 4°C, dehydrated and embedded in paraffin wax. Antigen retrieval was performed by boiling in high pH antigen unmasking solution (Vector Labs). Slides were blocked with 4% v/v donkey serum (abcam, ab7475) for 1.5 h. Primary antibodies in the following dilutions: anti-DOPA Decarboxylase antibody (ab3905) (1:500), goat-α-mouse ANF (1:100) in 0.01% Tween-20, 2% v/v donkey serum in PBS were dropped onto slides and incubated in a humidified chamber at 4°C for 16 h. Slides were washed 3X in 0.01% Tween-20 in PBS. Secondary antibodies Alexa Fluor 555 donkey-α-goat (Invitrogen) and Alexa Fluor 647 donkey-α-rabbit (Invitrogen) were diluted 1:300 in 0.01% Tween-20 PBS, dropped onto slides and incubated for 2 h at RT in the dark. Slides were washed 3X in PBS and mounted using ProLong® Gold Antifade Reagent with DAPI (Invitrogen).

Histological Analysis

Histological analysis was performed on the *Grb10KO* with a *lacZ* reporter construct at the 3′ end of exon 8 (Garfield et al., 2011). Gestating Grb10KO females were sacrificed at

18.5 days *post coitum* and the uterine horns isolated immediately. All animal experiments were approved and regulated by the university ethics committee at the University of Bath and conform to the guidelines from Directive 2010/63/EU of the European Parliament on the protection of animals used for scientific purposes. The reporter insertion ablates all isoforms of *Grb10* in mouse embryos and results in a null. Where *Grb10* expression is perturbed, lacZ protein expression occurs. In $Grb10^{+/KO}$ mice, tissue localization of *Grb10* is blue with X-gal staining. $Grb10^{+/KO}$ E15.5 embryos were fixed in 4% (w/v) PFA, cryoprotected in 30% sucrose and embedded in OCT. Sections were stained for paternal *Grb10* in 1 mg/ml X-gal diluted in stain base [30 mM $K_4Fe(CN)_6$ 30 mM $K_3Fe(CN)_6.3H_2O$, 2 mM $MgCl_2$, 0.01% (w/v) sodium deoxycholate, 0.02% (v/v) Igepal CA-630 in 0.1% PBS] for 18 h at 28°C and counterstained using nuclear fast red.

Samples were stained for DDC using VECTASTAIN® ABC kit (Vector Labs) with blocking in 5% skim milk. DDC antibody (Abcam, #3905) was used at a 1:500 dilution. Sections were counterstained with Harris' hematoxylin (30 s), and incubation in Scott's tap water (Fisher) for 1 min.

Morphological Analysis Using Episcopic Fluorescence Image Capture (EFIC)

Embryos and dissected neonatal mouse hearts were fixed and embedded in paraffin as for immunostaining. Measurements of embryos were adjusted to the crown rump length to account for differences in embryo size due to variation in the time of conception on a given day of gestation as is convention. For EFIC analysis (Rosenthal et al., 2004) sections were re-embedded in red aniline dyed wax and photographed using an EFIC system at Vanderbilt University, with a sectioning size of 5μm. Samples were photographed at a magnification of 20× using appropriate mCherry and GFP wavelengths. Images were quality controlled by visual inspection and rebuilt in 3D using Volocity™ image analysis software (Perkin Elmer) and virtually re-sectioned in a plane that bisected the mitral and aortic valve, with the measurements taken on this plane at the base of the papillary muscle to ensure samples were measured equally. All measurements were made blind with the identity of the samples only revealed prior to statistical analysis. Comparisons between sample groups were made using a Mann-Whitney test.

Microarray Analysis

Microarray libraries were generated as per manufacturers' instructions for Affymetrix Genechips™ on three Ddc^{WT} and four maternally deleted ($Ddc^{MAT\Delta}$) (making seven wild type samples) and four paternally deleted ($Ddc^{PAT\Delta}$) and one homozygous mutant ($Ddc^{\Delta\Delta}$) (providing five knockouts of *Ddc_exon1a* in embryonic heart) using two separate six-lane arrays. Raw probe signals were background-corrected using NEQC quartile-normalized and a linear model was fitted to compare the effects of different genotypes in LIMMA (Smyth, 2004). These data have been deposited in GEO, Accession number GSE87595.

Western Blotting

An E15.5 embryo carcass was macerated in 1 ml of RIPA buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% (w/v) sodium deoxycholate, 0.1% SDS, 1 mM PMSF, 1× protease inhibitor (Roche)] and centrifuged at 16,000 g for 20 min at 4°C. Total protein in the supernatant was measured using the BCA protein assay kit (Pierce) and stored at −20°C. The same protocol was used to extract protein from cultured cells without maceration. 20 μg in 25 μl of each sample was mixed 1:1 with 2× reducing buffer (62.5 mM Tris HCl pH 6.8, 2% (w/v) SDS, 6 M Urea, 2% (v/v) Igepal CA-630, 5% (v/v) β-mercaptoethanol, 0.02% (w/v) bromophenol blue, 4% glycerol and heated to 95°C for 5 min. Samples were electrophoresed alongside a Spectra™ multicolor protein ladder (Thermo Fisher Scientific) on a 12% polyacrylamide resolving gel: 12% polyacrylamide (National Diagnostics), 0.37M Tris:HCl pH 8.8, 0.1% SDS, 0.05% AMPS, 0.05% TEMED with a stacking gel (5% polyacrylamide, 0.12M Tris:HCl pH 6.8, 0.05% AMPS, 0.1% TEMED) at 100 V for 3 h in running buffer (0.1% SDS, 25 mM Tris, 208 mM glycine). Protein was transferred at 90 V for 2 h to a PVDF membrane (Bio-Rad) using western blot wet transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol. The membrane was blocked for 90 min in 5% powdered skimmed milk (Marvel) in 0.1% Tween-20 with PBS. Primary antibodies were diluted, rabbit-α-mouse DDC (1:1,000) in 5% milk in 0.1% Tween-20 with PBS and incubated with the membrane overnight at 4°C. Membranes were washed 3X in 0.1% Tween-20 with PBS for 15 min and incubated in peroxidase-conjugated goat anti-rabbit secondary antibody (DAKO P0448) diluted 1:2,000 in 5% milk in 0.1% Tween-20 with PBS for 1 h at RT. 3X washes were performed and protein detected using the ECL system (Amersham). Proteins were visualized by exposure to Fuji film developed on a Laser45 machine. For loading control, membranes were stripped by heating at 50°C for 30 min in stripping buffer (100 mM 2-Mecaptoethanol, 2% SDS, 62.5 mM Tris:HCl, pH6.7), washed 3X in 0.1% Tween-20 with PBS for 15 min, and re-probed using mouse-α-mouse Tubulin (abcam anti-alpha Tubulin antibody ab7291; 1:5,000) and rabbit-α-mouse Histone H3 (abcam H3 ab1791; 1:5,000).

Knockout Mice

Ddc knockout model generation was carried out by Lexicon Genetics Inc., United States and breeding, genotyping and tissue acquisition by the UC Davis mouse biology program. The mouse strain and cell lines are deposited as frozen embryos in the International Mouse Strain Resource[1] and listed in MGI as $Ddc^{Gt(OST129277)Lex}$ (B6;129S5-$^{DdcGt(neo)420Lex}$, ID# 11693–UCD). Homozygous null mice have a lethal phenotype and their number is lower than Mendelian expectations at E10.5 (for example, four heterozygous inter-crosses of this knock-out mouse resulted in 11 WT, 14 Heterozygous and 3 double knock-out embryos. Mendelian ratios would have expected numbers in line with 11 WT, 22, Heterozygous and 11 double knock-out embryos) non-Mendelian ratios were also observed at litters dissected at E15.5. Heterozygous mice did not show overt

[1] http://www.informatics.jax.org/external/ko/lexicon/2095.html

phenotypes in the Lexicon Genetics high-throughput phenotype screen which aimed to identify genes that when ablated, resulted in overt phenotypes in obesity and skeletal anomalies. This screen is acknowledged to be conflicted between studying individual lines of mice and screening many lines rapidly. Therefore, compromises were made in terms of phenotypic detail, making detailed analyses of heterozygotes essential at the individual gene/strain level. Here we examine the embryos in detail for cardiac phenotypes which were not scored in the Lexicon Genetics screen. The Grb10KO knockout mouse strain was that described in Garfield et al. (2011).

Human Tissue Acquisition

Twenty-five human heart and matched decidua from 4 to 13 weeks were provided by the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR)[2] from the Institute of Genetic Medicine, Newcastle and Institute of Child Health, London. Fifteen fetal heart samples and matched maternal cheek swabs were collected via an approved protocol from the Joint Research Ethics Committee of London, Camberwell St Giles, project ID 53717. Informed consent was obtained for the inclusion of these samples. The study was performed abiding by the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research.

Human embryo allele-specific assays by RT-PCR and Sanger sequencing. RNA was extracted from tissue and converted to cDNA using the RNAeasy^{TM} (Qiagen) and Superscript II^{TM} (Invitrogen) kits, as per manufacturers' instructions. gDNA was extracted using a DNAeasy kit^{TM} (Qiagen). SNPs were identified in the UCSC genome browser and amplified with primers:

DDC_13/15R:GGCATTTAGCCACATGACAA–59.5
DDC_13/15F:ATTCTGGGGCTTGTCTGCTT–61.2
DDC_1/4F:TGGAGAATCCCATCAAGGAG–60.0
DDC_1/1a/4R:CACAGTCTCCAGCTCTGTGC–59.8
DDC_1a/4F:GGACAGAGAGCAAGTCACTCC–59.0
DDC_1a/4F2:CTGTCACTGTGGAGAGGAGAG–57.6 and sequenced on an ABI 3730xl.

RESULTS

Ddc-Exon1a Is Predominantly Found in the Developing Mouse Heart

Fluorescence immunostaining reveals the cellular distribution of cardiac DDC protein in E15.5 hearts. Because *Ddc-Exon1a* is the only isoform expressed in the developing heart (Menheniott et al., 2008) the staining reflects its expression. Sections in the coronal plane show a four-chamber view of the heart (**Figure 1**) which were either co-stained for DDC and MF-20 (an antibody to the myocardial marker myosin heavy chain) (**Figures 1A,C**), or DDC and ANF (a marker of trabeculae, the complex meshwork of myocardial strands) (**Figures 1D–I**). DDC protein was present throughout the ventricular myocardium and inter-ventricular septum (IVS), the structure between the right and left ventricles (**Figures 1B,E,H**), but was absent from the atria and the trabecular layer, except where the trabeculae meet the compact myocardial layer (**Figures 1D–I**). In addition, DDC protein was not detected in the epicardium or endocardium. Expression in cardiac fibroblasts is not ruled out, but this cell population is a small component of the ventricular wall at this stage of development (Lajiness and Conway, 2014). Scattered DDC is detected throughout the compact myocardium of both ventricles, with protein present in the myocytes of the right ventricular apex, outflow tract, and right ventricular portion of the interventricular septum at E15.5, and this same pattern was observed at E18.5 (**Figures 1E,J–L**). All of these regions of the myocardium (RV apex, RV outflow tract, and RV portion of the interventricular septum) are derived primarily from progenitor cells of the secondary heart field and are particularly susceptible to abnormal development leading to congenital heart defects (Kelly, 2012).

Ddc_exon1a exists in an imprinting cluster along with *Grb10* and their imprinted expression is coordinated by the *Grb10* ICR. Paternal *Grb10,* which is expressed in a subset of cells in the heart, was detected using a *Grb10*-LacZ reporter transmitted through the paternal germline (Garfield et al., 2011). Its restricted expression appears punctate in the IVS and the atrio-ventricular septum (**Figure 2**) and is present in a small region next to the right ventricular myocardium suggestive of a coronary vessel. Paternally derived DDC_EXON1a protein, however, was more broadly evident throughout the myocardium (**Figure 2**). DDC_EXON1a protein and *Grb10* gene expression assays do not provide a direct comparison, but these data indicate that DDC_EXON1a and *Grb10* are not obviously present in the same cell types by visual inspection, therefore these genes are not likely to share tissue-specific regulatory elements.

Ddc_exon1a expression was also examined in whole brain (**Supplementary Figure 1**), because a number of imprinted genes exhibit tissue-specific imprinting in the brain (Ferrón et al., 2011) and because of the known switch to paternal expression of *Grb10* in neuronal cells. The assay may, however, be limited because the glia present in whole brain tissue samples express *Grb10* reciprocally from the maternal allele, which could confound an allelic-specific assay of mixed cell types (Yamasaki-Ishizaki et al., 2007). Allele-specific assays measure the height of sequencing peaks from parental alleles and here indicate that *Ddc_exon1a* is expressed from both parental alleles in whole brain, and in some sub-regions including the pre-optic area of the hypothalamus, the cerebellum and the brain stem (**Supplementary Figure 1**) consistent with published studies (Gregg et al., 2010a,b; DeVeale et al., 2012). We therefore examined data from a single cell type to complement this analysis. Transcriptomic data analysis in neural stem cells from C57Bl/6J × JF1 hybrids (Bouschet et al., 2017) can be utilized to assay allele-specificity by counting the number of aligned sequencing reads originating from each parental allele using SNPs between the two mouse strains. In neural stem cells there is low expression of *Ddc*, but a slight paternal bias of expression is detected at a number of different SNPs (albeit in the common part of the gene with *Ddc* canonical) (**Supplementary Figures 1C,D**).

[2]http://www.hdbr.org

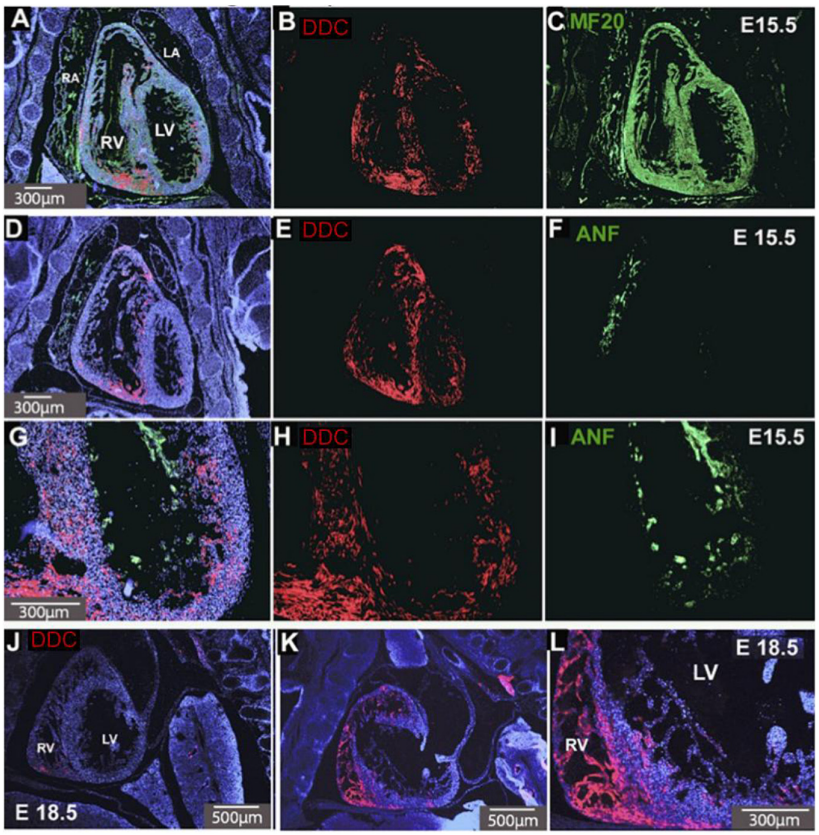


FIGURE 1 | DDC in E15.5 and E18.5 hearts. Coronal sections of the heart co-stained for DDC (red) and the myocardial marker MF-20 (green) **(A–C)** and DDC and a marker of trabeculae, ANF (green) **(D–I)**. DDC protein (red) is present throughout the ventricular myocardium and interventricular septum (IVS), the structure separating the right and left ventricles, **(B,E,H)**, but absent from the atria and the trabecular layer, except where the trabeculae meet the compact myocardial layer **(D–I)**. This is also seen in sagittal sections at E18.5 **(J–L)** where DDC is present in the myocardium, IVS, and right ventricular myocardium. Blue staining is DAPI nuclear counterstain **(A,D,G,J,K,L)**. RV Right Ventricle, LV Left Ventricle, RA Right Atrium, LA Left Atrium. Minimum number of hearts examined = 3.

Antisense transcripts are involved in imprinted gene regulation at several well characterized loci (Sleutels et al., 2002; Redrup et al., 2009). The *AK006690* transcript at this locus is annotated as transcribed in the antisense direction to *Ddc_exon1a* and its expression was confirmed in newborn brain, heart and liver (data not shown). *AK006690* was assayed for allele-specific expression in heart and brain at E13.5, E16.5/E15.5 and newborn stages in mouse reciprocal hybrids. The transcript was found to be expressed from both parental alleles in brain and heart at E13.5 with a bias toward expression from the paternal allele in later stages of development in heart (**Supplementary Figure 2**). The observed allelic expression bias also can have a genetic cause, for example the influence of a nearby SNP on the amplification efficiency, but reciprocal hybrid assays suggest that the parental expression in heart at later stages is from the paternal allele (**Supplementary Figure 2**).

Characterization of a *Ddc_exon1a* Deleted Mouse Model

Ddc_exon1a is expressed from the paternal allele in developing heart (Menheniott et al., 2008) therefore mice inheriting a null allele through the paternal line do not express *Ddc_exon1a* in this tissue. Four genotypes were assayed for expression; Ddc^{WT}, $Ddc^{MAT\,\Delta}$ (maternal deletion), $Ddc^{PAT\,\Delta}$ (paternal deletion), and $Ddc^{\Delta\,\Delta}$ (deletion on both alleles). Quantitative PCR showed diminished expression in $Ddc^{PAT\,\Delta}$ compared to Ddc^{WT} and $Ddc^{MAT\,\Delta}$ embryos (**Supplementary Figure 3**). There is a minor contribution from the maternal allele in the $Ddc^{PAT\,\Delta}$ genotype with the majority being derived from the paternal allele (**Supplementary Figure 3**), an observation that is consistent with imprinted gene expression (Horsthemke et al., 2011; Korostowski et al., 2011). DDC protein was not detected in $Ddc^{\Delta\,\Delta}$ animals (**Supplementary Figure 3**).

Morphological changes at key sites of cardiac *Ddc* expression, including the width of the IVS and the thickness of the compact layer at the apical region of the right ventricle were measured in the $Ddc^{PAT\,\Delta}$ heart. An episcopic system (Rosenthal et al., 2004) was used to eliminate distortions associated with sectioning at E15.5 in hearts from Ddc^{WT} ($n = 6$), and $Ddc^{PAT\,\Delta}$ ($n = 3$) animals which were all scored blind to the genotype. $Ddc^{PAT\,\Delta}$ embryos had a thinner compact layer in the right ventricle compared to Ddc^{WT} by 0.019 μm (**Figure 3**). A Mann-Whitney test did not reveal a difference in these measurements at a significance level

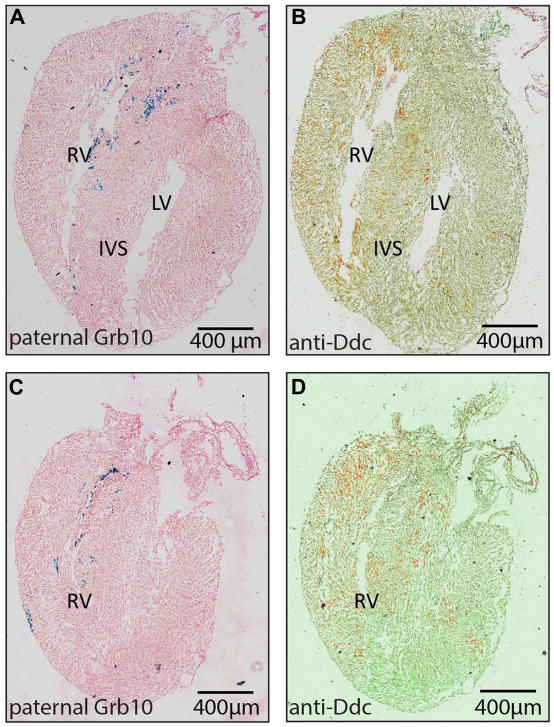


FIGURE 2 | Analysis of paternal *Grb10* expression and DDC protein staining in E18.5 heart serial sections. **(A,B)** are serial sections in the sagittal plane showing the right ventricle (RV) and the left ventricle (LV). **(C,D)** are serial sections showing only the RV where DDC staining is most abundant. **(A,C)** are stained with X-gal to reveal paternal *Grb10* expression in blue from the lacZ reporter construct and the counterstaining is pink. **(B,D)** are stained for DDC protein using DAB (brown) via immunohistochemistry, counterstaining is pale green. DDC protein distribution is extensive compared to *Grb10* which is restricted to the inter ventricular septum (IVS). Minimum number of hearts examined = 3, maximum number = 6.

of 5 % ($p = 0.0952$) with the number of mice examined but does suggest the possibility of thinning in the mutants. Thinning of the RV compact layer in the three ablated three ablated embryos could be the result of several different mechanisms (decreased myocyte proliferation, increased myocyte apoptosis, decreased progenitor cell expansion in the secondary heart field, or altered endocardial myocardial interactions) and further mechanistic resolution would be useful but is beyond the scope of this study. No statistically significant differences between Ddc^{WT} and $Ddc^{PAT\,\Delta}$ were observed for the septum thickness or overall embryo size measured by crown-rump length again measured blind to genotype.

Expression microarrays were performed between Ddc^{WT} and $Ddc^{\Delta\,\Delta}$ mice and the major difference detected was the *Ddc* gene itself (**Supplementary Table 1**). The modest impact on the transcriptome might be because the samples were heterogeneous or because the Ddc^{WT} and $Ddc^{MAT\,\Delta}$ were combined as were the $Ddc^{PAT\,\Delta}$ *and* $Ddc^{\Delta\,\Delta}$. Perturbations in molecular pathways could explain associated phenotypes and the ontology analysis (**Supplementary Table 1**) supports a role for DDC in cardiomyocyte growth and proliferation.

Imprinted Expression of *DDC_EXON1A* in Human Heart Tissues

The organization of the *Ddc/Grb10* locus is conserved between mouse and human where it is located on Chromosome 7 in the human genome (Hitchins et al., 2002). Studies have shown that *DDC* is expressed from both parental alleles in several tissues from six human fetuses (Hitchins et al., 2002) but heart had not been assayed. We sequenced for SNPs in 25 human fetal hearts to test for monoallelic and parent-of-origin-specific expression of *DDC_EXON1A*. A SNP was present in three informative samples, two displayed mono-allelic expression (**Figures 4A,B**), the third showed a biased expression (**Figure 4C**). A further 15 fetal heart samples were collected with matched maternal genomic DNA and these were sequenced for SNPs. A SNP was found in two informative samples, sample 11,886 showed biased expression from the paternal allele (**Figure 4D**) and 11,908 showed paternal expression (**Figure 4E**). Polymorphic imprinting patterns are consistent with findings at other human imprinted loci such as *IGF2* (Giannoukakis et al., 1996) where inter-individual variation in parental allele-specific expression/imprinting has been documented as well as more broadly at other imprinted loci across the genome (Zink et al., 2018).

DISCUSSION

Ddc_exon1a Imprinted Expression

Ddc_exon1a is paternally expressed in the developing mouse heart. Immunostaining of sectioned mouse hearts reveals strong signal in the region of the secondary heart field where progenitor cells go on to form the distal parts of the outflow tract and arterial trunks, right ventricle and interventricular system. Abnormal development in any of these components of cardiac development can result in congenital heart disease (Chaudhry et al., 2014; Kelly et al., 2014) and the importance of right ventricular abnormalities in the pathology of cardiovascular disease in the adult has recently been an intense are of investigation (Amsallem et al., 2018).

As is typical for imprinted genes, *Ddc_exon1a* exists in an imprinting cluster, sharing an imprinting control region (ICR) with the *Grb10* gene. The deletion of the ICR on the paternal allele in mouse heart results in the silencing of the active paternal *Ddc_exon1a* allele indicating that imprinted *Ddc_exon1a* expression in heart is governed by the ICR via a *cis*-acting mechanism. Deletion of the maternally inherited ICR does not alter expression of *Ddc_exon1a* in heart (Shiura et al., 2009) because the maternal allele is normally epigenetically silenced. Investigating the tissue distribution of *Ddc_exon1a* and *Grb10* is an important step for examining the regulatory relationship between these two clustered genes. Given that *Ddc_exon1a* is highly tissue-specific in its expression, the spatial distribution of these genes was examined in the developing heart and appeared to be non-overlapping. Paternal *Grb10* gene expression and DDC protein localization (**Figure 2**) suggests that paternal *Ddc_exon1a* and paternal *Grb10* imprinted gene expression is not coupled in the heart.

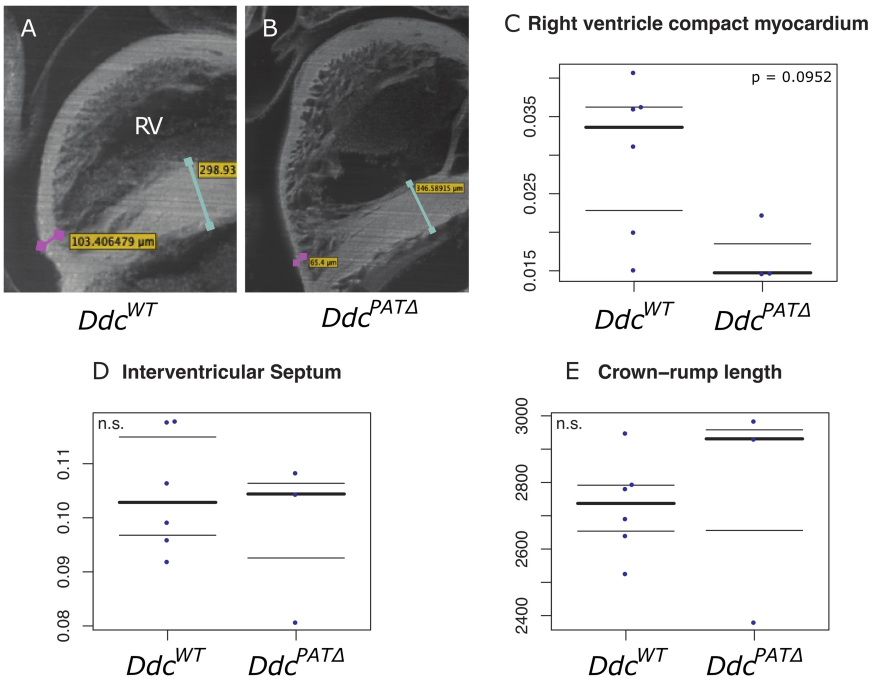


FIGURE 3 | Morphological analysis of *Ddc* knockout hearts. Episcopic fluorescence image capture measurements of wildtype Ddc^{WT} and $Ddc^{PAT\,\Delta}$ knockout E15.5 mouse heart regions using embedded, sequentially sectioned hearts built into a 3D model using the Velocity™ software. Visualization of the 3D ventricles depicting where the measurements were made in **(A)** Ddc^{WT} and **(B)** $Ddc^{PAT\,\Delta}$ embryos. The thickness of the right ventricle (RV) compact myocardium at the apical point parallel to the interventricular septum (IVS) is shown by the pink bar in **(A)** and pink bar in **(B)**. The IVS measured at the widest point is indicated by the turquoise bar in **(A)** and the turquoise bar in **(B)**. All measurements were adjusted to crown rump length to control for embryo size variation **(C–E)**. A thinning of the compact layer in the right ventricle is suggested **(C)** in knockout animal hearts compared to wild type with no change in the width of the IVS **(D)** or crown-rump length **(E)** but the total numbers of embryos studied was not sufficient to show statistical significance.

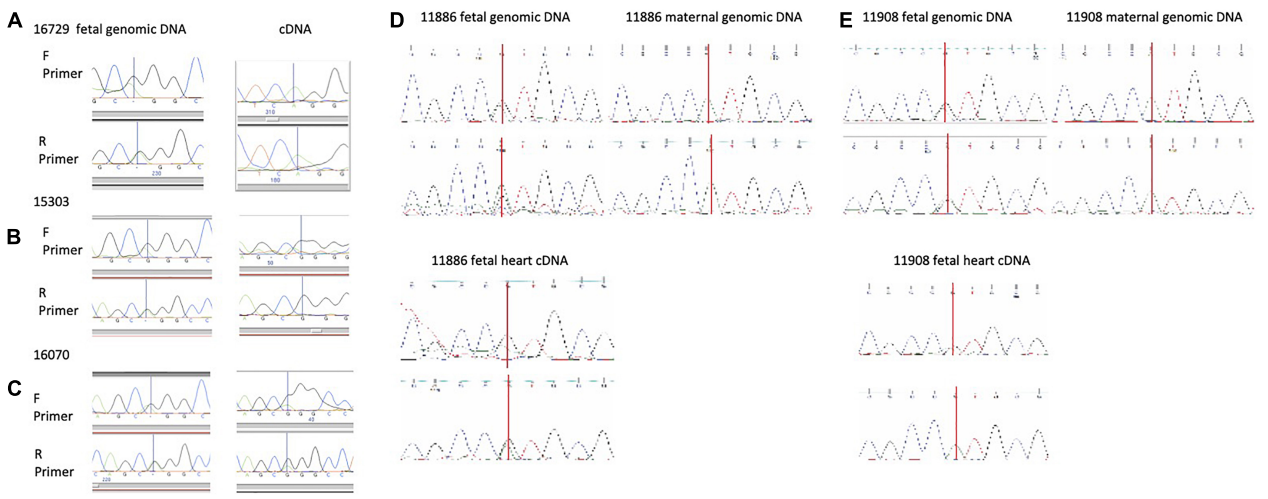


FIGURE 4 | (A,B) Human fetal heart allele-specific assays, informative samples from 25 hearts collected from a tissue bank. Individuals in **(A,B)** are mono allelic (one peak across the SNP) and the individual in **(C)** presented with a biallelic pattern of expression (2 peaks across the SNP). SNPs annotated in the UCSC genome browser were identified in these individuals by sequencing genomic (g)DNA in the forward (F) and reverse (R) directions but allelic origin could not be assigned without parental samples. The SNP is indicated by a vertical line. **(D,E)** Human fetal heart allele-specific assays, informative samples from 15 individual hearts with both fetal and maternal DNA samples. SNPs annotated in the UCSC genome browser were identified by sequencing with flanking primers in gDNA from the mother (single peak) and fetus (double peak) in the upper panels. Allele-specific assays amplifying and sequencing cDNA from fetal heart RNA are in the lower panels. Sample 11,886 has both peaks indicating expression from the maternal and paternal alleles whereas mono-allelic paternal expression of DDC_EXON1A is detected in 11,908.

Grb10 is also oppositely imprinted (paternally expressed) in the brain compared to most other tissues and this suggested that brain could also be a useful tissue in which to examine the imprinted expression of *Ddc_exon1a* for the identification of tissue-specific imprinting mechanisms. However, *Ddc_exon1a* is not imprinted in the brain of neonatal mice (**Supplementary Figure 2**) implying that the epigenetic control of *Ddc_exon1a* imprinting is unlikely to be co-ordinated with *Grb10* in brain, although paternal expression was reported in NSCs. The ICR seems only to influence tissue-specific imprinting of *Ddc_exon1a* in the heart. It is possible that at the individual cell type level, *Ddc_exon1a* could be imprinted in brain because *Grb10* exhibits imprinting in neurons, but the signal is masked by the maternal expression of *Grb10* in glia (Yamasaki-Ishizaki et al., 2007) and so there could be some cell-specific expression co-ordination (Tucci et al., 2019).

Antisense transcripts are involved in imprinted gene regulation at several well characterized loci (Sleutels et al., 2002). If the *AK006690* transcript was involved in the mechanism of imprinting at *Ddc_exon1a*, it would be predicted to be maternally expressed in heart based on imprinting mechanisms at other loci. However, we detected biallelic (at E15.5) or paternally biased expression (E16.5 and nb) which discounts an obvious mechanistic role for *AK006690* in the imprinting of *Ddc_exon1a*.

DDC Function in Heart

Homozygous null mice for *Ddc* die late in prenatal development (Eppig et al., 2012), likely due to a lack of neurotransmitter synthesis in the brain and CNS. However, mice harboring a deletion of the *Ddc_exon1a* allele inherited through the paternal line only ($Ddc^{Pat\,\Delta}$, are knockouts for *Ddc* in heart due to its imprinted status and the expression pattern of the *Ddc_Exon1A* isoform. In the small number of animals studied, compared to the wildtype littermates, $Ddc^{PAT\,\Delta}$ mice tend to have right ventricular hypoplasia of the myocardium at the region that exhibits the most abundant *Ddc* expression at E15.5 (**Figure 3** and **Supplementary Figure 3**) pointing to a role for DDC in myocardial growth. *Ddc* is expressed mainly in the myocardium of the right ventricle and IVS. These structures are derived from the secondary heart field (SHF) population of myocardial progenitors (Kelly, 2012) further delineating the SHF as a unique myocardial population distinct from the first heart field (FHF) and suggesting that *Ddc* plays a role in SHF ontogeny.

Gene expression differences between $Ddc^{PAT\,\Delta}$ and Ddc^{WT} hearts were found at *Ddc* itself, with only modest differences of other genes (**Supplementary Table 1**). DDC may not therefore function to directly mediate gene expression in the heart, but instead results in biochemical changes that influence local gene expression via feedback mechanisms. Of note, *Ddc* expression is not ubiquitous throughout the heart and is not expressed in all ventricular myocytes. There is also no detectable expression in other cardiac cell populations such as the endocardium or epicardium. This heterogeneity of expression, with expression limited to the septum and apical portion of the RV may mask changes in gene expression in $Ddc^{PAT\,\Delta}$ cells when pooled in bulk cell analyses such as microarrays with cell populations not affected by alterations in *Ddc* expression. Further evaluation of the impact of DDC deletion will require single cell transcriptomic analysis.

Human *DDC_EXON1A* Imprinting

DDC_EXON1A displays polymorphic monoallelic expression in the developing human heart, and where there is monoallelic expression, this is from the paternal allele as observed in the mouse. The NHGRI-EBI Catalog of published genome-wide association studies does not currently report mutations in DDC that relate to heart development or cardiomyopathy. However, hypermethylation of the GRB10 ICR in peripheral blood samples has recently been associated with congenital heart disease (Chang et al., 2021). The complex pattern of tissue-specific imprinted expression at this locus suggests it may warrant special consideration in genetic studies because *Ddc_exon1a* ablation has a mild effect on the developing heart and with a small effect size there could be moderately widespread ablation of this exon in human populations that presents a suitable genetic background for other mutations to cause developmental abnormalities.

In summary, *Ddc_exon1a* is a heart-specific imprinted isoform expressed from the paternally inherited allele regulated by differential DNA methylation at an ICR in the adjacent *Grb10* gene but not by the expression of *Grb10* itself. When ablated via gene knock-out in the heart, morphological changes are detected in small numbers of embryos. It is important to note that RV function was not assessed and it is reasonable to suspect that abnormal RV function might contribute to the late embryonic lethality observed in *Ddc* mutants. In humans, *DDC_EXON1A* gene expression has a paternal bias and is polymorphically imprinted, a finding common among imprinted genes in humans (Monk et al., 2006).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The microaray data have been deposited in GEO, accession number GSE87595.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Joint Research Ethics Committee of London, Camberwell St Giles, project ID 53717. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the *Ddc* knockout model generation [DdcGt(OST129277)Lex] [B6;129S5-DdcGt(neo)420Lex, ID# 11693–UCD] was carried out by Lexicon Genetics Inc., United States and breeding, genotyping and tissue acquisition by the UC Davis mouse biology program. Only frozen tissues were shipped out to the laboratory. Work involving *Grb10* animals was approved by the University of Bath Animal Welfare and Ethical Review Body and carried out under UK Home Office license (PPL 30/2839).

AUTHOR CONTRIBUTIONS

AP conducted experiments, interpreted data and contributed to writing the manuscript. SA, MF, MC, WP, MS, and BM conducted the experiments. MI collected human tissue samples. BM, NB, and RS designed and performed data analyses. MM-S and AW provided and collected *Grb10* transgenic tissue samples. HB conceived the EFIC analysis and provided EFIC facilities, supervised AP, and interpreted the heart sections and EFIC data. RO conceived and supervised the project, interpreted the data, and wrote the manuscript with AP. All authors contributed to the critical reading and editing of the manuscript.

FUNDING

This work was supported by the British Heart Foundation Project Grant (PG/13/35/30236) (RO) and PhD studentship FS/08/051/25748 (to RO undertaken by AP), the Wellcome Trust Project Grant (084358/Z/07/Z) (RO), the Medical Research Council Project Grant (G1001689) (RO), the National Institutes of Health (1R01 HL118386) (HB), and EFIC facility, the National Institute of Health (S10-RR27661) (HB).

ACKNOWLEDGMENTS

We thank Dr. Matt Arno and the KCL genomics facility for core equipment usage. We acknowledge support from the Department of Health via the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London for access to the genomics core facility at the time directed by Dr. Alka Saxena. We thank James Cain for help with the figures, Dr. Heba Saadeh for assistance with bioinformatics analysis of the expression microarray and Dr. Sabrina Böhm for technical support. We thank the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) (http://www.hdbr.org) from the Institute of Genetic Medicine, Newcastle and Institute of Child Health, London for human fetal tissues.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.676543/full#supplementary-material

REFERENCES

Amsallem, M., Mercier, O., Kobayashi, Y., Moneghetti, K., and Haddad, F. (2018). Forgotten no more: a focused update on the right ventricle in cardiovascular disease. *JACC Hear. Fail.* 6, 891–903. doi: 10.1016/j.jchf.2018.05.022

Barlow, D. P., and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6:a018382. doi: 10.1101/cshperspect.a018382

Bouschet, T., Dubois, E., Reynès, C., Kota, S. K., Rialle, S., Maupetit-Méhouas, S., et al. (2017). In Vitro corticogenesis from embryonic stem cells recapitulates the in vivo epigenetic control of imprinted gene expression. *Cereb. Cortex* 27, 2418–2433. doi: 10.1093/cercor/bhw102

Chang, S., Wang, Y., Xin, Y., Wang, S., Luo, Y., Wang, L., et al. (2021). DNA methylation abnormalities of imprinted genes in congenital heart disease: a pilot study. *BMC Med. Genomics* 14:4. doi: 10.1186/s12920-020-00848-0

Charalambous, M., Smith, F. M., Bennett, W. R., Crew, T. E., Mackenzie, F., and Ward, A. (2003). Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8292–8297. doi: 10.1073/pnas.1532175100

Chaudhry, B., Ramsbottom, S., and Henderson, D. J. (2014). Genetics of cardiovascular development. *Prog. Mol. Biol. Transl. Sci.* 124, 19–41. doi: 10.1016/B978-0-12-386930-2.00002-1

Cleaton, M. A. M., Edwards, C. A., and Ferguson-Smith, A. C. (2014). Phenotypic outcomes of imprinted gene models in mice: elucidation of pre- and postnatal functions of imprinted genes. *Annu. Rev. Genomics Hum. Genet.* 15, 93–126. doi: 10.1146/annurev-genom-091212-153441

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.* 8:e1002600. doi: 10.1371/journal.pgen.1002600

Edwards, C. A., and Ferguson-Smith, A. C. (2007). Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* 19, 281–289. doi: 10.1016/j.ceb.2007.04.013

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., and Richardson, J. E. (2012). The mouse genome database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* 40, D881–D886. doi: 10.1093/nar/gkr974

Ferrón, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., et al. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* 475, 381–385. doi: 10.1038/nature10229

Garfield, A. S., Cowley, M., Smith, F. M., Moorwood, K., Stewart-Cox, J. E., Gilroy, K., et al. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* 469, 534–538. doi: 10.1038/nature09651

Giannoukakis, N., Deal, C., Paquette, J., Kukuvitis, A., and Polychronakos, C. (1996). Polymorphic functional imprinting of the human IGF2 gene among individuals, in blood cells, is associated with H19 expression. *Biochem. Biophys. Res. Commun.* 220, 1014–1019. doi: 10.1006/bbrc.1996.0524

Gregg, C., Zhang, J., Butler, J. E., Haig, D., and Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* 329, 682–685. doi: 10.1126/science.1190831

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., et al. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329, 643–648. doi: 10.1126/science.1190830

Hitchins, M. P., Bentley, L., Monk, D., Beechey, C., Peters, J., Kelsey, G., et al. (2002). DDC and COBL, flanking the imprinted GRB10 gene on 7p12, are biallelically expressed. *Mamm. Genome* 13, 686–691. doi: 10.1007/s00335-002-3028-z

Horsthemke, B., Wawrzik, M., Groß, S., Lich, C., Sauer, B., Rost, I., et al. (2011). Parental origin and functional relevance of a de novo UBE3A variant. *Eur. J. Med. Genet.* 54, 19–24. doi: 10.1016/j.ejmg.2010.09.005

Ideraabdullah, F. Y., Vigneau, S., and Bartolomei, M. S. (2008). Genomic imprinting mechanisms in mammals. *Mutat. Res.* 647, 77–85. doi: 10.1016/j.mrfmmm.2008.08.008

Kelly, R. G. (2012). The second heart field. *Curr. Top. Dev. Biol.* 100, 33–65. doi: 10.1016/B978-0-12-387786-4.00002-6

Kelly, R. G., Buckingham, M. E., and Moorman, A. F. (2014). Heart fields and cardiac morphogenesis. *Cold Spring Harb. Perspect. Med.* 4, 1–11. doi: 10.1101/cshperspect.a015750

Kelsey, G., and Bartolomei, M. S. (2012). Imprinted Genes . . . and the number is? *PLoS Genet.* 8:e1002601. doi: 10.1371/journal.pgen.1002601.g001

Korostowski, L., Raval, A., Breuer, G., and Engel, N. (2011). Enhancer-driven chromatin interactions during development promote escape from silencing by a long non-coding RNA. *Epigenetics Chromatin* 4:21. doi: 10.1186/1756-8935-4-21

Lajiness, J. D., and Conway, S. J. (2014). Origin, development, and differentiation of cardiac fibroblasts. *J. Mol. Cell. Cardiol.* 70, 2–8. doi: 10.1016/j.yjmcc.2013.11.003

Lee, N. C., Shieh, Y. D., Chien, Y. H., Tzen, K. Y., Yu, I. S., Chen, P. W., et al. (2013). Regulation of the dopaminergic system in a murine model of aromatic L-amino acid decarboxylase deficiency. *Neurobiol. Dis.* 52, 177–190. doi: 10.1016/j.nbd.2012.12.005

Lewis, A., and Reik, W. (2006). How imprinting centres work. *Cytogenet. Genome Res.* 113, 81–89. doi: 10.1159/000090818

Madon-Simon, M., Cowley, M., Garfield, A. S., Moorwood, K., Bauer, S. R., and Ward, A. (2014). Antagonistic roles in fetal development and adult physiology for the oppositely imprinted Grb10 and Dlk1genes. *BMC Biol.* 12:771. doi: 10.1186/s12915-014-0099-8

Menheniott, T. R., Woodfine, K., Schulz, R., Wood, A. J., Monk, D., Giraud, A. S., et al. (2008). Genomic imprinting of Dopa decarboxylase in heart and reciprocal allelic expression with neighboring Grb10. *Mol. Cell. Biol.* 28, 386–396. doi: 10.1128/MCB.00862-07

Monk, D., Arnaud, P., Apostolidou, S., Hills, F. A., Kelsey, G., Stanier, P., et al. (2006). Limited evolutionary conservation of imprinting in the human placenta. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6623–6628. doi: 10.1073/pnas.0511031103

Plasschaert, R. N., and Bartolomei, M. S. (2015). Tissue-specific regulation and function of Grb10 during growth and neuronal commitment. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6841–6847. doi: 10.1073/pnas.1411254111

Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., et al. (2009). The long noncoding RNA Kcnq 1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* 136, 525–530. doi: 10.1242/dev.031328

Rosenthal, J., Mangal, V., Walker, D., Bennett, M., Mohun, T. J., and Lo, C. W. (2004). Rapid high resolution three dimensional reconstruction of embryos with episcopic fluorescence image capture. *Birth Defects Res. Part C Embryo Today Rev.* 72, 213–223. doi: 10.1002/bdrc.20023

Shiura, H., Nakamura, K., Hikichi, T., Hino, T., Oda, K., Suzuki-Migishima, R., et al. (2009). Paternal deletion of Meg1/Grb10 DMR causes maternalization of the Meg1/Grb10 cluster in mouse proximal Chromosome 11 leading to severe pre- and postnatal growth retardation. *Hum. Mol. Genet.* 18, 1424–1438. doi: 10.1093/hmg/ddp049

Sleutels, F., Zwart, R., and Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813. doi: 10.1038/415810a

Smith, C. M., Finger, J. H., Hayamizu, T. F., McCright, I. J., Xu, J., Berghout, J., et al. (2014). The mouse gene expression database (GXD): 2014 update. *Nucleic Acids Res.* 42, D818–D824. doi: 10.1093/nar/gkt954

Smith, F. M., Holt, L. J., Garfield, A. S., Charalambous, M., Koumanov, F., Perry, M., et al. (2007). Mice with a disruption of the imprinted Grb10 gene exhibit altered body composition, glucose homeostasis, and insulin signaling during postnatal life. *Mol. Cell. Biol.* 27, 5871–5886. doi: 10.1128/MCB.02087-06

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3:Article3. doi: 10.2202/1544-6115.1027

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., Bartolomei, M. S., Benvenisty, N., et al. (2019). Genomic imprinting and physiological processes in mammals. *Cell* 176, 952–965. doi: 10.1016/j.cell.2019.01.043

Yamasaki-Ishizaki, Y., Kayashima, T., Mapendano, C. K., Soejima, H., Ohta, T., Masuzaki, H., et al. (2007). Role of DNA methylation and histone H3 lysine 27 methylation in tissue-specific imprinting of mouse Grb10. *Mol. Cell. Biol.* 27, 732–742. doi: 10.1128/MCB.01329-06

Zink, F., Magnusdottir, D. N., Magnusson, O. T., Walker, N. J., Morris, T. J., Sigurdsson, A., et al. (2018). Insights into imprinting from parent-of-origin phased methylomes and transcriptomes. *Nat. Genet.* 50, 1542–1552. doi: 10.1038/s41588-018-0232-7

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.

Citation: Prickett AR, Montibus B, Barkas N, Amante SM, Franco MM, Cowley M, Puszyk W, Shannon MF, Irving MD, Madon-Simon M, Ward A, Schulz R, Baldwin HS and Oakey RJ (2021) Imprinted Gene Expression and Function of the Dopa Decarboxylase Gene in the Developing Heart. Front. Cell Dev. Biol. 9:676543. doi: 10.3389/fcell.2021.676543

frontiers in Cell and Developmental Biology

REVIEW
published: 05 August 2021
doi: 10.3389/fcell.2021.713878

Canonical and Non-canonical Genomic Imprinting in Rodents

*Hisato Kobayashi**

Department of Embryology, Nara Medical University, Kashihara, Japan

Genomic imprinting is an epigenetic phenomenon that results in unequal expression of homologous maternal and paternal alleles. This process is initiated in the germline, and the parental epigenetic memories can be maintained following fertilization and induce further allele-specific transcription and chromatin modifications of single or multiple neighboring genes, known as imprinted genes. To date, more than 260 imprinted genes have been identified in the mouse genome, most of which are controlled by imprinted germline differentially methylated regions (gDMRs) that exhibit parent-of-origin specific DNA methylation, which is considered primary imprint. Recent studies provide evidence that a subset of gDMR-less, placenta-specific imprinted genes is controlled by maternal-derived histone modifications. To further understand DNA methylation-dependent (canonical) and -independent (non-canonical) imprints, this review summarizes the loci under the control of each type of imprinting in the mouse and compares them with the respective homologs in other rodents. Understanding epigenetic systems that differ among loci or species may provide new models for exploring genetic regulation and evolutionary divergence.

Keywords: genomic imprinting, DNA methylation, non-canonical imprinting, histone modification, rodent, germline differentially methylated region, mouse genome, epigenetics

OPEN ACCESS

Edited by:
Anthony Isles,
Cardiff University, United Kingdom

Reviewed by:
Ferdinand Von Meyenn,
ETH Zürich, Switzerland
Courtney W. Hanna,
University of Cambridge, United Kingdom
Philippe Arnaud,
GReD, CNRS-Université Clermont Auvergne-INSERM, France

***Correspondence:**
Hisato Kobayashi
hiskobay@naramed-u.ac.jp

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

Received: *24 May 2021*
Accepted: *16 July 2021*
Published: *05 August 2021*

Citation:
Kobayashi H (2021) Canonical and Non-canonical Genomic Imprinting in Rodents. Front. Cell Dev. Biol. 9:713878. doi: 10.3389/fcell.2021.713878

RODENTS: SYMBOLIC MODELS IN BIOMEDICAL AND GENETIC RESEARCH

Rodents such as mice and rats are commonly used as representative laboratory animals. The genomes of these organisms have been progressed along with the human genome project; thus, the C57BL/6 mouse (*Mus musculus*) and Brown Norway rat (*Rattus norvegicus*) become the second and third mammals to have their genomes sequenced in 2002 and 2004, respectively (Waterston et al., 2002; Gibbs et al., 2004). Their genomes of approximately three billion base pairs each contains roughly the same number of genes as the human genome. Furthermore, almost all human genes associated with diseases have counterparts in the rodent genome, which appear highly conserved throughout mammalian evolution. Thus, these experimental rodents generally deepen our understanding of mammalian genetic and (epi-)genomic regulatory systems.

Mammals are diploid organisms arising from the fusion of two parental gametes, with each donating one set of autosomal chromosomes (19 autosomes in mice, 20 in rats, and 22 in humans) plus one set of sex chromosome (X or Y) to the offspring. According to Mendel's law, diploid

Abbreviations: DNMT, DNA methyltransferase; gDMR, germline differentially methylated region; ICR, imprinting control region; lncRNA, long non-coding RNA; LTR, long terminal repeat; piRNA, Piwi-interacting RNA; sDMR, secondary differentially methylated region.

cells contain parental copies of each autosomal gene, which are predicted to show the same transcription state. However, "genomic imprinting" is a form of non-Mendelian inheritance that results in parent-of-origin allele-specific gene expression of autosomal loci or of loci on the diploid X chromosome (only in extra-embryonic tissues in females). Polymorphism information between strains or individuals helps distinguish between paternal and maternal alleles. It was only natural that the phenomenon of genomic imprinting was discovered and well-investigated in mice, where nuclear transfer and genetic engineering technologies have always been developed and where numerous strains (polymorphisms between strains can serve as genetic markers of parent-of-origin in allele-specific analysis) have been established and maintained.

DISCOVERY OF GENOMIC IMPRINTING IN MAMMALS

In 1984, two laboratories published landmark papers that reported a new phenomenon in mammalian genome biology (Barton et al., 1984; McGrath and Solter, 1984). Both studies independently performed pronuclear transfer experiments from fertilized mouse eggs to produce androgenetic and gynogenetic embryos containing only sperm-derived or oocyte-derived chromosome sets. These "uniparental" embryos could not survive to term but could develop to some extent with sex-specific developmental abnormalities. Androgenetic embryos preferentially develop extra-embryonic and placental structures at the expense of embryo development. Conversely, gynogenetic embryos (or parthenogenetic embryos by artificial activation of oocytes) have poor growth of placental lineages and developmental arrest, possibly due to extra-embryonic defects. These opposite phenotypes underlie the functional differences in developmental genes in paternal and maternal genomes. The mice that were bred to have uniparental disomies, in which either single or partial chromosomes are inherited from only one parent, for individual chromosomes also show aberrant phenotypes, such as overgrowth, growth retardation, or abnormal behavior (Cattanach, 1986).

Nevertheless, not all chromosomes produce abnormalities when present as disomies, depending on which chromosome or part is made uniparental; however, those commonly lead to striking phenotypic differences. These investigations revealed the requirement of both maternal and paternal genomes for normal development, which was tied to an intriguing biological phenomenon called genomic imprinting. Uniparental inheritance of the genome or chromosome occurs spontaneously in humans, resulting in early pregnancy losses, like androgenetic and parthenogenetic conceptuses (hydatidiform moles and benign ovarian teratomas), or moderate to severe developmental disabilities, known as imprinting diseases (Linder et al., 1975; Kajii and Ohama, 1977; Wake et al., 1978; Kalish et al., 2014). Subsequent evolutionary and genetic studies of imprinted loci have shown that this phenomenon is present only in placental mammals among vertebrates.

The surprising finding of these studies was that mammalian genes could function differentially depending on whether they originated from the mother or father. Before the study of uniparental disomies, a "maternal-effect" locus called *Tme* (*T-associated maternal effect*) was identified on the proximal mouse chromosome 17 overlapping deletions of maternal-effect lethal mutants, like T^{hp} or T^{Tub2} (Johnson, 1974; Winking and Silver, 1984). The region was later revealed to be the locus of *Igf2r*, expressed exclusively from the maternally inherited allele; therefore, its expression is dependent on the "parent-of-origin." Simultaneously, the closely linked *H19* and *Igf2* genes, which are reciprocally imprinted, were identified in mouse chromosome 7; *H19* produces a long non-coding RNA (lncRNA) exclusively expressed from the maternal allele, and *Igf2* originates from the opposite allele. Interestingly, the opposite imprinting of *Igf2* and its scavenging receptor gene, *Igf2r*, demonstrates conflicting parental effects of growth promotion and growth restriction, which supports the classic "parent-offspring conflict theory" for the evolution of genomic imprinting (Trivers, 1974; Moore and Haig, 1991).

CANONICAL GENOMIC IMPRINTING IS MEDIATED BY MATERNAL OR PATERNAL DNA METHYLATION

The discovery of the first endogenous imprinted genes in 1991 (Barlow et al., 1991; Bartolomei et al., 1991; DeChiara et al., 1991), which were differentially expressed from the maternal and paternal alleles, sparked initial efforts to elucidate the mechanisms of imprint establishment, maintenance, and erasure that together control the timing and placement of genomic imprinting. One prominent candidate of the non-Mendelian system is epigenetic regulation, in which DNA methylation (mainly occurs in CpG dinucleotides) is the most studied mechanism and has been shown to play a key role in mouse models of genomic imprinting and fetal reprogramming. A strong link between DNA methylation and imprinting regulation has been indicated by the cases of imprinted transgenic mouse lines. In a few of these mice the foreign transgene becomes methylated in a parent-specific manner in the gamete, inherited with parent-of-origin specific methylation into the diploid cells of embryo, and subsequently, the modification is erased and reestablished upon passage through the germ line (Chaillet et al., 1991).

Allele-specific DNA methylation of imprinted regions, also known as imprinted germline differentially methylated regions (gDMRs), has been studied as the best candidate for the molecular mechanism of inheriting parental-specific imprints following fertilization. Because parental imprints must be established when the parental genomes can be distinguished, investigators assayed methylation acquisition during gametogenesis, when the maternal and paternal genomes are entirely separated and can be independently epigenetically modified. Paternal-specific methylation of the gDMRs at three imprinted loci (*H19* and subsequently discovered *Dlk1-Meg3* and *Rasgrf1*) is acquired prenatally in prospermatogonia before the onset of

meiosis in the male germline (Davis et al., 2000; Kato et al., 2007). In contrast, maternal-specific gDMR methylation occurs postnatally in growing oocytes, with different gDMRs (at least 21 maternal gDMRs have been identified in mice) that are methylated at a slightly different time during oocyte growth (Lucifero et al., 2004; Hiura et al., 2006). In both germlines, DNA methylation is established through the action of *de novo* DNA methyltransferase (DNMT) 3a and the accessory protein DNMT3L (Bourc'his et al., 2001; Hata et al., 2002; Kaneda et al., 2004). Although it is unclear how specific sequences are chosen for allele-specific DNA methylation in the germline, recent studies have demonstrated that histone modification across gDMR sequences provides an essential instructive step for DNMT proteins (**Figures 1A,B**). In oogenesis, the transcription-dependent histone marker H3K36me3 (trimethylation of histone H3 at lysine 36) guides DNA methylation over active gene bodies, leading to the establishment of all maternal methylation imprints (Kobayashi et al., 2012a; Veselovska et al., 2015; Xu et al., 2019). Transcription start sites in oocytes are often oocyte-specific (carried in part by retroviral promoters) and upstream of canonical promoters and imprinted DMRs, hence transcription-coupled DNA methylation spans these domains in an oocyte-specific manner (Chotalia et al., 2009; Brind'Amour et al., 2018). In fetal spermatogenesis, H3K36me2 (dimethylation of H3K36) shapes the gene body and intergenic DNA methylation and guides paternal methylation at the gDMRs (Shirane et al., 2020). Only in *Rasgrf1* gDMR, the Piwi-interacting RNA (piRNA) pathway and the rodent-specific DNMT3C are also responsible for the establishment of paternal DNA methylation (Watanabe et al., 2011; Barau et al., 2016). In addition to imprinted gDMRs, there are more than a thousand promoters or CpG islands on non-imprinted genes that are differentially methylated between oocytes and sperm; however, the vast majority lose their differential marks during epigenetic reprogramming events during early embryogenesis (Smallwood et al., 2011; Kobayashi et al., 2012a). After fertilization, the paternal genome is actively demethylated before the first DNA replication, whereas the maternal genome is passively demethylated throughout several rounds of DNA replication until the blastocyst stage. Imprinted gDMRs are protected from these erasure events by recruiting maintenance DNMT1 and accessory UHRF1 through the recognition of a methylated sequence motif by the zinc-finger proteins, ZFP57 and ZNF445, along with the interaction of TRIM28 with histone methyltransferases (Sharif et al., 2007; Hirasawa et al., 2008; Quenneville et al., 2011; Messerschmidt et al., 2012; Takahashi et al., 2019).

At the end of 2018, at least 260 coding and non-coding genes were found to be imprinted, and 24 imprinted gDMRs were identified in the mouse genome (Tucci et al., 2019). Many of these gDMRs act as imprinting control regions (ICRs) regulating the monoallelic expression of the neighboring solo imprinted gene and clusters of imprinted genes. The majority of maternal ICRs directly regulate a promoter for either a messenger RNA or a lncRNA by silencing one allele by DNA methylation. In contrast, paternal ICRs are not located at promoters but rather map to intergenic regions. However, the imprinting of gene clusters often involves locus-specific and complex molecular mechanisms, such as transcriptional silencing by an antisense transcript and allele-specific chromatin changes at target genes or *cis*-regulatory elements by the ICRs (Hark et al., 2000; Terranova et al., 2008; Latos et al., 2012). These imprinted genes under the control of ICRs also act as barriers to prevent mammalian embryos from parthenogenesis (Kono et al., 2004; Kawahara et al., 2007; Li et al., 2018). Thus, parent-of-origin specific DNA methylation, also called "canonical imprinting," is considered a primary imprint marker that directly or indirectly controls most imprinted genes, which are responsible for the abnormalities of uniparental disomies or embryos.

NON-CANONICAL IMPRINTING IS MEDIATED BY MATERNAL HISTONE MODIFICATION

Although DNA methylation has been known to specify imprinting, the possibility that histone modifications in the gametes could also determine imprinting has also been demonstrated (Okae et al., 2014). A subset of imprinted genes is specifically paternally expressed in the placenta but not imprinted in the embryo, and the establishment of a part of such imprinted genes is independent of oocyte-specific DNA methylation, as DNMT-deficiency in growing oocytes did not affect the imprinted paternal expression of these genes in the extra-embryonic lineage (Chen et al., 2019; Hanna et al., 2019). The key gametic imprinting mark of the "non-canonical" (DNA methylation-independent) imprinting is the repressive histone mark H3K27me3 (trimethylation of H3 at lysine 27) in the oocyte, which was found to transiently imprint several loci within pre-implantation (Inoue et al., 2017a). Furthermore, H2AK119ub1 (mono-ubiquitinated histone H2A at lysine 119) was highly colocalized with H3K27me3 in oocytes, which is equalized mainly at the two-cell stage but guides maternal H3K27me3 inheritance after fertilization (Chen et al., 2021; Mei et al., 2021). Thus, H2AK119ub1 and H3K27me3, which are catalyzed by the polycomb repressive complexes (PRC1 and PRC2), mediate maternal allele-specific silencing of at least seven imprinted genes, namely *Sfmbt2*, *Phf17*, *Gab1*, *Sall1*, *Platr20* (*5133400J02Rik*), *Smoc1*, and *Slc38a4*, in mice (**Figure 1C**), several of which have been previously shown to play important roles in placental function and development (Itoh et al., 2000; Miri et al., 2013; Matoba et al., 2019). Maternal H3K27me3 and H2AK119ub1 are not maintained beyond pre-implantation development (Hanna et al., 2019; Chen et al., 2021; Mei et al., 2021), and transition to a more permanent epigenetic state is required to preserve paternal expression during post-implantation development (Inoue et al., 2017a; Chen et al., 2019; Hanna et al., 2019). The long terminal repeats (LTRs) of endogenous retroviral elements can act as alternative promoters for non-canonical imprinted genes and paternal allele-specific accumulation of the active histone mark H3K4me3 (trimethylation of H3 at lysine 4) occurs at these LTR promoters (Hanna et al., 2019). Finally, these LTRs are methylated on the maternal allele in extra-embryonic tissues; thus, maternally inherited H3K27me3

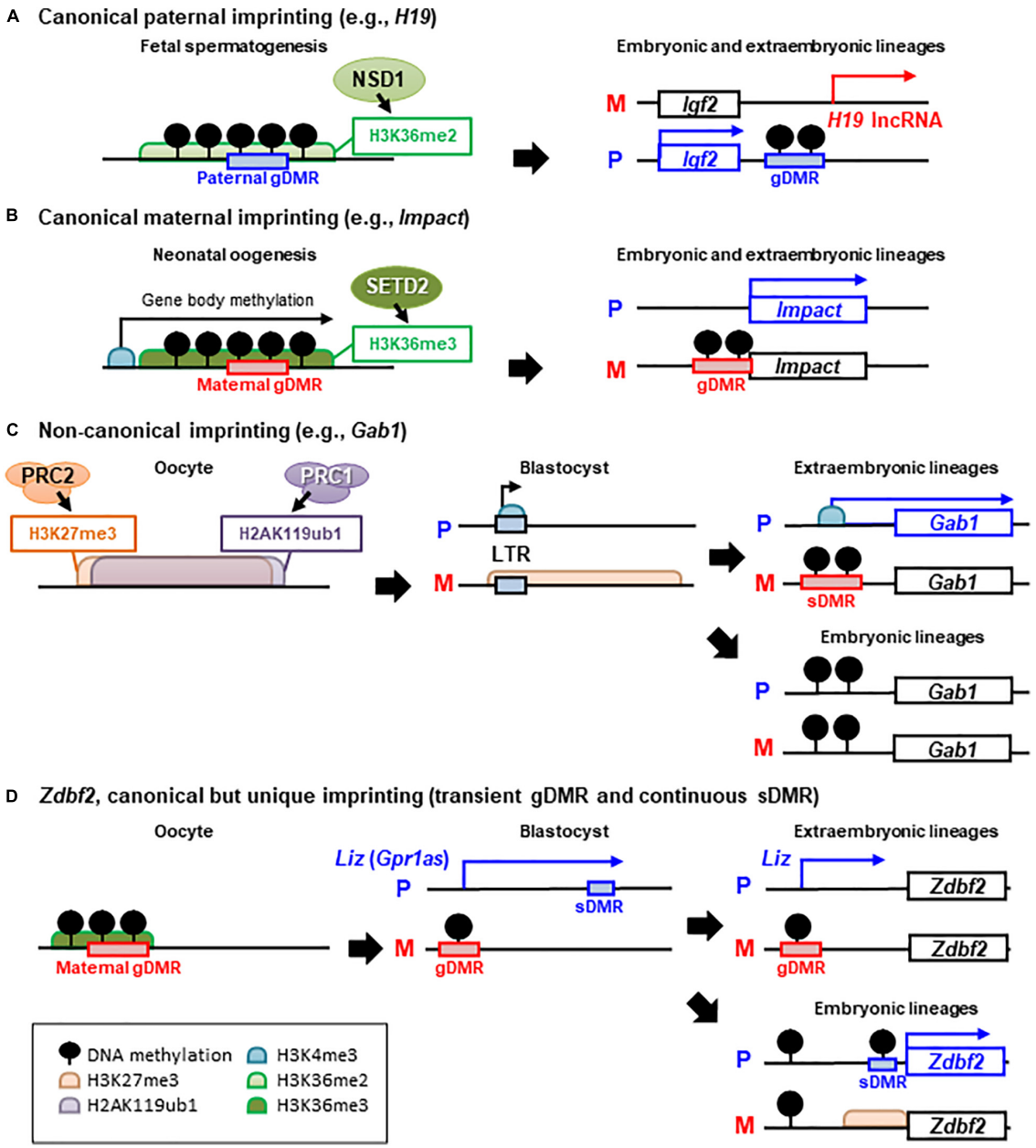


FIGURE 1 | Epigenetic mechanisms of canonical and non-canonical imprinted gene regulation. There are several examples of different epigenetic inheritance patterns between paternal (P) and maternal (M) alleles, that control paternally expressed protein-coding genes. **(A)** Paternal and **(B)** maternal germline differentially methylated region (gDMR)-mediated canonical imprinting. H3K36 methyltransferases NSD1 and SETD2 have been shown to establish H3K36me2 or H3K36me3 in pro-spermatogonia or oocytes and to be required for *de novo* DNA methylation at paternal or maternal gDMRs, respectively. In general, canonical imprinting is stably maintained throughout somatic life and mediates monoallelic silencing of imprinted genes or non-coding RNAs. **(C)** Non-canonical imprinting, such as *Gab1* loci. H2AK119ub1 guides maternal inheritance and zygotic deposition of H3K27me3, and thus, maternally inherited H3K27me3 is maintained until the blastocyst (pre-implantation) stage. Then, maternal H3K27me3 silences the LTR retrotransposon-derived alternative promoter, which becomes actively transcribed on only the paternal allele. Although maternal H3K27me3 is lost after implantation, maternal allele-specific DNA methylation is established as an imprinted sDMR in extra-embryonic tissues, and thus, monoallelic paternal expression of non-canonically imprinted LTRs and nearby protein-coding genes can be maintained. In the post-implantation epiblast, these ERVs are silenced by DNA methylation in both alleles, resulting in a loss-of-imprinting in somatic lineages (not shown in the figure). **(D)** *Zdbf2* locus is a unique example of secondary imprinting. Transient paternal allele-specific expression of a long isoform transcript of *Zdbf2* (*Liz*, also called *Gpr1as*) originated from the maternal gDMR, occurs in the pre-implantation embryo. *Liz* continues to be paternally expressed by the persistence of the maternal gDMR in extra-embryonic tissues. In the embryonic tissues, maternal gDMR is biallelically methylated and loses its imprinted status and *Liz* transcription; However, *Zdbf2* retains imprinted expression because of acquired paternal DNA methylation at the sDMR and active H3K4me3 at the *Zdbf2* promoter via traversing *Liz* transcription during gastrulation.

imprinting transitions to imprinted DNA methylation at the secondary DMRs (sDMRs) and can act as a long-term imprinting in placental linage.

Notably, oocyte-derived H3K27me3 also serves as a maternal imprint for the lncRNA Xist, triggering paternal X chromosome inactivation in mouse female pre-implantation embryos and extra-embryonic tissues (Inoue et al., 2017b). Like non-canonical imprinting at autosomal loci, X inactivation can be clonally inherited and suppress the entire chromosome through several epigenetic suppression pathways (Chen and Zhang, 2020). In addition to H3K27me3 imprinting, failure of X chromosome inactivation results in embryonic lethality, emphasizing the developmental importance of these interrelated processes. However, the functional and molecular relationship between H3K27me3-mediated non-canonical imprinting at autosomes and imprinted X chromosome inactivation or what distinguishes these strategies for biological diversity from DNA methylation-based canonical imprinting remains unresolved.

SECONDARY DMRS: A LESSON FROM *ZDBF2* IMPRINTED GENE

Unlike gDMRs, imprinted sDMRs acquire allele-specific DNA methylation during embryonic development, rather than inheriting it from germ cells. Although secondary DMRs do not function as primary imprinting markers, allele-specific methylation of these regions frequently corresponds to gene silencing in a tissue-specific manner, such as *Cdkn1c* (Fan et al., 2005; Wood et al., 2010). Although sDMRs may play a role in maintaining imprinted expression (John and Lefebvre, 2011; Kobayashi et al., 2012b), they remain untested in most regions. The majority of sDMRs at canonical imprinted loci have been identified to be located within the imprinted genes or clusters and acquire allele-specific methylation by the hierarchical regulation of the gDMRs (Stoger et al., 1993; Lopes et al., 2003; Yamasaki et al., 2005; Williamson et al., 2011; Mohammad et al., 2012; Greenberg et al., 2017; Saito et al., 2018). One mechanism across several imprinted loci is the presence of a monoallelic transcript from gDMR passing through regulatory elements such as promoters and CpG islands (Ferguson-Smith, 2011). Consequently, DNMT3B targets sites of transcriptional elongation (Baubec et al., 2015), resulting in the acquisition of DNA methylation along the transcribed allele. As not all DMRs are located within transcribed regions, there must also be alternative mechanisms to establish allelic methylation at secondary loci.

Differences in the acquisition of sDMRs between embryonic and extra-embryonic lineages have been observed across several canonical imprinted domains (Lewis et al., 2004; Sato et al., 2011; Duffie et al., 2014). In particular, the DMR dynamics observed at *Zdbf2* highlight epigenetic changes in these developmental processes (**Figure 1D**). *Zdbf2* is a canonical, but unique, imprinted gene with paternal expression and, paradoxically, a paternal DMR near its promoter [the paradoxical finding of the paternal DMR adjacent to a paternally expressed gene was later explained through serial experiments systematically ablating epigenetic modifiers (Greenberg et al., 2017)]. Early studies of *Zdbf2* suggested that paternal DMR might be a gDMR because the DMR is methylated in the sperm and not in oocytes (Kobayashi et al., 2009). However, subsequent studies in embryos showed that paternal DNA methylation was erased in pre-implantation embryos and reset secondarily during post-implantation development (Kobayashi et al., 2012b; Duffie et al., 2014). This paternal sDMR was established by the transient monoallelic expression of a long isoform of *Zbdf2* (*Liz*, also called *GPR1AS* in humans) originating from an upstream transcription start site, which is regulated by a maternal gDMR (Kobayashi et al., 2012b, 2013; Greenberg et al., 2017). Thus, *Liz*-induced sDMR can be maintained in embryonic lineage and lead to postnatal paternal expression of *Zdbf2*. Meanwhile, *Liz* transcription is lost with the subsequent monoallelic to biallelic DNA methylation switch of the upstream maternal gDMR in embryonic tissues; conversely, the maternal gDMR remains intact throughout the post-implantation epigenetic programming in extra-embryonic tissues (Kobayashi et al., 2013; Greenberg et al., 2017). Finally, the canonical *Zdbf2* promoter and exons remain silenced because of the incomplete establishment of the paternal sDMR, and the paternal expression of *Liz* continues throughout placental development (Greenberg et al., 2017). Thus far, it remains unclear why maternal gDMR persists in extra-embryonic tissues but not in embryos.

Paternal DNA methylation at the *Zdbf2* sDMR is required to prevent the accumulation of H3K27me3, thereby conferring an active chromatin state at the adjacent *Zdbf2* promoter (Greenberg et al., 2017). It is not clearly understood what controls allele-specific DNA methylation at the sDMRs of canonical and non-canonical imprinted loci. However, further investigations into sDMRs at both canonical and non-canonical imprinted loci will provide valuable suggestions on how reprogramming or preserving factors target imprinted epigenetic marks through post-implantation development.

CANONICAL AND NON-CANONICAL IMPRINTING IN THE OTHER RODENTS

Although mice are the primary research model used to study genomic imprinting, imprinted regions have been described in various mammals, including humans. Among the 24 gDMRs in mice, two paternal (*H19* and *Dlk1-Meg3*) and 16 maternal (*Gpr1as/Liz, Mcts2, Nnat, Nespas-Gnasxl, Gnas_exon1A, Peg10-Sgce, Mest, Nap1l5, Peg3, Snrpn, Inpp5f_v2, Kcnq1ot1, Plagl1, Grb10, Peg13*, and *Airn*) gDMRs were conserved between mice and humans (**Table 1**). Although some species-specific maternal gDMRs drive oocyte transcription initiation in lineage-specific LTR retrotransposons (Bogutz et al., 2019), many canonical imprinted loci are well conserved among species, and mice with deletions involving imprinted genes

TABLE 1 | List of identified canonically and non-canonically imprinted regions.

Type of imprinting	Mouse	Rat*[1]	Hamster*[1]	Human
Paternal gDMRs (canonical imprinting)	3 loci (*H19*, *Dlk1-Meg3*, *Rasgrf1*)	3 loci (*H19*, *Dlk1-Meg3*, *Rasgrf1*)	2 loci*[2] (*H19*, *Dlk1-Meg3*)	2 loci (*H19*, *DLK1-MEG3*)
Maternal gDMRs (canonical imprinting)	16 common and 5 mouse (rodent)-specific loci (*Fkbp6*, *Cdh15*, *Zrsr1*, *Slc38a4*, *Impact*)	*Igf2r*, *Peg3* (common) and *Impact* (rodent-specific)	6 common loci (*Peg3*, etc.) and *Impact* (rodent-specific)	16 common and numerous human-specific loci
Non-canonical imprinting	7 loci (*Sfmbt2*, *Smoc1*, *Gab1*, etc.)	*Sfmbt2*	*Smoc1*	5 loci (*FAM101A*, etc.)

*[1] *Basically, DNA methylation has not yet been well-analyzed in rats and hamsters. But imprinted regions that show imprinted expression of one or more homologous genes at each differentially methylated region (DMR) locus are listed.*
*[2] *There is no direct evidence that Rasgrf1 is imprinted in the hamster. However, Dnmt3C, which mediates Rasgrf1 imprinting, is present in the hamster genome (Barau et al., 2016).*

or ICRs are used as models for human imprinting diseases such as Prader–Willi, Angelman, Beckwith–Wiedemann, and Silver–Russell syndromes (Chang and Bartolomei, 2020). However, orthologs of non-canonical imprinted genes are not likely to be imprinted in humans. Preliminary studies in human embryos found five paternally expressed genes that may be regulated by maternal H3K27me3, but none of these have been reported to be imprinted in mice (Zhang et al., 2019). Thus, current studies to date do not provide any direct evidence for the existence of non-canonical imprinting in mammals other than mice. However, among these genes, *Sfmbt2* and *Smoc1* have been reported to show an expression biased toward one parental allele in rat and hamster placentas, respectively (Wang et al., 2011; Brekke et al., 2016). This evidence supports the hypothesis that non-canonical imprinting is conserved in rodents.

Although rats and hamsters are widely used for physiological, oncological, and other medical studies, mice have always been used as embryological and genetic studies models. In this situation, the number of imprinted genes identified in these rodents is limited compared to mice. However, because of the long history of laboratory animal research, numerous mouse, rat, and hamster strains have been established and maintained, and the genomes of some have been sequenced. It is possible to identify imprinting information from polymorphism information among strains (Hermsen et al., 2015).

It has already been shown that single or multiple genes are imprinted on the homologous regions of the three imprinted clusters (*H19*, *Dlk1-Meg3*, and *Rasgrf1*) that undergo paternal methylation imprinting in mice (Overall et al., 1997; Pearsall et al., 1999; Dietz et al., 2012). In addition, *Igf2r*, *Impact* [driven by a rodent-specific LTR (Bogutz et al., 2019)], and *Sfmbt2*, which are controlled by maternal imprinting in mice, are also expressed only from one parental allele in rats (Mills et al., 1998; Okamura et al., 2005; Miri et al., 2013). *Sfmbt2* is almost exclusively expressed in extra-embryonic tissues and is essential for the maintenance of trophoblast progenitors. Intriguingly, *Sfmbt2* contains a large cluster of microRNA (miRNA) genes within intron 10, and these miRNAs are also imprinted and essential for placental development (Inoue et al., 2017c). Notably, *Sfmbt2*, known to undergo non-canonical imprinting in mice, is also paternally expressed in the rat placenta in the presence of a large cluster of microRNAs (Wang et al., 2011). However, human, bovine, and pig *SFMBT2* are not imprinted and lack this block of microRNAs. These observations strengthen the argument for the recent evolution of *Sfmbt2*, in which the non-canonical imprint (and the block of miRNAs) drives its placental role in rodents.

In hamsters, reciprocal crosses between two dwarf hamsters (*Phodopus sungorus* and *Phodopus campbelli*) result in strong parent-of-origin effects on placental and embryonic growth (Brekke and Good, 2014). The expression of imprinted genes and related loss-of-imprinting has been evaluated to some extent in dwarf hamster hybrids (Brekke et al., 2016). Single-nucleotide variant-based allele-specific analysis of placental expressed genes identified 88 imprinted candidate genes in hamster autosomes. Among these, 18 genes overlapped between hamster and mice, including *Smoc1*, a non-canonical imprinted gene. Unexpectedly, *Smoc1* shows the opposite pattern of imprinting in hamster compared to mouse, with the maternal allele being expressed. This is similar to a report in human fibroblast cells, where SMOC1 showed maternal-allele specific expression (Santoni et al., 2017). While it is unclear whether this change is due to biological differences or false bias of allele-specific analysis, well-known examples of canonical paternally (*Dlk1*, *Igf2*, *Impact*, among others) and maternally (such as *H19*) expressed genes were also reidentified in hamster. *Smoc1* encodes a multi-domain secreted protein that may play a critical role in ocular and limb development (Okada et al., 2011). However, *Smoc1* is not likely associated with loss-of-imprinting in hybrid

hamsters, and its functional role in placental development remains unknown.

CONCLUSION

The discovery of a non-canonical imprinting mechanism mediated by histone modifications is an important finding that provides a new molecular mechanism for epigenetic transgenerational inheritance. In contrast, the diversity of canonical and non-canonical imprinting complicates our understanding of the underlying mechanisms and a better understanding of the differences among mammalian species that bridge the gap between humans and mice. For instance, the insertion of endogenous retroviral elements drives both canonical and non-canonical imprinting (Bogutz et al., 2019; Hanna et al., 2019). However, not all species-specific imprinted regions can be explained by this mechanism. Revealing the whole landscape of genomic imprinting in various rodents, such as rats and hamsters, and non-human primates would be a significant step forward in understanding the diversity of imprinting and epigenetic regulation systems.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was supported by the JSPS KAKENHI (Grant Number JP21H02382 to HK).

ACKNOWLEDGMENTS

I would like to thank Editage (http://www.editage.jp) for English language editing.

REFERENCES

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., et al. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* 354, 909–912. doi: 10.1126/science.aah5143

Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* 349, 84–87. doi: 10.1038/349084a0

Bartolomei, M. S., Zemel, S., and Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153–155. doi: 10.1038/351153a0

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

Bogutz, A. B., Brind'Amour, J., Kobayashi, H., Jensen, K. N., Nakabayashi, K., Imai, H., et al. (2019). Evolution of imprinting via lineage-specific insertion of retroviral promoters. *Nat. Commun.* 10:5674. doi: 10.1038/s41467-019-13662-9

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., and Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539. doi: 10.1126/science.1065848

Brekke, T. D., and Good, J. M. (2014). Parent-of-origin growth effects and the evolution of hybrid inviability in dwarf hamsters. *Evolution* 68, 3134–3148. doi: 10.1111/evo.12500

Brekke, T. D., Henry, L. A., and Good, J. M. (2016). Genomic imprinting, disrupted placental expression, and speciation. *Evolution* 70, 2690–2703. doi: 10.1111/evo.13085

Brind'Amour, J., Kobayashi, H., Richard Albert, J., Shirane, K., Sakashita, A., Kamio, A., et al. (2018). LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat. Commun.* 9, 3331. doi: 10.1038/s41467-018-05841-x

Cattanach, B. M. (1986). Parental origin effects in mice. *J. Embryol. Exp. Morphol.* 97, 137–150.

Chaillet, J. R., Vogt, T. F., Beier, D. R., and Leder, P. (1991). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell* 66, 77–83. doi: 10.1016/0092-8674(91)90140-t

Chang, S., and Bartolomei, M. S. (2020). Modeling human epigenetic disorders in mice: Beckwith-Wiedemann syndrome and Silver-Russell syndrome. *Dis. Model. Mech.* 13:dmm044123. doi: 10.1242/dmm.044123

Chen, Z., and Zhang, Y. (2020). Maternal H3K27me3-dependent autosomal and X chromosome imprinting. *Nat. Rev. Genet.* 21, 555–571. doi: 10.1038/s41576-020-0245-9

Chen, Z., Djekidel, M. N., and Zhang, Y. (2021). Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nat. Genet.* 53, 551–563. doi: 10.1038/s41588-021-00821-2

Chen, Z., Yin, Q., Inoue, A., Zhang, C., and Zhang, Y. (2019). Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. *Sci. Adv.* 5:eaay7246. doi: 10.1126/sciadv.aay7246

Chotalia, M., Smallwood, S. A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., et al. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev.* 23, 105–117. doi: 10.1101/gad.495809

Davis, T. L., Yang, G. J., Mccarrey, J. R., and Bartolomei, M. S. (2000). The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum. Mol. Genet.* 9, 2885–2894. doi: 10.1093/hmg/9.19.2885

DeChiara, T. M., Robertson, E. J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859. doi: 10.1016/0092-8674(91)90513-x

Dietz, W. H., Masterson, K., Sittig, L. J., Redei, E. E., and Herzing, L. B. (2012). Imprinting and expression of Dio3os mirrors Dio3 in rat. *Front. Genet.* 3:279. doi: 10.3389/fgene.2012.00279

Duffie, R., Ajjan, S., Greenberg, M. V., Zamudio, N., Escamilla Del Arenal, M., Iranzo, J., et al. (2014). The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes Dev.* 28, 463–478. doi: 10.1101/gad.232058.113

Fan, T., Hagan, J. P., Kozlov, S. V., Stewart, C. L., and Muegge, K. (2005). Lsh controls silencing of the imprinted Cdkn1c gene. *Development* 132, 635–644. doi: 10.1242/dev.01612

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* 12, 565–575. doi: 10.1038/nrg3032

Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., et al. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493–521. doi: 10.1038/nature02426

Greenberg, M. V., 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

Hanna, C. W., Perez-Palacios, R., Gahurova, L., Schubert, M., Krueger, F., Biggins, L., et al. (2019). Endogenous retroviral insertions drive non-canonical

imprinting in extra-embryonic tissues. *Genome Biol.* 20:225. doi: 10.1186/s13059-019-1833-x

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489. doi: 10.1038/35013106

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129, 1983–1993. doi: 10.1242/dev.129.8.1983

Hermsen, R., De Ligt, J., Spee, W., Blokzijl, F., Schafer, S., Adami, E., et al. (2015). Genomic landscape of rat strain and substrain variation. *BMC Genom.* 16:357. doi: 10.1186/s12864-015-1594-1

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., et al. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* 22, 1607–1616. doi: 10.1101/gad.1667008

Hiura, H., Obata, Y., Komiyama, J., Shirai, M., and Kono, T. (2006). Oocyte growth-dependent progression of maternal imprinting in mice. *Genes Cells* 11, 353–361. doi: 10.1111/j.1365-2443.2006.00943.x

Inoue, A., Jiang, L., Lu, F., Suzuki, T., and Zhang, Y. (2017a). Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* 547, 419–424. doi: 10.1038/nature23262

Inoue, A., Jiang, L., Lu, F., and Zhang, Y. (2017b). Genomic imprinting of Xist by maternal H3K27me3. *Genes Dev.* 31, 1927–1932. doi: 10.1101/gad.304113.117

Inoue, K., Hirose, M., Inoue, H., Hatanaka, Y., Honda, A., Hasegawa, A., et al. (2017c). The Rodent-Specific MicroRNA Cluster within the Sfmbt2 Gene Is Imprinted and Essential for Placental Development. *Cell. Rep.* 19, 949–956. doi: 10.1016/j.celrep.2017.04.018

Itoh, M., Yoshida, Y., Nishida, K., Narimatsu, M., Hibi, M., and Hirano, T. (2000). Role of Gab1 in heart, placenta, and skin development and growth factor- and cytokine-induced extracellular signal-regulated kinase mitogen-activated protein kinase activation. *Mol. Cell. Biol.* 20, 3695–3704. doi: 10.1128/mcb.20.10.3695-3704.2000

John, R. M., and Lefebvre, L. (2011). Developmental regulation of somatic imprints. *Differentiation* 81, 270–280. doi: 10.1016/j.diff.2011.01.007

Johnson, D. R. (1974). Hairpin-tail: a case of post-reductional gene action in the mouse egg. *Genetics* 76, 795–805.

Kajii, T., and Ohama, K. (1977). Androgenetic origin of hydatidiform mole. *Nature* 268, 633–634. doi: 10.1038/268633a0

Kalish, J. M., Jiang, C., and Bartolomei, M. S. (2014). Epigenetics and imprinting in human disease. *Int. J. Dev. Biol.* 58, 291–298. doi: 10.1387/ijdb.140077mb

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., et al. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900–903. doi: 10.1038/nature02633

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., et al. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* 16, 2272–2280. doi: 10.1093/hmg/ddm179

Kawahara, M., Wu, Q., Takahashi, N., Morita, S., Yamada, K., Ito, M., et al. (2007). High-frequency generation of viable mice from engineered bi-maternal embryos. *Nat. Biotechnol.* 25, 1045–1050. doi: 10.1038/nbt1331

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., et al. (2012a). 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

Kobayashi, H., Sakurai, T., Sato, S., Nakabayashi, K., Hata, K., and Kono, T. (2012b). Imprinted DNA methylation reprogramming during early mouse embryogenesis at the Gpr1-Zdbf2 locus is linked to long cis-intergenic transcription. *FEBS Lett.* 586, 827–833. doi: 10.1016/j.febslet.2012.01.059

Kobayashi, H., Yamada, K., Morita, S., Hiura, H., Fukuda, A., Kagami, M., et al. (2009). Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. *Genomics* 93, 461–472. doi: 10.1016/j.ygeno.2008.12.012

Kobayashi, H., Yanagisawa, E., Sakashita, A., Sugawara, N., Kumakura, S., Ogawa, H., et al. (2013). Epigenetic and transcriptional features of the novel human imprinted lncRNA GPR1AS suggest it is a functional ortholog to mouse Zdbf2linc. *Epigenetics* 8, 635–645. doi: 10.4161/epi.24887

Kono, T., Obata, Y., Wu, Q., Niwa, K., Ono, Y., Yamamoto, Y., et al. (2004). Birth of parthenogenetic mice that can develop to adulthood. *Nature* 428, 860–864. doi: 10.1038/nature02402

Latos, P. A., Pauler, F. M., Koerner, M. V., Senergin, H. B., Hudson, Q. J., Stocsits, R. R., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472. doi: 10.1126/science.1228110

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., et al. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat. Genet.* 36, 1291–1295. doi: 10.1038/ng1468

Li, Z. K., Wang, L. Y., Wang, L. B., Feng, G. H., Yuan, X. W., Liu, C., et al. (2018). Generation of Bimaternal and Bipaternal Mice from Hypomethylated Haploid ESCs with Imprinting Region Deletions. *Cell Stem Cell* 23:e664. doi: 10.1016/j.stem.2018.09.004

Linder, D., Mccaw, B. K., and Hecht, F. (1975). Parthenogenic origin of benign ovarian teratomas. *N. Engl. J. Med.* 292, 63–66. doi: 10.1056/NEJM197501092920202

Lopes, S., Lewis, A., Hajkova, P., Dean, W., Oswald, J., Forne, T., et al. (2003). Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum. Mol. Genet.* 12, 295–305. doi: 10.1093/hmg/ddg022

Lucifero, D., Mann, M. R., Bartolomei, M. S., and Trasler, J. M. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum. Mol. Genet.* 13, 839–849. doi: 10.1093/hmg/ddh104

Matoba, S., Nakamuta, S., Miura, K., Hirose, M., Shiura, H., Kohda, T., et al. (2019). Paternal knockout of Slc38a4/SNAT4 causes placental hypoplasia associated with intrauterine growth restriction in mice. *Proc. Natl. Acad. Sci. U. S. A.* 116, 21047–21053. doi: 10.1073/pnas.1907884116

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

Mei, H., Kozuka, C., Hayashi, R., Kumon, M., Koseki, H., and Inoue, A. (2021). H2AK119ub1 guides maternal inheritance and zygotic deposition of H3K27me3 in mouse embryos. *Nat. Genet.* 53, 539–550. doi: 10.1038/s41588-021-00820-3

Messerschmidt, D. M., De Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B. B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* 335, 1499–1502. doi: 10.1126/science.1216154

Mills, J. J., Falls, J. G., De Souza, A. T., and Jirtle, R. L. (1998). Imprinted M6p/Igf2 receptor is mutated in rat liver tumors. *Oncogene* 16, 2797–2802. doi: 10.1038/sj.onc.1201801

Miri, K., Latham, K., Panning, B., Zhong, Z., Andersen, A., and Varmuza, S. (2013). The imprinted polycomb group gene Sfmbt2 is required for trophoblast maintenance and placenta development. *Development* 140, 4480–4489. doi: 10.1242/dev.096511

Mohammad, F., Pandey, G. K., Mondal, T., Enroth, S., Redrup, L., Gyllensten, U., et al. (2012). Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development* 139, 2792–2803. doi: 10.1242/dev.079566

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)90230-N

Okada, I., Hamanoue, H., Terada, K., Tohma, T., Megarbane, A., Chouery, E., et al. (2011). SMOC1 is essential for ocular and limb development in humans and mice. *Am. J. Hum. Genet.* 88, 30–41. doi: 10.1016/j.ajhg.2010.11.012

Okae, H., Matoba, S., Nagashima, T., Mizutani, E., Inoue, K., Ogonuki, N., et al. (2014). RNA sequencing-based identification of aberrant imprinting in cloned mice. *Hum. Mol. Genet.* 23, 992–1001. doi: 10.1093/hmg/ddt495

Okamura, K., Sakaki, Y., and Ito, T. (2005). Comparative genomics approach toward critical determinants for the imprinting of an evolutionarily conserved gene Impact. *Biochem. Biophys. Res. Commun.* 329, 824–830. doi: 10.1016/j.bbrc.2005.02.048

Overall, M., Bakker, M., Spencer, J., Parker, N., Smith, P., and Dziadek, M. (1997). Genomic imprinting in the rat: linkage of Igf2 and H19 genes and opposite parental allele-specific expression during embryogenesis. *Genomics* 45, 416–420. doi: 10.1006/geno.1997.4933

Pearsall, R. S., Plass, C., Romano, M. A., Garrick, M. D., Shibata, H., Hayashizaki, Y., et al. (1999). A direct repeat sequence at the Rasgrf1 locus and imprinted expression. *Genomics* 55, 194–201. doi: 10.1006/geno.1998.5660

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* .44, 361–372. doi: 10.1016/j.molcel.2011.08.032

Saito, T., Hara, S., Kato, T., Tamano, M., Muramatsu, A., Asahara, H., et al. (2018). A tandem repeat array in IG-DMR is essential for imprinting of paternal allele at the Dlk1-Dio3 domain during embryonic development. *Hum. Mol. Genet.* 27, 3283–3292. doi: 10.1093/hmg/ddy235

Santoni, F. A., Stamoulis, G., Garieri, M., Falconnet, E., Ribaux, P., Borel, C., et al. (2017). Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. *Am. J. Hum. Genet.* 100, 444–453. doi: 10.1016/j.ajhg.2017.01.028

Sato, S., Yoshida, W., Soejima, H., Nakabayashi, K., and Hata, K. (2011). Methylation dynamics of IG-DMR and Gtl2-DMR during murine embryonic and placental development. *Genomics* 98, 120–127. doi: 10.1016/j.ygeno.2011.05.003

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

Shirane, K., Miura, F., Ito, T., and Lorincz, M. C. (2020). NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nat. Genet.* 52, 1088–1098. doi: 10.1038/s41588-020-0689-z

Smallwood, S. A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., et al. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* 43, 811–814. doi: 10.1038/ng.864

Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H., et al. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* 73, 61–71. doi: 10.1016/0092-8674(93)90160-r

Takahashi, N., Coluccio, A., Thorball, C. W., Planet, E., Shi, H., Offner, S., et al. (2019). ZNF445 is a primary regulator of genomic imprinting. *Genes Dev.* 33, 49–54. doi: 10.1101/gad.320069.118

Terranova, R., Yokobayashi, S., Stadler, M. B., Otte, A. P., Van Lohuizen, M., Orkin, S. H., et al. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell.* 15, 668–679. doi: 10.1016/j.devcel.2008.08.015

Trivers, R. L. (1974). Parent-Offspring Conflict. *Am. Zool.* 14, 249–264. doi: 10.1093/icb/14.1.249

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., and Erice Imprinting, G. (2019). Genomic Imprinting and Physiological Processes in Mammals. *Cell* 176, 952–965. doi: 10.1016/j.cell.2019.01.043

Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit-Mehouas, S., et al. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* 16:209. doi: 10.1186/s13059-015-0769-z

Wake, N., Takagi, N., and Sasaki, M. (1978). Androgenesis as a cause of hydatidiform mole. *J. Natl. Cancer Inst.* 60, 51–57. doi: 10.1093/jnci/60.1.51

Wang, Q., Chow, J., Hong, J., Smith, A. F., Moreno, C., Seaby, P., et al. (2011). Recent acquisition of imprinting at the rodent Sfmbt2 locus correlates with insertion of a large block of miRNAs. *BMC Genom.* 12:204. doi: 10.1186/1471-2164-12-204

Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., et al. (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* 332, 848–852. doi: 10.1126/science.1203919

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562. doi: 10.1038/nature01262

Williamson, C. M., Ball, S. T., Dawson, C., Mehta, S., Beechey, C. V., Fray, M., et al. (2011). Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. *PLoS Genet.* 7:e1001347. doi: 10.1371/journal.pgen.1001347

Winking, H., and Silver, L. M. (1984). Characterization of a recombinant mouse T haplotype that expresses a dominant lethal maternal effect. *Genetics* 108, 1013–1020.

Wood, M. D., Hiura, H., Tunster, S. J., Arima, T., Shin, J. Y., Higgins, M. J., et al. (2010). Autonomous silencing of the imprinted Cdkn1c gene in stem cells. *Epigenetics* 5, 214–221. doi: 10.4161/epi.5.3.11275

Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'amour, J., Bogutz, A. B., et al. (2019). SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* 51, 844–856. doi: 10.1038/s41588-019-0398-7

Yamasaki, Y., Kayashima, T., Soejima, H., Kinoshita, A., Yoshiura, K., Matsumoto, N., et al. (2005). Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. *Hum. Mol. Genet.* 14, 2511–2520. doi: 10.1093/hmg/ddi255

Zhang, W., Chen, Z., Yin, Q., Zhang, D., Racowsky, C., and Zhang, Y. (2019). Maternal-biased H3K27me3 correlates with paternal-specific gene expression in the human morula. *Genes Dev.* 33, 382–387. doi: 10.1101/gad.323105.118

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

ORIGINAL RESEARCH
published: 23 September 2021
doi: 10.3389/fcell.2021.717555

Systematic Analysis of Monoallelic Gene Expression and Chromatin Accessibility Across Multiple Tissues in Hybrid Mice

Weizheng Liang[1,2,3†], Xudong Zou[2,3*†], Guipeng Li[2,3,4], Shaojie Zhou[2,3], Chi Tian[2,3] and Bernhard Schaefke[2,3,4*‡]*

[1] Harbin Institute of Technology, Harbin, China, [2] Shenzhen Key Laboratory of Gene Regulation and Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China, [3] Department of Biology, Southern University of Science and Technology, Shenzhen, China, [4] Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China

In diploid eukaryotic organisms, both alleles of each autosomal gene are usually assumed to be simultaneously expressed at similar levels. However, some genes can be expressed preferentially or strictly from a single allele, a process known as monoallelic expression. Classic monoallelic expression of X-chromosome-linked genes, olfactory receptor genes and developmentally imprinted genes is the result of epigenetic modifications. Genetic-origin-dependent monoallelic expression, however, is caused by *cis*-regulatory differences between the alleles. There is a paucity of systematic study to investigate these phenomena across multiple tissues, and the mechanisms underlying such monoallelic expression are not yet fully understood. Here we provide a detailed portrait of monoallelic gene expression across multiple tissues/cell lines in a hybrid mouse cross between the *Mus musculus* strain C57BL/6J and the *Mus spretus* strain SPRET/EiJ. We observed pervasive tissue-dependent allele-specific gene expression: in total, 1,839 genes exhibited monoallelic expression in at least one tissue, and 410 genes in at least two tissues. Among these 88 are monoallelic genes with different active alleles between tissues, probably representing genetic-origin-dependent monoallelic expression. We also identified six autosomal monoallelic genes with the active allele being identical in all eight tissues, which are likely novel candidates of imprinted genes. To depict the underlying regulatory mechanisms at the chromatin layer, we performed ATAC-seq in two different cell lines derived from the F1 mouse. Consistent with the global expression pattern, cell-type dependent monoallelic peaks were found, and a higher proportion of C57BL/6J-active peaks were observed in both cell types, implying possible species-specific regulation. Finally, only a small part of monoallelic gene expression could be explained by allelic differences in chromatin organization in promoter regions, suggesting that other distal elements may play important roles in shaping the patterns of allelic gene expression across tissues.

Keywords: monoallelic, allelic, gene expression, chromatin accessibility, *cis* regulation, hybrid mice

OPEN ACCESS

Edited by:
Lan Jiang,
Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences (CAS), China

Reviewed by:
Muhammad Tariq,
Lahore University of Management Sciences, Pakistan
Emily S. Wong,
Victor Chang Cardiac Research Institute, Australia

****Correspondence:***
Weizheng Liang
11749306@mail.sustech.edu.cn
Xudong Zou
zou.xudong@foxmail.com
Bernhard Schaefke
b.schaefke@protonmail.com

† These authors have contributed equally to this work and share first authorship

‡ Present address:
Bernhard Schaefke,
The Brain Cognition and Brain Disease Institute, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (CAS); Shenzhen-Hong Kong Institute of Brain Science – Shenzhen Fundamental Research Institutions, Shenzhen, China

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

Received: *31 May 2021*
Accepted: *01 September 2021*
Published: *23 September 2021*

INTRODUCTION

Protein-coding information stored in DNA is first transcribed to mRNA and then translated into polypeptide chains. Knowing how these processes are regulated is critical for the understanding of development and evolution. Indeed, divergence in gene expression is considered a major cause of phenotypic differences between species (King and Wilson, 1975). Transcriptional regulation is mediated by the interaction between *cis*-regulatory elements (e.g., promoters and enhancers) and *trans*-factors (e.g., transcription factors (TFs)). Whereas *cis*-elements are usually located within or nearby a single target gene whose gene expression they regulate, *trans*-factors can be located on different chromosomes and potentially influence the expression of several often distal target genes. Besides quantitative trait loci (QTL) mapping analyses, which require large sample sizes and inform about distal and proximal elements affecting gene expression differences, F1 hybrid studies are another widely applied and more straightforward approach to distinguish between *cis* and *trans* acting regulatory components (Wittkopp et al., 2004; Gao et al., 2015; Hou et al., 2015; Xiao et al., 2016; Xu et al., 2017). With two alleles sharing the same *trans* environment, allelic differences in the F1 hybrid can be directly interpreted as *cis*-regulatory divergence (Gao et al., 2015; Hou et al., 2015; Xiao et al., 2016). By comparing these allelic-specific variations with the differences between parental strains or species, the *trans*-component of gene expression differences can be estimated (Wittkopp et al., 2004, 2008; Goncalves et al., 2012; Wong et al., 2017). The F1 hybrid approach has been used to study *cis* and *trans* regulatory divergence contributing to differences in gene expression between strains of the same species or closely related species in many model organisms, including yeast (Tirosh et al., 2009; Emerson et al., 2010; Schaefke et al., 2013), Drosophila (Wittkopp et al., 2004) and mouse (Goncalves et al., 2012). F1 hybrid studies of different *Mus musculus* subspecies revealed pervasive *cis*-regulatory differences but comparatively few *trans*-regulatory differences (Goncalves et al., 2012; Crowley et al., 2015). However, the interplay of these two kinds of elements shaping the regulatory patterns of gene expression divergence across tissues in mammals has not been fully understood.

The most extreme case of an allelic-biased expression pattern is monoallelic expression (when a gene is only transcribed from one of the two parental alleles). Classic monoallelic expression of X-chromosome-linked genes, olfactory receptor genes and developmentally imprinted genes is the result of epigenetic modifications (Chess, 2013, 2016; Eckersley-Maslin et al., 2014; Gendre et al., 2014). Genetic-origin-dependent monoallelic expression, in contrast, is caused by *cis*-regulatory differences between the alleles (Ohishi et al., 2020), and cases of non-random allele-dependent X-chromosome inactivation have also been described (Orstavik et al., 1995; Thorvaldsen et al., 2012; Calaway et al., 2013; Jones, 2014). However, the tissue-dependence of these phenomena is rather underexplored, and the mechanisms underlying asymmetric expression are not yet fully understood.

Here we provide a detailed portrait of monoallelic gene expression across multiple tissues/cell lines in a hybrid mouse model and allelic chromatin accessibility patterns in two different cell lines. We focus on depicting tissue-dependent allele-specific gene expression patterns and the underlying regulatory mechanisms at the chromatin accessibility layer. We observed pervasive tissue-dependent allele-specific gene expression and chromatin accessibility patterns. In total, 1,839 genes exhibited monoallelic expression in at least one tissue, and 410 in at least two tissues. We identified six autosomal monoallelic genes with the active allele being identical in all eight tissues, which are likely novel candidates of imprinted genes. Also, we found 88 monoallelic genes with different active alleles between tissues. Only a small part of monoallelic gene expression could be explained by allelic chromatin structural differences in promoter regions, suggesting that other distal elements or differential TF binding without divergence in chromatin remodeling may play important roles in shaping the patterns of allelic gene expression across tissues.

MATERIALS AND METHODS

RNA-Seq Data of F1 Hybrid Mouse

RNA-seq raw data of F1 hybrid mice containing six organs, embryonic stem cells (ESCs), and fibroblasts were obtained from previous studies (Gao et al., 2015; Zou et al., 2021). For each tissue/cell type, raw sequencing data of two biological replicates were downloaded. Samples from heart, kidney and cortex were sequenced with paired-end reads of 101 bp length. Samples from spleen, lung and ESCs were sequenced with paired-end reads of 76 bp length. The sequencing depth for each biological replicate was 240–260 million reads per sample, except for the two ESC samples, for which we obtained 175 million and 202 million reads, respectively.

Assay for Transposase-Accessible Chromatin-Seq Library Construction and Sequencing

The ATAC-seq libraries of F1-ESCs and F1-fibroblasts, each with three biological replicates, were prepared as previously described with minor modifications (Corces et al., 2017; Liang et al., 2021). Briefly, 50,000 fresh cells were lysed in lysis buffer for 10 min on ice to prepare the nuclei. Immediately after lysis, nuclei were spun at 500 g for 10 min to remove the supernatant. Nuclei were then incubated with the Tn5 transposase (Vazyme) in tagmentation buffer at 37°C for 30 min. After tagmentation, PCR was performed to amplify the library for 12 cycles under the following PCR conditions: 72°C for 3 min; 98°C for 30 s; and thermocycling at 98°C for 15 s, 60°C for 30 s, and 72°C for 40 s; followed by 5 min at 72°C. After the PCR reaction, libraries were purified with DNA purification beads (Vazyme). The libraries were sent to Annoroad for sequencing and 2 × 150 bp reads were obtained.

RNA-Seq Data Processing

The reference *M. musculus* genome (mm10) and gene annotation of the C57BL/6J strain were downloaded from the

Ensemble database[1] (version: GRCm38, release 74). SNVs and insertions/deletions (indels) between C57BL/6J and SPRET/EiJ were downloaded from the Mouse Genome Project.[2]

The vcf2diploid tool (version 0.2.6) in the AlleleSeq pipeline (Rozowsky et al., 2011) was used to construct the SPRET/EiJ genome by incorporating the SNVs and indels into the C57BL/6J genome. The chain file between the two genomes was also reported as an output, which was further used with g2gtools to convert SPRET/EiJ coordinates to C57BL/6J coordinates.

To ensure that RNA-seq reads from all tissues have the same length, we trimmed 25 bp from the 3′ end of the 101 bp reads. We aligned the RNA-seq reads to the C57BL/6J reference genome and SPRET/EiJ genome separately with HISAT2 (version 2.0.1) with parameters -p 12 -k 2 –reorder –no-softclip ("softclip" was not allowed when mapping in order to avoid junction reads to be cut off; the "reorder" parameter was used to ensure that the reads order of the mate pairs in the HISAT2 output is consistent with the order of reads of the input file for efficient assignment). Reads were assigned to the genome with less mapping edit distance. The reads with equal mapping distance to both genomes were assigned as common reads. Genomic alignment coordinates of the reads that were assigned to SPRET/EIJ were then converted to the corresponding locations in the C57BL/6J reference genome using the g2gtools software (version 0.1.29). The bias of allelic reads assignment in favor of the reference genome (C57BL/6J) was low, ranging from 0.1% in kidney to 3.5% in fibroblasts (S), which indicates that our strategy for allelic reads assignment is reliable for allelic gene expression estimation.

Gene Expression Level Quantification

After reads alignment and allelic reads assignment, uniquely mapped reads of each allele were chosen and fed into featureCounts (v1.6.0) for gene expression quantification. Only both ends of a read pair concordantly mapped were counted (by "-B" and "-C"). Raw read counts were then normalized as transcripts per kilobase million (TPM).

Identifying Allelic Differential Genes

Divergent and monoallelic genes were detected following the pipeline in **Supplementary Figure 1**. At first, protein-coding genes with TPM no greater than 1 (not allelic) in both biological replicates, and genes located in X, Y, and mitochondrial chromosomes, as well as known imprinting genes were removed. The remaining autosome protein-coding genes were kept for divergent and monoallelic genes analysis. To make sure the differential analysis between alleles are supported by enough SNVs, we selected genes with 5 or more allele informative SNVs (covered by more than 20 reads) between alleles in all annotated exons. Paired-sample *t*-test on count of reads cross SNVs in a genes were performed and BH adjust *p*-values were obtained. The log2 transformed fold change were calculated between alleles based on summed up reads cross all SNVs. A gene was defined as allelic differential gene (ADE) if LFC greater than 1 and adjusted *p*-value less than 0.05.

Assay for Transposase-Accessible Chromatin-Seq Data Processing

2 × 150 bp paired-end reads were first trimmed to remove adapter sequences using Trim Galore v0.6.4 (Krueger, 2016) (–cores 4 –paired –nextera –length 50). Cleaned reads were aligned to the C57BL/6J reference genome and SPRET/EiJ genome separately with g Bowtie2 (Langmead and Salzberg, 2012) (version 2.4.1) with parameters -p 8 -X 2000. Reads mapped to the mitochondrial genome and low mapping quality reads (MAPQ < 10) were filtered out using custom scripts. Picard (v2.12.1) was then used to sort the reads and remove duplicates. Reads were assigned to the two mouse genomes with less mapping edit distance. Only reads which could be assigned unambiguously (allelic reads or allele informative reads) to either of the two genomes were kept for further analysis. Genomic alignment coordinates of the reads that were assigned to SPRET/EIJ were then converted to the corresponding locations in the C57BL/6J reference genome using the g2gtools software (version 0.1.29).

Reads of both alleles in each of the six samples were merged as input for MACS2 to call consensus peaks (-f BAMPE -g mm –keep-dup all –nomodel –nolambda -B). And to ensure reproducibility, only peaks detected by IDR (version 2.0.4.2, with parameter –idr-threshold 0.05) in all three replicates were used for further analysis.

After obtaining consensus peaks, allelic read counts for each peak in each sample were analyzed by featureCounts v1.6.4 (Liao et al., 2014) (with parameters: -F SAF -p -B -C -T 4). Differential peaks between alleles were detected using the R package DESeq2 (Love et al., 2014) (under R version 4), and peak annotation was analyzed with the R package ChIPseeker (Yu et al., 2015) (under R version 4).

Gene Annotation

Gene type, exons, and transcription start site (TSS) annotations were extracted from the gene annotation file of the mouse reference genome mm10 downloaded from the Ensembl website.[3]

Filtering

Non-coding genes were firstly removed from our gene list, and to ensure reliable downstream analysis, X-chromosomal genes and known imprinted genes[4] were analyzed separately and only autosomal genes with TPM ≥ 1.0 in both replicates remain. Since allelic reads assignment is dependent on *cis*-variants between alleles, to avoid bias of reads assignment, we further filtered out genes with less than 5 informative SNVs (covered by 20 reads).

Principal Component Analysis

Principal component analysis (PCA) was performed on all protein-coding genes after filtering. Allelic reads count of each

[1] ftp://ftp.ensembl.org

[2] http://www.sanger.ac.uk/

[3] http://ensembl.org/

[4] http://www.geneimprint.com/site/genes-by-species

gene in each sample were normalized as counts per million (CPM). Then the normalized count matrix was fed into the R prcomp function to run the PCA analysis, and the first two components were used for sample visualization.

Allelic Differential Genes and Monoallelic Expression Gene

For each gene, allelic reads covering informative SNVs were summed up, then a logarithm transformed fold change (LFC) between alleles was calculated as in the following Eq. 1:

$$\mathrm{LFC} = \log_2((BL6+1)/(SPR+1)), \quad (1)$$

(1) where *BL6* means informative allelic reads from the C57BL/6J allele, and *SPR* means informative allelic reads from SPRET/EiJ allele. Allelic differential genes (ADE) are defined as those with absolute LFC equal or greater than 1, and two-sample paired *t*-test *p*-value < 0.05.

We also defined a *p* score (calculated as the proportion of BL6 allelic reads, Eq. 2) to distinguish monoallelic expression genes (MAE). As defined in a previous study (AV Gendrel, Development cell, 2014), Genes with *p* score >0.85 or *p* score <0.15 were defined as MAEs.

$$\text{p score} = BL6/(BL6+SPR), \quad (2)$$

where *BL6* means informative allelic reads from the C57BL/6J allele, and *SPR* means reads from the SPRET/EiJ allele.

Reproducibility of Monoallelic Genes Between Biological Replicates

We used the same cutoff of *p* score (>0.85 or <0.15) as above to define monoallelic patterns in each of the two biological replicates in each tissue. And we calculated the proportion of consistent patterns between the two replicates as shown in **Supplementary Table 5**. We observed high consistency between replicates for all tissues/cell lines.

Replication Rate of Monoallelic Genes Between Tissues

To define the replication rate of monoallelic genes between tissues, we calculate a replication rate, similar to the Jaccard Index, which represents the proportion of intersection in the union (Eq. 3). For any two tissues, the replication rate was defined as:

$$\text{replication rate} = \bigcap MAEs / \bigcup MAEs, \quad (3)$$

Where $\bigcap MAEs$ means monoallelically expressed genes in one tissue, also monoallelically expressed in the other tissue with the same preferred allele; $\bigcup MAEs$ means the union set of MAE genes between the two tissues.

d Score

We calculated a *d* score for each peak based on the previous definition (Xu et al., 2017). We treated each fibroblast cell line as if derived from a single clone, based on this, peaks of one allele on the X-chromosome would mostly be inactive with a few peaks escaped. Therefore, we compared the distribution of *d* scores in X-inactive peaks and X-escaped peaks, and set the cross-site where the *d* score equals 0.35 as the threshold for defining monoallelic peak.

RESULTS

Autosomal Monoallelic Gene Expression Is Pervasive Across Tissues

To identify allelic differentially expressed genes in the mouse genome, we performed RNA-seq of six different organs (cerebral cortex, heart, kidney, liver, lung, and spleen) and two cell types (ESCs and fibroblasts) from a highly divergent F1 hybrid cross between the house mouse *M. musculus* (C57BL/6J) and the Algerian mouse *Mus spretus* (SPRET/EiJ) which was generated for previous studies (Gao et al., 2015) in our lab (**Figure 1A**). Data of two biological replicates of each organ and cell type were used. After read mapping (see section "Materials and Methods"), uniquely mapped read pairs were assigned to each allele based on "edit distance." PCA analysis shows that the samples are clustered together firstly by tissue type or cell type and then by species (**Figure 1B**), which is consistent with previous studies (Barbosa-Morais et al., 2012; Merkin et al., 2012). To accurately estimate genes with allelic differential expression (ADE genes) and monoallelic expression (MAE genes), only protein-coding genes in autosomes with at least 5 allele-informative SNVs (see section "Materials and Methods") were used and an in-house pipeline based on allele-informative SNVs was designed to identify ADE and MAE genes (**Supplementary Figure 1** and section "Materials and Methods"). After filtering, 15,469 protein-coding genes (~68.5% of protein-coding genes in the genome), in total, remained for downstream allelic gene expression analysis. The numbers of genes expressed in the eight tissues/cell types are similar, with an average number of 11,258, the highest number of 12,632 in lung, and the lowest number of 9,805 in liver (**Supplementary Table 1**). In contrast to this, the numbers of ADE genes we identified across tissues vary significantly. In cerebral cortex, we identified only 680 (5.48%) ADE genes which were less than half of the number observed in fibroblasts (13.92%, **Supplementary Table 1**). This is consistent with previous reports that brain is one of the most conserved organs between species (Zheng-Bradley et al., 2010). Similar to a previous study (Andergassen et al., 2017), we also observed bias toward higher expression of the C57BL/6J allele in ADE genes (**Supplementary Table 1**). To check whether the bias of allelic gene expression is caused by technical issues, we compared the log fold change (LFC) of allelic gene expression between tissues/cell types. If the allelic bias is largely caused by technical issues, we would observe similar correlations among different tissue pairs. As shown in **Supplementary Figure 2A**, the Spearman's correlation coefficient between tissues ranges from 0.25 (between ESC and liver) to 0.55 (between lung and spleen). Such a big variation indicates that the number of ADE genes biased toward C57BL/6J may mainly represent a biological phenomenon rather than technical bias.

In each tissue, on average 27% of ADE genes show monoallelic expression (**Figure 1C**, **Supplementary Table 1**, and section

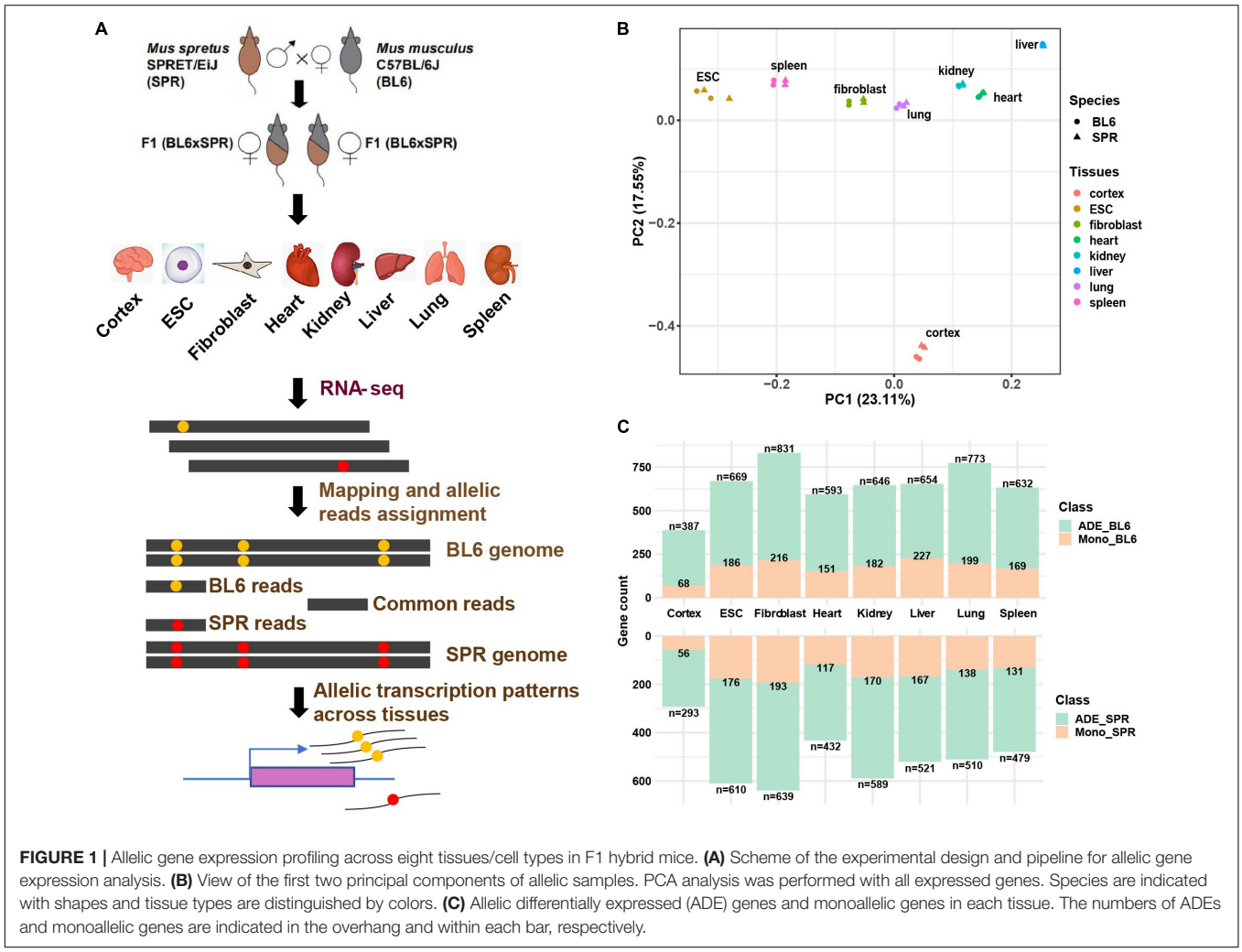


FIGURE 1 | Allelic gene expression profiling across eight tissues/cell types in F1 hybrid mice. **(A)** Scheme of the experimental design and pipeline for allelic gene expression analysis. **(B)** View of the first two principal components of allelic samples. PCA analysis was performed with all expressed genes. Species are indicated with shapes and tissue types are distinguished by colors. **(C)** Allelic differentially expressed (ADE) genes and monoallelic genes in each tissue. The numbers of ADEs and monoallelic genes are indicated in the overhang and within each bar, respectively.

"Materials and Methods"). Cerebral cortex contains fewer (∼18%) monoallelic genes than other tissues, while liver contains the highest percentage (∼33%) of monoallelic genes. Except for liver (394), the highest numbers of MAE genes were found in fibroblasts (409, 27.8%) and ESCs (362, 28.3%). These might partially represent clonally fixed random monoallelic expression, which cannot be easily detected in less homogenous tissues. Moreover, previous studies have shown that in hybrids the *M. spretus* X-chromosome is less likely to be inactivated than the *M. musculus domesticus* X-chromosome (Calaway et al., 2013). Our data are consistent with this prediction, with average *p* scores (see section "Materials and Methods") for X-chromosomal genes ranging between 0.37 in heart and 0.49 in kidney.

Monoallelic Genes Are Mostly Under Tissue-Dependent Regulation

Cis-regulatory divergence of monoallelic genes is also shaped by *trans*-factors in different tissues. To study this *cis-trans* interplay, we analyzed the tissue-dependent patterns of monoallelic gene expression (**Figure 2A**). Among the 1,839 MAE genes, most (1,429) are tissue specific, 404 genes are MAE in 2–7 tissues, and only 6 genes are monoallelic across all eight tissues. We also compared the replication rate of monoallelic genes between tissues by calculating a Jaccard index (see section "Materials and Methods"), for which also the direction of the expression bias is considered. As shown in **Figure 2B**, the replication rates for most tissue pairs are less than 20%, with a mean replication rate of 16.7%, and the maximal replication rate of 29.1% between spleen and lung. In comparison, the 125 known imprinted genes show higher consistency of allelic preference between tissues, as expected. Replication rates for these genes between tissues are mostly greater than 60%, and closely related tissues have higher replication rates; for example, the highest replication rate is 0.85 between lung and spleen (**Supplementary Figure 2B**). These results indicate that *cis*-regulatory monoallelic gene expression is pervasively tissue-dependent. Additionally, we found that the monoallelic status transitions mostly occur between monoallelic ("Mono" in **Figure 2C**) and non-divergent ("Non-Div" in **Figure 2C**), which comprised 57% of between tissue patterns, followed by the transition between divergent-but-not-monoallelic and monoallelic (∼25%, **Figure 2C**), and

the remaining 18% between tissue patterns are the transitions between BL6-monoallelic and SPR-monoallelic. To further study the tissue-dependent patterns of monoallelic genes, we focused on genes expressed in two or more tissues/cell types. This group in total contains 13,855 genes, which we classified into six groups: (G1) genes are non-divergent in all expressing tissues; (G2) genes are divergent in at least one tissue but not monoallelic in any tissue; (G3) genes are monoallelic in only one expressing tissue; (G4) genes are monoallelic in two or more tissues and the BL6 allele is the active allele only; (G5) genes are monoallelic in two or more tissues and the SPR allele is the active allele only; (G6) genes are monoallelic in two or more tissues and with different active allele in different tissues. As shown in **Figure 2D**, except for the 8,967 (64.7%) non-divergent genes (G1), 3,306 genes (23.9%) belong to "G2," 1,172 genes (8.5%) to "G3," and 181 and 141 genes belong to "G4" and "G5," respectively. In addition, we found 88 genes in "G6," which have different active alleles in different tissues.

The *cis*-regulatory divergence of gene expression between species could be functional (adaptive evolution) or it could be noise caused by molecular error, as postulated before as "error hypothesis" (Zhang, 2018). Previous studies in alternative polyadenylation have found that most *cis*-regulatory divergence between species is noise caused by molecular error (Xu and Zhang, 2018). To test whether different tissue-dependent *cis*-regulatory monoallelic genes are under different selection constraint, we compared dN/dS ratios of genes among the six groups. As shown in **Figure 2E**, the dN/dS ratios of non-divergent genes are lower (G1 in **Figure 2E**) than those of divergent genes, no matter whether monoallelic or not. Among the genes with monoallelic patterns, those monoallelic in only one tissue have lower dN/dS ratios than those monoallelic in two or more tissues and with the same active allele between tissues. Interestingly, although statistically not significant, monoallelic genes with different active allele in different tissues have lower dN/dS ratios than those with the same active allele between tissues (**Figure 2E**), suggesting more complex regulatory patterns under this small set of genes.

To further explore the tissue-dependent *cis*-regulatory divergence and its underlying regulatory mechanisms, we put the focus on two cell lines (ESCs and fibroblasts), and performed Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) in samples of these two cell lines from our F1 hybrid mice (described in the next section). Comparing ESCs and fibroblasts (**Figure 2F**), there are 118 fibroblast-specific monoallelic genes, 85 ESC-specific monoallelic genes, and 53 genes monoallelic in both cell types. We confirmed with Sanger sequencing that allelic expression of the gene encoding apolipoprotein E (ApoE) is biased only in fibroblasts but not in ESC (**Supplementary Figures 3A,B**). And more interestingly, among the 53 genes monoallelic in both cell types, nine (∼17%) genes had different dominant alleles (with opposite direction of divergence) in the two cell lines. Again, by validating with Sanger sequencing (data not shown), two genes were confirmed, one is Msln with dominant BL6 allele in ESC and dominant SPR allele in fibroblasts. The other is Epb41l3 which, in contrast, has an active SPR allele in ESC and active BL6 allele in fibroblasts (**Supplementary Figures 3A,B**).

Allelic Chromatin Accessibility Patterns in F1 ESCs and Fibroblasts

To understand tissue-dependent ADE patterns on the level of chromatin organization, we performed ATAC-seq experiments on six samples obtained from cultured ESC (three biological replicates) and fibroblasts (three biological replicates, see section "Materials and Methods"). Both replicates showed good correlation for ESCs and fibroblasts (**Supplementary Figure 6**). After processing the sequencing data (**Supplementary Table 3**) by following the pipeline described in **Supplementary Figure 4**, we identified 47,498 and 55,699 reproducible peaks in ESC and fibroblasts, respectively (**Supplementary Table 4**). When checking allelic read counts in each peak, fibroblast presents 4,247 (∼8.9%, **Supplementary Table 3**) allelic divergent peaks (ADP) which is nearly two times higher than ADPs in ESC (∼5.1%, **Supplementary Table 3**). To further identify monoallelic peaks, we calculated a *d* score (see section "Materials and Methods") as defined before (Xu et al., 2017) for each peak, and use X-chromosomal peaks in fibroblast to determine the threshold of *d* score as 0.35 (**Figure 3A** and section "Materials and Methods") for monoallelic peak identification. Based on this threshold, we identified 2,699 (∼5.7%, **Figure 3B**) monoallelic peaks in fibroblast, which is almost two times of the proportion in ESC (1,712 peaks, ∼3.1%, **Supplementary Figure 5A**). In addition to the differences in total number of monoallelic peaks between the two cell types, the ratio of components (C57BL/6J-active peaks and SPRET/EiJ-active peaks) is also different between ESC and fibroblasts. Among the 1,712 monoallelic peaks in ESC, 62.5% of them are C57BL/6J-active peaks and only 37.5% are SPRET/EiJ-active peaks, while in fibroblast, the two proportions are 55.9 and 44.1%, respectively (**Supplementary Table 3**). Since in the case of genetic-origin-dependent monoallelic expression the divergences between alleles are caused by *cis*-variants, we supposed that the SNV density in monoallelic peaks should be greater than that of non-monoallelic peaks. Indeed, as shown in **Figure 3C**, the median number of SNVs in monoallelic peaks in fibroblast cells is 1.89 per 100 base pairs, which is significantly higher than that in non-monoallelic peaks (with 1.56 median number of SNVs per 100 base pairs) in fibroblast. This is also observed in ESC (**Supplementary Figure 5B**). As shown in **Figure 3D**, most of the peaks are cell-type specific, only 24.4% of the identified peaks are shared between ESC and fibroblasts, while the others are either ESC-specific (42.8%) or fibroblast-specific (32.9%). And the monoallelic peaks are more likely to be found in cell-type specific peaks (**Figure 3E**). For those shared peaks, we also compared their divergence patterns between ESC and fibroblasts. Unlike the patterns of allelic gene expression (shown in **Figure 2E**), the bias of monoallelic peaks between tissues is much bigger (**Figure 3F**). Only 52 ESC-dependent monoallelic peaks were found, compared to 309 fibroblast-dependent monoallelic peaks. Among the 79 monoallelic peaks in both cell types, only

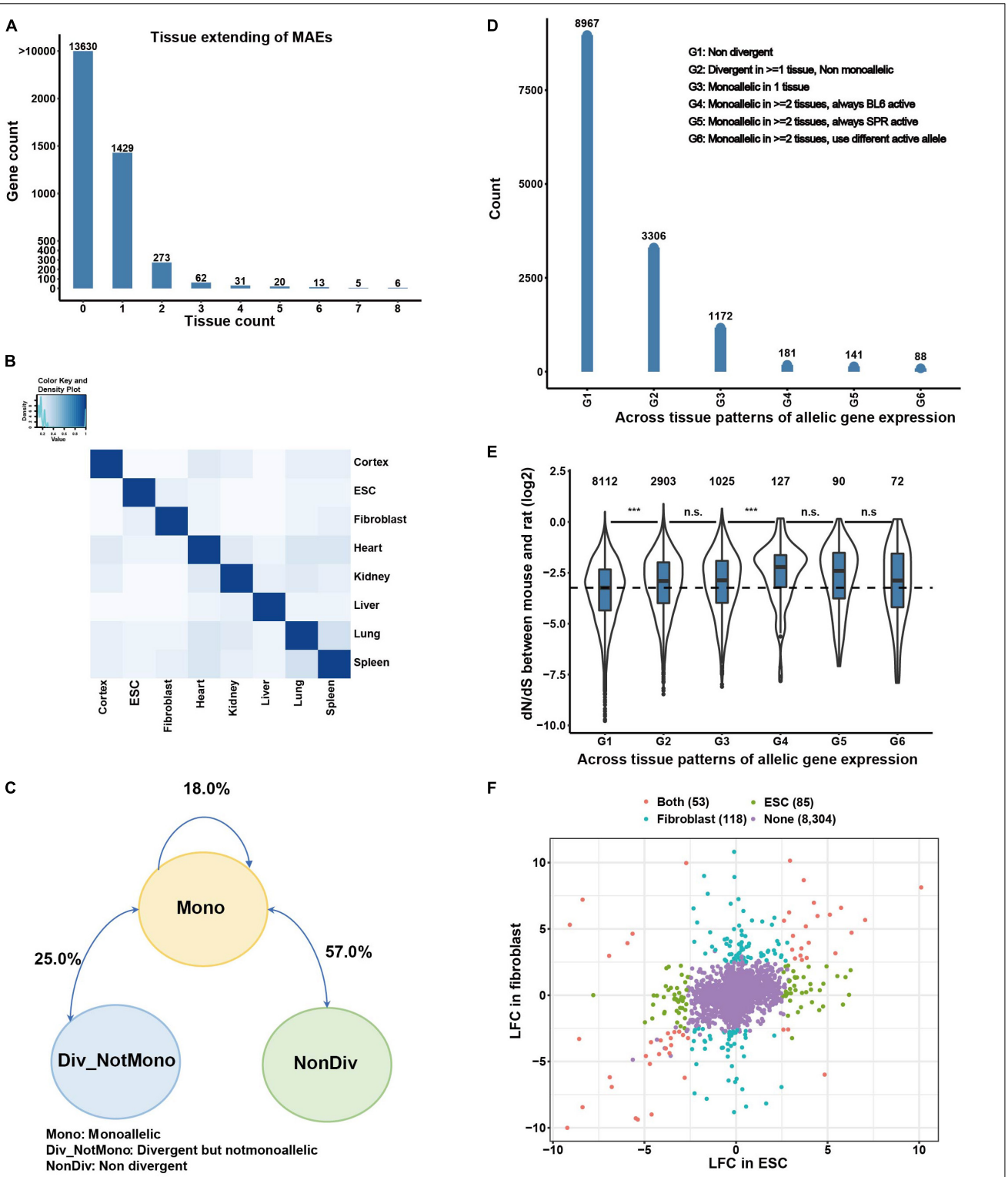


FIGURE 2 | Monoallelic gene expression is mostly tissue-dependent. **(A)** Tissue distribution of monoallelic expressed genes. Zero in x axis means those genes are not MAEs in any of the analyzed tissues/cell types. **(B)** Replication rate of MAEs between tissues. Replication rates are measured by Jaccard Index (see section "Materials and Methods"). **(C)** Transition between different regulatory patterns. **(D)** Genes are classified into six groups based on their monoallelic patterns across tissues. G1 means the genes are non-divergent across tissues; G2 means the genes are divergent in more than one tissue but not monoallelic; G3 means the genes are monoallelic in only one tissue; G4 means the genes are monoallelic in more than one tissue and always BL6 allele active; G5 means the genes are monoallelic in more than one tissue and always SPR allele active; G6 means the genes are monoallelic in more than one tissue but with different active alleles. **(E)** dN/dS ratios of genes in different classes. **(F)** Comparison of monoallelic gene expression patterns between ESCs and fibroblasts.

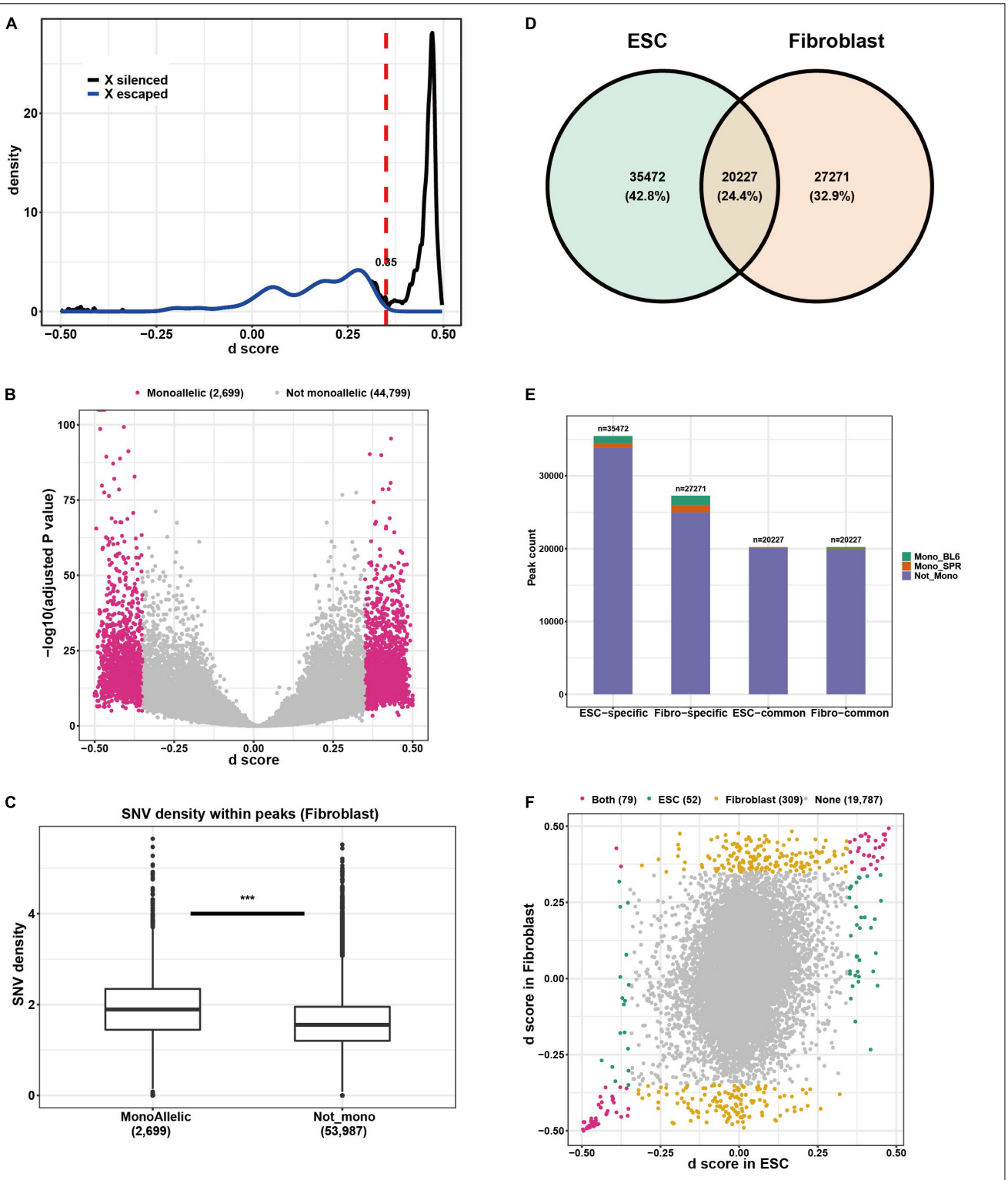


FIGURE 3 | Quantification of allelic chromatin accessibility in ESCs and fibroblasts and cell-type dependent patterns. **(A)** *d* score of X-chromosome peaks in fibroblast. The blue line describes the distribution of *d* score of X-escaped peaks, black line indicates the *d* score distribution of X-silenced peaks. **(B)** Monoallelic peaks detected in fibroblast cells. **(C)** SNV density comparison between monoallelic peaks and non-monoallelic peaks. **(D)** Overlapped peaks between ESC and fibroblast. **(E)** Monoallelic peaks in cell type specific peaks and in peaks shared by the two cell types. **(F)** Cell type-dependent patterns of monoallelic peaks.

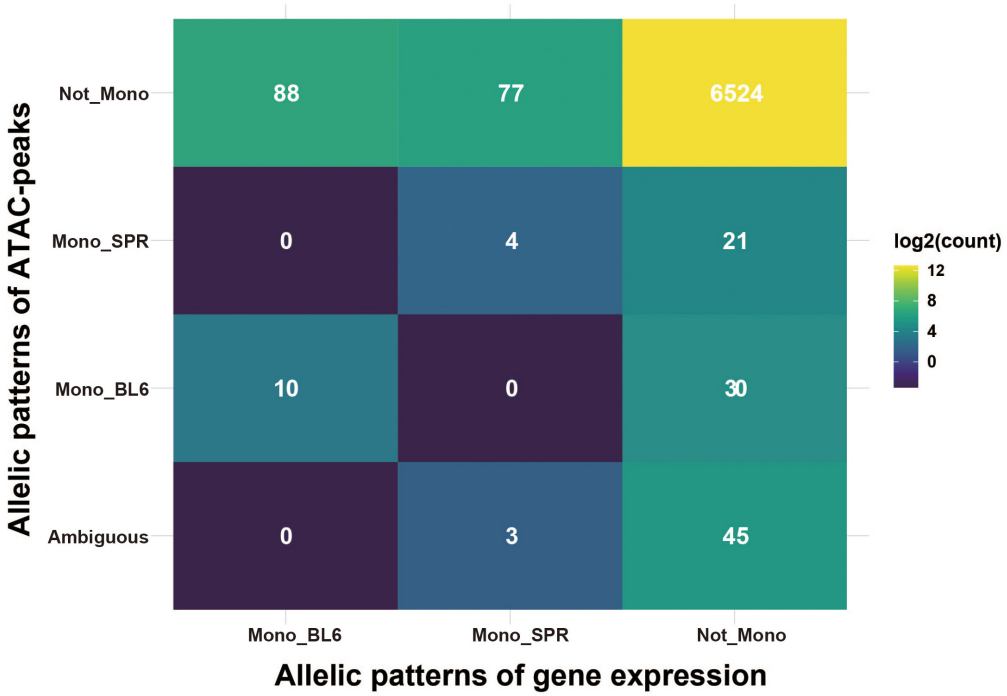


FIGURE 4 | Integrated patterns of allelic gene expression and ATAC-peaks. The heatmap presents the integrated allelic patterns of gene expression and ATAC-peaks in its promoter regions (2.5 kb upstream and 0.5 kb downstream of TSS). Gene expression was classified into three classes: (1) Mono_BL6, means genes monoallelically expressed in C57BL/6J allele; (2) Mono_SPR, means genes monoallelically expressed in SPRETUS allele; (3) Not_Mono, means the genes are not monoallelic genes. ATAC-peaks are classified similarly with one more class indicated as "ambiguous" which means there are both "Mono_BL6" peaks and "Mono_SPR" peaks in that gene. Numbers in each cell indicate the count of overlapped genes. The color filled in cells are scaled by log2 of count.

2 (∼2.5%) have different active alleles, which is much less than on the transcriptional level (17%).

Integration of Allelic Gene Expression and Allelic Assay for Transposase-Accessible Chromatin-Peaks

To see the relationship of allelic patterns between transcription level (gene expression) and chromatin accessibility level (ATAC-peak), we integrated the two kinds of data by annotating peaks to promoter regions (2.5 kb upstream and 0.5 kb downstream of TSS) of target genes. Among the 10,559 genes expressed in fibroblasts, 6,802 (64.4%) contain at least one ATAC-peak in the promoter region. Interestingly, only 14 genes had consistent allelic patterns between gene expression and ATAC-peaks (**Figure 4**). In ESC, such cases are even fewer (three genes, **Supplementary Figure 5C**). This, on one hand, indicates that elements at promoter regions have limited contributions to allelic regulation of transcription, as reported before, distal elements like enhancers could play an important role. On the other hand, the allelic divergence at the transcription level may be invisible at the chromatin level. This is possible if the *cis* variants change the motif of one *trans* factor to another one without affecting chromatin organization. A previous study found that *cis*-regulatory mutations are more likely to change the binding motif of one transcription factor to that of another one than completely abolishing transcription factor binding (Payne et al., 2018), suggesting the plausibility of this mechanism.

DISCUSSION

In diploid eukaryotic organisms, the two alleles of each gene are generally expressed at similar levels. However, monoallelic gene expression occurs in various types and can be regulated by differential mechanisms involving genetic, epigenetic and/or stochastic elements. Classic monoallelic expression of X-chromosome-linked genes, olfactory receptor genes and developmentally imprinted genes has been documented elsewhere (Cedar and Bergman, 2008; Li and Sasaki, 2011; Schulz and Heard, 2013; Tucci et al., 2019). But the genetic-origin-dependent case including its regulatory mechanisms and especially evolutionary conservation among tissues is less studied until now. Here, we applied allelic RNA-seq and ATAC-seq to a highly polymorphic mouse hybrid F1 system crossed between the *M. musculus* strain C57BL/6J and the *M. spretus* strain SPRET/EiJ which possess the largest evolutionary distance to date in mouse to explore this case. Our study provides a detailed portrait of allelic gene expression including monoallelic genes, across multiple tissues/cell lines

in a hybrid mouse model and allelic chromatin accessibility patterns in two different cell lines. We focus on depicting tissue-dependent allele-specific gene expression patterns and the underlying regulatory mechanisms at the chromatin accessibility layer. We observed pervasive tissue-dependent allele-specific gene expression and chromatin accessibility patterns. Cortex exhibited the fewest allele-specific expression differences while fibroblasts showed the most, which is consistent with previous results that the brain is one of the most conserved organs with regards to expression patterns (Zheng-Bradley et al., 2010). We identified six autosomal monoallelic genes with the active allele being identical in all eight tissues, resembling the patterns found for known imprinted genes and therefore likely to be novel candidates of imprinted genes. In addition, we found 88 monoallelic genes with different active alleles between tissues, which likely represent cases of genetic-origin-dependent MAE rather than random MAE. As shown in multiple previous studies, random monoallelic expression can be reliably detected in clonal cell lines (Eckersley-Maslin et al., 2014; Gendre et al., 2014). However, it is unlikely that the F1 hybrid tissues in our study are derived from a single clonal cell line. If this were the case, the X chromosome inactivation pattern would also show a biased pattern. Here, it is not the case, as shown in **Supplementary Figure 7**, the *p* scores (ranges from 0 to 1.0) calculated as the proportion of BL6 reads (*p* score close to 1.0 or close 0 indicates allele-specific expression, see details in section "Materials and Methods") of the six F1 tissues are close to 0.5. In addition, we also observed high consistency between the two biological replicates for identified monoallelic genes (**Supplementary Table 5**), which further supports the mono allelic expression pattern was unlikely due to random inactivation of one allele. In contrast to the F1 tissues, the F1 ES cells and fibroblasts used in this study are clonal cell lines, and as expected we observe a *p* score in ES cells of about 0.5 (because X chromosome is not inactivated in ES cells) and a *p* score in fibroblasts of nearly 1.0. Therefore, some random monoallelic expression might be present in these two cell lines. However, as the number of monoallelic genes detected in these two cell lines is similar to that in the tissues (**Figure 1C**), we think even here the monoallelic expression is largely non-random. Our study is limited to identifying putative candidates for these two classes of MAE, as clonally fixed random MAE can only be detected in monoclonal cell lines, but not in any of the solid tissues, and single-cell experiments would be needed for detecting dynamic random MAE in the future. Interestingly, we found that for autosomal genes the C57BL/6J (maternal) allele is slightly more likely to be expressed at higher levels. Future studies should address the question whether this is a genome-wide genetic-origin-dependent (strain-specific) effect or a parent-of-origin effect comparable to that found by Crowley et al. (2015), who showed a global bias toward the paternal allele in *M. musculus* subspecies hybrids, and whether the preference of the maternal allele we found is unique to the interspecific cross used here.

We also elucidated the possible causal relationship between differential chromatin accessibility and gene expression. We observed that fibroblast cells had more monoallelic ATAC peaks than ESCs, suggesting ESCs are more conserved at the chromatin layer between these two strains. This finding agrees with the expectation that stronger selective constraints act on gene regulation in this early developmental stage than in fibroblasts. We also found that cell type dependent patterns similar to those at the gene expression level were also prominent at the chromatin layer. Additionally, C57BL/6J-active peaks were more prevalent than SPRET/EiJ-active peaks in both cell types, which corresponds to our data on the gene expression level, implying a potential parent-of-origin or strain-specific regulatory mechanism which needs future exploration. Finally, the *cis* regulatory mechanisms can partially account for the existence of monoallelic peaks, as the SNP density is higher in monoallelic peaks for both cell types. However, only a small part of monoallelic gene expression patterns could be explained by allelic chromatin structural patterns in promoter regions, suggesting that other distal elements may play important roles in shaping the patterns of allelic gene expression across tissues or that *cis*-regulatory mutations can change gene expression without affecting chromatin organization.

As reviewed in da Rocha and Gendrel (2019), DNA methylation, one major driver controlling gene expression, plays an important role in X chromosome inactivation and imprinting, but it is not a common epigenetic signature at loci with random monoallelic expression. Therefore, in the future, methods combining chromatin, DNA, RNA, protein and also single cell omics techniques will help to understand the interplay of hypermethylation and other molecular mechanisms for the regulation of different kinds of monoallelic expression, including genetic-origin-dependent monoallelic expression, and their relevance in speciation and phylogeny as well as health and disease.

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 below: Gene Expression Omnibus GSE176259, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176259.

AUTHOR CONTRIBUTIONS

BS, WL, and XZ developed the concept of the project and wrote the manuscript. WL performed the experiments. XZ analyzed the data with the help of BS, SZ, GL, and CT. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Shenzhen Key Laboratory of Gene Regulation and Systems Biology (ZDSYS202008

11144002008) and National Natural Science Foundation of China (Grant No. 31900431).

ACKNOWLEDGMENTS

The authors acknowledge the Center for Computational Science and Engineering of SUSTech for the support on computational resource and also acknowledge the SUSTech Core Research Facilities for technical supports.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.717555/full#supplementary-material

REFERENCES

Andergassen, D., Dotter, C. P., Wenzel, D., Sigl, V., Bammer, P. C., Muckenhuber, M., et al. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *Elife* 6:e25125.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593. doi: 10.1126/science.1230612

Calaway, J. D., Lenarcic, A. B., Didion, J. P., Wang, J. R., Searle, J. B., McMillan, L., et al. (2013). Genetic architecture of skewed X inactivation in the laboratory mouse. *PLoS Genet.* 9:e1003853. doi: 10.1371/journal.pgen.1003853

Cedar, H., and Bergman, Y. (2008). Choreography of Ig allelic exclusion. *Curr. Opin. Immunol.* 20, 308–317. doi: 10.1016/j.coi.2008.02.002

Chess, A. (2013). Random and non-random monoallelic expression. *Neuropsychopharmacology* 38, 55–61. doi: 10.1038/npp.2012.85

Chess, A. (2016). Monoallelic gene expression in mammals. *Annu. Rev. Genet.* 50, 317–327.

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

Crowley, J. J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I. K., Kim, Y., et al. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* 47, 353–360. doi: 10.1038/ng.3222

da Rocha, S. T., and Gendrel, A. V. (2019). The influence of DNA methylation on monoallelic expression. *Essays Biochem.* 63, 663–676. doi: 10.1042/ebc20190034

Eckersley-Maslin, M. A., Thybert, D., Bergmann, J. H., Marioni, J. C., Flicek, P., and Spector, D. L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* 28, 351–365. doi: 10.1016/j.devcel.2014.01.017

Emerson, J. J., Hsieh, L. C., Sung, H. M., Wang, T. Y., Huang, C. J., Lu, H. H. S., et al. (2010). Natural selection on cis and trans regulation in yeasts. *Genome Res.* 20, 826–836. doi: 10.1101/gr.101576.109

Gao, Q. S., Sun, W., Ballegeer, M., Libert, C., and Chen, W. (2015). Predominant contribution of cis-regulatory divergence in the evolution of mouse alternative splicing. *Mol. Syst. Biol.* 11:816. doi: 10.15252/msb.20145970

Gendre, A. V., Attia, M., Chen, C. J., Diabangouaya, P., Servant, N., Barillot, E., et al. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* 28, 366–380. doi: 10.1016/j.devcel.2014.01.016

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., et al. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res.* 22, 2376–2384. doi: 10.1101/gr.142281.112

Hou, J., Wang, X., McShane, E., Zauber, H., Sun, W., Selbach, M., et al. (2015). Extensive allele-specific translational regulation in hybrid mice. *Mol. Syst. Biol.* 11:825. doi: 10.15252/msb.156240

Jones, J. (2014). Nonrandom X chromosome inactivation detection. *Curr. Protoc. Hum. Genet.* 80, 9.7.1–7.

King, M. C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116. doi: 10.1126/science.1090005

Krueger, F. (2016). *Trim Galore*. Babraham: Babraham Bioinformatics.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, Y. F., and Sasaki, H. (2011). Genomic imprinting in mammals: its life cycle, molecular mechanisms and reprogramming. *Cell Res.* 21, 466–473. doi: 10.1038/cr.2011.15

Liang, W., Li, G., Cui, H., Wang, Y., Wei, W., Sun, S., et al. (2021). Evolutionary analysis of transcriptional regulation mediated by Cdx2 in rodents. *bioRxiv* [Preprint]. doi: 10.1101/2021.03.01.433326

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

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.

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* 338, 1593–1599. doi: 10.1126/science.1228186

Ohishi, H., Yeung, W. K. A., Unoki, M., Ichiyanagi, K., Fukuda, K., Maenohara, S., et al. (2020). Characterization of genetic-origin-dependent monoallelic expression in mouse embryonic stem cells. *Genes Cells* 25, 54–64. doi: 10.1111/gtc.12736

Orstavik, R. E., Tommerup, N., Eiklid, K., and Orstavik, K. H. (1995). Non-random X chromosome inactivation in an affected twin in a monozygotic twin pair discordant for Wiedemann-Beckwith syndrome. *Am. J. Med. Genet.* 56, 210–214. doi: 10.1002/ajmg.1320560219

Payne, J. L., Khalid, F., and Wagner, A. (2018). RNA-mediated gene regulation is less evolvable than transcriptional regulation. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3481–E3490.

Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., et al. (2011). AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* 7:522. doi: 10.1038/msb.2011.54

Schaefke, B., Emerson, J. J., Wang, T. Y., Lu, M. Y., Hsieh, L. C., and Li, W. H. (2013). Inheritance of gene expression level and selective constraints on trans- and cis-regulatory changes in yeast. *Mol. Biol. Evol.* 30, 2121–2133. doi: 10.1093/molbev/mst114

Schulz, E. G., and Heard, E. (2013). Role and control of X chromosome dosage in mammalian development. *Curr. Opin. Genet. Dev.* 23, 109–115. doi: 10.1016/j.gde.2013.01.008

Thorvaldsen, J. L., Krapp, C., Willard, H. F., and Bartolomei, M. S. (2012). Nonrandom X chromosome inactivation is influenced by multiple regions on the murine X chromosome. *Genetics* 192:1095. doi: 10.1534/genetics.112.144477

Tirosh, I., Reikhav, S., Levy, A. A., and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324, 659–662. doi: 10.1126/science.1169766

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., and Grp, E. I. (2019). Genomic imprinting and physiological processes in mammals. *Cell* 176, 952–965.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature* 430, 85–88. doi: 10.1038/nature02698

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* 40, 346–350. doi: 10.1038/ng.77

Wong, E. S., Schmitt, B. M., Kazachenka, A., Thybert, D., Redmond, A., Connor, F., et al. (2017). Interplay of cis and trans mechanisms driving transcription factor binding and gene expression evolution. *Nat. Commun.* 8:1092.

Xiao, M. S., Zhang, B., Li, Y. S., Gao, Q., Sun, W., and Chen, W. (2016). Global analysis of regulatory divergence in the evolution of mouse alternative polyadenylation. *Mol. Syst. Biol.* 12, 890. doi: 10.15252/msb.20167375

Xu, C., and Zhang, J. (2018). Alternative polyadenylation of mammalian transcripts is generally deleterious, not adaptive. *Cell Syst.* 6, 734–742.e4.

Xu, J., Carter, A. C., Gendrel, A. V., Attia, M., Loftus, J., Greenleaf, W. J., et al. (2017). Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nat. Genet.* 49, 377–386. doi: 10.1038/ng.3769

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. doi: 10.1093/bioinformatics/btv145

Zhang, J. Z. (2018). Neutral theory and phenotypic evolution. *Mol. Biol. Evol.* 35, 1327–1331. doi: 10.1093/molbev/msy065

Zheng-Bradley, X., Rung, J., Parkinson, H., and Brazma, A. (2010). Large scale comparison of global gene expression patterns in human and mouse. *Genome Biol.* 11:R124.

Zou, X., Schaefke, B., Li, Y., Jia, F., Sun, W., Li, G., et al. (2021). *Mammalian Splicing Divergence is Shaped by Drift, Buffering in Trans and a Scaling Law*.

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.

Citation: Liang W, Zou X, Li G, Zhou S, Tian C and Schaefke B (2021) Systematic Analysis of Monoallelic Gene Expression and Chromatin Accessibility Across Multiple Tissues in Hybrid Mice. Front. Cell Dev. Biol. 9:717555. doi: 10.3389/fcell.2021.717555

frontiers in Cell and Developmental Biology

REVIEW
published: 23 September 2021
doi: 10.3389/fcell.2021.740937

X-Chromosome Inactivation and Autosomal Random Monoallelic Expression as "Faux Amis"

Vasco M. Barreto[1], Nadiya Kubasova[1], Clara F. Alves-Pereira[2] and Anne-Valerie Gendrel[3]**

[1] Chronic Diseases Research Centre, CEDOC, Nova Medical School, Lisbon, Portugal, [2] Department of Genetics, Smurfit Institute of Genetics, Trinity College Dublin, University of Dublin, Dublin, Ireland, [3] Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal

X-chromosome inactivation (XCI) and random monoallelic expression of autosomal genes (RMAE) are two paradigms of gene expression regulation where, at the single cell level, genes can be expressed from either the maternal or paternal alleles. X-chromosome inactivation takes place in female marsupial and placental mammals, while RMAE has been described in mammals and also other species. Although the outcome of both processes results in random monoallelic expression and mosaicism at the cellular level, there are many important differences. We provide here a brief sketch of the history behind the discovery of XCI and RMAE. Moreover, we review some of the distinctive features of these two phenomena, with respect to when in development they are established, their roles in dosage compensation and cellular phenotypic diversity, and the molecular mechanisms underlying their initiation and stability.

Keywords: X-chromosome inactivation, random monoallelic expression, epigenetic silencing, LINE-1 elements, cellular diversity, stochasticity, dosage compensation

OPEN ACCESS

Edited by:
Jin Xu,
Sun Yat-sen University, China

Reviewed by:
Bernhard Payer,
Centre for Genomic Regulation (CRG), Spain
Jacqueline Mermoud,
University of Marburg, Germany

****Correspondence:***
Vasco M. Barreto
vasco.barreto@nms.unl.pt
Anne-Valerie Gendrel
anne.gendrel@medicina.ulisboa.pt

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

Received: *13 July 2021*
Accepted: *30 August 2021*
Published: *23 September 2021*

Citation:
Barreto VM, Kubasova N, Alves-Pereira CF and Gendrel A-V (2021) X-Chromosome Inactivation and Autosomal Random Monoallelic Expression as "Faux Amis". Front. Cell Dev. Biol. 9:740937. doi: 10.3389/fcell.2021.740937

INTRODUCTION

In diploid organisms, the two alleles of a gene are usually expressed. However, the expression levels of each allele are not necessarily equal, and allelic imbalance (AI) in transcript levels can occur due to genetic differences in the regulatory sequences or the stability of the transcripts. There are, however, special cases not explained by in *cis* differences in the sequence of the alleles. These have been lumped under the umbrella term "monoallelic expression." In a broad sense, all genetic expression is epigenetic, but if we use a conservative definition of epigenetics to include all heritable (during mitosis or meiosis) changes in gene expression that occur without any changes in the underlying DNA sequence, then monoallelic expression becomes the poster child of epigenetics. Known cases of monoallelic expression include genomic imprinting, X-chromosome inactivation (XCI), and random monoallelic autosomal expression (RMAE). Genomic imprinting affects all cells of an organism the same way, i.e., it is always the same allele that is expressed, depending on the parent of origin (Barton et al., 1984; McGrath and Solter, 1984; Surani et al., 1984). The fate (expression or silencing) is defined during the formation of the gametes in the progenitor. Thus,

Abbreviations: AI, allelic imbalance; IL, interleukin; iPSC, induced pluripotent stem cell; LINE-1, long interspersed nuclear element-1; OR, olfactory receptor; PAR, pseudoautosomal region; RMAE, random monoallelic autosomal expression; SNP, single nucleotide polymorphism; TCR, T-cell receptor; Xa, active X; XCI, X-chromosome inactivation; Xi, inactive X.

despite the associated fascinating molecular mechanisms, evolutionary theory (Haig, 1993), and relevance for development and diseases (Ferguson-Smith and Bourc'his, 2018), genomic imprinting is merely a case of transgenerational gene expression that is reset each generation during the formation of the oocyte and sperm cells. XCI and RMAE differ from genomic imprinting because they give rise to mosaicism: in the same organism, some cells express the maternal allele and other cells express the paternal allele. Over the last decades, this common feature has recurrently tempted many to draw parallels between XCI and RMAE, both in reviews or opinion pieces [e.g., (Efstratiadis, 1995; Goldmit and Bergman, 2004; Chess, 2016; Gendrel et al., 2016)] and original articles [e.g., (Mostoslavsky et al., 2001; Pereira et al., 2003)]. But much like the confusion created by false cognates or "faux amis" between two languages, the parallels between two phenomena often prevent us from seeing the obvious and meaningful differences; parallels can illuminate but also deceive. Thus, here we propose to critically evaluate the relevance of the parallels drawn between XCI and RMAE, and expose their key differences at the cellular and molecular levels.

HISTORICAL BACKGROUND

XCI and RMAE were described in the same decade. X-chromosome inactivation, also named "Lyonisation," was first proposed in 1961 by mouse geneticist Mary Lyon in a short report with no figures, where she laid the fundamental principles of XCI based solely on mouse genetics and earlier cytological evidence (Lyon, 1961). A few years later, individual B cells were shown to express only one immunoglobulin allele, both for the heavy and the kappa chains (Cebra and Goldstein, 1965; Pernis et al., 1965), two autosomal genes. Retrospectively, these three papers were seminal, but Lyon's work immediately created a new field, whereas the RMAE of antigen receptors remained essentially a pet subject for a niche of scientists.

Mary Lyon was examining the inheritance and the phenotype of different mutations in X-linked genes affecting coat color in mice. She observed that heterozygous females had mosaic or variegated phenotypes, with patches of normal and mutant color, unlike males. This, coupled with the knowledge that female mice with only one X chromosome were viable and fertile (Welshons and Russell, 1959) and that female cells exhibit one condensed X chromosome in their nuclei (Ohno et al., 1959; Ohno and Hauschka, 1960), led her to the XCI hypothesis. The key principles underlying this hypothesis were the genetic inactivation of the X chromosome of either paternal or maternal origin, the early inactivation during embryogenesis, and the clonal inheritance of the inactive state through cell division (Lyon, 1961). Soon after, other scientists correlated the genetic observations made by Mary Lyon with experimental studies, such as the presence of two red blood cell populations or protein variants associated with mutations in the *G6pd* X-linked gene in female cells (Beutler et al., 1962; Davidson et al., 1963). In 1962, Mary Lyon published a much longer report focusing on human X-linked syndromes, providing evidence that XCI is present in other mammals, such as humans, and is the basis for dosage compensation between the sex chromosomes (Lyon, 1962). XCI is still often referred nowadays to as the "Lyon hypothesis," although it should be considered as a fully established law (Gendrel and Heard, 2011).

The finding that the immunoglobulin chains are expressed monoallelically at the cellular level predates the discovery of the mechanism of V(D)J recombination that sets apart the antigen receptor genes (Hozumi and Tonegawa, 1976), including the immunoglobulin and T-cell receptor genes. Over the years, it was found that monoallelic expression – more commonly described in this literature as "allelic exclusion" – is a feature of most antigen receptor genes only partly explained by the relatively high frequency of non-productive sequences (with frameshifts leading to premature stop codons) generated by V(D)J recombination and that the percentage of cells with monoallelic expression varies considerably depending on the antigen receptor gene [reviewed in Vettermann and Schlissel (2010)].

The collection of genes under allelic expression expanded beyond the antigen receptor genes only in the 1990s. The olfactory receptor (OR) genes form the largest gene family in mammals; in the mouse, there are 1,296 OR genes (Zhang and Firestein, 2002). Remarkably, each neuron expresses only one gene and, taking advantage of the OR gene polymorphisms found in *Mus musculus* x *M. spretus* F1 mice, it was found that, in a given neuron, only one of the two alleles of the chosen expressed gene is transcribed (Chess et al., 1994). Soon after this finding, using allele-specific antibodies, the *Ly49* genes of natural killer cells were also shown to display a monoallelic expression pattern in mice (Held et al., 1995) and a few years later, a pheromone receptor, similar to the OR genes, was shown to display RMAE in the neurons of the accessory olfactory system (Rodriguez et al., 1999).

The end of the 1990s would mark the beginning of a short controversy on the expression patterns of the interleukins (ILs). In T cells, *IL-2* was reported to be monoallelically expressed in T cells (Hollander et al., 1998), whereas for *IL-4* the expression was described as biallelic or monoallelic, depending on the clone (Bix and Locksley, 1998) or the strength of the signal delivered through the TCR (Rivière et al., 1998). Four subsequent studies on *IL-2* reached different conclusions: whereas *IL-2* was found to be biallelically expressed in human T cell clones (Chiodetti et al., 2000; Bayley et al., 2003), in mice heterozygous for an *IL-2-GFP* transgene (Naramura et al., 1998) and in single-cell RT-PCR experiments (Rhoades et al., 2000), most cells expressed both *IL-2* alleles but there were also single expressors. Overall, the data for *IL-2* seem consistent with the data for *IL-4*: under optimal and continuous stimulation, cells will tend to express both alleles, but at suboptimal levels of expression, there may be cells expressing only one of the alleles.

The list of genes under monoallelic expression grew slowly until the mid-2000s based on additional reports focused on single genes [e.g., (Endo et al., 1995); **Table 1**], but in 2007, the application of genome-wide transcriptomics to a collection of human lymphoblast clonal cell lines revealed that over 5% of expressed genes display patterns of random monoallelic

TABLE 1 | List of autosomal genes under random monoallelic expression reported in studies focused on single genes.

Gene	Cell type/tissue	Species	*In vitro/in vivo*	Year	References
Immunoglobulin receptor genes	B and T lymphocytes	rabbit mouse	*in vivo*	1965 1976 1985	Cebra and Goldstein, 1965; Pernis et al., 1965; Hozumi and Tonegawa, 1976; Goverman et al., 1985
Olfactory receptor (OR) genes	sensory neurons	mouse	*in vivo*	1994	Chess et al., 1994
HUMARA (human androgen receptor) gene	colonic crypts	human	*in vivo*	1995	Endo et al., 1995
Ly49 receptor genes	natural killer cells	mouse	*in vivo*	1995	Held et al., 1995
Interleukin genes (*IL2*, *IL4*, *IL5*, *IL10*, *IL13*)	T cells	mouse	*in vitro*	1998, 2000, 2006	Bix and Locksley, 1998; Hollander et al., 1998; Kelly and Locksley, 2000; Calado et al., 2006
Pax5	early progenitors and mature B cells	mouse	*in vitro*	1999	Nutt et al., 1999
VRi2	sensory neurons of the vomeronasal system	mouse	*in vivo*	1999	Rodriguez et al., 1999
Nubp2, Igfals, and *Jsap1*	bone marrow stromal cells and hepatocytes	mouse	*in vitro*	2001	Sano et al., 2001
Variable lymphocyte receptors (VLRs) genes	lymphocytes	lamprey	*in vivo*	2004	Pancer et al., 2004
Protocadherin genes	Purkinje cells	mouse human	*in vitro/in vivo*	2002 2005 2006	Tasic et al., 2002; Wang et al., 2002; Esumi et al., 2005; Kaneko et al., 2006
Tlr4	B cells	mouse	*in vitro*	2003	Pereira et al., 2003
KIR genes	natural killer cells	human	*in vitro*	2003	Chan et al., 2003
Cd4	CD4 + lymphocytes	mouse	*in vitro*	2004	Capparelli et al., 2004
p120 catenin	pre-B clonal cell lines lymphoblastoid lines	mouse human	*in vitro*	2005	Gimelbrant et al., 2005
Gfap (*glial fibrillary acidic protein*)	cortical astrocytes	mouse	*in vitro*	2008	Takizawa et al., 2008
rDNA loci	lymphoblasts	human	*in vitro*	2009	Schlesinger et al., 2009
Krt12	limbal stem cells	mouse	*in vivo*	2010	Hayashi et al., 2010
IGF2BP1	B cells	human	*in vitro*	2011	Thomas et al., 2011
ASAR6	P175 cell line (derived from HTD114 fibrosarcoma cell line)	human	*in vitro*	2011	Stoffregen et al., 2011
Cubilin	renal proximal tubules and small intestine	mouse	*in vivo*	2013	Aseem et al., 2013
ASAR15	P268 cell line (derived from HTD114 fibrosarcoma cell line)	human	*in vitro*	2015	Donley et al., 2015
Gata3	hematopoietic stem cells and early T-cell progenitors	mouse	*in vitro/in vivo*	2015	Ku et al., 2015
FOXP2	B lymphoblastoid cell lines and clonal T-cell lines	human	*in vitro/in vivo*	2015	Adegbola et al., 2015
Bcl11b	T cells	mouse	*in vitro/in vivo*	2018	Ng et al., 2018

expression across the collection of clones (Gimelbrant et al., 2007). Over the subsequent years, several independent reports on clonal cell lines confirmed that the number of genes under random monoallelic expression was higher than previously thought (**Table 2**).

Technological progress allowing transcriptomics at the single-cell level revealed that stochastic bursts of transcription occurring independently at the allelic level may lead to the presence of RNA from only one of the alleles at a given time (Deng et al., 2014). We will not cover these cases because RMAE due to transcriptional bursting is not expected to be stable over time and is not observed at the clonal level.

HOMOGENEOUS VERSUS HETEROGENEOUS PHENOMENA

XCI is usually perceived as a homogeneous process. It affects an entire chromosome leading to silencing of nearly all genes, with a few notable exceptions, and therefore sets the basis for a

TABLE 2 | A summary of reports based on genome-wide transcriptomics analysis in different cell types.

Cell type	Experimental assay	Species	Genotypes	% of RMAE	Number of clones analyzed	References
Lymphoblastoid cells (*in vitro*)	SNP-sensitive microarrays	Human	NA	5-10	12	Gimelbrant et al., 2007
		Mouse	129S X CAST; Balb/c X C57BL/6J	15.6	11	Zwemer et al., 2012
Fibroblasts (*in vitro*)	SNP-sensitive microarrays	Mouse	129S X CAST	2.1	2	Zwemer et al., 2012
	RNA-seq	Mouse	CAST X 129S	0.52-1.9	6	Pinter et al., 2015
Neural stem cells (*in vitro*)	SNP-sensitive microarrays	Human	NA	1.4-2.0	9	Jeffries et al., 2012
	RNA-seq	Mouse	C57BL/6 X JF1	2.4	4	Li et al., 2012
	SNP-sensitive microarrays	Human	NA	0.63	3	Jeffries et al., 2016
	RNA-seq	Mouse	C57BL/6 X JF1	4.6	4	Branciamore et al., 2018
Neural progenitor cells from embryonic stem cells (*in vitro*)	RNA-seq	Mouse	C57BL/6 X CAST	3.0	6	Eckersley-Maslin et al., 2014
			129S X CAST	2.5	8	Gendrel et al., 2014
Embryonic stem cells (*in vitro*)	RNA-seq	Mouse	C57BL/6 X CAST	0.5	6	Eckersley-Maslin et al., 2014
iPSC (*in vitro*)	SNP-sensitive microarrays	Human	NA	0.88	2	Jeffries et al., 2016
Neural stem cells from iPSC (*in vitro*)	SNP-sensitive microarrays	Human	NA	0.65-0.84	2	Jeffries et al., 2016
Astrocyte-like cells (*in vitro*)	RNA-seq	Mouse	C57BL/6 X JF1	6.4	4	Branciamore et al., 2018

robust monoallelic expression of these genes. The inactivation is established in all cells in a random manner early during embryogenesis and is then stably inherited during mitotic cell divisions throughout life; all cells therefore carry an inactive X (Xi) and active X (Xa) chromosome, and females are mosaic individuals with cell populations expressing genes from either the maternal or the paternal X chromosome (**Figure 1**). Most genes that are subject to XCI stay stably repressed during development and adulthood, and rarely become biallelically expressed, except under specific circumstances discussed below (Galupa and Heard, 2018).

In clear contrast with XCI, which has defined physical boundaries (the X chromosome), a precise timing during development, a phylogenetic association with female marsupial and placental mammals, and a master player (the *Xist* long non-coding RNA), genes under RMAE are scattered throughout autosomal chromosomes, are expressed at different times and in different tissues, can be biallelically expressed and are found in animals other than mammals, including jawless vertebrates (Pancer et al., 2004), trypanosomes (Borst, 2002), and perhaps even in diatom algae (Hoguin et al., 2021) (**Figure 1**). Additionally, there is no evidence that they are regulated by a single factor and no clear molecular signature that could suggest the existence of a common mechanism regulating RMAE has emerged. Although the first examples of genes under RMAE were of cell surface receptors, which remain an over-represented class, this group is diverse in terms of function.

TIMING

XCI is initiated during early embryogenesis in mammals. In mice, random XCI starts around the time of implantation (E5.5) and is complete by E6.5 in all cells in embryonic tissues (Mak et al., 2004). In humans, the timing of XCI has been complicated to address owing to obvious difficulties to access relevant material. Nevertheless, studies showed that random XCI is initiated around the implantation stage and, compared to mice, appears to be a much more gradual process during the first four weeks of embryonic development (Tang et al., 2015; Zhou et al., 2019). This process is absolutely essential, as failure to induce XCI results in early embryonic lethality at around day 10 of development for mouse embryos (Takagi and Abe, 1990).

Whereas XCI is established early in development, RMAE can occur early or late, depending on when the gene is first expressed. For instance, in the case of OR genes, the critical events leading to RMAE, namely the stochastic expression of a single OR allele from a pool of OR genes silenced with heterochromatic marks, occurs in the maturing olfactory sensory neurons of the mouse olfactory epithelium (Magklara et al., 2011). However, for most genes under RMAE, it is unknown if the choice of which allele to express is pre-determined in progenitor cells long before the gene becomes expressed. In 2001, that possibility was suggested for the antigen receptor genes based on the finding that these genes replicate asynchronously (Mostoslavsky et al.,

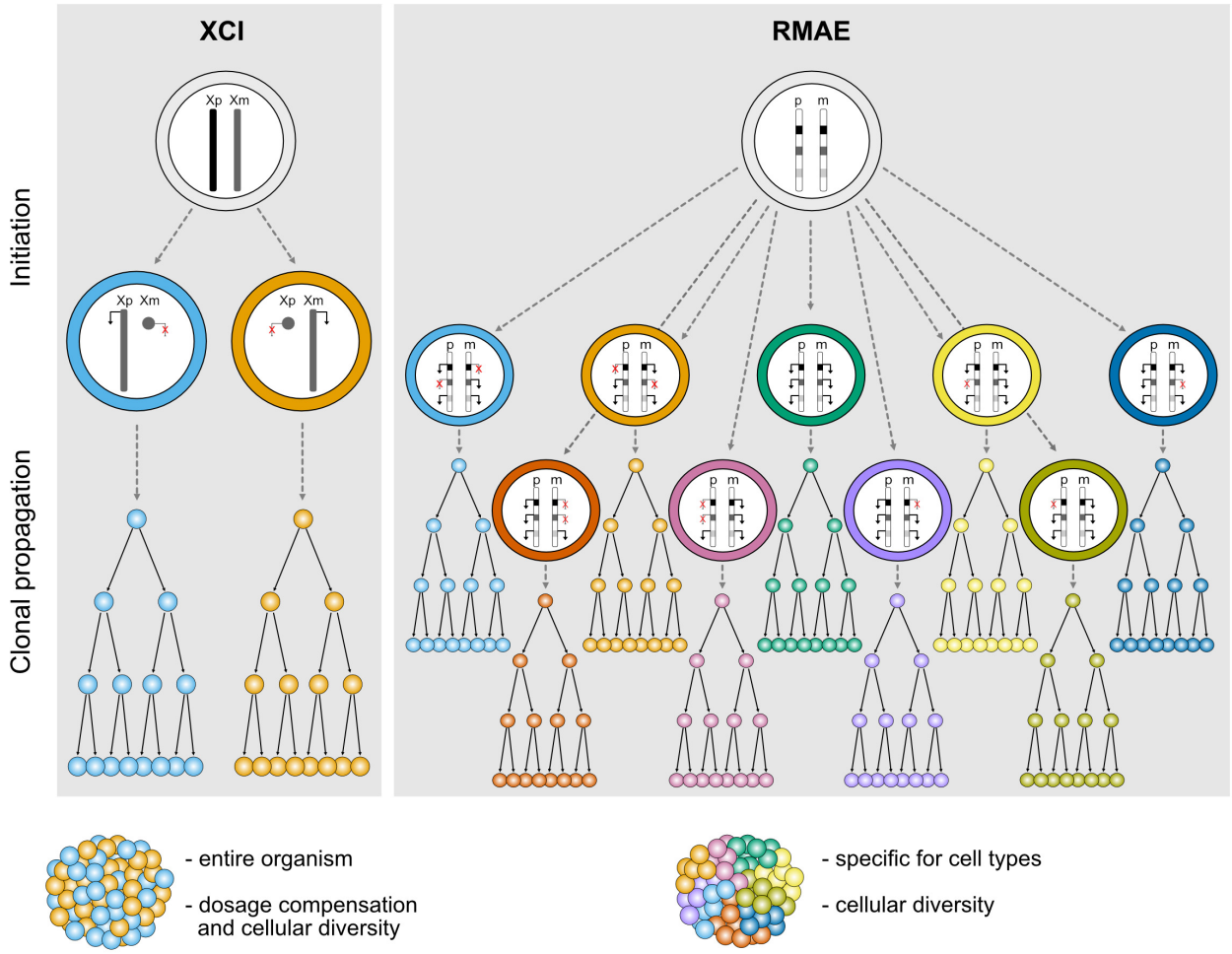


FIGURE 1 | Schematic representation illustrating the features of different types of random monoallelic expression: X-chromosome inactivation (XCI) and random monoallelic autosomal expression (RMAE). Xp, X chromosome of paternal origin, Xm, X chromosome of maternal origin; p, paternal autosome, m, maternal autosome.

2001). Asynchronous replication is a feature of monoallelically expressed genes because, typically, transcribed genes (or alleles) undergo replication before silenced genes (or alleles). The authors drew parallels between XCI and the pattern of asynchronous replication in autosomal genes in their study. In both processes, the epigenetic marks are erased in the morula, re-established around the time of implantation randomly and then clonally maintained. However, in subsequent publications from the same group and others, these parallels fell apart (Farago et al., 2012; Alves-Pereira et al., 2014). Notably, in mice reconstituted with a single hematopoietic stem cell, it was shown that the immunoglobulin heavy chain alleles rearrange independently, i.e., without any evidence for an epigenetic mark (Alves-Pereira et al., 2014). Whether such mark is eventually established later in development and before V(D)J rearrangement is an open question. In the kappa light chain, this pre-determination has been proposed (Farago et al., 2012), but for the heavy chain no evidence was found (Alves-Pereira et al., 2014). In any case, the advantage of such clonal pre-determination long before the genes are expressed is not obvious.

ROLE

The main purpose of XCI is to enable dosage compensation of X-linked genes products to correct for the imbalance between XX females and XY males in mammals (Graves, 2016). The lethality resulting from failure of XCI is a consequence of the absence of dosage compensation. However, it remains unclear whether dosage compensation is critical for all X-linked genes or only a fraction of them. It is also not known whether compensation of dosage-sensitive genes is necessary in all tissues and all developmental stages. Transcriptome analysis of pre-implantation embryos and differentiating embryonic stem cells indicate that absence of XCI leads to failure to exit the pluripotent state, aberrant expression of extra-embryonic factors, and inappropriate expression of developmental genes,

which leads to compromised development and differentiation, hence early lethality (Schulz et al., 2014; Chen et al., 2016; Borensztein et al., 2017).

Besides dosage compensation, XCI is also able to generate a significant level of biological diversity both within and between female individuals (**Figure 1**). This was exemplified by a study that built topographic maps of XCI mosaicism at single cell resolution, using female mice carrying X-linked fluorescent reporters on each X chromosome (Wu et al., 2014). The authors observed that some organs or tissues are particularly prone to deviations from the expected 50:50 inactivation ratio. In particular, the brain stands out as one organ where the diversity conferred by XCI could have an important functional impact for the stimulus-response amplitude of neuronal networks, especially when considering heterozygosity for an X-linked gene expressed in the brain (Wu et al., 2014). Given the level of genetic variation on the human genome including the X chromosome, XCI could generate a remarkable level of intra- and inter-individual differences in the human central nervous system.

RMAE is thought to have evolved exclusively to increase the biological (or phenotypic) diversity at the cellular level. Assuming polymorphisms within a gene, heterozygous cells with a biallelic pattern of expression will be phenotypically identical, whereas partial RMAE will produce three types of cells within the organism: single paternal allele-expressing cells, single maternal-expressing cells, and biallelic expressing cells (**Figure 1**). The most exuberant cases of phenotypic diversity are found in the OR and antigen receptor genes. In the former, RMAE is coupled with the selection of a single gene from the largest gene family for expression at the single-cell level, leading to the generation of hundreds or thousands of different sensory olfactory neurons, depending on the species (Niimura et al., 2014). In the latter, thousands of allelic forms are generated during the organism's life by V(D)J recombination that will be monoallelically expressed to ensure the single cell-single receptor rule, thus facilitating the processes of negative and positive selection that shape the immune repertoires. The potential phenotypic diversity for the average gene served by only two alleles is much lower than that of OR and antigen receptor genes, but for phenotypes determined by multiple genes under RMAE there is a considerable combinatory potential (3^n phenotypes, where n is the number of relevant polymorphic genes under RMAE). However, outside of OR and antigen receptor genes, the importance of RMAE-driven phenotypic diversity remains to be demonstrated, and it is a complicated problem to tackle experimentally. As explained below, particular *cis*-regulatory sequences play a role in RMAE. Thus, a feasible approach would be to replace these sequences with regular promoters, but even in this case the interpretation of the data would not be clear-cut because the expression of multiple receptors at the surface of the cell would decrease the density of any particular receptor compared to its density in a cell with RMAE. Since the manipulation of master epigenetic regulators is unlikely to be sufficiently specific, a proof of principle will probably be obtained using CRISPR/Cas tools that allow epigenetic manipulations at the allele-specific level. An alternative approach would be to generate aggregation chimeras of cells expressing different receptors or other relevant proteins for the quantitative response to be tested. In the absence of such data, other hypotheses can be raised, such as a role for RMAE in dosage compensation (Gendrel et al., 2016). However, the finding that genes under RMAE have increased genetic diversity (polymorphisms) in humans compared to biallelically expressed genes remains a powerful indication that RMAE evolved to increase phenotypic diversity at the cellular level (Savova et al., 2016).

MOLECULAR MECHANISMS

Stochasticity

One has to recognize that stochasticity is the key feature common to XCI and RMAE: how come identical or quasi-identical sequences (the X chromosomes or autosomal alleles) sharing the same nuclear environment undergo completely opposite fates (expression or silencing)? This stochastic component is at the core of the appeal these phenomena have to biologists, but there is a critical difference between XCI and RMAE worth mentioning. It has been proposed that each individual X chromosome has an independent probability to be inactivated that is directly proportional to the X: ploidy ratio. Selection in favor of cells keeping one active X chromosome per diploid genome eliminates cells with two inactive X or two active X chromosomes (Monkhorst et al., 2008; Sousa et al., 2018). Thus, XCI involves the inactivation of one X chromosome and counterselection at the cellular level. In contrast, the stochastic component in OR genes, antigen receptor genes, and possibly other autosomal genes under RMAE involves the activation of alleles in a default state of silencing, and cell counterselection is not thought to play a major role in shaping the pattern of monoallelic expression; in fact, B lymphocytes genetically engineered to express two different immunoglobulin heavy chains at the surface were shown to be fit and able to generate a normal B cell compartment (Sonoda et al., 1997).

Feedback

A key aspect of RMAE in antigen receptor genes is the feedback mechanism that prevents the recombination of the second allele once the protein encoded by the first allele to rearrange productively is expressed at the surface. When the exon encoding the transmembrane domain of the immunoglobulin chain is disrupted, the cell is no longer able to trigger this feedback mechanism and the second allele is given the chance to recombine (Kitamura and Rajewsky, 1992). A similar mechanism has been described for the beta chain of the TCR gene (Aifantis et al., 1997) and the OR genes (Serizawa et al., 2003). The overall picture, then, is the coupling of a stochastic process of gene activation that is sufficiently slow for negative feedback mechanisms to act, preventing further rearrangements (antigen receptors) or gene activation (OR genes). Because the feedback mechanism implies a time-window during which the two alleles can be activated, it also explains the generation of biallelic cells; a slow feedback mechanism will produce many biallelic cells, whereas biallelic cells are rare when the time-window is narrow. However, it is not clear whether such feedbacks are involved for

other genes under RMAE and additional mechanisms have been described, which we discuss below.

Epigenetics

The process of XCI can be divided in two distinct stages: initiation and maintenance. During the initiation phase, XCI is dependent on the expression of the long non-coding RNA *Xist*, which induces transcriptional silencing in *cis* and ultimately coats the entire inactive X chromosome (Loda and Heard, 2019). However, *Xist* is no longer essential for the maintenance of XCI, as deletion of *Xist* in somatic cells in culture does not lead to Xi reactivation (Brown and Willard, 1994). Following *Xist* accumulation on the Xi, one of the first observable events is the formation of a 3D silent nuclear compartment excluding RNA polymerase II and transcription factors, likely to be important for *Xist* spreading and the initiation of gene silencing (Chaumeil et al., 2006; Clemson et al., 2006; Pandya-Jones et al., 2020). *Xist* interacts with several RNA-binding proteins, in particular SPEN, which acts as a bridge between *Xist* RNA and repressor complexes that mediates the removal of histone modifications associated with active genes (H3K27ac, H3K9ac, H4ac), a crucial early step for the initiation of gene silencing (Żylicz et al., 2019; Dossin et al., 2020). Following this, a number of chromosome-wide chromatin changes occur on the Xi to lock in the silenced state, such as deposition of repressive histone modifications (H2AK119Ub and H3K27me3) mediated by the Polycomb repressive complexes 1 and 2 [reviewed in Boeren and Gribnau (2021)]. The late or maintenance phase is characterized by a switch to late replication timing, incorporation of the histone variant macroH2A and DNA methylation of X-linked gene promoter regions by the DNA methyltransferase Dnmt3b [reviewed in Strehle and Guttman (2020)]. These changes ensure the stable and heritable silencing of the majority of genes on the Xi, over hundreds of cell divisions.

By definition, epigenetic modifications, namely histone modifications and DNA methylation, and non-coding RNAs, have been shown to be associated with genes under RMAE (Gendrel et al., 2016). However, unlike XCI, there is no master regulator, and several scenarios have been reported, such the initial repression of both alleles followed by activation (e.g., OR genes and murine *Ly49* genes) or the initial activation of both alleles followed by the inactivation of one allele (e.g., human *KIR* genes).

Long Interspersed Nuclear Element-1

One puzzling question in XCI has been the nature of the X-linked *cis*-acting elements important for the binding and spreading of *Xist* along the X chromosome, prior to gene silencing. Because of the higher density of long interspersed nuclear element-1 (LINE-1) retrotransposons in the X chromosome compared to autosomes (Boyle et al., 1990; Ross et al., 2005), with her so-called "repeat hypothesis," Mary Lyon postulated that these sequences could act as booster elements for the spreading of the inactive signal along the chromosome and efficient silencing (Lyon, 1998). However, we now know that *Xist* does not bind directly LINE-1 sequences nor associate with LINE-1-enriched regions. *Xist* rather exploits the 3D conformation of the X chromosome to spread first to sites that are spatially proximal to the *Xist* gene at the onset of XCI and is then found enriched over gene-dense regions that are depleted of LINE-1 sequences (Engreitz et al., 2013; Simon et al., 2013). Yet, studies of *Xist* spreading on autosomal chromatin in X:autosome translocations (Sharp et al., 2002; Popova et al., 2006) or using *Xist* transgenes on autosomes (Chow et al., 2010; Tang et al., 2010; Loda et al., 2017) all show a good correlation between LINE-1 density, efficiency of spreading, and gene silencing. These observations suggest that LINE-1 elements may contribute to the process of XCI, either by facilitating gene silencing locally in some regions, reinforcing the long-term maintenance of XCI and/or influencing heterochromatin reorganization. This, however, remains to be formally tested using, for example, functional approaches to perturb LINE-1 expression or enrichment on the X chromosome.

Whether genes under RMAE have a DNA sequence signature remains unclear. It has been proposed that these genes are surrounded by an increased density of LINE-1 elements, which are evolutionarily more recent and less truncated than the LINE-1 elements around biallelically expressed genes (Allen et al., 2003). How exactly LINE-1 elements could contribute to RMAE is not known, but their association to RMAE would be one of the few potential parallels with XCI. However, the overlap between predicted RMAE genes based on the presence of LINE-1 elements and the collection of genes under RMAE generated by a genome-wide approach is not statistically significant (Allen et al., 2003; Gimelbrant et al., 2007).

It has been known for decades that the X chromosome and autosomal genes under RMAE replicate asynchronously (Taylor, 1960; Chess et al., 1994) and that this mitotically stable pattern is established early in development. Asynchronous replication was even found to be a property of autosomal chromosomes (Singh et al., 2003), reinforcing the parallel with the X chromosome. Whether the asynchronous replication of autosomes is absolutely stable is not clear, and it has been found that RMAE is not coordinated at the chromosome level, i.e., the alleles from different genes under RMAE on the same chromosome can be active or silent (Gimelbrant et al., 2007). However, an autosomal gene named asynchronous replication and autosomal RNA on chromosome 6 (*ASAR6*) was shown to encode a non-coding RNA under RMAE, which when expressed leads to the silencing of nearby alleles and remains associated with the chromosome from which it is expressed. Moreover, the disruption of this locus results in delayed replication timing and reactivation of previously silent alleles of nearby genes (Stoffregen et al., 2011; Donley et al., 2013). There is an obvious parallel with *XIST*, which is also monoallelically expressed, silences most of the genes on the X chromosome in *cis* and, when deleted, also alters replication timing (Diaz-Perez et al., 2005). Interestingly, a LINE-1 retrotransposon located within *ASAR6*, in antisense orientation, was then found to control the replication timing (Platt et al., 2018). This constitutes one of the most solid evidence that LINE-1 elements could be involved in the spreading of inactivation also on autosomal chromosomes. Another locus, *ASAR15*, displays features similar with *ASAR6* (Donley et al., 2015). However, it is not known how frequent this type of regional silencing occurs on autosomes, and, unlike *XIST*, the *ASAR6* RNA does not

seem to coat the entire chromosome 6 and is not expressed in adult tissues (Stoffregen et al., 2011). A thorough and granular reappreciation of the impact of LINE-1 elements in RMAE would be welcomed.

Bivalent Promoters

Several histone modifications influence gene expression, including H3K4me3 and H3K27me3, which are associated with gene activation and repression, respectively. Although active and repressive histone marks are typically imagined as being mutually exclusive, in 2006 two groups reported the existence of regulatory regions – named bivalent domains – that have both (Azuara et al., 2006; Bernstein et al., 2006). Genes with bivalent promoters in embryonic stem cells are expressed at low levels but thought to be poised for rapid activation upon differentiation cues. Interestingly, about 80% of the genes under RMAE in differentiated cells, identified by transcriptomics or the presence of activation and repression histone marks on different alleles, were found to have bivalent promoters in precursor cells (Nag et al., 2013) (**Figure 2**). Thus, the rapid and timely activation ensured by bivalent promoters seems to increase the probability of RMAE, as if the alleles resolve their status stochastically, leading to cells that activate only the paternal or maternal allele and cells that activate both (Nag et al., 2013).

Bidirectional Promoters and Other Switches

Two cases of RMAE dissected in considerable detail are the murine *Ly49* receptor genes of natural killer cells and the human *KIR* genes [reviewed in Anderson (2014)]. Both rely on *cis* probabilistic bidirectional promoter switches that produce sense and antisense transcripts associated with expression and silencing, respectively (**Figure 2**). In the case of the murine *Ly49* receptor genes, the default condition is silencing. Transcription starts if the sense non-coding transcripts of a distal bidirectional promoter (Pro1) activate a downstream promoter (Pro2); the antisense transcripts of the bidirectional promoter do not lead to gene activation (Saleh et al., 2002, 2004). In the case of the *KIR* genes, the default condition is activation and the role of the stochastic switch, located close to the ATG start codon, is to produce a sense transcript that correlates with the maintenance of the activation state or an antisense piRNA that silences the allele. The murine *Ly49* and the human *KIR* genes illustrate how a probabilistic bidirectional promoter can create a mitotically stable asymmetry between two alleles. Whether these are two exceptional cases or examples of a frequent solution to generate RMAE has not been addressed. Notably, divergently transcribed gene pairs represent more than 10% of the human genes (Trinklein et al., 2004; Yang et al., 2007), and thorough

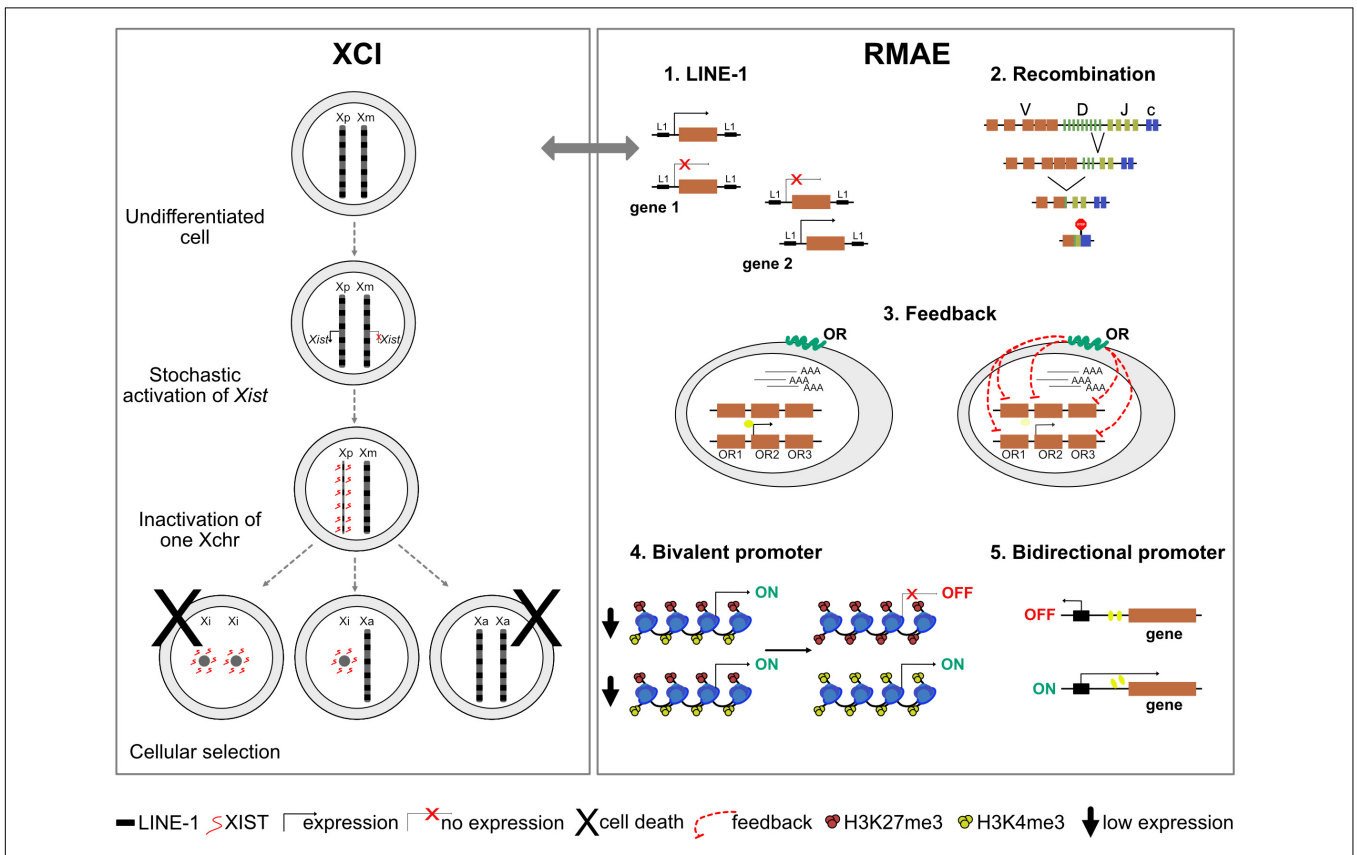


FIGURE 2 | A schematic representation of mechanisms responsible for X-chromosome inactivation (XCI) and mechanisms possibly responsible for random monoallelic autosomal expression (RMAE). The gray arrow represents a potential parallel between XCI and RMAE associated with LINE-1 (L1) elements.

analyses of tissue-specific sense and antisense transcripts from the same locus [e.g., (Hu et al., 2014)] could reveal additional candidate genes to be under RMAE due to bidirectional promoter or other complex arrangements of regulatory sequences leading to genetic switches.

STABILITY

Once established, XCI is believed to be extremely stable and irreversible. Genes that are subject to XCI rarely show reactivation and biallelic expression, as silencing is maintained through multiple layers of epigenetic control. However, there are some exceptions and some genes can be expressed from both the Xa and the Xi. This is the case for genes that have a Y-linked homolog, including genes from the pseudoautosomal regions (PAR1 and PAR2, short regions of homology between the X and Y chromosomes, which undergo recombination during meiosis), for which there is no requirement for dosage compensation. Several genes not located in the PAR regions have also retained a functional Y paralog and would thus appear not to require dosage compensation. However, other genes do not have a Y-linked copy yet still have the ability to escape XCI (Berletch et al., 2010). In some cases, this may be due to a highly controlled process permitting escape where the gene product is required in increased dose, while in other cases, it may be due to leaky or inefficient XCI.

Escape from X inactivation is rather limited in the mouse, with around 3% of genes displaying such behavior in somatic cells (Yang et al., 2010; Pinter et al., 2012). In humans, the situation is different, as 15% of genes (excluding the PAR) have been reported to escape XCI (Carrel and Willard, 2005). Intriguingly, an additional 10% of X-linked genes appear to show heterogeneous inactivation and escape, varying between lineages and from one individual to another (Carrel and Willard, 2005). Such candidates seem to display accessible promoter regions on the Xi (Kucera et al., 2011), suggesting that they may be poised for expression in some cell lineages and that the Xi allele becomes active under specific circumstances. In the mouse, lineage-specific escape has also been found, for example in the case of the *Atrx* gene, which is fully inactivated in embryonic tissues but escapes inactivation in specific subsets of extraembryonic cells (Garrick et al., 2006). The Atrx protein is actually enriched on the Xi in extraembryonic tissues (Baumann and De La Fuente, 2009; Sarma et al., 2014), suggesting that its escape from XCI might occur in a regulated manner in tissues where a higher dose of the protein is necessary (Corbel et al., 2013).

Interestingly, some of the phenotypes observed in Turner (X0) syndrome patients are believed to be due, in part, to the reduced expression levels of escapees given the lack of the second X chromosome (Berletch et al., 2010). This indicates that expression of a double dose is essential for some X-linked genes and that escape for these genes is a highly controlled process. Escape or reactivation of genes from the Xi can also occur more sporadically, but it is currently unknown whether this is caused by inefficient XCI or associated with a controlled mechanism.

Sporadic reactivation of genes from the Xi has been observed in non-pathological contexts, in specific tissues (Gendrel et al., 2014) or lineages (Wang et al., 2016), during aging (Migeon et al., 1988; Sharp et al., 2000) and also in disease contexts (Youness et al., 2021). In normal contexts, both the brain and the lymphoid lineage appear to stand as exceptions. In the brain, the *Mecp2* gene, which is associated with Rett syndrome, was shown to display biallelic expression in a significant proportion of neural stem cells in the subventricular zone in the neonatal brain of inbred female mice (Gendrel et al., 2014). This could be indicative of a certain relaxation of epigenetic control of the Xi in these cells at least for this gene or a need for an increased dose of the protein, given that MeCP2 is a highly abundant protein in the brain (Skene et al., 2010). Moreover, *Xist* conditional deletion in adult mice leads to a global erasure of repressive histone modifications from the Xi, and more importantly, a mild loss of dosage compensation in 2-5% of neurons (Adrianse et al., 2018), highlighting again the peculiarity of the brain and neuronal lineages. Other studies have also reported partial Xi reactivation following *Xist* conditional deletion in adult tissues (Yildirim et al., 2013; Yang et al., 2016, 2020).

In the female lymphoid lineage, the maintenance of XCI is atypical and it has been hypothesized that this could predispose females to autoimmunity (Wang et al., 2016; Syrett et al., 2019). It was shown that both human and mouse naive B and T lymphocytes miss the typical *XIST*/*Xist* RNA domain within the nuclei. Instead, *Xist* shows an unusual and dispersed pattern, associated with a structure lacking some of the canonical hallmarks of heterochromatin of the Xi, such as H3K27me3, H2AK119ub1 and macroH2A. However, this state appears transient as both *Xist* and repressive histone marks are relocalized to the Xi upon B/T cell activation (Savarese et al., 2006; Wang et al., 2016; Syrett et al., 2019). This state was shown to be correlated with modest biallelic expression and increased expression of X-linked immunity-related genes (Wang et al., 2016; Souyris et al., 2018). The role of reactivation/increased expression of these genes and whether this is a cause of the atypical maintenance of XCI during lymphocyte differentiation remain unclear. However, enhanced expression of these genes could contribute to higher susceptibility of females to autoimmune disorders, such as systemic lupus erythematosus if not properly regulated (Youness et al., 2021).

Concerning RMAE, by definition all the examples of monoallelic expression discovered in clonal cell lines are mitotically stable. However, most reports have focused on cells that are also phenotypically stable, i.e., cells that during the period of the study do not go through major steps of differentiation. Thus, it remains to be addressed if RMAE is as stable as XCI, which is known to keep the status of the X chromosomes established early in development even after hundreds of cell divisions and extensive differentiation. Unfortunately, the antigen receptor and OR genes, which have been thoroughly investigated over decades, do not shed much light on this issue. The monoallelic expression pattern of antigen receptor genes is in part established by the process of V(D)J recombination and in developing lymphocytes, when recombination is active, the second allele is given the chance to

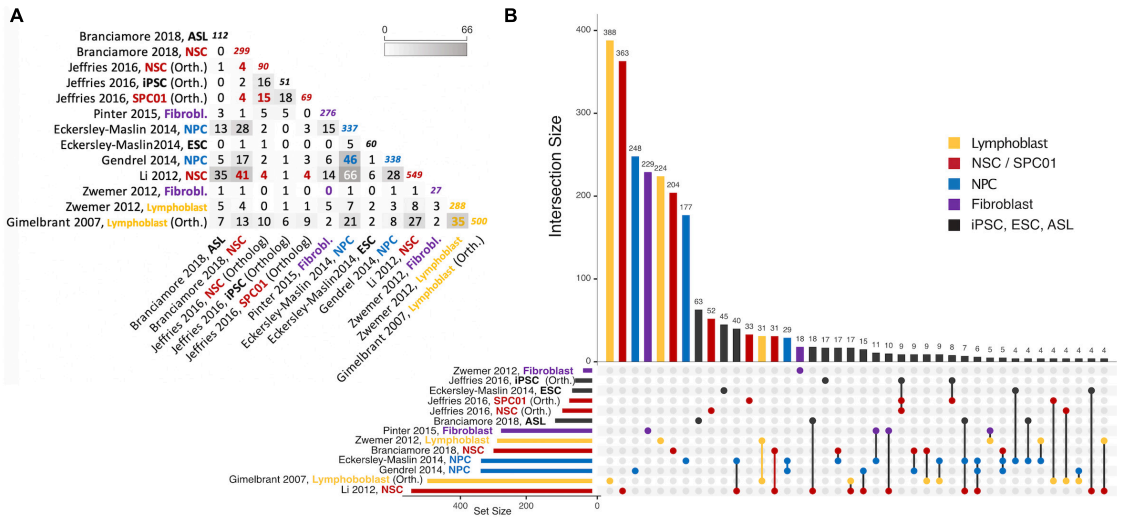


FIGURE 3 | Intersections of autosomal gene collections identified as random monoallelically expressed in the genome-wide studies described in **Table 2** (except Jeffries et al., 2012, which is not publicly available). **(A)** Half-matrix showing all pairwise intersections. **(B)** Upset plot (Conway et al., 2017) showing gene collection intersections of the same studies. The lower part of the panel has a horizontal bar plot showing the number of elements on each study collection, and a right section with a dot matrix. Each dot represents unique gene intersections, i.e., each gene is represented only once in the dot matrix. The upper vertical bar plot is related to the dot matrix, showing the number of unique genes in each intersection (for instance, there are 500 MAE genes in the Gimelbrant et al. (2007) dataset, but only 388 of those are uniquely present in that dataset; similarly, the Li et al. (2012) dataset shares more than 40 MAE genes with Eckerley-Maslin (2014) NPC dataset, but those 40 are uniquely shared between those two sets). Intersections of size smaller than 4 are not represented. For a complete description of the intersections and gene listing, see the **Supplementary File** provided with this review. ASL, Astrocyte-like cells; NSC, Neural stem cells; NPC, Neural progenitor cells; ESC, Embryonic stem cells; SPC01, Clonal Neural stem cells (before epigenetic reprogramming); iPSC, induced Pluripotent stem cells after epigenetic reprogramming of SPC01. Note that "NPC" on Jeffries et al. (2016) are derived from iPSC. Colors represent instances where a different cell/tissue type was studied more than once. To obtain intersections, gene ids were manually curated for immediate inconsistencies (e.g., gene name-to-date conversions when data was originally provided in microsoft excel format). All gene sets were then parsed with the gprofiler2 R package (Raudvere et al., 2019) for gene id consistency, using transcript ids as query whenever possible, and ENSEMBL gene ids as target (performed July 12th, 2021). Orthology conversion (from human to mouse) was performed with the same package for datasets involving human data. For Gimelbrant et al. (2007) and Zwemer et al. (2012) gene collections, MAE classes I, II and III were used to retrieve RMAE genes, and for Gendrel et al. (2014), the "NPC_random_catalog" classification was retrieved as RMAE.

recombine if the rearrangement of the first allele did not lead to the production of a receptor. In other words, the stability is not achieved before the expression of the receptor on the surface. Furthermore, the kappa immunoglobulin undergoes a process of receptor editing during which it can replace at its surface one protein by the protein encoded by the other allele (Casellas et al., 2001, 2007). It is only in mature lymphocytes that the pattern of monoallelic expression is stable, because the process of V(D)J recombination is permanently shut down and the silenced allele is epigenetically repressed and repositioned in the nucleus. With respect to OR genes, the patterns of monoallelic expression are stable, but it should be kept in mind that the cells are post-mitotic and terminally differentiated (Monahan and Lomvardas, 2015).

EXCEPTIONS

In mammalian females, under normal physiological conditions cases of two active X chromosomes are only found in undifferentiated cells and primordial germ cells before meiosis entry. All other cells have only one active X chromosome, because the double X dosage interferes with differentiation (Schulz et al., 2014). In contrast, biallelic expression of genes under RMAE is common and ranges from rare cells, such as in the case of the immunoglobulin heavy chain (Barreto and Cumano, 2000), to biallelic populations as frequent as the monoallelic ones (Gimelbrant et al., 2007). In the case of genes under RMAE with a low frequency of biallelic expression, these exceptions could correspond to the rare cases in which the two alleles become activated within the time-window allowed, before a negative feedback is triggered. Cases with a sizable population of biallelic expression could result from a relatively high individual probability of allele activation if the fitness of the cell is not compromised by the dual expression.

CONCLUDING REMARKS

X-chromosome inactivation (XCI) is a well-established specific silencing mechanism that ensures dosage compensation between the sexes in marsupial and placental mammals. At the heart of this process lies the long non-coding RNA *Xist*, which is capable of orchestrating structural changes and recruiting chromatin and repressor complexes to ensure transcriptional gene silencing at the level of an entire chromosome, early in development. XCI has clear implications in disease, as illustrated by the Turner (X0) and Klinefelter (XXY) syndromes, as well as the severe phenotypes or

lethality in males and variable phenotypes in females associated with X-linked disorders (e.g., Duchenne muscular dystrophy, hemophilia, and Rett syndrome). In contrast, RMAE evolved independently in a wide range of organisms beyond mammals, mostly to increase phenotypic diversity at the cellular level. RMAE lacks a master regulator and various mechanisms can establish it at different times during cellular differentiation. The collection of target genes encompass many cell types, showing some degree of overlap (**Figure 3** and **Supplementary File**), but have been reported to be largely tissue-specific (Gendrel et al., 2016). Finally, the extent of the bias in monoallelic expression varies widely amongst RMAE genes. In addition, there is so far no obvious link between the RMAE of autosomal genes and disease, although a number of RMAE genes are associated with autosomal dominant diseases. X-chromosome inactivation and RMAE are essentially different phenomena that share the stochastic component and perhaps the asymmetric silencing of chromosomal regions dependent on the presence of LINE-1 elements. But it is not known whether and how exactly LINE-1 elements boost XCI and if a similar process explains a considerable fraction of genes under RMAE.

AUTHOR CONTRIBUTIONS

VMB and A-VG wrote the manuscript. NK and CFA-P prepared the tables and figures. All authors contributed to the review content.

FUNDING

The work of NK and VMB was funded by iNOVA4Health – UIDB/Multi/04462/2020 and UIDP/Multi/04462/2020, a program financially supported by Fundação para a Ciência e Tecnologia (FCT)/Ministério da Educação e Ciência through national funds, and the FCT grant PTDC/BEX-BCM/5900/2014. CFA-P has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 752806. A-VG was supported by Fundação para a Ciência e Tecnologia (FCT), Portugal, through an assistant research contract (CEECIND/02085/2018) and the project grant PTDC/MED-OUT/4301/2020 IC&DT.

ACKNOWLEDGMENTS

We thank the reviewers for providing constructive comments to this review.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.740937/full#supplementary-material

REFERENCES

Adegbola, A. A., Cox, G. F., Bradshaw, E. M., Hafler, D. A., Gimelbrant, A., and Chess, A. (2015). Monoallelic expression of the human FOXP2 speech gene. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6848–6854. doi: 10.1073/pnas.1411270111

Adrianse, R. L., Smith, K., Gatbonton-Schwager, T., Sripathy, S. P., Lao, U., Foss, E. J., et al. (2018). Perturbed maintenance of transcriptional repression on the inactive X-chromosome in the mouse brain after Xist deletion. *Epigenetics Chromatin* 11:50. doi: 10.1186/s13072-018-0219-8

Aifantis, I., Buer, J., von Boehmer, H., and Azogui, O. (1997). Essential role of the pre-T cell receptor in allelic exclusion of the T cell receptor beta locus. *Immunity* 7, 601–607. doi: 10.1016/s1074-7613(00)80381-7

Allen, E., Horvath, S., Tong, F., Kraft, P., Spiteri, E., Riggs, A. D., et al. (2003). High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9940–9945. doi: 10.1073/pnas.1737401100

Alves-Pereira, C. F., de Freitas, R., Lopes, T., Gardner, R., Marta, F., Vieira, P., et al. (2014). Independent recruitment of Igh alleles in V(D)J recombination. *Nat. Commun.* 5:5623. doi: 10.1038/ncomms6623

Anderson, S. K. (2014). Probabilistic bidirectional promoter switches: noncoding RNA takes control. *Mol. Ther. Nucleic Acids* 3:e191. doi: 10.1038/mtna.2014.42

Aseem, O., Barth, J. L., Klatt, S. C., Smith, B. T., and Argraves, W. S. (2013). Cubilin expression is monoallelic and epigenetically augmented via PPARs. *BMC Genomics* 14:405. doi: 10.1186/1471-2164-14-405

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

Barreto, V., and Cumano, A. (2000). Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *J. Immunol.* 164, 893–899. doi: 10.4049/jimmunol.164.2.893

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

Baumann, C., and De La Fuente, R. (2009). ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. *Chromosoma* 118, 209–222. doi: 10.1007/s00412-008-0189-x

Bayley, J.-P., Bakker, A. M., Kaijzel, E. L., Wierenga, E. A., van der Pouw-Kraan, T. C. T. M., Huizinga, T. W. J., et al. (2003). Analysis of allelic expression patterns of IL-2, IL-3, IL-4, and IL-13 in human CD4+ T cell clones. *Eur. J. Immunol.* 33, 2142–2148. doi: 10.1002/eji.200323976

Berletch, J. B., Yang, F., and Disteche, C. M. (2010). Escape from X inactivation in mice and humans. *Genome Biol.* 11, 213–217. doi: 10.1186/gb-2010-11-6-213

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

Beutler, E., Yeh, M., and Fairbanks, V. F. (1962). The normal human female as a mosaic of X-chromosome activity: studies using the gene for C-6-PD-deficiency as a marker. *Proc. Natl. Acad. Sci. U.S.A.* 48, 9–16. doi: 10.1073/pnas.48.1.9

Bix, M., and Locksley, R. M. (1998). Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. *Science* 281, 1352–1354. doi: 10.1126/science.281.5381.1352

Boeren, J., and Gribnau, J. (2021). Xist-mediated chromatin changes that establish silencing of an entire X chromosome in mammals. *Curr. Opin. Cell Biol.* 70, 44–50. doi: 10.1016/j.ceb.2020.11.004

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., et al. (2017). Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat. Struct. Mol. Biol.* 24, 226–233. doi: 10.1038/nsmb.3365

Borst, P. (2002). Antigenic variation and allelic exclusion. *Cell* 109, 5–8. doi: 10.1016/s0092-8674(02)00711-0

Boyle, A. L., Ballard, S. G., and Ward, D. C. (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome

karyotyping by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U.S.A.* 87, 7757–7761. doi: 10.1073/pnas.87.19.7757

Branciamore, S., Valo, Z., Li, M., Wang, J., Riggs, A. D., and Singer-Sam, J. (2018). Frequent monoallelic or skewed expression for developmental genes in CNS-derived cells and evidence for balancing selection. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10379–E10386. doi: 10.1073/pnas.1808652115

Brown, C. J., and Willard, H. F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* 368, 154–156. doi: 10.1038/368154a0

Calado, D. P., Paixão, T., Holmberg, D., and Haury, M. (2006). Stochastic monoallelic expression of IL-10 in T cells. *J. Immunol.* 177, 5358–5364. doi: 10.4049/jimmunol.177.8.5358

Capparelli, R., Costabile, A., Viscardi, M., and Iannelli, D. (2004). Monoallelic expression of mouse Cd4 gene. *Mamm. Genome* 15, 579–584. doi: 10.1007/s00335-004-2351-y

Carrel, L., and Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434, 400–404. doi: 10.1038/nature03479

Casellas, R., Shih, T. A., Kleinewietfeld, M., Rakonjac, J., Nemazee, D., Rajewsky, K., et al. (2001). Contribution of receptor editing to the antibody repertoire. *Science* 291, 1541–1544. doi: 10.1126/science.1056600

Casellas, R., Zhang, Q., Zheng, N.-Y., Mathias, M. D., Smith, K., and Wilson, P. C. (2007). Igkappa allelic inclusion is a consequence of receptor editing. *J. Exp. Med.* 204, 153–160. doi: 10.1084/jem.20061918

Cebra, J. J., and Goldstein, G. (1965). Chromatographic purification of tetramethylrhodamine-immune globulin conjugates and their use in the cellular localization of rabbit gamma-globulin polypeptide chains. *J. Immunol.* 95, 230–245.

Chan, H.-W., Kurago, Z. B., Stewart, C. A., Wilson, M. J., Martin, M. P., Mace, B. E., et al. (2003). DNA methylation maintains allele-specific KIR gene expression in human natural killer cells. *J. Exp. Med.* 197, 245–255. doi: 10.1084/jem.20021127

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* 20, 2223–2237. doi: 10.1101/gad.380906

Chen, G., Schell, J. P., Benitez, J. A., Petropoulos, S., Yilmaz, M., Reinius, B., et al. (2016). Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. *Genome Res.* 26, 1342–1354. doi: 10.1101/gr.201954.115

Chess, A. (2016). Monoallelic gene expression in mammals. *Annu. Rev. Genet.* 50, 317–327. doi: 10.1146/annurev-genet-120215-035120

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78, 823–834. doi: 10.1016/s0092-8674(94)90562-2

Chiodetti, L., Barber, D. L., and Schwartz, R. H. (2000). Biallelic expression of the IL-2 locus under optimal stimulation conditions. *Eur. J. Immunol.* 30, 2157–2163. doi: 10.1002/1521-4141(2000)30:18<2157::aid-immu2157>3.0.co;2-g

Chow, J. C., Ciaudo, C., Fazzari, M. J., Mise, N., Servant, N., Glass, J. L., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* 141, 956–969. doi: 10.1016/j.cell.2010.04.042

Clemson, C. M., Hall, L. L., Byron, M., McNeil, J., and Lawrence, J. B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7688–7693. doi: 10.1073/pnas.0601069103

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

Corbel, C., Diabangouaya, P., Gendrel, A.-V., Chow, J. C., and Heard, E. (2013). Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells. *Development* 140, 861–872. doi: 10.1242/dev.087429

Davidson, R. G., Nitowsky, H. M., and Childs, B. (1963). Demonstration of two populations of cells in the human female heterozygous for glucose-6-phosphate dehydrogenase variants. *Proc. Natl. Acad. Sci. U.S.A.* 50, 481–485. doi: 10.1073/pnas.50.3.481

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196. doi: 10.1126/science.1245316

Diaz-Perez, S., Ouyang, Y., Perez, V., Cisneros, R., Regelson, M., and Marahrens, Y. (2005). The element(s) at the nontranscribed Xist locus of the active X chromosome controls chromosomal replication timing in the mouse. *Genetics* 171, 663–672. doi: 10.1534/genetics.105.043026

Donley, N., Smith, L., and Thayer, M. J. (2015). ASAR15, A cis-acting locus that controls chromosome-wide replication timing and stability of human chromosome 15. *PLoS Genet.* 11:e1004923. doi: 10.1371/journal.pgen.1004923

Donley, N., Stoffregen, E. P., Smith, L., Montagna, C., and Thayer, M. J. (2013). Asynchronous replication, mono-allelic expression, and long range Cis-effects of ASAR6. *PLoS Genet.* 9:e1003423. doi: 10.1371/journal.pgen.1003423

Dossin, F., Pinheiro, I., Żylicz, J. J., Roensch, J., Collombet, S., Le Saux, A., et al. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* 578, 455–460. doi: 10.1038/s41586-020-1974-9

Eckersley-Maslin, M. A., Thybert, D., Bergmann, J. H., Marioni, J. C., Flicek, P., and Spector, D. L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* 28, 351–365. doi: 10.1016/j.devcel.2014.01.017

Efstratiadis, A. (1995). Epigenetics. A new whiff of monoallelic expression. *Curr. Biol.* 5, 21–24. doi: 10.1016/s0960-9822(95)00007-8

Endo, Y., Sugimura, H., and Kino, I. (1995). Monoclonality of normal human colonic crypts. *Pathol. Int.* 45, 602–604. doi: 10.1111/j.1440-1827.1995.tb03509.x

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341:1237973. doi: 10.1126/science.1237973

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., et al. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* 37, 171–176. doi: 10.1038/ng1500

Farago, M., Rosenbluh, C., Tevlin, M., Fraenkel, S., Schlesinger, S., Masika, H., et al. (2012). Clonal allelic predetermination of immunoglobulin-κ rearrangement. *Nature* 490, 561–565. doi: 10.1038/nature11496

Ferguson-Smith, A. C., and Bourc'his, D. (2018). The discovery and importance of genomic imprinting. *Elife* 7:e42368. doi: 10.7554/eLife.42368

Galupa, R., and Heard, E. (2018). X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. *Annu. Rev. Genet.* 52, 535–566. doi: 10.1146/annurev-genet-120116-024611

Garrick, D., Sharpe, J. A., Arkell, R., Dobbie, L., Smith, A. J. H., Wood, W. G., et al. (2006). Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. *PLoS Genet.* 2:e58. doi: 10.1371/journal.pgen.0020058

Gendrel, A.-V., Attia, M., Chen, C.-J., Diabangouaya, P., Servant, N., Barillot, E., et al. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* 28, 366–380. doi: 10.1016/j.devcel.2014.01.016

Gendrel, A.-V., and Heard, E. (2011). Fifty years of X-inactivation research. *Development* 138, 5049–5055. doi: 10.1242/dev.068320

Gendrel, A.-V., Marion-Poll, L., Katoh, K., and Heard, E. (2016). Random monoallelic expression of genes on autosomes: parallels with X-chromosome inactivation. *Semin. Cell Dev. Biol.* 56, 100–110. doi: 10.1016/j.semcdb.2016.04.007

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science* 318, 1136–1140. doi: 10.1126/science.1148910

Gimelbrant, A. A., Ensminger, A. W., Qi, P., Zucker, J., and Chess, A. (2005). Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. *J. Biol. Chem.* 280, 1354–1359. doi: 10.1074/jbc.M411283200

Goldmit, M., and Bergman, Y. (2004). Monoallelic gene expression: a repertoire of recurrent themes. *Immunol. Rev.* 200, 197–214. doi: 10.1111/j.0105-2896.2004.00158.x

Goverman, J., Minard, K., Shastri, N., Hunkapiller, T., Hansburg, D., Sercarz, E., et al. (1985). Rearranged beta T cell receptor genes in a helper T cell clone

specific for lysozyme: no correlation between V beta and MHC restriction. *Cell* 40, 859–867. doi: 10.1016/0092-8674(85)90345-9

Graves, J. A. M. (2016). Evolution of vertebrate sex chromosomes and dosage compensation. *Nat. Rev. Genet.* 17, 33–46. doi: 10.1038/nrg.2015.2

Haig, D. (1993). Genetic conflicts in human pregnancy. *Q. Rev. Biol.* 68, 495–532. doi: 10.1086/418300

Hayashi, Y., Call, M. K., Liu, C.-Y., Hayashi, M., Babcock, G., Ohashi, Y., et al. (2010). Monoallelic expression of Krt12 gene during corneal-type epithelium differentiation of limbal stem cells. *Invest. Ophthalmol. Vis. Sci.* 51, 4562–4568. doi: 10.1167/iovs.10-5331

Held, W., Roland, J., and Raulet, D. H. (1995). Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. *Nature* 376, 355–358. doi: 10.1038/376355a0

Hoguin, A., Rastogi, A., Bowler, C., and Tirichine, L. (2021). Genome-wide analysis of allele-specific expression of genes in the model diatom Phaeodactylum tricornutum. *Sci. Rep.* 11:2954. doi: 10.1038/s41598-021-82529-1

Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., et al. (1998). Monoallelic expression of the interleukin-2 locus. *Science* 279, 2118–2121. doi: 10.1126/science.279.5359.2118

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3628–3632. doi: 10.1073/pnas.73.10.3628

Hu, H. Y., He, L., and Khaitovich, P. (2014). Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes. *BMC Genomics* 15:457. doi: 10.1186/1471-2164-15-457

Jeffries, A. R., Perfect, L. W., Ledderose, J., Schalkwyk, L. C., Bray, N. J., Mill, J., et al. (2012). Stochastic choice of allelic expression in human neural stem cells. *Stem Cells* 30, 1938–1947. doi: 10.1002/stem.1155

Jeffries, A. R., Uwanogho, D. A., Cocks, G., Perfect, L. W., Dempster, E., Mill, J., et al. (2016). Erasure and reestablishment of random allelic expression imbalance after epigenetic reprogramming. *RNA* 22, 1620–1630. doi: 10.1261/rna.058347.116

Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., et al. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J. Biol. Chem.* 281, 30551–30560. doi: 10.1074/jbc.M605677200

Kelly, B. L., and Locksley, R. M. (2000). Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns. *J. Immunol.* 165, 2982–2986. doi: 10.4049/jimmunol.165.6.2982

Kitamura, D., and Rajewsky, K. (1992). Targeted disruption of mu chain membrane exon causes loss of heavy-chain allelic exclusion. *Nature* 356, 154–156. doi: 10.1038/356154a0

Ku, C.-J., Lim, K.-C., Kalantry, S., Maillard, I., Engel, J. D., and Hosoya, T. (2015). A monoallelic-to-biallelic T-cell transcriptional switch regulates GATA3 abundance. *Genes Dev.* 29, 1930–1941. doi: 10.1101/gad.265025.115

Kucera, K. S., Reddy, T. E., Pauli, F., Gertz, J., Logan, J. E., Myers, R. M., et al. (2011). Allele-specific distribution of RNA polymerase II on female X chromosomes. *Hum. Mol. Genet.* 20, 3964–3973. doi: 10.1093/hmg/ddr315

Li, S. M., Valo, Z., Wang, J., Gao, H., Bowers, C. W., and Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS One* 7:e31751. doi: 10.1371/journal.pone.0031751

Loda, A., Brandsma, J. H., Vassilev, I., Servant, N., Loos, F., Amirnasr, A., et al. (2017). Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nat. Commun.* 8:690. doi: 10.1038/s41467-017-00528-1

Loda, A., and Heard, E. (2019). Xist RNA in action: past, present, and future. *PLoS Genet.* 15:e1008333. doi: 10.1371/journal.pgen.1008333

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* 190, 372–373. doi: 10.1038/190372a0

Lyon, M. F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. *Am. J. Hum. Genet.* 14, 135–148.

Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* 80, 133–137. doi: 10.1159/000014969

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., et al. (2011). An epigenetic signature for monoallelic olfactory receptor expression. *Cell* 145, 555–570. doi: 10.1016/j.cell.2011.03.040

Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P., et al. (2004). Reactivation of the paternal X chromosome in early mouse embryos. *Science* 303, 666–669. doi: 10.1126/science.1092674

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

Migeon, B. R., Axelman, J., and Beggs, A. H. (1988). Effect of ageing on reactivation of the human X-linked HPRT locus. *Nature* 335, 93–96. doi: 10.1038/335093a0

Monahan, K., and Lomvardas, S. (2015). Monoallelic expression of olfactory receptors. *Annu. Rev. Cell. Dev. Biol.* 31, 721–740. doi: 10.1146/annurev-cellbio-100814-125308

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F., and Gribnau, J. (2008). X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. *Cell* 132, 410–421. doi: 10.1016/j.cell.2007.12.036

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., et al. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature* 414, 221–225. doi: 10.1038/35102606

Nag, A., Savova, V., Fung, H.-L., Miron, A., Yuan, G.-C., Zhang, K., et al. (2013). Chromatin signature of widespread monoallelic expression. *Elife (Cambridge)* 2:e01256. doi: 10.7554/eLife.01256

Naramura, M., Hu, R. J., and Gu, H. (1998). Mice with a fluorescent marker for interleukin 2 gene activation. *Immunity* 9, 209–216. doi: 10.1016/s1074-7613(00)80603-2

Ng, K. K., Yui, M. A., Mehta, A., Siu, S., Irwin, B., Pease, S., et al. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *Elife* 7:e37851. doi: 10.7554/eLife.37851

Niimura, Y., Matsui, A., and Touhara, K. (2014). Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* 24, 1485–1496. doi: 10.1101/gr.169532.113

Nutt, S. L., Vambrie, S., Steinlein, P., Kozmik, Z., Rolink, A., Weith, A., et al. (1999). Independent regulation of the two Pax5 alleles during B-cell development. *Nat. Genet.* 21, 390–395. doi: 10.1038/7720

Ohno, S., and Hauschka, T. S. (1960). Allocycly of the X-chromosome in tumors and normal tissues. *Cancer Res.* 20, 541–545.

Ohno, S., Kaplan, W. D., and Kinosita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. *Exp. Cell Res.* 18, 415–418. doi: 10.1016/0014-4827(59)90031-x

Pancer, Z., Amemiya, C. T., Ehrhardt, G. R. A., Ceitlin, J., Gartland, G. L., and Cooper, M. D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430, 174–180. doi: 10.1038/nature02740

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W. R., Damianov, A., et al. (2020). A protein assembly mediates Xist localization and gene silencing. *Nature* 587, 145–151. doi: 10.1038/s41586-020-2703-0

Pereira, J. P., Girard, R., Chaby, R., Cumano, A., and Vieira, P. (2003). Monoallelic expression of the murine gene encoding Toll-like receptor 4. *Nat. Immunol.* 4, 464–470. doi: 10.1038/ni917

Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J. Exp. Med.* 122, 853–876. doi: 10.1084/jem.122.5.853

Pinter, S. F., Colognori, D., Beliveau, B. J., Sadreyev, R. I., Payer, B., Yildirim, E., et al. (2015). Allelic imbalance is a prevalent and tissue-specific feature of the mouse transcriptome. *Genetics* 200, 537–549. doi: 10.1534/genetics.115.176263

Pinter, S. F., Sadreyev, R. I., Yildirim, E., Jeon, Y., Ohsumi, T. K., Borowsky, M., et al. (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res.* 22, 1864–1876. doi: 10.1101/gr.133751.111

Platt, E. J., Smith, L., and Thayer, M. J. (2018). L1 retrotransposon antisense RNA within ASAR lncRNAs controls chromosome-wide replication timing. *J. Cell Biol.* 217, 541–553. doi: 10.1083/jcb.201707082

Popova, B. C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T. B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7706–7711. doi: 10.1073/pnas.0602021103

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

Rhoades, K. L., Singh, N., Simon, I., Glidden, B., Cedar, H., and Chess, A. (2000). Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. *Curr. Biol.* 10, 789–792. doi: 10.1016/s0960-9822(00)00565-0

Rivière, I., Sunshine, M. J., and Littman, D. R. (1998). Regulation of IL-4 expression by activation of individual alleles. *Immunity* 9, 217–228. doi: 10.1016/s1074-7613(00)80604-4

Rodriguez, I., Feinstein, P., and Mombaerts, P. (1999). Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell* 97, 199–208. doi: 10.1016/s0092-8674(00)80730-8

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., et al. (2005). The DNA sequence of the human X chromosome. *Nature* 434, 325–337. doi: 10.1038/nature03440

Saleh, A., Davies, G. E., Pascal, V., Wright, P. W., Hodge, D. L., Cho, E. H., et al. (2004). Identification of probabilistic transcriptional switches in the Ly49 gene cluster: a eukaryotic mechanism for selective gene activation. *Immunity* 21, 55–66. doi: 10.1016/j.immuni.2004.06.005

Saleh, A., Makrigiannis, A. P., Hodge, D. L., and Anderson, S. K. (2002). Identification of a novel Ly49 promoter that is active in bone marrow and fetal thymus. *J. Immunol.* 168, 5163–5169. doi: 10.4049/jimmunol.168.10.5163

Sano, Y., Shimada, T., Nakashima, H., Nicholson, R. H., Eliason, J. F., Kocarek, T. A., et al. (2001). Random monoallelic expression of three genes clustered within 60 kb of mouse t complex genomic DNA. *Genome Res.* 11, 1833–1841. doi: 10.1101/gr.194301

Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., et al. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* 159, 869–883. doi: 10.1016/j.cell.2014.10.019

Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M., and Wutz, A. (2006). Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. *Mol. Cell. Biol.* 26, 7167–7177. doi: 10.1128/MCB.00810-06

Savova, V., Chun, S., Sohail, M., McCole, R. B., Witwicki, R., Gai, L., et al. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat. Genet.* 48, 231–237. doi: 10.1038/ng.3493

Schlesinger, S., Selig, S., Bergman, Y., and Cedar, H. (2009). Allelic inactivation of rDNA loci. *Genes Dev.* 23, 2437–2447. doi: 10.1101/gad.544509

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

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., et al. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* 302, 2088–2094. doi: 10.1126/science.1089122

Sharp, A., Robinson, D., and Jacobs, P. (2000). Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Hum. Genet.* 107, 343–349. doi: 10.1007/s004390000382

Sharp, A. J., Spotswood, H. T., Robinson, D. O., Turner, B. M., and Jacobs, P. A. (2002). Molecular and cytogenetic analysis of the spreading of X inactivation in X;autosome translocations. *Hum. Mol. Genet.* 11, 3145–3156. doi: 10.1093/hmg/11.25.3145

Simon, M. D., Pinter, S. F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S. K., et al. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* 504, 465–469. doi: 10.1038/nature12719

Singh, N., Ebrahimi, F. A. W., Gimelbrant, A. A., Ensminger, A. W., Tackett, M. R., Qi, P., et al. (2003). Coordination of the random asynchronous replication of autosomal loci. *Nat. Genet.* 33, 339–341. doi: 10.1038/ng1102

Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., et al. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* 37, 457–468. doi: 10.1016/j.molcel.2010.01.030

Sonoda, E., Pewzner-Jung, Y., Schwers, S., Taki, S., Jung, S., Eilat, D., et al. (1997). B cell development under the condition of allelic inclusion. *Immunity* 6, 225–233. doi: 10.1016/s1074-7613(00)80325-8

Sousa, E. J., Stuart, H. T., Bates, L. E., Ghorbani, M., Nichols, J., Dietmann, S., et al. (2018). Exit from naive pluripotency induces a transient X chromosome inactivation-like state in males. *Cell Stem Cell* 22, 919–928.e6. doi: 10.1016/j.stem.2018.05.001

Souyris, M., Cenac, C., Azar, P., Daviaud, D., Canivet, A., Grunenwald, S., et al. (2018). TLR7 escapes X chromosome inactivation in immune cells. *Sci. Immunol.* 3:eaa8855. doi: 10.1126/sciimmunol.aap8855

Stoffregen, E. P., Donley, N., Stauffer, D., Smith, L., and Thayer, M. J. (2011). An autosomal locus that controls chromosome-wide replication timing and mono-allelic expression. *Hum. Mol. Genet.* 20, 2366–2378. doi: 10.1093/hmg/ddr138

Strehle, M., and Guttman, M. (2020). Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Curr. Opin. Cell Biol.* 64, 139–147. doi: 10.1016/j.ceb.2020.04.009

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

Syrett, C. M., Paneru, B., Sandoval-Heglund, D., Wang, J., Banerjee, S., Sindhava, V., et al. (2019). Altered X-chromosome inactivation in T cells may promote sex-biased autoimmune diseases. *JCI Insight* 4:1290. doi: 10.1172/jci.insight.126751

Takagi, N., and Abe, K. (1990). Detrimental effects of two active X chromosomes on early mouse development. *Development* 109, 189–201. doi: 10.1242/dev.109.1.189

Takizawa, T., Gudla, P. R., Guo, L., Lockett, S., and Misteli, T. (2008). Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev.* 22, 489–498. doi: 10.1101/gad.1634608

Tang, W. W. C., Dietmann, S., Irie, N., Leitch, H. G., Floros, V. I., Bradshaw, C. R., et al. (2015). A unique gene regulatory network resets the human germline epigenome for development. *Cell* 161, 1453–1467. doi: 10.1016/j.cell.2015.04.053

Tang, Y. A., Huntley, D., Montana, G., Cerase, A., Nesterova, T. B., and Brockdorff, N. (2010). Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation. *Epigenetics Chromatin* 3, 10–12. doi: 10.1186/1756-8935-3-10

Tasic, B., Nabholz, C. E., Baldwin, K. K., Kim, Y., Rueckert, E. H., Ribich, S. A., et al. (2002). Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol. Cell* 10, 21–33. doi: 10.1016/s1097-2765(02)00578-6

Taylor, J. H. (1960). Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. *J. Biophys. Biochem. Cytol.* 7, 455–464. doi: 10.1083/jcb.7.3.455

Thomas, B. J., Rubio, E. D., Krumm, N., Broin, P. O., Bomsztyk, K., Welcsh, P., et al. (2011). Allele-specific transcriptional elongation regulates monoallelic expression of the IGF2BP1 gene. *Epigenetics Chromatin* 4:14. doi: 10.1186/1756-8935-4-14

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Res.* 14, 62–66. doi: 10.1101/gr.1982804

Vettermann, C., and Schlissel, M. S. (2010). Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol. Rev.* 237, 22–42. doi: 10.1111/j.1600-065X.2010.00935.x

Wang, J., Syrett, C. M., Kramer, M. C., Basu, A., Atchison, M. L., and Anguera, M. C. (2016). Unusual maintenance of X chromosome inactivation predisposes female lymphocytes for increased expression from the inactive X. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2029–E2038. doi: 10.1073/pnas.1520113113

Wang, X., Su, H., and Bradley, A. (2002). Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. *Genes Dev.* 16, 1890–1905. doi: 10.1101/gad.1004802

Welshons, W. J., and Russell, L. B. (1959). The Y-chromosome as the bearer of male determining factors in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* 45, 560–566. doi: 10.1073/pnas.45.4.560

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., et al. (2014). Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. *Neuron* 81, 103–119. doi: 10.1016/j.neuron.2013.10.051

Yang, F., Babak, T., Shendure, J., and Disteche, C. M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* 20, 614–622. doi: 10.1101/gr.103200.109

Yang, L., Kirby, J. E., Sunwoo, H., and Lee, J. T. (2016). Female mice lacking Xist RNA show partial dosage compensation and survive to term. *Genes Dev.* 30, 1747–1760. doi: 10.1101/gad.281162.116

Yang, L., Yildirim, E., Kirby, J. E., Press, W., and Lee, J. T. (2020). Widespread organ tolerance to Xist loss and X reactivation except under chronic stress in the gut. *Proc. Natl. Acad. Sci. U.S.A.* 117, 4262–4272. doi: 10.1073/pnas.1917203117

Yang, M. Q., Koehly, L. M., and Elnitski, L. L. (2007). Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes. *PLoS Comput. Biol.* 3:e72. doi: 10.1371/journal.pcbi.0030072

Yildirim, E., Kirby, J. E., Brown, D. E., Mercier, F. E., Sadreyev, R. I., Scadden, D. T., et al. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* 152, 727–742. doi: 10.1016/j.cell.2013.01.034

Youness, A., Miquel, C.-H., and Guéry, J.-C. (2021). Escape from X chromosome inactivation and the female predominance in autoimmune diseases. *Int. J. Mol. Sci.* 22:1114. doi: 10.3390/ijms22031114

Zhang, X., and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* 5, 124–133. doi: 10.1038/nn800

Zhou, Q., Wang, T., Leng, L., Zheng, W., Huang, J., Fang, F., et al. (2019). Single-cell RNA-seq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis. *Mol. Reprod. Dev.* 86, 871–882. doi: 10.1002/mrd.23162

Zwemer, L. M., Zak, A., Thompson, B. R., Kirby, A., Daly, M. J., Chess, A., et al. (2012). Autosomal monoallelic expression in the mouse. *Genome Biol.* 13:R10. doi: 10.1186/gb-2012-13-2-r10

Żylicz, J. J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S. T., et al. (2019). The implication of early chromatin changes in X chromosome inactivation. *Cell* 176, 182–197.e23. doi: 10.1016/j.cell.2018.11.041

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

REVIEW
published: 01 October 2021
doi: 10.3389/fcell.2021.737681

Asynchronous Replication Timing: A Mechanism for Monoallelic Choice During Development

Yehudit Bergman[1]†, Itamar Simon[2]† and Howard Cedar[1]†*

[1] Department of Developmental Biology and Cancer Research, Hebrew University Hadassah Medical School, Jerusalem, Israel, [2] Department of Microbiology and Molecular Genetics, Hebrew University Hadassah Medical School, The Institute for Medical Research Israel-Canada (IMRIC), Jerusalem, Israel

Developmental programming is carried out by a sequence of molecular choices that epigenetically mark the genome to generate the stable cell types which make up the total organism. A number of important processes, such as genomic imprinting, selection of immune or olfactory receptors, and X-chromosome inactivation in females are dependent on the ability to stably choose one single allele in each cell. In this perspective, we propose that asynchronous replication timing (ASRT) serves as the basis for a sophisticated universal mechanism for mediating and maintaining these decisions.

Keywords: epigenetic regulation, chromatin accessibility, embryonal stem cells, DNA replication, genomic imprinting, X-chromosome inactivation

OPEN ACCESS

Edited by:
Jin Xu,
Sun Yat-sen University, China

Reviewed by:
Joseph Mauro Calabrese, University of North Carolina at Chapel Hill, United States
Simão Teixeira da Rocha, University of Lisbon, Portugal

***Correspondence:**
Yehudit Bergman
yehudit.bergman@mail.huji.ac.il

† These authors have contributed equally to this work

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

Received: *07 July 2021*
Accepted: *14 September 2021*
Published: *01 October 2021*

Citation:
Bergman Y, Simon I and Cedar H (2021) Asynchronous Replication Timing: A Mechanism for Monoallelic Choice During Development. Front. Cell Dev. Biol. 9:737681. doi: 10.3389/fcell.2021.737681

MONOALLELIC EXPRESSION

Although the mammalian genome has a diploid composition, many genes are regulated in a monoallelic manner. The most common form of this phenomenon is characterized by skewed allelic expression with some cells exhibiting preferential transcription for the paternal allele, some favoring the maternal allele, while other cells in this same population express this gene either biallelically or not at all (Guo et al., 2005; Savol et al., 2017; Branciamore et al., 2018; Galupa and Heard, 2018). A second type of monoallelic expression (MAE) is characterized by defined regions of the genome that are actually developmentally programmed to choose between the two alleles on the basis of stable differential marks. A classic example of this phenomenon is genomic imprinting, where a single parental allele, either the maternal or paternal, is programmed by the gametes to be transcribed in somatic cells of the offspring, while the other allele is silent. This group includes many genes, such as *Igf2* and *Snrpn*, which appear to play some role in early embryonic growth control and have been found to be involved in a number of genetic diseases (Hanna and Kelsey, 2021).

Other genome domains exhibit a random pattern of MAE with some cells selecting the maternal allele, while others choose to express the paternal copy. These regions are enriched for receptor gene clusters involved in defining cell identity by mediating interactions between the cell and its environment. This includes many of the gene arrays that make up the foundation for the immune system, olfaction and cell positioning during development (Chess et al., 1994; Rodriguez, 2013;

Levin-Klein and Bergman, 2014; Chess, 2016). One of the main features of all these developmentally programmed domains is that they replicate in an asynchronous manner, one allele being copied earlier than the other during S phase, thus providing a mark that can distinguish between the two alleles (Cedar and Bergman, 2008). A similar pattern is seen in X-chromosome inactivation in somatic cells (Avner and Heard, 2001). In this perspective, we will attempt to understand how this epigenetic process is established during development, and in this way, explain the basic mechanism underlying stable allelic choice, both imprinted and random.

REPLICATION TIMING

Due to the large size of the mammalian genome, its replication is not only extended over time, but is apparently also carried out by an organized and carefully regulated program (Goren and Cedar, 2003; Marchal et al., 2019). One of the most outstanding aspects of this process involves temporal control, with some regions of the genome undergoing DNA replication in early S phase, while others replicate late. By labeling cells with BrdU, one can actually visualize these regions as alternating chromosomal bands, representing replication time zones with an average size of about 1 Mb that colocalize with the structurally determined G banding pattern (Hand, 1978). Strikingly, this organization is also correlated to gene expression, with housekeeping and other active genes replicating early, while heterochromatin and inactive genes largely replicate in late S (Schübeler et al., 2002; Greenberg et al., 2020). In keeping with this picture, the early zones have been found to be in a relatively accessible DNaseI sensitive configuration (Kerem et al., 1984), while the late regions have a more closed structure and are localized to nuclear lamina associated domains (LADs) (Heun et al., 2001; Vogel et al., 2007; Guelen et al., 2008; van Steensel and Belmont, 2017). Furthermore, many replication time zones are regulated in a tissue or developmental-specific pattern, replicating late in most cell types, but switching to early replication in keeping with its expression profile (Holmquist, 1987; Siefert et al., 2017). Replication timing is also correlated with many important epigenetic features within the genome architecture (Rhind and Gilbert, 2013; Reverón-Gómez et al., 2018; Escobar et al., 2019). In keeping with this, several studies have provided more direct evidence that replication timing itself plays a key role in orchestrating and maintaining epigenetic states (Zhang et al., 2002; Klein et al., 2021).

ASYNCHRONOUS REPLICATION TIMING

While most regions of the genome have a fixed replication time, with both alleles being equally recognized by the trans-acting factors that govern replication timing control, there are several categories of genes that replicate in an asynchronous manner, with one allele being marked for replication early in S phase and the other, for late replication. The most striking example of this phenomenon is the X-chromosome in female somatic cells, where one copy replicates in early S, while the other copy replicates later, as demonstrated by *in situ* S-phase-specific BrdU labeling (Latt, 1973) as well as whole genome DNA sequence analysis (Koren and McCarroll, 2014; Blumenfeld et al., 2021). In keeping with this, genes on the late chromosome are generally inactive and have a non-accessible chromatin structure, characterized by DNA-methylated promoters, as well as a variety of inactivating histone modifications and variants (Mohandas et al., 1981; Jeppesen and Turner, 1993; Heard, 2004; Żylicz and Heard, 2020; Boeren and Gribnau, 2021). The actual inactivation process in the early embryo appears to take place stochastically in each individual cell, either on the paternal or maternal X chromosome and this decision is then stably maintained through future cell divisions (Heard, 2004; Sahakyan et al., 2017). Genomically imprinted gene regions represent a second category subject to asynchronous replication timing (ASRT), as determined by FISH, but in this case, it is always the same allele that is early replicating, apparently because of predetermined epigenetic events that occur in the individual gametes (Simon et al., 1999; Goren and Cedar, 2003).

In addition to these classic examples, a large number of autosomal chromosome regions (1–2 Mb in length) have been found to replicate asynchronously. This was originally documented using fluorescence *in situ* hybridization (FISH) to visualize specific gene regions in diploid cells growing in culture (Selig et al., 1992). In this assay one can visualize both copies of any particular gene region in interphase cells. In nuclei that have not yet replicated this region, one observes two single hybridization dots, representing the two alleles. After replication and subsequent segregation, however, these loci exhibit double dots. For a large percentage of the genome, both alleles are synchronized, with all nuclei exhibiting either two single or two double signals. At some loci, however, one observes a large percentage of nuclei with one allele showing a single dot (not yet replicated) and the other having a double dot (already replicated), indicating that this region replicates asynchronously (Kitsberg et al., 1993). This FISH assay encompasses two aspects of DNA replication, differential time of DNA synthesis in S-phase, as well as the time of visual chromatid segregation, suggesting that asynchronous loci are essentially characterized by allele-differential "chromosomal replication," with structure and segregation being an important, often dominant, part of this process (Azuara et al., 2003; Rivera-Mulia et al., 2018; Blumenfeld et al., 2021).

PRINCIPLES OF ALLELIC CHOICE

Asynchronous replication timing was originally observed for select genome loci containing olfactory receptor (Chess et al., 1994) or immune system (Mostoslavsky et al., 2001) gene arrays, regions clearly associated with monoallelic behavior. In both cases, each individual cell must be able to choose one allele out of the two available in order to ensure production of only one unique receptor for presentation on the cell surface. The observation of allelic asynchrony suggested that replication timing may somehow serve as a way to distinguish between

the two alleles, thereby providing a simple epigenetic mark for directing allelic choice.

Use of the *Igκ* locus as a prototype provides an excellent system for better understanding ASRT and its role in monoallelic choice. During B-cell lineage formation, each cell has an equal chance of choosing the paternal, or alternatively the maternal allele for making the *Igκ* light chain (Cedar and Bergman, 2011). It was originally postulated that the decision for which allele undergoes rearrangement is completely stochastic and is mediated in trans by nuclear protein factors, with the first allele to bind the factor undergoing rearrangement (Mostoslavsky et al., 2004). The light-chain produced from this reaction would then be capable of preventing rearrangement on the other allele through feedback inhibition. This "first come, first served" mechanism has also been proposed for other cases of allelic choice, such as that seen in early embryonic random X-chromosome inactivation (Penny et al., 1996; Mutzel and Schulz, 2020).

Although this trans-acting concept provided a reasonable explanation for the choosing process, FISH replication timing experiments raised the possibility that the choice of allele may actually be pre-determined, since the initial rearrangement always occurs on the early allele in mature B-cells regardless of its parental identity (**Figure 1**; Farago et al., 2012). Furthermore, it was demonstrated that the active allele is specifically associated with other basic structural marks, such as preferential chromatin accessibility, DNA undermethylation and localization away from the nuclear periphery, all properties that are thought to be acquired prior to the actual rearrangement step (Mostoslavsky et al., 1998, 2001; Goldmit et al., 2002, 2005; Ji et al., 2003). In pre-B cells, for example, the *Igκ* locus was found to already replicate asynchronously and when clones were prepared from single cells, it was shown by allele-specific FISH analysis that in some clones the maternal locus replicates early in every cell, while in other clones, early replication occurs consistently on the paternal allele, suggesting that the two alleles are structurally distinct from each other even prior to rearrangement. Interestingly, each clone shows an allelic pattern of chromatin accessibility and when differentiated to B-cells *in vitro*, produces the light chain antibody almost exclusively from the early allele (**Figure 1**; Farago et al., 2012). These experiments clearly indicate that the choosing process involves recognition of predetermined allelic marking. This same type of mechanism is probably also used for other immune system gene arrays, such as the immunoglobulin heavy chain, the T-cell receptor β locus, NK receptors as well as cytokines and their receptors, all of which have been shown to undergo asynchronous replication timing (Chess, 1998; Mostoslavsky et al., 2001; Guo et al., 2005).

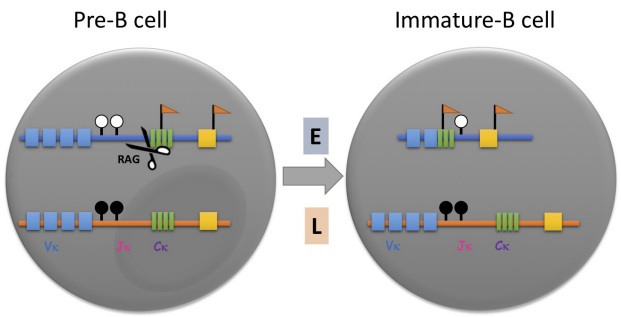


FIGURE 1 | Asynchronous replication timing (ASRT) and its role in immunoglobulin allelic exclusion. In pre-B cells prior to *Igκ* rearrangement, the two unrearranged alleles, one of which replicates early (E, blue) and the other of which replicates later (L, orange), are marked with different epigenetic marks. The early replicating allele is hypomethylated at the DNA level (white lollipops) and enriched with hyperacetylated histones and methylation of histone H3 lysine 4 (flags). The late replicating allele is hypoacetylated, CpG methylated (black lollipops) and sequestered within heterochromatin. Once the pre-B cell senses a signal for rearrangement, the early replicating allele binds trans-acting factors such as B-cell-specific transcription factor, Pax5, and the rearrangement machinery (RAG), which binds to H3K4me3, thereby rendering it susceptible to rearrangement and progression to the stage of immature B cell. Thus, it is almost always the early replicating allele that undergoes rearrangement first. Vκ, variable gene segments; Jκ, joining gene segments; Cκ, constant region.

Another example of allelic choice can be observed in the olfactory system where each olfactory neuron must choose a single receptor gene copy from amongst 1,000 different gene sequences that are organized as arrays within multiple asynchronous replicating domains scattered over the genome. While the choice of one specific gene sequence is apparently mediated by a single olfactosome enhancer element on chromosome 14 that can only engage one receptor gene at a time (Lomvardas et al., 2006; Markenscoff-Papadimitriou et al., 2014), there must still be a mechanism to ensure that only one of the two allelic olfactosome loci is utilized, and it is possible that this selection process is directed by its pre-existing asynchronous replication-timing mark. Interestingly, this same type of allelic non-homologous chromatin contact has also been observed at other ASRT domains, perhaps suggesting that this structure may be a general feature of allelic choice (Maass et al., 2019).

Since the FISH assay must be carried out using individual specific probes, it was previously possible to identify only a relatively small number of asynchronous replicating regions, but recent studies utilizing allelically marked hybrid pre-B cell clones have succeeded in carrying out genome-wide quantitative DNA sequence analysis of S phase cells, thus enabling the discovery of almost 150 new regions of the genome in which one allele replicates prior to the other. In each clone, some sites show early replication of the maternal allele, while others are in the opposite orientation and, in contrast to what had been observed previously, these loci are widely distributed over many different chromosomes (Blumenfeld et al., 2021). At all of these regions, the early replicating allele is preferentially more accessible (as determined by ATAC-Seq), including gene regions that are not actively expressed in these cells (e.g., olfactory receptors), suggesting that this represents an independent epigenetic mark that may be found in a wide variety of different cell types and that both ASRT and monoallelic accessible chromatin structure exist prior to expression, at a stage when it may actually be involved in the allelic choice process itself. In addition to these important structural features, genomic analysis also revealed new, potentially mono-allelic gene functions located preferentially in ASRT domains, including the taste receptors, the

vomeronasal receptors, as well as chemokines and their receptors that are used for chemotaxis, all of which are organized as gene arrays (Blumenfeld et al., 2021). It is interesting in this regard that the olfactory receptor system evidently plays a dual role as an odor receptor, as well as a guiding element that directs specific neurons to their proper location in the olfactory bulb (Mombaerts et al., 1996).

DEVELOPMENT

A number of different studies have noted that, in general, the ASRT pattern can be detected in a wide variety of different cell types independently of whether these regions are actually expressed, thus suggesting that the establishment of ASRT must take place during very early development. This concept is also supported by the observation that all of the known ASRT loci also replicate asynchronously in early embryonic stem cells (ESCs) (Gribnau et al., 2003; Alexander et al., 2007; Blumenfeld et al., 2021).

Early studies *in vivo* were influential in elucidating the developmental timing of this process by showing that loci associated with ASRT start off in the early embryo by replicating synchronously, as observed in cells of the morula (6–16 cells), blastula (∼60 cells) and inner cell mass (ICM) (Mostoslavsky et al., 2001; Shufaro et al., 2010). Thus, the actual process of establishing this mark must occur during the transition to implantation stage. In an attempt to mimic this process *in vitro* and thereby decipher its mechanism, ES cells were converted to a more-naive pre-implantation stage by culturing them in 2i medium and this was sufficient to revert all ASRT loci to a synchronous replication pattern. Furthermore, subsequent removal of the 2i medium quickly restored these cells to their original state, with these loci already becoming asynchronous in the first division cycle after transfer (Masika et al., 2017).

When analyzed in detail, this ASRT initiation event turns out to be very interesting, since for each individual ASRT locus it is always one specific parental allele that is chosen to be early during the first round. For some sites, the paternal allele is set up as early, while for other sites, it is the maternal allele that replicates early and this serves to establish a fixed coordinated pattern of parallel and anti-parallel ASRT loci across the genome (Masika et al., 2017; Blumenfeld et al., 2021). This pre-determined pattern indicates that at every ASRT domain, each parental allele must already be marked in the gamete in a manner that will allow it to dictate whether to undergo early or late replication during the first cycle of ASRT at the time of implantation. Thus, the information for distinguishing between the alleles is inherently encoded by epigenetic tags derived from the individual homozygous gametes and thus does not actually involve making a stochastic decision between two equal alleles. Although the identity of these marks is not known, ES cells lacking all DNA methylation were unable to generate this asynchrony, suggesting that this early marking process may, in some way, involve DNA methylation (Masika et al., 2017) in conjunction with histone marks, as has been shown to be the case for genomic imprinting (Nakamura et al., 2007).

Once this initial orientation pattern is established, subsequent cell divisions still perpetuate the asynchronous state, but each cycle is then subject to allele switching, so that all loci set up as maternal early will generate daughter cells characterized by paternal early replication, while all loci generated as paternal early will undergo a complete switch to maternal early (Masika et al., 2017). This automatic switching behavior essentially preserves the original parallel or antiparallel relationship between the many ASRT loci in the genome and thus sets up a bimodal population containing two "enantiomeric" cell types, each having a mirror image ASRT orientation profile (**Figure 2**; Blumenfeld et al., 2021). Switching continues throughout early stem-cell-like stages, until commitment comes into play at the time of definitive differentiation when cells then begin to clonally maintain each "enantiomer" separately. In line with this notion, it is likely that all tissues in the body are constituted from a mixture of two mirror image ASRT states.

During lymphoid development in the immune system, for example, hematopoietic stem cells (HSCs) and multi-potent precursor cells (MPPs) are still in the allele switching mode, but progression to Common Lymphoid Precursors (CLPs) is accompanied by commitment to one specific direction, which is then clonally maintained during subsequent stages of lineage commitment (Farago et al., 2012). Taken together, these findings indicate that switching represents a form of plasticity that preserves the potential for stem cells to use either one of two fixed options. Upon differentiation, they lose this plasticity and become committed to targeting one allele. In the immune system, it is this allelic clonality that actually allows the formation of memory cells with the potential to mount an antibody defense to specific antigens.

MECHANISMS OF ALLELIC CHOICE

The classical way of thinking about "allelic choice" usually entails interactions between trans-acting factors in the nucleus and identical cis-acting sequences that compete with one another. This model assumes that both alleles have the same probability of engagement, making it difficult to choose only one and then maintain this decision for extended periods of time. Kinetically, this type of mechanism would also require low concentrations of the trans-acting protein factor, perhaps combined with a feedback regulatory loop that can quickly prevent the other allele from being activated, a pathway which has been shown to exist in the immune system (Coleclough, 1983; Yancopoulos and Alt, 1986; Gorman and Alt, 1998; Liang et al., 2004). Probability considerations predict that low concentrations of the activating factor could indeed bring about targeting of one allele prior to the other, but this model also predicts that in many cells, neither allele may get activated, a situation which is not appropriate for carrying out programmed developmental decisions.

Early developmental programming of ASRT represents an excellent alternative solution to the problem of choosing, by providing a stable epigenetic mark in cis that distinguishes between two almost-identical alleles in the same cell, making one more accessible than the other. This structural difference is set up early in the embryo and then maintained in all cell types where

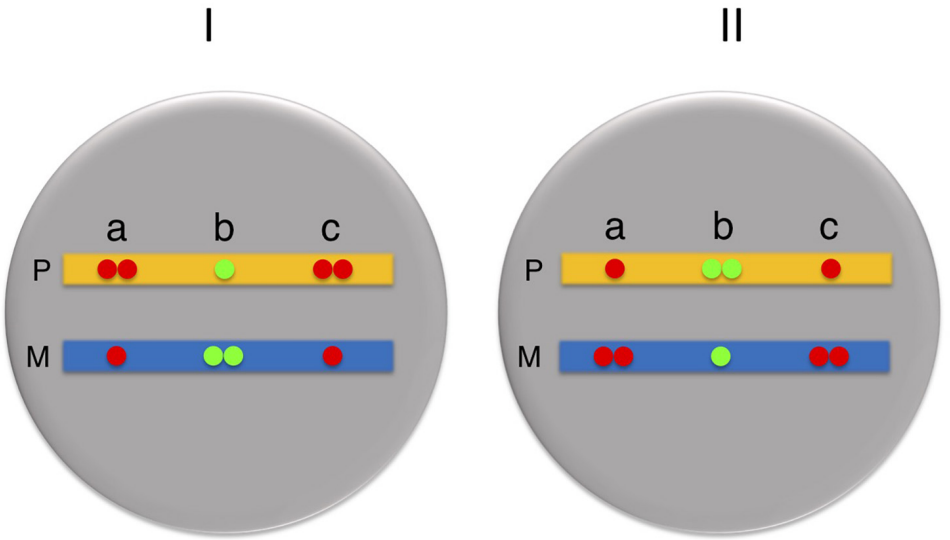


FIGURE 2 | Orientation of asynchronous regions. All cells in the body are in one of two mirror-image states (I or II) with regard to their Paternal (P, yellow) and Maternal (M, blue) alleles. Probes a and c always replicate in a parallel manner, with both being early on the same allele, while probes a and b always replicate in an antiparallel manner both in cell I and cell II.

it can provide a template for preferentially activating one allele as opposed to the other. Thus, the choice of allele is a built-in part of replication-time-directed genome structure (Klein et al., 2021), waiting to be utilized in a specific manner in individual cell types. There is thus no need for "choosing" in trans over and over again in each cell, since the decision process itself has already been pre-coded in cis in all somatic cells. This may be accomplished in a very simple and, in fact, fool-proof manner by pre-marking each allele separately in the gametes (i.e., in cells carrying only one allele) by an, as yet, unknown mechanism. This information is then employed to set up differential asynchronous replication in the implantation embryo.

It is still a mystery how the asynchronous replication state can be maintained through cell division and replication. Unlike most structural features defined by fixed epigenetic marks such as DNA methylation, maintenance of ASRT is complicated by the fact that this property can exist in either a switching or committed mode. For this reason, we suggest that replication timing may itself be an epigenetic feature that has an inherent mechanism that allows it to be autonomously perpetuated. A great deal of evidence indicates that the time of replication for each locus is set up during the G1 stage of the cell cycle and this is accomplished by the recruitment of protein complexes at all the coordinated replication origins in a given time zone (Goren and Cedar, 2003). This marking system provides information that determines the time of replication in S. In the case of ASRT, one allele is marked for early replication and the other for late. Since S-phase progression is associated with programmed changes in nuclear environment, each allele will encounter a different set of trans-acting factors, which may then mark the newly replicated allele as having been copied in either early or late S. A major candidate for this type of regulation is histone H3/H4 acetylation, which has already been shown to modify nucleosomes at the replication fork in an S-phase specific manner (Zhang et al., 2002; Lande-Diner et al., 2009). Following replication and cell division, this temporal-dependent feature can then be used for re-establishing the allele-specific time of replication during the next cell cycle, thus maintaining ASRT (**Figure 3**).

It should be noted that in all of the developmental systems where ASRT may play a role in allelic choice, ASRT does not seem to be a key element in the process of gene inactivation or activation itself, with this being accomplished by a variety of many different mechanisms that may include DNA methylation, histone modification, ncRNA, and others (Żylicz and Heard, 2020; Boeren and Gribnau, 2021). ASRT would simply serve as a means to mark the two alleles differently, thus enabling these other factors to operate on only one of the two copies. From this perspective, the underlying function of ASRT is the process of "allelic choice" itself.

BIOLOGICAL FUNCTION OF ASRT-BASED MONOALLELIC EXPRESSION

In order to put ASRT into a more biological perspective, it is worthwhile considering the potential functions of monoallelic choice and the possible molecular mechanisms that could mediate this process. From a careful analysis of the genes located in asynchronously replicating domains, it emerges that many of these regions include gene arrays, each of which contain a variety of alternate receptor genes that make up a reservoir from which each cell can uniquely choose one for presentation on the cell surface. Because the genome is diploid, this process would not only require the stochastic selection of a single receptor

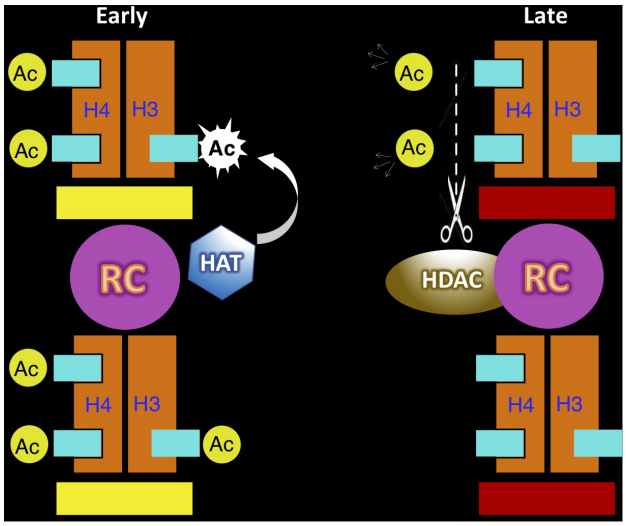


FIGURE 3 | Autonomous maintenance of asynchronous replication timing (ASRT). This model demonstrates how histone acetylation (Ac) at the replication origin may serve as a mark for the autonomous maintenance of replication timing on individual alleles. In cells set up to replicate one allele early (yellow bar) and the other late (red bar), the early allele is marked by acetylation of both histone H3 and H4 at the replication origin, while the late replicating allele lacks histone acetylation. During the next cycle the acetylated allele (left) will be recognized to undergo replication in early S, thus generating two copies of the DNA template, with the original nucleosome remaining on one copy **(bottom)** and a newly made nucleosome **(top)** placed on the other. Since histone H4 becomes acetylated immediately after its synthesis in the cytoplasm, this new nucleosome is already acetylated at these sites. Acetylation of histone H3 is carried out by a histone acetylase (HAT) that is associated with the replication machinery, but only in early S phase. Thus, following early replication the origin on both chromatids is now packaged with nucleosomes marked for early replication in the next cell cycle. The other allele in the same cell (right) is marked for late replication at the origin and is packaged with nucleosomes lacking histone acetylation. This allele is recognized for replication as the cell passes through early S phase, but will then undergoes replication during late S. The original nucleosome will remain with one of two DNA copies (bottom), while the other DNA template will get packaged with a new nucleosome, already acetylated on histone H4. During late S, the replication complex (RC) contains a cell-cycle-dependent histone deacetylase (HDAC) that can remove these H4 acetyl groups, thus guaranteeing that both chromatids will be packaged with un-acetylated histones and effectively regenerate the mark that directs late replication in the next cell cycle. It should be noted that this mechanism may also be able to accommodate the stem-cell switching mode of ASRT by automatically switching the acetylation state on each allele at the end of S phase in every cell cycle. As opposed to all other mechanisms for epigenetic maintenance which are carried out by "copying" specific marks, replication timing memory is time-based and takes advantage of differential cell-cycle properties.

gene within an array, but may also be dependent on a reliable mechanism for ensuring that only a single one of the two alleles is actually activated for transcription and it is likely that this choice is mediated by the ASRT-associated chromosomal and nuclear structural features that essentially make these loci epigenetically "monoploid." Indeed, because ASRT operates at the regional as opposed to local level, it is capable of carrying out a form of epigenetic regulation that is uniquely appropriate for controlling large gene arrays. Taken together, this developmental system provides a mechanism for the stable and reliable programming of choices within the immune, sensory and motility systems by defining cell identity.

HOLISTIC MODEL

Allelic choice by means of asynchronous replication timing may represent a subset of general strategies that utilize genomic imprinting. Extensive research on the mechanisms involved in imprinting have indicated that DNA methylation plays a prominent role by marking gene sequences in one of the gametes, thereby designating this allele as being inactive (Li et al., 1993). Because DNA methylation can be maintained autonomously at every cell division (Cedar and Bergman, 2012), this early generated mark is then remembered in cis throughout development, thereby perpetuating a decision that was initially made at a stage when both alleles were completely separated from each other. Independently of being epigenetically marked by DNA methylation (Gribnau et al., 2003), imprinted genes are clustered within asynchronous replication timing domains (May et al., 2008; Shufaro et al., 2010), with all tested cases showing a paternal early pattern (Simon et al., 1999), regardless of their expression profile. Another example of non-random allelic silencing is the paternal specific X-chromosome inactivation that takes place in extraembryonic tissues of the female mammal and in all cell types of marsupials (Migeon et al., 1989; Samollow et al., 1995). In the mouse, it has been demonstrated that this choice is associated with differential early replication of the paternal allele (Takagi and Sasaki, 1975).

It appears that random asynchronous replication timing and its association with monoallelic choice has many of the features associated with genomic imprinting (**Figure 4**). In both of these basic processes, regulation occurs at a regional level, involves allele differential replication and is faithfully maintained throughout development. As opposed to imprinting, which has a fixed parental orientation pattern, random ASRT allows the selection of either the maternal or paternal allele, but the strategy used to initially establish the differential state is essentially very similar in that it involves an early developmental choice of one fixed allele to be early replicating. The decision itself is actually initiated in the individual gametes, at a stage where there is only one allele, which is then epigenetically marked to dictate either early or late replication when ASRT is set up in the early embryo. This mechanism thus provides a simple and sophisticated system for avoiding having to choose between two identical alleles in a single cell. Following this initial step, the only difference between random ASRT and imprinting is the subsequent introduction of a switching mechanism, making it possible to get exclusive expression from either the maternal or paternal allele.

It is worthwhile noting that random X-inactivation in female embryos also occurs at about the time of implantation and generates some cells in which the maternal X is inactivated and others in which it is the paternal, with the inactive chromosome always being differentially late replicating (Mlynarczyk-Evans et al., 2006). It is very possible that this process is also part of

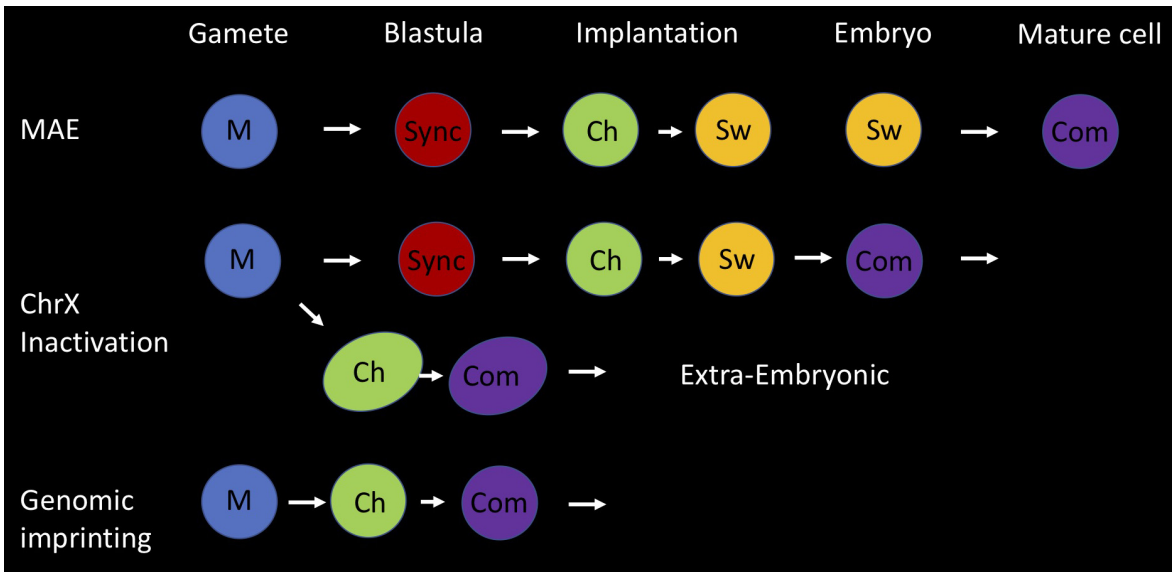


FIGURE 4 | Holistic model for asynchronous replication timing (ASRT). Three different forms of developmentally based monoallelic expression (MAE) are associated with regional asynchronous replication timing; genomic imprinting, inactivation of one X chromosome in female animals and random MAE of autosomal genes. Here we present an integrative developmental model for the establishment of allelic choice with its basic building blocks. This process is accomplished in three seminal steps; marking (M), choice (Ch), and commitment (Com). In all cases, the two alleles are marked (M) differentially in the gametes, where each parental allele is separated from its partner. One is marked as being of maternal origin and the other as being of paternal origin. For random MAE, the two alleles replicate synchronously (sync) in cells of the early pre-implantation embryo. Asynchronous replication is initiated at the stage of implantation, when one allele is chosen (Ch) to replicate early and the other late, based on the epigenetic mark (M) derived from the gametes. In all subsequent replication cycles, the two alleles switch (Sw) their time of replication in S phase, but each locus still retains its parallel or anti-parallel orientation relative to other ASRT loci in the genome. Switching continues in cells of the embryo until they commit (Com) to one fixed parental replication pattern. Imprinted gene regions are initially marked in the gametes, but immediately adopt allele-specific asynchronous replication timing in the early embryo and become committed (Com) to this fixed pattern in all cells, without going through a stage of switching. The inactivation of one copy of ChrX in female embryos utilizes both these pathways. After marking in the gametes, they establish an imprinted pattern of ASRT during formation of the extra-embryonic tissues, while they replicate synchronously in embryonic cells before setting up ASRT at the time of implantation and then undergoing switching to enable random X inactivation and commitment, post implantation.

the random allelic choice system that occurs on select autosomal regions. In keeping with this idea, it has been demonstrated that the two X-chromosomes actually replicate asynchronously in ES cells with an allelic pattern that is not preserved in single-cell clones, suggesting that this entire chromosome may be subject to allelic switching similar to what occurs in autosomal ASRT, thus explaining how one X in each cell is chosen for inactivation in the early embryo (Gribnau et al., 2005; Mlynarczyk-Evans et al., 2006).

Taken together, this suggests that X-chromosome inactivation constitutes a general prototype for both random and non-random allelic choice (**Figure 4**). One allele, the paternal, is initially marked for early replication and retains this memory to set up imprinted X inactivation in cells destined for the formation of extra-embryonic tissues. Further pre-implantation embryonic stages become subject to switching, which then serves as the basis for random inactivation that, in this case, becomes clonally committed shorty after implantation. Although the main role of ASRT on the X-chromosome presumably involves dosage compensation, it should be noted that at least at one locus this mechanism serves the more general function of defining single-cell specificity. The red and green pigment genes for color vision located in a small array on chromosome X are individually activated by a common enhancer sequence that can only choose one at a time (Wang et al., 1992). In males, where there is only a single X chromosome, this decision allows the generation of unique pigment cells (either red or green) in the retinal cone. In females, however, where there are two X-chromosomes, the production of two different pigments in the same cell is prevented by X-inactivation, in a manner that is very similar to the function of ASRT in many autosomal loci.

Monoallelic expression appears to constitute a fundamental aspect of mammalian biology and development, which by its very nature must utilize epigenetic regulation. In this perspective, we have proposed that asynchronous replication timing plays a unique role in the establishment and maintenance of allelic choice. Specific regions in the genome become differentially marked in the individual gametes and this feature is then used in the embryo as a blueprint for setting up structural allelic differences that are maintained in all cells of the body, where it can, if needed, enable allelic choice. Because this system is essentially based on preserving a "difference" between the alleles with an option to switch their identity, it can serve as a mechanism for both genomic imprinting as well as random MAE, processes that underlie both dosage compensation as well as the determination of cell identity (**Figure 4**).

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by research grants from the Israel Science Foundation (grants #734/13 and #1228/18 to YB, grant #282/16 to HC, grants #184/16 and #1283/21 to IS), ISF-NSFC (grant #2555/16 to IS), the Israel Cancer Research Foundation (IS and grant #211410 to YB, grant #210910 to HC), the Binational Science Foundation (grant #2100289 to YB, grant #2019688 to IS), the Emanuel Rubin Chair in Medical Sciences (YB), the Israel Center of Excellence Program (grant #1796/12 to YB), the German Israeli Foundation (grant #1424 to YB), and the Rosetrees Foundation (HC).

ACKNOWLEDGMENTS

We thank our students and collaborators who participated in this research, especially Hagit Masika, Marganit Farago, Britny Blumenfeld, Lamia Halaseh, and Rachel Rapoport, the members of the Core Research Facility in the Hebrew University School of Medicine; Dan Lehmann and Eleonora Medvedev for assistance with the FACS analysis; Idit Shiff, Abed Nasereddin, and Alexia Azoulay for generating genomic data.

REFERENCES

Alexander, M. K., Mlynarczyk-Evans, S., Royce-Tolland, M., Plocik, A., Kalantry, S., Magnuson, T., et al. (2007). Differences between homologous alleles of olfactory receptor genes require the polycomb Group protein Eed. *J. Cell Biol.* 179, 269–276. doi: 10.1083/jcb.200706053

Avner, P., and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. *Nat. Rev. Genet.* 2, 59–67. doi: 10.1038/35047580

Azuara, V., Brown, K. E., Williams, R. R. E., Webb, N., Dillon, N., Festenstein, R., et al. (2003). Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication. *Nat. Cell Biol.* 5, 668–674. doi: 10.1038/ncb1006

Blumenfeld, B., Masika, H., Farago, M., Yehuda, Y., Halaseh, L., Vardi, O., et al. (2021). Chromosomal coordination and differential structure of asynchronous replicating regions. *Nat. Commun.* 12:1035.

Boeren, J., and Gribnau, J. (2021). Xist-mediated chromatin changes that establish silencing of an entire X chromosome in mammals. *Curr. Opin. Cell Biol.* 70, 44–50. doi: 10.1016/j.ceb.2020.11.004

Branciamore, S., Valo, Z., Li, M., Wang, J., Riggs, A. D., and Singer-Sam, J. (2018). Frequent monoallelic or skewed expression for developmental genes in CNS-derived cells and evidence for balancing selection. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10379–E10386.

Cedar, H., and Bergman, Y. (2008). Choreography of Ig allelic exclusion. *Curr. Opin. Immunol.* 20, 308–317. doi: 10.1016/j.coi.2008.02.002

Cedar, H., and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. *Nat. Rev. Immunol.* 11, 478–488. doi: 10.1038/nri2991

Cedar, H., and Bergman, Y. (2012). Programming of DNA methylation patterns. *Annu. Rev. Biochem.* 81, 97–117. doi: 10.1146/annurev-biochem-052610-091920

Chess, A. (1998). Expansion of the allelic exclusion principle? *Science* 279, 2067–2068. doi: 10.1126/science.279.5359.2067

Chess, A. (2016). Monoallelic gene expression in mammals. *Annu. Rev. Genet.* 50, 317–327. doi: 10.1146/annurev-genet-120215-035120

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78, 823–834. doi: 10.1016/s0092-8674(94)90562-2

Coleclough, C. (1983). Chance, necessity and antibody gene dynamics. *Nature* 303, 23–26. doi: 10.1038/303023a0

Escobar, T. M., Oksuz, O., Saldaña-Meyer, R., Descostes, N., Bonasio, R., and Reinberg, D. (2019). Active and repressed chromatin domains exhibit distinct nucleosome segregation during DNA replication. *Cell* 179, 953–963. doi: 10.1016/j.cell.2019.10.009

Farago, M., Rosenbluh, C., Tevlin, M., Fraenkel, S., Schlesinger, S., Masika, H., et al. (2012). Clonal allelic predetermination of immunoglobulin-κ rearrangement. *Nature* 490, 561–565. doi: 10.1038/nature11496

Galupa, R., and Heard, E. (2018). X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. *Annu. Rev. Genet.* 52, 535–566. doi: 10.1146/annurev-genet-120116-024611

Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., et al. (2005). Epigenetic ontogeny of the Igk locus during B cell development. *Nat. Immunol.* 6, 198–203. doi: 10.1038/ni1154

Goldmit, M., Schlissel, M., Cedar, H., and Bergman, Y. (2002). Differential accessibility at the kappa chain locus plays a role in allelic exclusion. *EMBO J.* 21, 5255–5261. doi: 10.1093/emboj/cdf518

Goren, A., and Cedar, H. (2003). Replicating by the clock. *Nat. Rev. Mol. Cell. Biol.* 4, 25–32. doi: 10.1038/nrm1008

Gorman, J. R., and Alt, F. W. (1998). Regulation of immunoglobulin light chain isotype expression. *Adv. Immunol.* 69, 113–181. doi: 10.1016/S0065-2776(08)60607-0

Greenberg, A., Blumenfeld, B., and Simon, I. (2020). The delicate relationship between DNA replication timing and gene expression. *Curr. Opin. Syst. Biol.* 19, 8–15. doi: 10.1016/j.coisb.2020.06.001

Gribnau, J., Hochedlinger, K., Hata, K., Li, E., and Jaenisch, R. (2003). Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. *Genes Dev.* 17, 759–773. doi: 10.1101/gad.1059603

Gribnau, J., Luikenhuis, S., Hochedlinger, K., Monkhorst, K., and Jaenisch, R. (2005). X chromosome choice occurs independently of asynchronous replication timing. *J. Cell Biol.* 168, 365–373. doi: 10.1083/jcb.200405117

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951. doi: 10.1038/nature06947

Guo, L., Hu-Li, J., and Paul, W. E. (2005). Probabilistic regulation in TH2 cells accounts for monoallelic expression of IL-4 and IL-13. *Immunity* 23, 89–99. doi: 10.1016/j.immuni.2005.05.008

Hand, R. (1978). Eucaryotic DNA: organization of the genome for replication. *Cell* 15, 317–325. doi: 10.1016/0092-8674(78)90001-6

Hanna, C. W., and Kelsey, G. (2021). Features and mechanisms of canonical and noncanonical genomic imprinting. *Genes Dev.* 35, 821–834. doi: 10.1101/gad.348422.121

Heard, E. (2004). Recent advances in X-chromosome inactivation. *Curr. Opin. Cell Biol.* 16, 247–255. doi: 10.1016/j.ceb.2004.03.005

Heun, P., Laroche, T., Raghuraman, M. K., and Gasser, S. M. (2001). The positioning and dynamics of origins of replication in the budding yeast nucleus. *J. Cell Biol.* 152, 385–400. doi: 10.1083/jcb.152.2.385

Holmquist, G. P. (1987). Role of replication time in the control of tissue-specific gene expression. *Am. J. Hum. Genet.* 40, 151–173.

Jeppesen, P., and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic

marker for gene expression. *Cell* 74, 281–289. doi: 10.1016/0092-8674(93)90419-Q

Ji, Y., Zhang, J., Lee, A. I., Cedar, H., and Bergman, Y. (2003). A multistep mechanism for the activation of rearrangement in the immune system. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7557–7562. doi: 10.1073/pnas.0932635100

Kerem, B. S., Goitein, R., Diamond, G., Cedar, H., and Marcus, M. (1984). Mapping of DNAase I sensitive regions on mitotic chromosomes. *Cell* 38, 493–499. doi: 10.1016/0092-8674(84)90504-X

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., et al. (1993). Allele-specific replication timing of imprinted gene regions. *Nature* 364, 459–463. doi: 10.1038/364459a0

Klein, K. N., Zhao, P. A., Lyu, X., Sasaki, T., Bartlett, D. A., Singh, A. M., et al. (2021). Replication timing maintains the global epigenetic state in human cells. *Science* 372, 371–378. doi: 10.1126/science.aba5545

Koren, A., and McCarroll, S. A. (2014). Random replication of the inactive X chromosome. *Genome Res.* 24, 64–69. doi: 10.1101/gr.161828.113

Lande-Diner, L., Zhang, J., and Cedar, H. (2009). Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. *Mol. Cell* 34, 767–774. doi: 10.1016/j.molcel.2009.05.027

Latt, S. A. (1973). Microfluorometric detection of deoxyribonucleic acid replication in human metaphase chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 70, 3395–3399. doi: 10.1073/pnas.70.12.3395

Levin-Klein, R., and Bergman, Y. (2014). Epigenetic regulation of monoallelic rearrangement (allelic exclusion) of antigen receptor genes. *Front. Immunol.* 5:625. doi: 10.3389/fimmu.2014.00625

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365. doi: 10.1038/366362a0

Liang, H.-E., Hsu, L.-Y., Cado, D., and Schlissel, M. S. (2004). Variegated transcriptional activation of the immunoglobulin kappa locus in pre-b cells contributes to the allelic exclusion of light-chain expression. *Cell* 118, 19–29. doi: 10.1016/j.cell.2004.06.019

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* 126, 403–413. doi: 10.1016/j.cell.2006.06.035

Maass, P. G., Barutcu, A. R., and Rinn, J. L. (2019). Interchromosomal interactions: a genomic love story of kissing chromosomes. *J. Cell Biol.* 218, 27–38. doi: 10.1083/jcb.201806052

Marchal, C., Sima, J., and Gilbert, D. M. (2019). Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* 20, 721–737. doi: 10.1038/s41580-019-0162-y

Markenscoff-Papadimitriou, E., Allen, W. E., Colquitt, B. M., Goh, T., Murphy, K. K., Monahan, K., et al. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell* 159, 543–557. doi: 10.1016/j.cell.2014.09.033

Masika, H., Farago, M., Hecht, M., Condiotti, R., Makedonski, K., Buganim, Y., et al. (2017). Programming asynchronous replication in stem cells. *Nat. Struct. Mol. Biol.* 24, 1132–1138. doi: 10.1038/nsmb.3503

May, A., Reifenberg, K., Zechner, U., and Haaf, T. (2008). Asynchronous replication dynamics of imprinted and non-imprinted chromosome regions in early mouse embryos. *Exp. Cell Res.* 314, 2788–2795. doi: 10.1016/j.yexcr.2008.07.009

Migeon, B. R., Jan de Beur, S., and Axelman, J. (1989). Frequent derepression of G6PD and HPRT on the marsupial inactive X chromosome associated with cell proliferation in vitro. *Exp. Cell Res.* 182, 597–609. doi: 10.1016/0014-4827(89)90262-0

Mlynarczyk-Evans, S., Royce-Tolland, M., Alexander, M. K., Andersen, A. A., Kalantry, S., Gribnau, J., et al. (2006). X chromosomes alternate between two states prior to random X-inactivation. *PLoS Biol.* 4:e159. doi: 10.1371/journal.pbio.0040159

Mohandas, T., Sparkes, R. S., and Shapiro, L. J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 211, 393–396. doi: 10.1126/science.6164095

Mombaerts, P., Wang, F., Dulac, C., Chao, S. K., Nemes, A., Mendelsohn, M., et al. (1996). Visualizing an olfactory sensory map. *Cell* 87, 675–686. doi: 10.1016/S0092-8674(00)81387-2

Mostoslavsky, R., Alt, F. W., and Rajewsky, K. (2004). The lingering enigma of the allelic exclusion mechanism. *Cell* 118, 539–544. doi: 10.1016/j.cell.2004.08.023

Mostoslavsky, R., Singh, N., Kirillov, A., Pelanda, R., Cedar, H., Chess, A., et al. (1998). Kappa chain monoallelic demethylation and the establishment of allelic exclusion. *Genes Dev.* 12, 1801–1811. doi: 10.1101/gad.12.12.1801

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., et al. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature* 414, 221–225. doi: 10.1038/35102606

Mutzel, V., and Schulz, E. G. (2020). Dosage sensing, threshold responses, and epigenetic memory: a systems biology perspective on random X-chromosome inactivation. *Bioessays* 42:e1900163. doi: 10.1002/bies.201900163

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

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature* 379, 131–137. doi: 10.1038/379131a0

Reverón-Gómez, N., González-Aguilera, C., Stewart-Morgan, K. R., Petryk, N., Flury, V., Graziano, S., et al. (2018). Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication. *Mol. Cell* 72, 239–249. doi: 10.1016/j.molcel.2018.08.010

Rhind, N., and Gilbert, D. M. (2013). DNA replication timing. *Cold Spring Harb. Perspect. Biol.* 5:a010132. doi: 10.1101/cshperspect.a010132

Rivera-Mulia, J. C., Dimond, A., Vera, D., Trevilla-Garcia, C., Sasaki, T., Zimmerman, J., et al. (2018). Allele-specific control of replication timing and genome organization during development. *Genome Res.* 28, 800–811. doi: 10.1101/gr.232561.117

Rodriguez, I. (2013). Singular expression of olfactory receptor genes. *Cell* 155, 274–277. doi: 10.1016/j.cell.2013.09.032

Sahakyan, A., Plath, K., and Rougeulle, C. (2017). Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160363. doi: 10.1098/rstb.2016.0363

Samollow, P. B., Robinson, E. S., Ford, A. L., and Vandeberg, J. L. (1995). Developmental progression of Gpd expression from the inactive X chromosome of the Virginia opossum. *Dev. Genet.* 16, 367–378. doi: 10.1002/dvg.1020160410

Savol, A. J., Wang, P. I., Jeon, Y., Colognori, D., Yildirim, E., Pinter, S. F., et al. (2017). Genome-wide identification of autosomal genes with allelic imbalance of chromatin state. *Plos One* 12:e0182568. doi: 10.1371/journal.pone.0182568

Schübeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., and Groudine, M. (2002). Genome-wide DNA replication profile for *Drosophila* melanogaster: a link between transcription and replication timing. *Nat. Genet.* 32, 438–442. doi: 10.1038/ng1005

Selig, S., Okumura, K., Ward, D. C., and Cedar, H. (1992). Delineation of DNA replication time zones by fluorescence in situ hybridization. *EMBO J.* 11, 1217–1225. doi: 10.1002/j.1460-2075.1992.tb05162.x

Shufaro, Y., Lacham-Kaplan, O., Tzuberi, B.-Z., McLaughlin, J., Trounson, A., Cedar, H., et al. (2010). Reprogramming of DNA replication timing. *Stem Cells* 28, 443–449. doi: 10.1002/stem.303

Siefert, J. C., Georgescu, C., Wren, J. D., Koren, A., and Sansam, C. L. (2017). DNA replication timing during development anticipates transcriptional programs and parallels enhancer activation. *Genome Res.* 27, 1406–1416. doi: 10.1101/gr.218602.116

Simon, I., Tenzen, T., Reubinoff, B. E., Hillman, D., McCarrey, J. R., and Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. *Nature* 401, 929–932. doi: 10.1038/44866

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* 256, 640–642. doi: 10.1038/256640a0

van Steensel, B., and Belmont, A. S. (2017). Lamina-associated domains: links with chromosome architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780–791. doi: 10.1016/j.cell.2017.04.022

Vogel, M. J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* 2, 1467–1478. doi: 10.1038/nprot.2007.148

Wang, Y., Macke, J. P., Merbs, S. L., Zack, D. J., Klaunberg, B., Bennett, J., et al. (1992). A locus control region adjacent to the human red and green visual pigment genes. *Neuron* 9, 429–440. doi: 10.1016/0896-6273(92)90181-C

Yancopoulos, G. D., and Alt, F. W. (1986). Regulation of the assembly and expression of variable-region genes. *Annu. Rev. Immunol.* 4, 339–368. doi: 10.1146/annurev.iy.04.040186.002011

Zhang, J., Xu, F., Hashimshony, T., Keshet, I., and Cedar, H. (2002). Establishment of transcriptional competence in early and late S phase. *Nature* 420, 198–202. doi: 10.1038/nature01150

Żylicz, J. J., and Heard, E. (2020). Molecular mechanisms of facultative heterochromatin formation: an X-chromosome perspective. *Annu. Rev. Biochem.* 89, 255–282. doi: 10.1146/annurev-biochem-062917-012655

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: 13 October 2021
doi: 10.3389/fcell.2021.702897

A Comprehensive Characterization of Monoallelic Expression During Hematopoiesis and Leukemogenesis *via* Single-Cell RNA-Sequencing

Ruiqing Fu[1,2,3], Pengfei Qin[4], Xianghui Zou[3], Zhangli Hu[1,2], Ni Hong[4], Yun Wang[1]* and Wenfei Jin[4]**

[1] Shenzhen Key Laboratory of Microbiology and Gene Engineering, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, [2] Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province, College of Optoelectronic Engineering, Shenzhen University, Shenzhen, China, [3] School of Food Engineering and Biotechnology, Hanshan Normal University, Chaozhou, China, [4] Shenzhen Key Laboratory of Gene Regulation and Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China

OPEN ACCESS

Edited by:
Jin Xu,
Sun Yat-sen University, China

Reviewed by:
Srimonta Gayen,
Indian Institute of Science (IISc), India
Xu Xiaojiang,
National Institute of Environmental Health Sciences (NIEHS), United States

***Correspondence:**
Ni Hong
hongn@mail.sustech.edu.cn
Yun Wang
yunw@szu.edu.cn
Wenfei Jin
jinwf@sustech.edu.cn

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

Received: 30 April 2021
Accepted: 13 September 2021
Published: 13 October 2021

Citation:
Fu R, Qin P, Zou X, Hu Z, Hong N, Wang Y and Jin W (2021) A Comprehensive Characterization of Monoallelic Expression During Hematopoiesis and Leukemogenesis via Single-Cell RNA-Sequencing. Front. Cell Dev. Biol. 9:702897. doi: 10.3389/fcell.2021.702897

Single-cell RNA-sequencing (scRNA-seq) is becoming a powerful tool to investigate monoallelic expression (MAE) in various developmental and pathological processes. However, our knowledge of MAE during hematopoiesis and leukemogenesis is limited. In this study, we conducted a systematic interrogation of MAEs in bone marrow mononuclear cells (BMMCs) at single-cell resolution to construct a MAE atlas of BMMCs. We identified 1,020 constitutive MAEs in BMMCs, which included imprinted genes such as *MEG8*, *NAP1L5*, and *IRAIN*. We classified the BMMCs into six cell types and identified 74 cell type specific MAEs including *MTSS1*, *MOB1A*, and *TCF12*. We further identified 114 random MAEs (rMAEs) at single-cell level, with 78.1% single-allele rMAE and 21.9% biallelic mosaic rMAE. Many MAEs identified in BMMCs have not been reported and are potentially hematopoietic specific, supporting MAEs are functional relevance. Comparison of BMMC samples from a leukemia patient with multiple clinical stages showed the fractions of constitutive MAE were correlated with fractions of leukemia cells in BMMCs. Further separation of the BMMCs into leukemia cells and normal cells showed that leukemia cells have much higher constitutive MAE and rMAEs than normal cells. We identified the leukemia cell-specific MAEs and relapsed leukemia cell-specific MAEs, which were enriched in immune-related functions. These results indicate MAE is prevalent and is an important gene regulation mechanism during hematopoiesis and leukemogenesis. As the first systematical interrogation of constitutive MAEs, cell type specific MAEs, and rMAEs during hematopoiesis and leukemogenesis, the study significantly increased our knowledge about the features and functions of MAEs.

Keywords: single-cell RNA sequencing, constitutive monoallelic expression, random monoallelic expression, bone marrow mononuclear cells, leukemia

INTRODUCTION

Mammalian genomes including human genome are diploid, with one haploid inherited from mother and the other inherited from father. Although it is usually assumed that genes are expressed from both alleles of the diploid genome, some genes are expressed from only one allele, which is called monoallelic expression (MAE) (Eckersley-Maslin and Spector, 2014a; Reinius and Sandberg, 2015; Chess, 2016; Han et al., 2020). One kind of the most studied MAE is genomic imprinting, in which either the paternal or the maternal allele of imprinted genes is expressed. The parental-origin-specific MAEs of imprinted genes have been demonstrated to play an important role in embryonic development (Reik and Walter, 2001; Ferguson-Smith, 2011). However, the constitutive MAEs such as genomic imprinting only account for a small fraction of total MAEs. Random MAEs (rMAEs), that stochastically determine one allele to be transcribed and lead to different cells of the organism expressing different alleles, are much prevalent (Gimelbrant et al., 2007; Deng et al., 2014; Reinius and Sandberg, 2015; Chess, 2016). The earliest reported rMAE was random X-chromosome inactivation that was described >60 years ago (Lyon, 1961). X-chromosome inactivation mainly balances X-chromosome gene dosages between male and female, which carry one and two copies of X-chromosome, respectively (Lyon, 1986). In contrast to chromosome-wide rMAE caused by random X-chromosome inactivation, autosomal rMAE on immunoglobulins and odorant receptors has been well studied in the past decades (Pernis et al., 1965; Hozumi and Tonegawa, 1976; Chess et al., 1994). A lot of autosomal rMAEs interspersing over the genome was detected in recent decade (Gimelbrant et al., 2007; Deng et al., 2014; Reinius and Sandberg, 2015). Recent studies showed that a considerable proportion of the rMAE should be attributed to RNA transcriptional bursting, which describes the switching kinetics of the two alleles expressing periodically (Kim and Marioni, 2013; Choi et al., 2019; Larsson et al., 2019; Ochiai et al., 2020). However, the genome-wide landscape of autosomal rMAE in hematopoiesis is largely unexplored.

Single-cell RNA-sequencing (scRNA-seq) provides a unique opportunity to analyze rMAE genome wide (Gimelbrant et al., 2007; Deng et al., 2014; Borel et al., 2015). Tools have been designed to perform rMAE analysis originally for full-length scRNA-seq data, simultaneously dealing with the transcriptional bursting, e.g., SCALE (Jiang et al., 2017) and scBase (Choi et al., 2019), but they are not suitable for 3′-scRNA-seq data. Analyses of rMAE in different cell lineages/types suggest that rMAE is established during development (Eckersley-Maslin et al., 2014b; Gendrel et al., 2014). However, the reported fractions of autosomal rMAE in human genome are quite different from study to study, ranging from 5 to 76.4% (Gimelbrant et al., 2007; Deng et al., 2014; Borel et al., 2015; Kim et al., 2015; Reinius et al., 2016). The contradictions between these studies may be caused by different cell lineages/types and false positives of rMAE identifications in these studies. Several studies have explored the relationships between MAE and tumor (Meehan et al., 2007; Walker et al., 2012; Polson et al., 2013; Al Seraihi et al., 2018; Silcock et al., 2019). For example, MAE of *TP53* was observed in mutated brain tumors while not in healthy tissues, indicating MAE potentially is associated with tumor progression (Walker et al., 2012). However, these studies only analyzed a very limited number of cells and did not conduct systematic analysis on MAE. In order to systematically characterize the MAEs during hematopoiesis and leukemogenesis, we identified and analyzed the constitutive MAEs, cell type specific MAEs, and rMAEs using large scale scRNA-seq data.

MATERIALS AND METHODS

Sample Information

The sample information and scRNA-seq data have been described in our recent study (Qin et al., 2021). In short, bone marrow mononuclear cells (BMMCs) were collected from a boy diagnosed with acute lymphoblastic leukemia (ALL) separately at four clinical time points, i.e., diagnosis, refractory, complete remission, and relapse. In addition, the whole-genome sequencing (WGS) data were generated from the boy's saliva sample and BMMC samples from the four time points, except the complete remission stage (Zhang et al., 2018).

Identification of Genomic Single-Nucleotide Variant and Filtering

Reads from WGS data were trimmed using cutadapt (Martin, 2011), and then mapped to the hg38 human reference genome with BWA (Li and Durbin, 2010). We used CNVnator (Abyzov et al., 2011) to call copy number variations (CNVs) in each of the samples, with default parameters. GATK best practice pipeline (McKenna et al., 2010; DePristo et al., 2011) was applied to process the duplicate-marked raw reads to analysis-ready mapped reads. HaplotypeCaller mode of the GATK was performed for each of the samples and then joint calling was conducted across the samples. Low-quality (QUAL $\leq$ 30) single nucleotide variants (SNVs) were removed and only autosomal bi-allelic SNVs were kept. To avoid the *cis*-influence from CNVs, we removed the SNVs located in the detected CNV regions for each sample. We also removed the SNVs that were not in dbSNP (v147). Finally, we removed the putative somatic mutations. According to the empirical data, a SNV was identified as a somatic mutation if its UMI count and percentage of the alternative-allele (alt-allele) were not larger than 10 and less than 40%, respectively.

Single-Cell RNA-Sequencing Data Process and Cell-Type Inference

The scRNA-seq raw data were processed following 10X Genomics workflow, using Cell Ranger (suite 2), with hg38 human reference genome. The basic transcriptomic analyses have been described in our recent study (Qin et al., 2021), namely, filtering cells, inferring major cell types, and identifying the cell states (i.e., normal cells or leukemia cells) in BMMCs.

The identified SNVs in WGS data were examined in mapped reads of scRNA-seq data, as well as the information of cell barcode and UMI in matched reads. Thus, it yielded the allelic UMI counts for each given SNV for each cell. The variant allele frequency (VAF) of alt-allele can be estimated directly by calculating the

fraction of UMIs of alt-allele. When the reads were extracted from the bam files, including both WGS data and scRNA-seq data, only the ones with a Phred score larger than 30 at the given SNV position were kept for further calculation.

Dimension Reduction and T-Distribution Stochastic Neighbor Embedding Projection

Dimension reduction was performed by principal component analysis (PCA) and visualized by t-distribution stochastic neighbor embedding (tSNE), following our previous study (Qin et al., 2021). The cells were colored accordingly to the inferred cell types, sample stages, or cell states. When displaying the expression pattern, highlighted cells were colored according to the allele expression, with their size scaled to $\log_{10}$ of the UMI counts.

Identification of Monoallelic Expressions

Cells from each sample, each cell type, or each subpopulation (e.g., Norm) were pooled together to detect the constitutive MAEs, in a way that the common concerns for the scRNA-seq data, e.g., allelic drop-outs (ADOs), noise, and sparseness, were largely alleviated or canceled out (Borel et al., 2015; Castel et al., 2015). To increase the statistical power and reduce the false positives, SNVs observed in at least 10 cells were used for further analyses. We first identified the SNVs showing significantly biased allele expression against the expected balanced expression (by χ^2-test). We further defined the SNVs showing serious deviation, in which UMI fractions of the minor allele were <5%, as constitutive MAE while other SNVs showing mild biased allelic expression were defined as allelic imbalanced expression (AIM). The constitutive MAEs in BMMCs were excluded from the cell-type-specific MAEs.

To detect random MAE (rMAE) at single-cell level, we only consider the SNV supported by >5 UMIs in a cell (i.e., "qualified" cell), thus the observed MAE of a SNV was not caused by chance, under an assumption of the binomial process ($p < 0.05$). This criterion leads to exclusion of a lot of SNVs and cells, leaving the SNVs possibly representing moderately and highly expressed genes, which are less affected by the technical variations (Deng et al., 2014; Kim et al., 2015; Zhao et al., 2017; Fan et al., 2018; Stamoulis et al., 2019) and undergoing relative fast transcriptional bursting (Kim and Marioni, 2013; Reinius and Sandberg, 2015; Stamoulis et al., 2019). A SNV was identified as single-cell MAE if its UMI of the minor allele was less than 1 or less than 5% of its total UMI counts of the two alleles, following the previous study (Reinius et al., 2016). The rMAE was defined as MAE excluding the constitutive MAEs and cell type specific MAEs. The fraction of rMAEs per cell was calculated by rMAE number dividing by the number of SNVs passed the "5-UMI" criterion. The cell fraction of a rMAE was measured by the proportion of the cells that monoallelically expressed the certain allele among the qualified cells.

Permutation of Random MAEs

To address the contribution of randomness in the observed rMAE, we permuted the observed alleles of each SNV across observed cells to calculate the expected proportion of single-cell rMAEs. More specifically, for each SNV, we pooled the allele UMIs across the observed cells together, from which allele UMIs were sampled into each cell according to its original count. Then, we used the same criterion to identify the expected rMAEs in cells. The same procedure was used to test the significance of biallelic mosaic rMAEs in balanced expressed SNVs, the two alleles of which were not significantly biased in pooled cells ($p > 0.05$; χ^2-test). All the permutations in the analysis were done by 1,000 times.

Detection of Leukemia-Specific Monoallelic Expressions

Pairwise comparisons were conducted to detect the leukemia-specific MAEs among three cell subpopulations (i.e., Norm, preR.Leuk, and postR.Leuk). For each pair (e.g., preR.Leuk comparing with Norm), we first selected the MAEs only in the test cells (e.g., preR.Leuk), and then tested if two alleles of each MAE were expressed with significant difference between the two cell subpopulations, by Fisher's exact test ($p < 0.05$). For detection of the leukemia-differentiated rMAEs in single cells, we only included the rMAEs that were shared between the comparing pairs. Cell numbers of the rMAEs and non-MAEs in each cell subpopulation were pair-wise compared by Fisher's exact test ($p < 0.05$).

Annotation and Enrichment Analysis

The SNVs were annotated by ANNOVAR (Wang et al., 2010) with relevant databases and assigned to genes according to their locations within the gene region. The gene enrichment analyses were performed by Metascape with default parameters and background gene set (Zhou et al., 2019)[1]. For cell type specific MAEs, the genes that were expressed in cells of the corresponding cell type were chosen as the background gene set, e.g., B cells.

Statistical Analysis

All the statistical analyses in the study were conducted in R, and if not specified, the Fisher's exact test was applied. When it was necessary, the BH method (Benjamini and Hochberg, 1995) was used for multiple test corrections.

Data Availability Statement

Publicly available datasets were analyzed in this study. These data can be found here: https://ngdc.cncb.ac.cn/, *HRA000084* and *CRA000588*. The code used in this study has been deposited in https://github.com/faculty/MonoAlleleExpr.

RESULTS

Identification of Constitutive Monoallelic Expression in Bone Marrow Mononuclear Cells

The BMMCs were obtained from a boy diagnosed with acute lymphoblastic leukemia (Qin et al., 2021). The BMMCs from

[1]https://metascape.org

the boy at complete remission are treated as normal BMMCs for analyzing MAE during hematopoiesis. After series of quality control, 7,016 cells were left for further analyses. The boy's saliva sample was used for WGS (Zhang et al., 2018). SNVs were identified in WGS data using GTAK (McKenna et al., 2010; DePristo et al., 2011). We further filtered out SNVs by the following three conditions: (1) SNVs in CNV regions; (2) SNVs not in dbSNP database; and (3) SNVs detected in less than five cells. Finally, we obtained 83,174 SNVs for MAE analyses, with a median number of 287 SNVs per cell (**Supplementary Table 1**). For each SNV, the allele that is the same as the reference is called a ref-allele, while the other allele is called an alt-allele. The number of UMI was used to represent the expression level of each allele.

The distribution of variant allele frequency (VAF) estimated by UMI fraction across all cells was almost symmetrically centered in 0.5 (**Figure 1A** and **Supplementary Figure 1**). There are increased SNVs at both tails of the VAF distribution, suggesting the existence of biased allelic expression. We further separate the biased allelic expression into mildly biased allelic expressions [allelic imbalanced expression (AIM)] ($p < 0.05$; χ^2-test) and strongly biased allelic expression with UMI fractions of the minor allele <5%. The strongly biased allelic expressions are constitutive MAEs across the BMMCs, accounting for 2.18% of the SNVs (**Figure 1B** and **Supplementary Table 2**). The constitutive MAEs contained several imprinted genes, such as *MEG8*, *NAP1L5*, and *IRAIN*; e.g., rs143537461 (C/A) located on imprinted gene *MEG8*, while only ref-allele (C) is exclusively expressed in BMMCs (**Figure 1C**). In addition to the imprinted genes, most of the detected constitutive MAEs are novel, indicating the existence of many hematopoiesis specific MAEs. For example, *RPS14*, showing strong constitutive MAE of reference allele (**Figure 1D**), is associated with hematopoiesis, particularly erythropoiesis (Wang et al., 2014; Schneider et al., 2016). *BRD2*, showing strong constitutive MAE of alterative allele (**Figure 1E**), is located in the MHC class II region and regulates the expression of many genes involved in immune pathways (Wang et al., 2021). GO enrichment analysis of constitutive MAE showed the immune relevant functional categories are significantly enriched (**Figure 1F**); e.g., "immune response-regulating signaling pathway" ($p = 2.78e^{-8}$) and "adaptive immune system" ($p = 1.27e^{-5}$).

Constitutive Monoallelic Expressions in Major Cell Types of Bone Marrow Mononuclear Cells

We classified the BMMCs into six major cell types: T cells (38.33%; *CD3D*, *CD3E*, and *CD3G*), B cells (35.31%; *CD79A*, *CD79B*, and *CD19*), natural killer (NK) cells (9.21%; *FCGR3A* and *NCAM1*), myelocytes/monocytes (Mye/Mono; 8.55%; *LYZ*, *CD14*, and *CD68*), erythroid cells (Ery; 5.97%; *HBB* and *HBA2*), and hematopoietic stem and progenitor cells (HSPC; 2.64%; *CD34* and *AVP*) (**Figure 2A**). We then identified the constitutive MAEs in each of the six hematopoietic cell types. Interestingly, the majority of constitutive MAEs identified in each cell type were overlapped with that in BMMCs (**Figure 2B** and **Supplementary Table 3**), indicating MAEs are either conserved during development or highly shared between/among different cell types. These cell type shared constitutive MAEs include *HLA-DQB2* (B cells), *IL32* (T cells), and *SERPINA1* (Mye/Mono). For example, *SERPINA1*, identified as constitutive MAEs in BMMCs and only expressed in Mye/Mono (**Figure 2C**), participates in the monocyte recruitment and proinflammatory activation (Moraga et al., 2001; Janciauskiene et al., 2007). *IL32*, identified as constitutive MAEs in BMMCs, T cells, and NK cells (**Supplementary Figure 2A**), is a cytokine involved in inflammation and cancer development.

Monoallelic expressions that are identified in specific cell types but not in the BMMCs constitutive MAEs are called as cell type specific MAEs. There are only a few MAEs shared among these cell types (**Figure 2D**). GO enrichment analysis of B cell specific MAEs showed that they were enriched in the immune process including "TNF-α signaling pathway" ($p = 1.66e^{-3}$) and "positive regulation of NF-κB transcription factor activity" ($p = 6.25e^{-3}$) (**Figure 2E**). These cell type specific MAEs include *HLA-DRB5* (HSPC and Mye/Mono), *ZNF83* (Ery), *NUP210* (T cells), *MTSS1* (NK cells), *MOB1A* (Mye/Mono), and *RFTN1* and *TCF12* (B cells). For example, *MTSS1*, showing NK cell specific MAE (**Figure 2F**), is a tumor suppressor gene in leukemia (Yu et al., 2012; Schemionek et al., 2016) and plays an important role in the development of B cells (Yu et al., 2012). *MOB1A*, showing Mye/Mono cell specific MAE (**Figure 2G**), involves in the regulation of organ size and tumor growth by enhancing apoptosis. *TCF12*, showing B cell specific MAE (**Figure 2H**), is a transcription factor that regulates gene expression during hematopoiesis. *HLA-DRB5*, which plays an important role in antigen presentation, shows HSPCs and Mye/Mono cell specific MAE (**Supplementary Figure 2B**). *NUP210*, as a cell-intrinsic regulator of TCR signaling and T cell homeostasis (Borlido et al., 2018), shows T cell specific MAE (**Supplementary Figure 2C**).

Identification of Random MAEs at Single-Cell Level

The scRNA-seq is a powerful approach to systematically analyze rMAEs in BMMCs. After strict quality control, we identified 114 rMAEs in BMMCs at single-cell level, accounting for 20–40% of the highly expressed genes (**Figure 3A**), giving rise to 7.29% SNVs showed rMAE per cell (**Figure 3B**), which is a little lower than other studies (Deng et al., 2014; Reinius et al., 2016; Savova et al., 2016), possibly due to our strict criteria (see "MATERIALS AND METHODS"). It is interesting to examine to which extend the observed rMAEs could be explained "by chance". We permuted (1,000 times) the alleles of each SNV by sampling from the pooled UMIs across all cells, which resulted in 3.25% SNVs showing rMAE per cell on average (**Figure 3B**). Therefore, more than half (55.39%) of the rMAEs in real data were not observed by chance.

The rMAEs were further divided into single-allele rMAE and biallelic mosaic rMAE, with percentages of 78.1% and 21.9%, respectively (**Figures 3C,D**). The fractions of cells showing rMAEs vary a lot among different single-allele rMAEs (**Figures 3C,D**). Further investigation showed that most of the

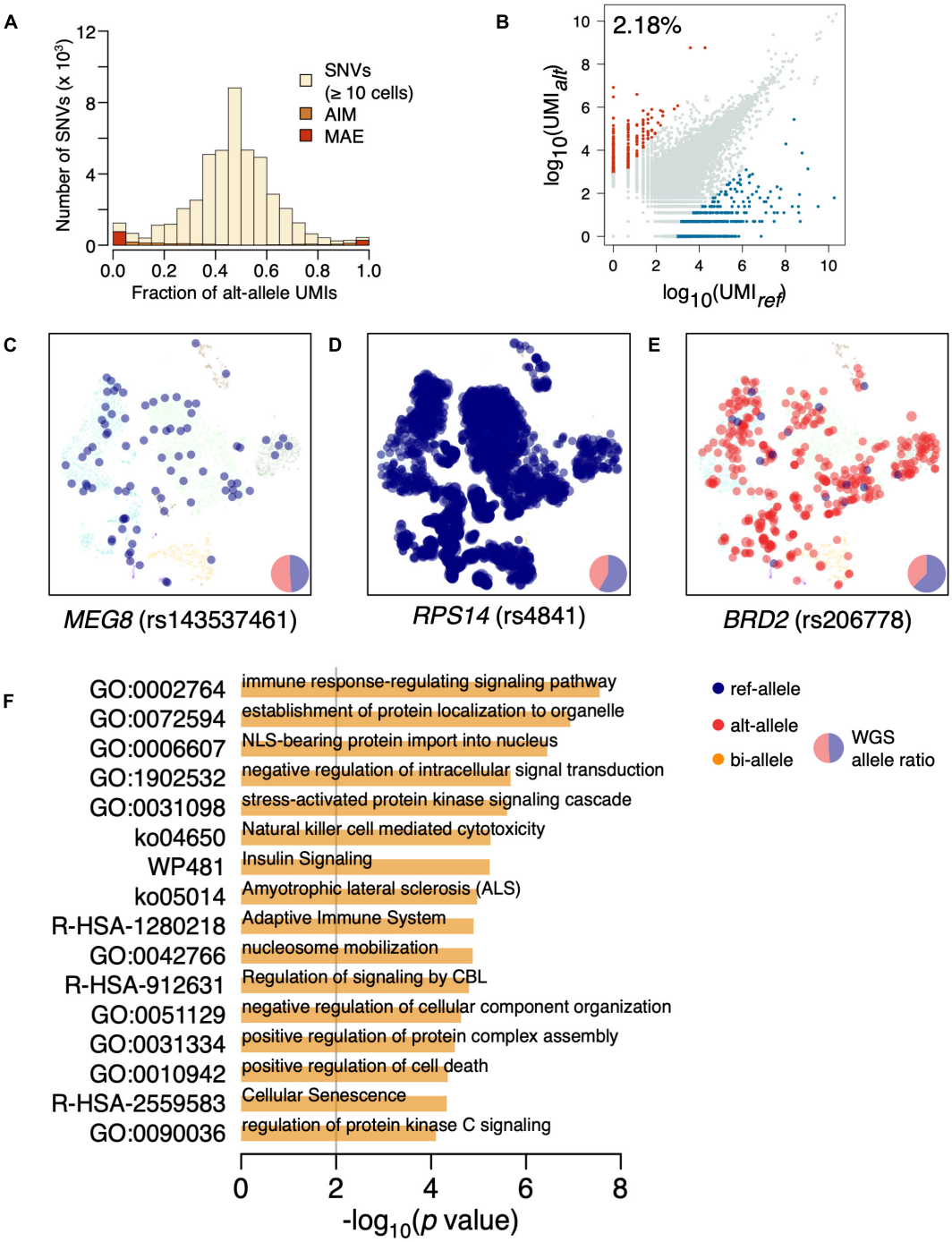


FIGURE 1 | Identification of constitutive MAEs in BMMCs. **(A)** Histogram of VAF in BMMCs estimated by UMI counts. AIM expression and MAE are highlighted. **(B)** Identification of constitutive MAEs, with blue representing MAEs of reference allele while red representing MAEs of the alternative allele. The percentage of the constitutive MAE is shown on the top left of the plot. **(C)** tSNE projection of BMMCs (background), colored by expressed allele of *MEG8*. The size of the highlighted dot is scaled to $\log_{10}$ of UMI count. The pie chart in the bottom right shows the genetic allelic ratio of the two alleles from whole-genome sequencing (WGS) data. **(D)** tSNE projection of BMMCs, colored by expressed allele of *RPS14*. **(E)** tSNE projection of BMMCs, colored by expressed allele of *BRD2*. **(F)** GO enrichment analysis of the constitutive MAEs in BMMCs.

rMAEs of high cell fractions, which accounted for 36.0% of the single-allele rMAEs, were observed in only one qualified cell, but with UMI counts ranging from 6 to as high as 285 (**Figure 3D**). The single-allele rMAEs include *CD52* (rs1071849), *SNHG5* (rs1059307), and *HLA-B* (rs2769), e.g., 9.2% of the cells show rMAE on *CD52* (**Figure 3E**). For biallelic mosaic rMAE,

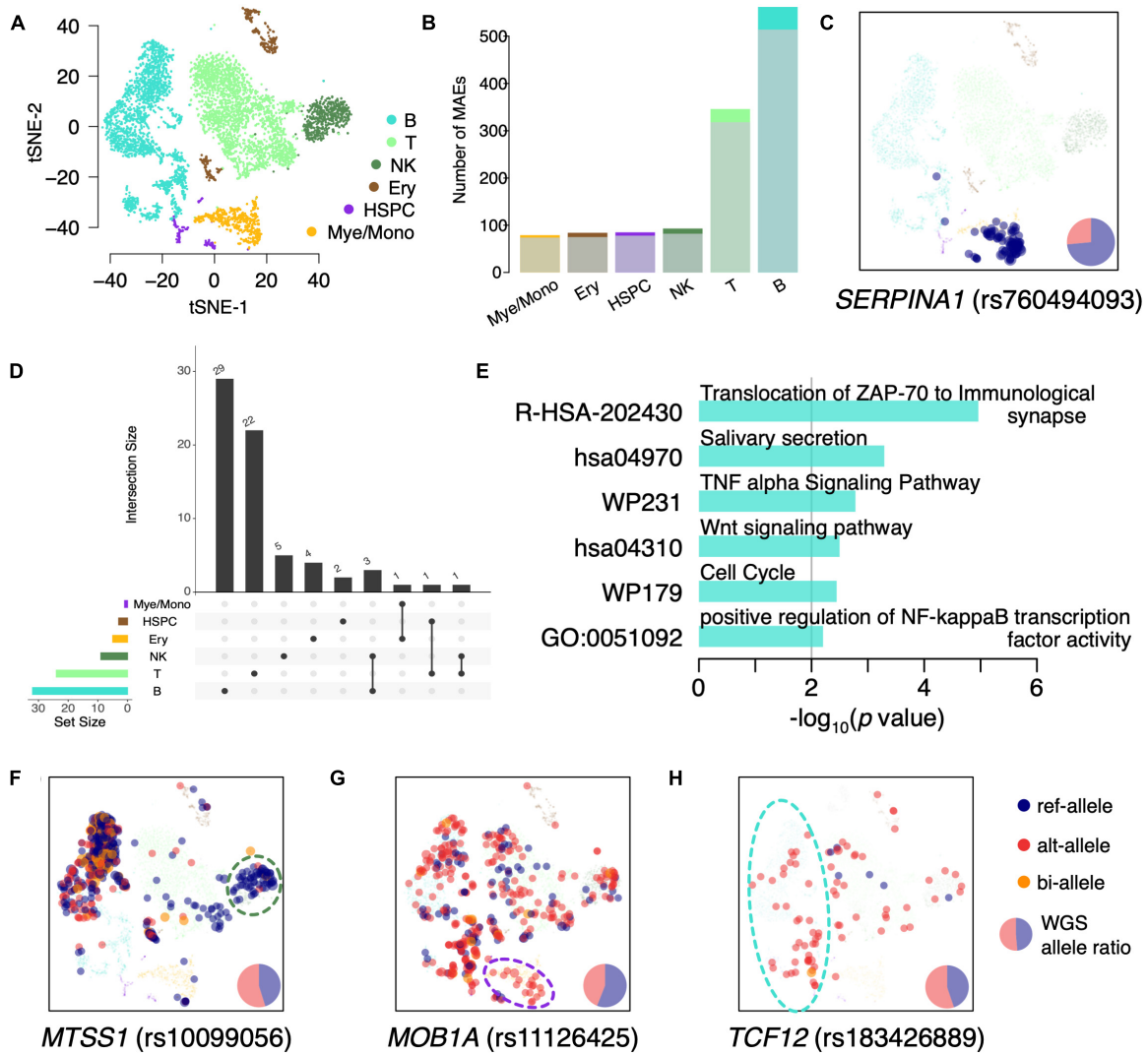


FIGURE 2 | Identification of cell type specific MAEs in BMMCs. **(A)** tSNE projection of BMMCs, colored by inferred major cell types. **(B)** Detected number of MAEs in each major cell type. The shadowed (bottom) and bright (top) colors indicate the constitutive MAEs in BMMCs and cell type specific MAEs, respectively. **(C)** tSNE projection of BMMCs, colored by expressed allele of *SERPINA1*. **(D)** The upset plot of the cell type specific MAEs. **(E)** GO enrichment analysis of the B cell specific MAEs. **(F–H)** tSNE projection of BMMCs, colored by the expressed allele of *MTSS1* **(F)**, *MOB1A* **(G)**, and *TCF12* **(H)**.

the fraction of the cells showing rMAE is low or intermediate thus has not been detected in constitutive MAEs (**Figures 3C,E**). For example, the fractions of cells showing ref-allele rMAE and alt-allele rMAE at *ELK2AP* (rs1059713) are 16.2% and 11.7%, respectively (**Figure 3E**). The fractions of cells showing ref-allele rMAE and alt-allele rMAE at *CLEC11A* (rs13866) are 18.8% and 25%, respectively (**Figure 3E**). Furthermore, we permutated alleles of the biallelic mosaic rMAE and got rMAEs ranging from 2 to 18, which is significantly less than that of empirical value ($n = 19$) ($p < 0.001$; permutation test) (**Figure 3F**), indicating that the biallelic mosaic rMAEs were not observed by chance.

We then interrogated the rMAE by the cell type. As the largest cell group, we detected 59 rMAEs in B cells, while there were 34 rMAEs in the smallest cell group (HSPCs), which was the same as that in T cells ($n = 34$). Among the 114 rMAEs, about one-third ($n = 39$) was shared by at least two cell types (**Figure 3G**). Comparing with constitutive MAEs, the rMAEs are more shared between cell types, indicating they are less cell type specific or the stochasticity to increase the cell heterogeneity, despite that they represent a range of highly expressed genes.

Leukemia Cells Showing Increased Constitutive Monoallelic Expressions and Random MAEs

In addition to analyzing "normal" BMMCs, the BMMCs at diagnosis, refractory, and relapse of the same boy were analyzed for studying the changes of MAEs in leukemia (**Supplementary Table 1**). Interestingly, analysis of the four samples showed that fractions of leukemia cells were correlated

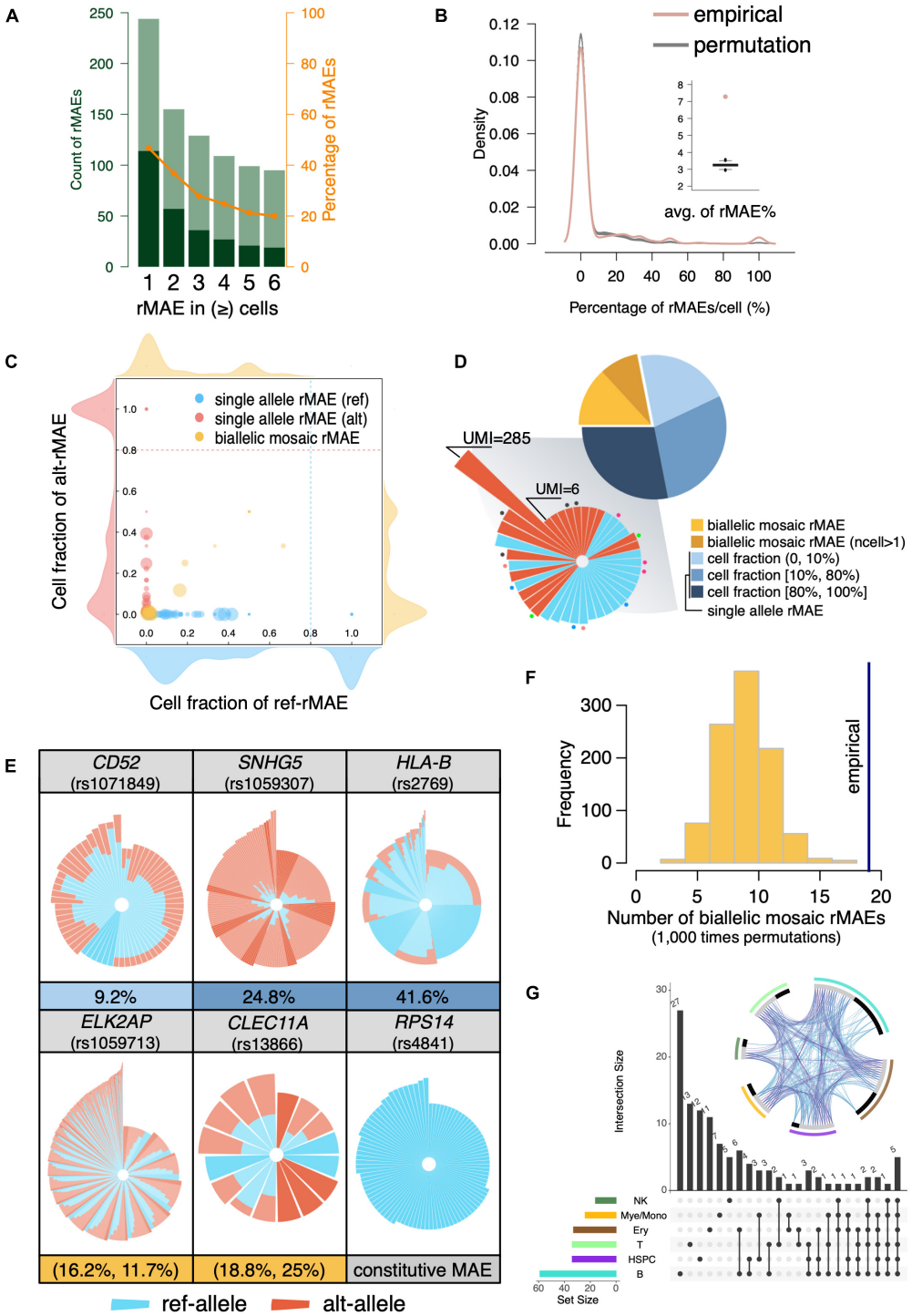


FIGURE 3 | Identification of rMAEs and their features. **(A)** The identified rMAEs with different thresholds of the observed number of cells. The non-rMAEs and rMAEs are represented by olive and dark green, respectively. The orange line shows the percentage of rMAEs in the analyzed SNVs. **(B)** The percentage of rMAEs per cell (red), comparing with permutations of alleles (gray). The inset shows the average of the percentages of rMAEs per cell. **(C)** The rMAEs, including single-allele rMAE (red/blue) and biallelic mosaic rMAE (yellow), are detected in different fractions of cells. The density plots around the scatter plot show the density of rMAEs of different cell fractions with the corresponding allele. **(D)** The pie chart shows the fractions of different types of rMAEs. The single-allele rMAEs of high cell fraction are shown in circular stacked bars in the shadowed sector. Each bar represents a cell with stacked UMI counts (log_{10}) of ref-allele (blue) and alt-allele (red) of a rMAE. The dots with the same color indicate the bars (cells) of the same rMAE, while bars without a dot mean that they are of different rMAEs. **(E)** Circular stacked profile of several rMAEs at the single-cell level. Each bar represents a cell with stacked UMI counts (log_{10}) of ref-allele (blue) and alt-allele (red). Cells are ordered by the total

(Continued)

FIGURE 3 | (Continued)
UMI counts and the fraction of the ref-allele. Highlighted bars are cells showing MAE. The fraction of cells showing MAE is shown under the circular stacked profile. A constitutive MAE (*RPS14*) is also displayed as a control. **(F)** The observed biallelic mosaic rMAEs (vertical line) are significantly more than that by allelic permutation ($p < 0.001$). **(G)** Sharing of the rMAEs among different cell types. The circos plot shows the shared genes (purple line) and pathways (blue line) among different cell types. Gray bar indicates the genes that are shared by other cell types, and black bar indicates the genes that are unique to the corresponding cell type.

with the fractions of constitutive MAEs (**Supplementary Figure 3** and **Figure 4A**). We further identified the rMAEs in each cell of the four samples and found that "normal" BMMCs showed the lowest fraction of rMAEs and BMMCs at relapse showed the highest fraction of rMAEs, while the other two leukemia samples showed intermediate values (**Figure 4B**). Analyses of constitutive MAEs and rMAEs manifested that leukemia samples showed increased MAEs, thus we expected much stronger MAEs in leukemia cells since the normal cells in leukemia samples may not contribute to the increased MAEs.

After PCA, we clustered these BMMCs into normal cells, leukemia cells before relapse (preR.Leuk), and leukemia cells after relapse (postR.Leuk). Normal cells were further separated into normal cells before relapse (Norm) and normal cells after relapse (postR.Norm) (**Supplementary Figure 4** and **Figure 4C**). There are 2.41% and 2.90% SNVs showing constitutive MAE in Norm and postR.Norm, respectively, while 3.35% and 3.11% in preR.Leuk and postR.Leuk, respectively (**Figure 4D**), thus leukemia cells have increased constitutive MAEs comparing with normal cells. The rMAE per cell between Norm and postR.Norm is not significantly different (**Figure 4E**). The leukemia cells from both preR.Leuk and postR.Leuk showed significantly higher levels of rMAE per cell, with postR.Leuk showing the highest value (**Figure 4E**). The results showed that separating the leukemia cells from normal cells in the leukemia samples made their difference more pronounced.

Analyzing the Leukemia Cell-Specific Monoallelic Expressions

Since we found that leukemia cells showed increased MAE, it would be more interesting to identify the leukemia cell-specific MAEs that potentially play an important role in leukemogenesis and leukemia development. Although leukemia cells showed increased MAEs comparing with normal cells, we only detected a few constitutive MAEs showed significant differences between normal cells and leukemia cells (**Figure 5A**), which indicated that most of the MAE changes between normal cells and leukemia cells are weak. GO enrichment analysis showed that immune-associated categories were commonly shared by the differentiated MAEs among Norm, preR.Leuk, and postR.Leuk, e.g., "IL-4 production" and "positive regulation of I-κB kinase/NF-κB signaling". The postR.Leuk-specific MAEs were enriched in "histone H3-K9 modification" and "mitotic cell cycle checkpoint" and "apoptosis", comparing with Norm and preR.Leuk, respectively (**Figure 5B**). Among these leukemia cell-specific MAEs, *RPSAP58* (rs78322935) and *TRG-AS1* (rs4373430) only expressed one allele in leukemia cells (**Figure 5C**). *TRG-AS1* is a lncRNA and regulates cancer progression by interacting with other microRNAs (Xie et al., 2019; He et al., 2020; Sun et al., 2020). We further identified the relapse-specific constitutive MAEs, which include *ACER3*, *TCL6*, and *TFDP2* (**Figure 5D**). *ACER3* coregulates cell proliferation and survival with *ACER2* (Hu et al., 2010) and plays an important role in leukemia development (Chen et al., 2016); while *TCL6* is associated with clinical outcomes of B-cell acute lymphoblastic leukemia patients (Cuadros et al., 2019); *TFDP2* plays core roles in apoptosis and cell proliferation (Korz et al., 2002). Altogether, most of the significantly changed MAEs were involved in immune pathways and regulation of cell proliferation, thus could explain the association between increased MAEs and the dysfunction in leukemia cells.

We also identified the significantly different rMAEs among Norm, preR.Leuk, and postR.Leuk (**Figure 5E**). Among these different rMAEs, 62.8% postR.Leuk showed rMAE at *PTMA* (rs12415), which is significantly higher than that of Norm (∼41.4%) and preR.Leuk (∼38.8%) (**Figure 5F**). Notably, *PTMA* is associated with lymphocyte proliferation and apoptosis in leukemia (Gómez-Márquez et al., 1989; Fan et al., 2006), thus finding the change of rMAE on *PTMA* has a lot of implications. The *HLA-DRB1* (rs35445101) shows high reference allele rMAE in Norm (∼70.8%) and preR.Leuk (∼94.7%), while postR.Leuk maintains the lowest reference allele rMAE (35.7%) among these cell populations (**Figure 5F**). *HLA-DRB1* plays a central role in antigen presentation and the decreased reference allele rMAE may impact its function.

DISCUSSION

Mammalian genomes are diploid, we usually just assume both alleles are equally expressed and did not consider the differences between the bialleles (Jin et al., 2012; Han et al., 2020). In this way, most studies only analyzed the average gene expression profile of the two alleles, even though MAE has been discovered during analyses of X-chromosome inactivation in the 1960s (Lyon, 1986), partially because most people do not realize the prevalence of MAE. Large-scale interrogations of MAEs have demonstrated MAEs were widespread in mammalian cells (Gimelbrant et al., 2007; Zwemer et al., 2012; Deng et al., 2014; Gendrel et al., 2014; Savova et al., 2016). The advance of scRNA-seq provides new biological insight on MAE, although most studies only used hundreds of cells (Deng et al., 2014; Borel et al., 2015; Kim et al., 2015; Reinius et al., 2016). Taking advantage of high throughput scRNA-seq with about 31,000 single-cell transcriptomes from the same individual, this study provides a fine scale landscape of MAE in hematopoiesis, at sample level, cell type level, and single-cell level. In addition to the known imprinted genes, we detected a lot of novel MAEs in BMMCs. As a cross validation, we found more than three quarters of the constitutive MAEs were reproducible in the bulk RNA sequencing of the same

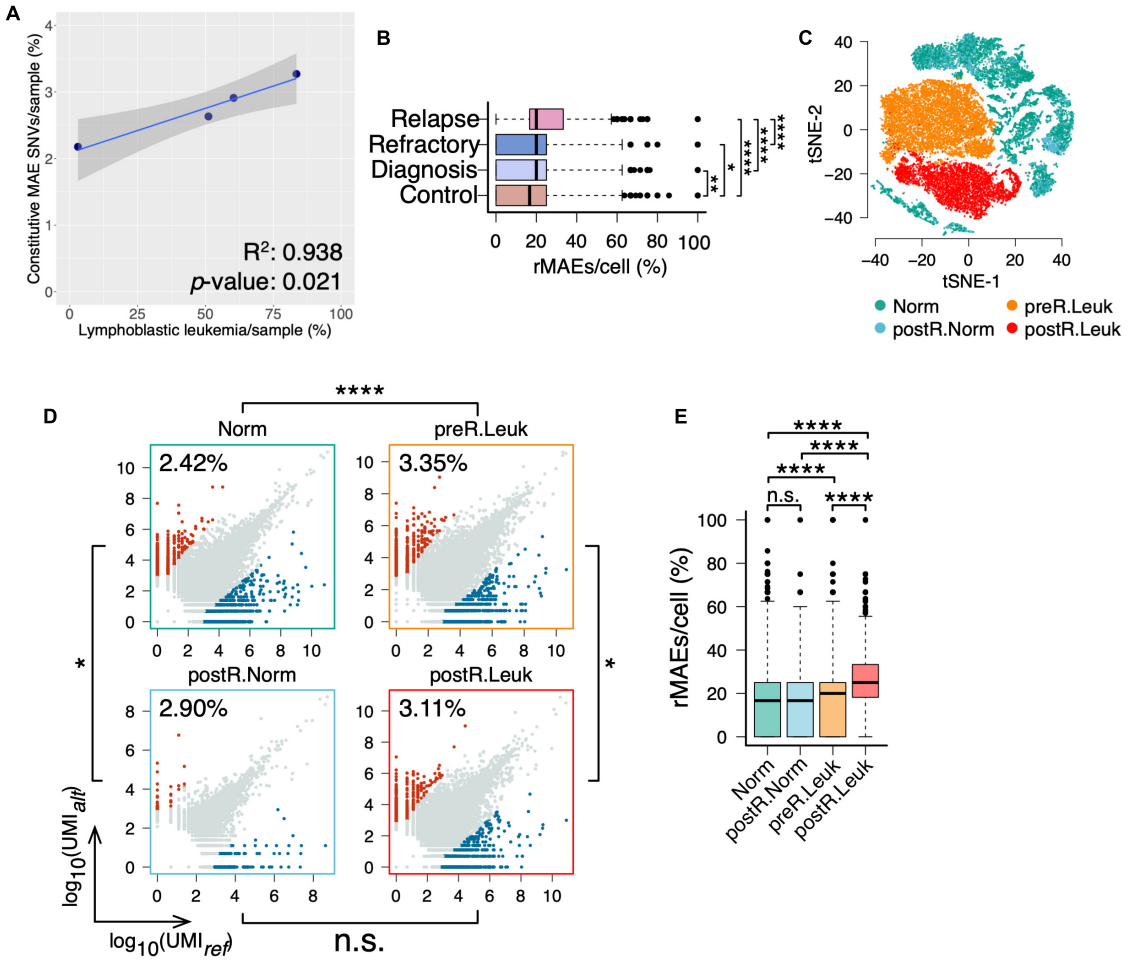


FIGURE 4 | Leukemia cells showing increased constitutive MAEs and rMAEs. **(A)** The fractions of constitutive MAEs are correlated with the fractions of leukemia cells in BMMCs. **(B)** The fractions of rMAE per cell in BMMCs from control, diagnosis, refractory, and relapse. **(C)** tSNE projection of all BMMCs from the boy diagnosis with leukemia, colored by the inferred cell states, namely, leukemia cells before relapse (preR.Leuk), non-leukemia cells before relapse (Norm), leukemia cells after relapse (postR.Leuk), and non-leukemia cells after relapse (postR.Norm). **(D)** Constitutive MAEs of Norm cells, preR.Leuk cells, postR.Norm cells, and postR.Leuk cells. **(E)** The fractions of rMAE per cell of Norm cells, preR.Leuk cells, postR.Norm cells, and postR.Leuk cells.*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$, and ****$p < 0.001$.

individual. The MAEs are associated with immune functions, which may indicate that the diversity of immunity is attributed to MAE.

We detected a considerable number of rMAE at single-cell level. Interestingly, a cell can stochastically express either of the two alleles thus leading to different cells expressing different alleles, which is called biallelic mosaic rMAEs. With a small but significant number, such genes are presumably increasing the cellular heterogeneity when the two alleles are different. Meanwhile, the biallelic mosaic rMAEs might be caused by transcriptional bursting thus it is the outcome of this important periodic switching kinetics. Furthermore, we observed much higher MAE levels in leukemia cells than that in normal cells, indicating the association between MAE and leukemogenesis. Leukemia-specific MAEs, including *TCL6*, *TFDP2*, and *PTMA*, are reported to be associated with tumorigenesis and cell proliferation. It is interesting to detect the *TCL6* in leukemia-specific MAEs, since it was recently reported that low *TCL6* levels were associated with poor survival of B-cell ALL patient, through a link between *TCL6*, *TCL1B*, and the *AKT1* pathway (Cuadros et al., 2019). The monoallelic expression may be indicative of insufficient dosage or expression deficiency of *TCL6* in our sample, who was also a B-cell ALL patient. Another interesting gene would be *PTMA*, which shows significantly higher proportion of MAE cells in the relapsed sample, and studies demonstrated that, though in other types of cancers, it can predict recurrence and poor prognosis (Ha et al., 2015; Chen et al., 2018). The observation that a higher level of MAE was in line with altered epigenetic regulations of leukemia (Miles et al., 2020; Waanders et al., 2020). MAE is also highly mediated by epigenetics, such as DNA methylation and histone modifications (Eckersley-Maslin and Spector, 2014a;

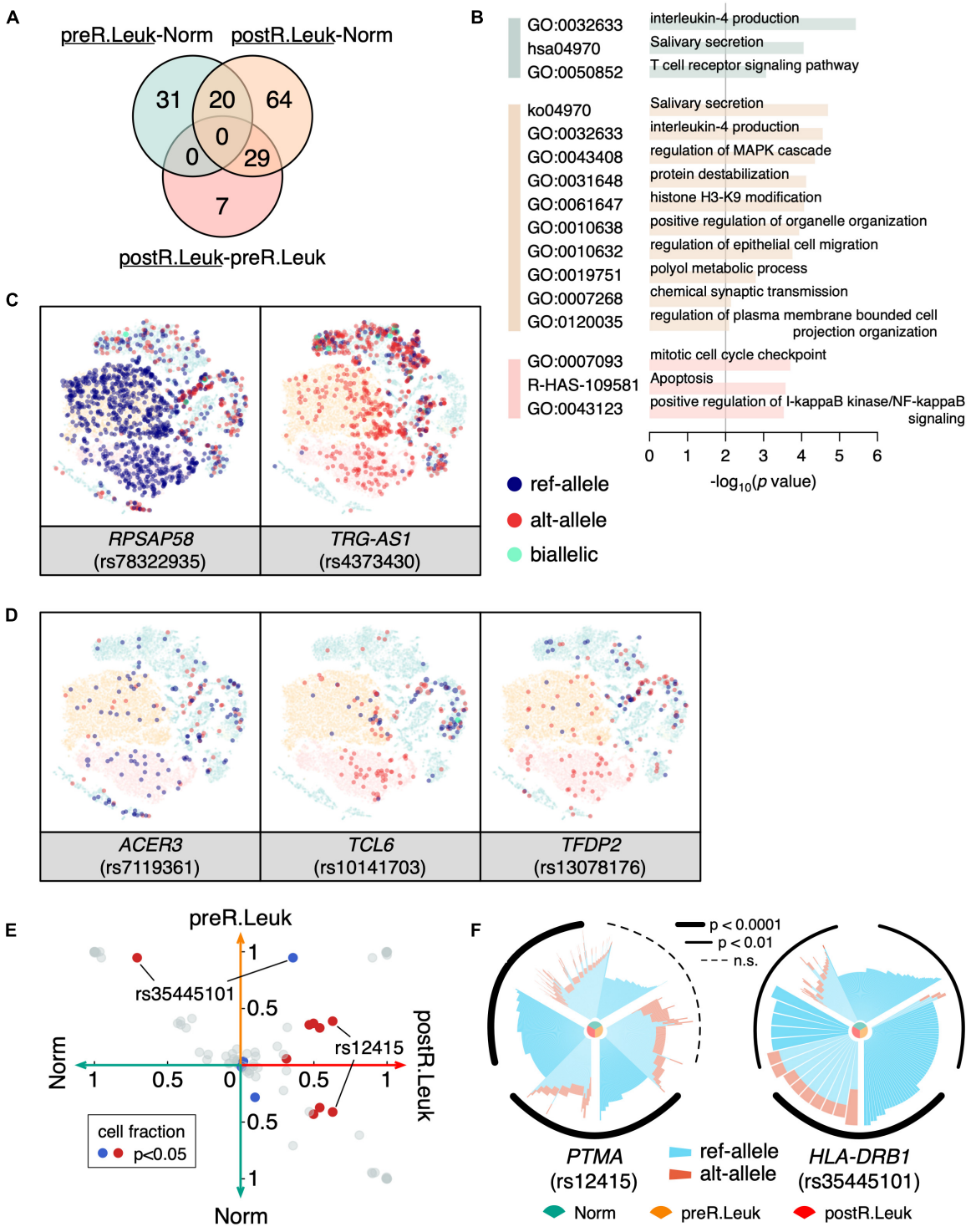


FIGURE 5 | Analyzing the leukemia cell-specific MAEs. **(A)** Venn diagram shows constitutive MAEs specific to preR.Leuk and postR.Leuk (underlined), after pairwise comparisons. The green circle indicates preR.Leuk-specific MAEs, comparing with Norm, the orange circle indicates postR.Leuk-specific MAEs, comparing with Norm, while the red circle indicates postR.Leuk-specific MAEs, comparing with preR.Leuk. **(B)** GO enrichment analysis of the leukemia cell-specific MAEs. The colors match with that in **(A)**. **(C)** *RPSAP58* and *TRG-AS1* show Leukemia specific MAEs. **(D)** *ACER3*, *TCL6*, and *TFDP2* show postR.Leuk-specific MAEs. **(E)** Pairwise comparison of the detected rMAEs among Norm, preR.Leuk, and postR.Leuk. Each dot represents a rMAE, and the axis indicates the cell fraction of each rMAE, within the corresponding cell sub-population. Significantly biased ($p < 0.05$) rMAE in terms of the cell fraction is highlighted in red/blue. **(F)** Significantly differentiated rMAE among Norm cells, preR.Leuk cells, and postR.Leuk cells. An increased number of cells showed MAE of ref-allele on *PTMA* (rs12415) in postR.Leuk. *HLA-DRB1* (rs35445101) shows increased MAE in preR.Leuk but deceased in postR.Leuk.

Reinius and Sandberg, 2015), and interestingly, we found an enrichment for "histone H3-H9 modification" in relapsed leukemia cells (**Figure 5B**).

In summary, as far as we know, this is the first systematic study on MAEs in human BMMCs using scRNA-seq and analyzed MAE in three layers including sample level, cell type level, and single-cell level. We found increased MAEs (both constitutive and random) in leukemia cells by comparing with normal cells, indicating the association between MAE and leukemogenesis. Particularly, these leukemia-associated MAEs may be the epigenetically therapeutic targets of leukemia.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://ngdc.cncb.ac.cn/, HRA000084 and CRA000588.

AUTHOR CONTRIBUTIONS

RF conducted major work of data analysis and interpretation. PQ performed process of the scRNA-seq data. NH collected and integrated the data. WJ conceived and supervised the project. YW interpreted the results and co-supervised the project. ZH and XZ coordinated the cooperation. WJ, RF, and YW drafted the manuscript, with critical revisions by NH, ZH, and XZ. All authors contributed to manuscript revision, read and approved the final manuscript for publication.

FUNDING

This study was supported by the National Key R&D Program of China (2018YFC1004500), National Natural Science Foundation of China (81872330), Shenzhen Innovation Committee of Science and Technology (JCYJ20170817111841427 and ZDSYS20200811144002008), Shenzhen Science and Technology Program (KQTD20180411143432337), and Center for Computational Science and Engineering, Southern University of Science and Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank the members of Jin Lab for their discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.702897/full#supplementary-material

Supplementary Figure 1 | Q-Q plot of VAF of in BMMCs, estimated by UMI counts.

Supplementary Figure 2 | tSNE projection of BMMCs (background), colored by the expressed allele of *IL32* **(A)**, *HLA-DRB5* **(B)**, and *NUP210* **(C)**. The size of the highlighted dot is scaled to $\log_{10}$ of its UMI count, and the pie chart in the bottom right shows the genetic allelic ratio of the two alleles from WGS data.

Supplementary Figure 3 | Identification of constitutive MAEs in diagnosis **(A)**, refractory **(B)**, and relapse **(C)**, with blue representing MAE of the reference allele while red representing MAE of the alternative allele. The percentage of the SNVs with constitutive MAE is shown on the top left of the plot.

Supplementary Figure 4 | tSNE projection of all BMMCs from the boy diagnosed with leukemia, colored by sampling times, namely control, diagnosis, refractory, and relapse.

REFERENCES

Abyzov, A., Urban, A. E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 21, 974–984. doi: 10.1101/gr.114876.110

Al Seraihi, A. F., Rio-Machin, A., Tawana, K., Bodor, C., Wang, J., Nagano, A., et al. (2018). GATA2 monoallelic expression underlies reduced penetrance in inherited GATA2-mutated MDS/AML. *Leukemia* 32, 2502–2507. doi: 10.1038/s41375-018-0134-9

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* 57, 289–300.

Borel, C., Ferreira, P. G., Santoni, F., Delaneau, O., Fort, A., Popadin, et al. (2015). Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.* 96, 70–80. doi: 10.1016/j.ajhg.2014.12.001

Borlido, J., Sakuma, S., Raices, M., Carrette, F., Tinoco, R., Bradley, L. M., et al. (2018). Nuclear pore complex-mediated modulation of TCR signaling is required for naive CD4(+) T cell homeostasis. *Nat. Immunol.* 19, 594–605. doi: 10.1038/s41590-018-0103-5

Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16:195. doi: 10.1186/s13059-015-0762-6

Chen, C., Yin, Y., Li, C., Chen, J., Xie, J., Lu, Z., et al. (2016). ACER3 supports development of acute myeloid leukemia. *Biochem. Biophys. Res. Commun.* 478, 33–38. doi: 10.1016/j.bbrc.2016.07.099

Chen, K., Xiong, L., Yang, Z., Huang, S., Zeng, R., and Miao, X. (2018). Prothymosin-alpha and parathymosin expression predicts poor prognosis in squamous and adenosquamous carcinomas of the gallbladder. *Oncol. Lett.* 15, 4485–4494. doi: 10.3892/ol.2018.7824

Chess, A. (2016). Monoallelic gene expression in mammals. *Annu. Rev. Genet.* 50, 317–327. doi: 10.1146/annurev-genet-120215-035120

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78, 823–834. doi: 10.1016/s0092-8674(94)90562-2

Choi, K., Raghupathy, N., and Churchill, G. A. (2019). A Bayesian mixture model for the analysis of allelic expression in single cells. *Nat. Commun.* 10:5188. doi: 10.1038/s41467-019-13099-0

Cuadros, M., Andrades, A., Coira, I. F., Balinas, C., Rodriguez, M. I., Alvarez-Perez, J. C., et al. (2019). Expression of the long non-coding RNA TCL6 is associated with clinical outcome in pediatric B-cell acute lymphoblastic leukemia. *Blood Cancer J.* 9:93. doi: 10.1038/s41408-019-0258-9

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196. doi: 10.1126/science.1245316

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806

Eckersley-Maslin, M. A., and Spector, D. L. (2014a). Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet.* 30, 237–244. doi: 10.1016/j.tig.2014.03.003

Eckersley-Maslin, M. A., Thybert, D., Bergmann, J. H., Marioni, J. C., Flicek, P., and Spector, D. L. (2014b). Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* 28, 351–365. doi: 10.1016/j.devcel.2014.01.017

Fan, J., Lee, H. O., Lee, S., Ryu, D. E., Lee, S., Xue, C., et al. (2018). Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* 28, 1217–1227. doi: 10.1101/gr.228080.117

Fan, Y. Z., Chang, H., Yu, Y., Liu, J., and Wang, R. (2006). Thymosin alpha1 suppresses proliferation and induces apoptosis in human leukemia cell lines. *Peptides* 27, 2165–2173. doi: 10.1016/j.peptides.2006.03.012

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* 12, 565–575. doi: 10.1038/nrg3032

Gendrel, A. V., Attia, M., Chen, C. J., Diabangouaya, P., Servant, N., Barillot, E., et al. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* 28, 366–380. doi: 10.1016/j.devcel.2014.01.016

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science* 318, 1136–1140. doi: 10.1126/science.1148910

Gómez-Márquez, J., Segade, F., Dosil, M., Pichel, J. G., Bustelo, X. R., and Freire, M. (1989). The expression of prothymosin α gene in T lymphocytes and leukemic lymphoid cells is tied to lymphocyte proliferation. *J. Biol. Chem.* 264, 8451–8454. doi: 10.1016/s0021-9258(18)81807-5

Ha, S. Y., Song, D. H., Hwang, S. H., Cho, S. Y., and Park, C.-K. (2015). Expression of prothymosin α predicts early recurrence and poor prognosis of hepatocellular carcinoma. *Hepatobiliary Pancreat. Dis. Int.* 14, 171–177. doi: 10.1016/s1499-3872(14)60326-x

Han, Z., Cui, K., Placek, K., Hong, N., Lin, C., Chen, W., et al. (2020). Diploid genome architecture revealed by multi-omic data of hybrid mice. *Genome Res.* 30, 1097–1106. doi: 10.1101/gr.257568.119

He, S., Wang, X., Zhang, J., Zhou, F., Li, L., and Han, X. (2020). TRG-AS1 is a potent driver of oncogenicity of tongue squamous cell carcinoma through microRNA-543/Yes-associated protein 1 axis regulation. *Cell Cycle* 19, 1969–1982. doi: 10.1080/15384101.2020.1786622

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3628–3632. doi: 10.1073/pnas.73.10.3628

Hu, W., Xu, R., Sun, W., Szulc, Z. M., Bielawski, J., Obeid, L. M., et al. (2010). Alkaline ceramidase 3 (ACER3) hydrolyzes unsaturated long-chain ceramides, and its down-regulation inhibits both cell proliferation and apoptosis. *J. Biol. Chem.* 285, 7964–7976. doi: 10.1074/jbc.M109.063586

Janciauskiene, S. M., Nita, I. M., and Stevens, T. (2007). Alpha1-antitrypsin, old dog, new tricks. alpha1-antitrypsin exerts in vitro anti-inflammatory activity in human monocytes by elevating cAMP. *J. Biol. Chem.* 282, 8573–8582. doi: 10.1074/jbc.M607976200

Jiang, Y., Zhang, N. R., and Li, M. (2017). SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol.* 18:74. doi: 10.1186/s13059-017-1200-8

Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B., et al. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res.* 22, 519–527. doi: 10.1101/gr.124784.111

Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A., and Marioni, J. C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* 6:8687. doi: 10.1038/ncomms9687

Kim, J. K., and Marioni, J. C. (2013). Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.* 14:R7. doi: 10.1186/gb-2013-14-1-r7

Korz, C., Pscherer, A., Benner, A., Mertens, D., Schaffner, C., Leupolt, E., et al. (2002). Evidence for distinct pathomechanisms in B-cell chronic lymphocytic leukemia and mantle cell lymphoma by quantitative expression analysis of cell cycle and apoptosis-associated genes. *Blood* 99, 4554–4561. doi: 10.1182/blood.v99.12.4554

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., et al. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* 565, 251–254. doi: 10.1038/s41586-018-0836-1

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus L.*). *Nature* 190, 372–373. doi: 10.1038/190372a0

Lyon, M. F. (1986). X chromosomes and dosage compensation. *Nature* 320:313. doi: 10.1038/320313b0

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17:10. doi: 10.14806/ej.17.1.200

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meehan, M., Melvin, A., Gallagher, E., Smith, J., McGoldrick, A., Moss, C., et al. (2007). Alpha-T-catenin (CTNNA3) displays tumour specific monoallelic expression in urothelial carcinoma of the bladder. *Genes Chromosomes Cancer* 46, 587–593. doi: 10.1002/gcc.20443

Miles, L. A., Bowman, R. L., Merlinsky, T. R., Csete, I. S., Ooi, A. T., Durruthy-Durruthy, R., et al. (2020). Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* 587, 477–482. doi: 10.1038/s41586-020-2864-x

Moraga, F., Lindgren, S., and Janciaskiene, S. (2001). Effects of noninhibitory alpha-1-antitrypsin on primary human monocyte activation in vitro. *Arch. Biochem. Biophys.* 386, 221–226. doi: 10.1006/abbi.2000.2211

Ochiai, H., Hayashi, T., Umeda, M., Yoshimura, M., Harada, A., Shimizu, Y., et al. (2020). Genome-wide kinetic properties of transcriptional bursting in mouse embryonic stem cells. *Sci. Adv.* 6:eaaz6699. doi: 10.1126/sciadv.aaz6699

Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J. Exp. Med.* 122, 853–876. doi: 10.1084/jem.122.5.853

Polson, E. S., Lewis, J. L., Celik, H., Mann, V. M., Stower, M. J., Simms, M. S., et al. (2013). Monoallelic expression of TMPRSS2/ERG in prostate cancer stem cells. *Nat. Commun.* 4:1623. doi: 10.1038/ncomms2627

Qin, P., Pang, Y., Hou, W., Fu, R., Zhang, Y., Wang, X., et al. (2021). Integrated decoding hematopoiesis and leukemogenesis using single-cell sequencing and its medical implication. *Cell Discov.* 7:2. doi: 10.1038/s41421-020-00223-4

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32. doi: 10.1038/35047554

Reinius, B., Mold, J. E., Ramskold, D., Deng, Q., Johnsson, P., Michaelsson, J., et al. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* 48, 1430–1435. doi: 10.1038/ng.3678

Reinius, B., and Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat. Rev. Genet.* 16, 653–664. doi: 10.1038/nrg3888

Savova, V., Chun, S., Sohail, M., McCole, R. B., Witwicki, R., Gai, L., et al. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat. Genet.* 48, 231–237. doi: 10.1038/ng.3493

Schemionek, M., Herrmann, O., Reher, M. M., Chatain, N., Schubert, C., Costa, I. G., et al. (2016). Mtss1 is a critical epigenetically regulated tumor suppressor in CML. *Leukemia* 30, 823–832. doi: 10.1038/leu.2015.329

Schneider, R. K., Schenone, M., Ferreira, M. V., Kramann, R., Joyce, C. E., Hartigan, C., et al. (2016). Rps14 haploinsufficiency causes a block in erythroid differentiation mediated by S100A8 and S100A9. *Nat. Med.* 22, 288–297. doi: 10.1038/nm.4047

Silcock, L., Almabrazi, H., Mokrab, Y., Jithesh, P., Al-Hashmi, M., James, N., et al. (2019). Monoallelic expression in melanoma. *J. Transl. Med.* 17:112. doi: 10.1186/s12967-019-1863-x

Stamoulis, G., Garieri, M., Makrythanasis, P., Letourneau, A., Guipponi, M., Panousis, N., et al. (2019). Single cell transcriptome in aneuploidies reveals mechanisms of gene dosage imbalance. *Nat. Commun.* 10:4495. doi: 10.1038/s41467-019-12273-8

Sun, X., Qian, Y., Wang, X., Cao, R., Zhang, J., Chen, W., et al. (2020). LncRNA TRG-AS1 stimulates hepatocellular carcinoma progression by sponging miR-4500 to modulate BACH1. *Cancer Cell Int.* 20:367. doi: 10.1186/s12935-020-01440-3

Waanders, E., Gu, Z., Dobson, S. M., Antić, Ž, Crawford, J. C., Ma, X., et al. (2020). Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. *Blood Cancer Discov.* 1, 96–111. doi: 10.1158/0008-5472.Bcd-19-0041

Walker, E. J., Zhang, C., Castelo-Branco, P., Hawkins, C., Wilson, W., Zhukova, N., et al. (2012). Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res.* 72, 636–644. doi: 10.1158/0008-5472.CAN-11-2266

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

Wang, L., Luo, J., Nian, Q., Xiao, Q., Yang, Z., and Liu, L. (2014). Ribosomal protein S14 silencing inhibits growth of acute myeloid leukemia transformed from myelodysplastic syndromes via activating p53. *Hematology* 19, 225–231. doi: 10.1179/1607845413Y.0000000127

Wang, N., Wu, R., Tang, D., and Kang, R. (2021). The BET family in immunity and disease. *Signal Transduct. Target Ther.* 6:23. doi: 10.1038/s41392-020-00384-4

Xie, H., Shi, S., Chen, Q., and Chen, Z. (2019). LncRNA TRG-AS1 promotes glioblastoma cell proliferation by competitively binding with miR-877-5p to regulate SUZ12 expression. *Pathol. Res. Pract.* 215:152476. doi: 10.1016/j.prp.2019.152476

Yu, D., Zhan, X. H., Zhao, X. F., Williams, M. S., Carey, G. B., Smith, E., et al. (2012). Mice deficient in MIM expression are predisposed to lymphomagenesis. *Oncogene* 31, 3561–3568. doi: 10.1038/onc.2011.509

Zhang, Y., Gao, Y., Zhang, H., Zhang, J., He, F., Hnizda, A., et al. (2018). PDGFRB mutation and tyrosine kinase inhibitor resistance in Ph-like acute lymphoblastic leukemia. *Blood* 131, 2256–2261. doi: 10.1182/blood-2017-11-817510

Zhao, D., Lin, M., Pedrosa, E., Lachman, H. M., and Zheng, D. (2017). Characteristics of allelic gene expression in human brain cells from single-cell RNA-seq data analysis. *BMC Genomics* 18:860. doi: 10.1186/s12864-017-4261-x

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6

Zwemer, L. M., Zak, A., Thompson, B. R., Kirby, A., Daly, M. J., Chess, A., et al. (2012). Autosomal monoallelic expression in the mouse. *Genome Biol.* 13:R10. doi: 10.1186/gb-2012-13-2-r10

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

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fcell.2021.751154

Ectopic Splicing Disturbs the Function of *Xist* RNA to Establish the Stable Heterochromatin State

*Ruka Matsuura[1†], Tatsuro Nakajima[2†], Saya Ichihara[1] and Takashi Sado[1,3]**

[1] Department of Advanced Bioscience, Graduate School of Agriculture, Kindai University, Nara, Japan, [2] Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan, [3] Agricultural Technology and Innovation Research Institute, Kindai University, Nara, Japan

Non-coding *Xist* RNA plays an essential role in X chromosome inactivation (XCI) in female mammals. It coats the X chromosome in cis and mediates the recruitment of many proteins involved in gene silencing and heterochromatinization. The molecular basis of how *Xist* RNA initiates chromosomal silencing and what proteins participate in this process has been extensively studied and elucidated. Its involvement in the establishment and maintenance of the X-inactivated state is, however, less understood. The $Xist^{\Delta IVS}$ allele we previously reported is peculiar in that it can initiate XCI but fails to establish the inactive state that is stably maintained and, therefore, may provide an opportunity to explore how *Xist* RNA contributes to establish a robust heterochromatin state. Here we demonstrate that ectopic splicing taking place to produce $Xist^{\Delta IVS}$ RNA disturbs its function to properly establish stable XCI state. This finding warrants the potential of $Xist^{\Delta IVS}$ RNA to provide further insight into our understanding of how *Xist* RNA contributes to establish sustainable heterochromatin.

Keywords: *Xist* RNA, embryonic stem cells (ESCs), X chromosome inactivation, heterochromatin, gene silencing, RNA-FISH

OPEN ACCESS

Edited by:
Louis Lefebvre, University of British Columbia, Canada

Reviewed by:
Joseph Mauro Calabrese, University of North Carolina at Chapel Hill, United States
Stefan Pinter, University of Connecticut Health Center, United States

***Correspondence:**
Takashi Sado
tsado@nara.kindai.ac.jp

† These authors have contributed equally to this work

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

Received: *31 July 2021*
Accepted: *20 September 2021*
Published: *14 October 2021*

Citation:
Matsuura R, Nakajima T, Ichihara S and Sado T (2021) Ectopic Splicing Disturbs the Function of Xist RNA to Establish the Stable Heterochromatin State. Front. Cell Dev. Biol. 9:751154. doi: 10.3389/fcell.2021.751154

INTRODUCTION

A subset of long non-coding RNA (lncRNA) associates with chromatin and regulates chromatin state. *Xist* RNA is one of the most extensively studied lncRNAs (Borsani et al., 1991; Brockdorff et al., 1991), which becomes upregulated from one of the two X chromosomes during the early development of female mammals (Kay et al., 1993) and induces X chromosome inactivation (XCI) to compensate for the dosage difference of X-linked genes between the sexes by associating with the X chromosome it originates from (Lyon, 1961; Clemson et al., 1996; Penny et al., 1996; Marahrens et al., 1997). In the mouse, XCI is imprinted in the extraembryonic lineages (Takagi and Sasaki, 1975), which contributes to the placenta and a part of the extraembryonic membranes, whereas it takes place in a random fashion to either X in the embryonic lineage, giving rise to whole tissues of the fetus (Monk and Harper, 1979). *Xist* RNA plays a pivotal role in both types of XCI by serving as a scaffold for recruiting many proteins involved in heterochromatin formation and maintenance.

Female mouse embryonic stem cells (ESCs) have been extensively used for the study of XCI as they carry two active X chromosomes, one of which undergoes chromosome-wide silencing with monoallelic upregulation of *Xist* and its association with the chromosome in cis upon induction of differentiation (Panning and Jaenisch, 1996). Functional domains of *Xist* RNA have been identified

by studying ESCs introduced with a transgene expressing mutant *Xist* RNA with a deletion of various conserved domains among species. Deletion of the most proximal conserved repeat called the A-repeat not only totally compromises the silencing function of *Xist* RNA in the context of differentiating ESCs and developing embryos (Wutz et al., 2002; Sakata et al., 2017) but also abolishes an interaction of SPEN with *Xist* RNA (Chu et al., 2015; Moindrot et al., 2015; Monfort et al., 2015). A more recent study showed that the B-repeat interacts with hnRNPK to recruit PGF3/5-PRC1 to establish Polycomb-mediated chromosomal silencing (Pintacuda et al., 2017). Genetic screening and a more comprehensive proteomic analysis using ESCs have revealed many proteins that interact with *Xist* RNA (Chu et al., 2015; Moindrot et al., 2015; Pandya-Jones et al., 2020).

We previously reported a partial loss-of-function allele of *Xist*, $Xist^{IVS}$, which initiates XCI in the extraembryonic tissues upon paternal transmission, but the XCI state initiated by $Xist^{IVS}$ is not stable, resulting in derepression of X-inactivated genes in the extraembryonic tissues (Hoki et al., 2011). The embryos carrying the paternally derived $Xist^{IVS}$ eventually die at the midgestation stage, at least, partly due to malfunction of the placenta. $Xist^{IVS}$ was generated by introducing the 0.9-kb second intron of human γ-globin gene (IVS2) 0.9 kb downstream of the major transcription start site of *Xist* as a control allele for the analysis of antisense regulation by *Tsix* (Ohhata et al., 2008). When this allele is transcribed, the introduced IVS2 sequence is spliced out to produce essentially the same transcript as that transcribed from the wild-type allele except for the presence of an additional 16-base insertion derived from the targeting vector at the splicing junction. Although it was unknown whether or not the presence of such small insertion affects the function of *Xist* RNA, it was no doubt that the $Xist^{IVS}$ allele is not fully functional. Unlike other mutant alleles of *Xist*, $Xist^{IVS}$ is peculiar in that its RNA product can initiate appreciable levels of XCI in the extraembryonic tissues, which supports early postimplantation development, but fail to establish the stable XCI state. The $Xist^{IVS}$ allele would, therefore, provide a unique opportunity to understand the molecular basis of how *Xist* RNA contributes to establishing the heritable robust heterochromatin, which allows stable maintenance of the X-inactivated state essential for embryonic development and healthy life after birth.

In this study, we addressed how $Xist^{IVS}$ behaves in the embryonic tissue, in which random XCI takes place, and the impact of the presence of the 16-base insertion in *Xist* RNA on its function. The results demonstrate that the $Xist^{IVS}$ allele is not upregulated in the embryonic tissue in contrast to that in the extraembryonic tissues, suggesting some difference in the mechanisms for upregulation of *Xist* between the embryonic and extraembryonic tissues. In addition, forced expression of $Xist^{IVS}$ RNA from the newly generated $Xist^{CAGIVS}$ allele and another allele, $Xist^{CAG16in}$, which produces the transcript containing the same 16-base insertion at the same position as $Xist^{IVS}$ RNA without splicing, in differentiating ESCs demonstrates that the 16-base insertion per se does not affect the function of *Xist* RNA, indicating that splicing in the former transcript to remove the IVS2 sequence brings the qualitative difference between these transcripts. We discuss the potential of the $Xist^{IVS}$ allele, which could facilitate our further understanding of how *Xist* RNA contributes to establish robust heterochromatin state of the inactive X chromosome.

MATERIALS AND METHODS

Mice

$Xist^{IVS}$ and $Xist^{1lox}$ mice were described elsewhere (Sado et al., 2005; Hoki et al., 2011). The JF1 strain was maintained in-house and C57BL/6J strain purchased from Japan SLC, Inc. (Shizuoka, Japan). All mice were maintained and used in accordance with the Guidelines for the Care and Use of Laboratory Animals of Kindai University (KDAS-26-0006).

Cells and Culture Condition

Embryonic stem cells used in this study were established from E3.5 blastocysts according to Ying et al. (2008). Blastocysts were cultured on four-well plates containing feeder cells with N2B27 medium supplemented with 1 μM of PD0325901 (Cayman Chemical, Ann Arbor, MI, United States) and 3 μM of CHIR99021 (Cayman) as well as 1,000 U/ml of LIF (Nacalai USA, San Diego, CA, United States). Outgrowths of undifferentiated cells were dissociated with TrypLE (Invitrogen, Carlsbad, CA, United States) and seeded on a four-well plate containing feeder cells. Many ESC colonies that appeared in several days were dissociated by TrypLE and seeded on a 35-mm dish containing feeder cells as passage 1 (P1). Blastocysts used for the establishment of ESCs homozygous for $Xist^{IVS}$ were prepared from a cross between females heterozygous for $Xist^{IVS}$ and males hemizygous for $Xist^{IVS}$. Those used for the establishment of F1 hybrid female ESCs were prepared from a cross between JF1 (*Mus m. molossinus*) females and C57BL/6J (B6) males. JB4, one of the established ESCs, was adapted to grow on a gelatin-coated dish with conventional ESC medium containing 2i (DMEM containing 15% FBS, 1× non-essential amino acids, 1× penicillin/streptomycin, 0.1 mM β-mercaptoethanol, 1,000 U/ml of LIF, 1 μM of PD0325901, 3 μM of CHIR99021) for some passages. For induction of differentiation, ESCs were cultured with N2B27 medium on gelatin-coated dishes for up to 7 days.

Construction of Targeting Vectors

For construction of a targeting vector for *Eif2s3x*, a 938-bp fragment present downstream of a termination codon of *Eif2s3x* was amplified by PCR using primers, Eif2s3x-F (5′-actctgtaga caaggctggc-3′) and Eif2s3x-R (5′-TTCTGTAGGGAGAATTGG CC-3′), on B6 genomic DNA and cloned into pBluescriptII-SK(+), in which the SpeI site present in MCS had been destroyed. This plasmid was linearized at the SpeI site present in the *Eif2s3x* fragment, and an IRES neo cassette was subsequently cloned in an appropriate orientation to generate p3′EifIRESneo. For construction of a targeting vector for *Hprt*, a 763-bp fragment containing an ATG start codon of *Hprt* was amplified by

PCR using primers, gHprt-F1 (5′-agacgacagagggcctgggggctgc-3′) and gHprt-R1 (5′-ttgtagagctgggcctctcccagga-3′), on JF1 genomic DNA. This plasmid was used for inverse PCR using primers Hprt-invR(Spe) (5′-gggaaacttactagtCGGCAAAAAGC GGTCTGAGGAGGAAGC-3′) and Hprt-invF(Pst) (5′-gggaaact tctgcagCGACCCGCAGTCCCAGCGTCGTGgtga-3′), and the amplified fragment was circulized by self-ligation. The resultant plasmid was digested by *Pst*I and *Spe*I and used to clone a CAG-Zeo-pA cassette to produce pHprt_invF/R-Zeo.

For construction of a targeting vector to generate the $Xist^{CAGIVS2lox}$ allele, a 0.9-kb IVS fragment amplified on genomic DNA carrying the $Xist^{IVS}$ allele by PCR according to Ohhata et al. (2008) was cloned into a targeting vector, pCAG-CΔM20, which was used for replacing the endogenous *Xist* promoter with the CAG promoter (Amakawa et al., 2015), to generate pCAG-CΔM20-IVS#5. For construction of a targeting vector to generate the $Xist^{CAG16in2lox}$ allele, an extra 16-base double-stranded DNA $Xist^{IVS}$ RNA was introduced in pCAG-CΔM20 at the unique *Xho*I site to generate pCAG16in#3.

A 20-bp double-stranded DNA fragment corresponding to the respective specific guide RNA sequence was cloned into pX330 (Add genes) linearized by BbsI to generate pX330-EifgRNA1, pHprt-gRNA, and pX330-Xist(-20). sgRNA sequences are as follows: Eif2s3x (5′-ATTTATAGCTGCTACTAGTA-3′), Hprt (5′-TGACGGGACCGGTCGGCTCG-3′), and Xist (5′-GATCAGTTAAAGGCGTGCAA-3′).

Establishment of Hybrid Female Embryonic Stem Cells Stably Maintaining Two X Chromosomes

The *Eif2s3x* locus, which is known to escape XCI, and the *Hprt* locus were selected for the site to introduce an IRESneo cassette and a CAGzeo-pA cassette, respectively. JB4 ESCs were used at P8 to introduce an IRESneo cassette at the *Eif2s3x* locus, and 5×10^5 cells were transfected with 1 μg of pX330-EifgRNA1 and 1 μg of p3′EifIRESneo using FuGENE HD Transfection Reagent (Promega, Madison, WI, United States) and seeded on two gelatin-coated 60-mm dishes. Twenty-four hours later, selection was started with conventional ES medium containing 2i/LIF and 200 μg/ml of G418 for 8 days. Of the 40 colonies isolated, 12 turned out to harbor expected homologous recombination at one of the two *Eif2s3x* alleles. One of the 12 clones, termed JB4/EI7, which were confirmed to have an IRESneo cassette integrated on the X chromosome derived from B6, was used for targeting a CAGzeo-pA cassette at the *Hprt* locus on the X chromosome derived from JF1. Five hundred thousand cells were transfected with 1 μg of pHprt-gRNA and 1 μg of pHprt_invF/R-Zeo, as above. Selection was carried out using conventional ES medium containing 2i/LIF and 25 μg/ml of Zeocin for 10 days. Of the 40 colonies screened by PCR, one was found to harbor a correct homologous recombination at the *Hprt* locus on the X chromosome derived from JF1 and termed JB4/EI7HZ2. JB4/EI7HZ2 retained two X chromosomes in 99% of the population after 70 days of culture in the presence of G418 and Zeocin.

Generation of Embryonic Stem Cells Carrying Either $Xist^{CAGIVS2lox}$ or $Xist^{CAG16in2lox}$

Five hundred thousand JB4/EI7HZ2 cells were transfected with 1 μg of pX330-Xist(-20) and 1 μg of either pCAG-CΔM20-IVS#5 or pCAG16in#3, as above. Selection was carried out using conventional ES medium containing 2i/LIF and 2 μg/ml of puromycin for 7–10 days. Twelve of 48 and 24 of 96 colonies isolated turned out to harbor $Xist^{CAGIVS2lox}$ and $Xist^{CAG16in2lox}$, respectively. To identify which of the B6 or JF1 alleles was targeted, a relevant region containing a single-nucleotide polymorphism (SNP) was amplified by PCR and analyzed by restriction digestion and sequencing. A floxed PacECFP-pA cassette was removed by transient expression of Cre recombinase to convert $Xist^{CAGIVS2lox}$ and $Xist^{CAG16in2lox}$ into $Xist^{CAGIVS}$ and $Xist^{CAG16in}$, respectively.

RNA-FISH and Whole-Mount RNA-FISH

RNA-FISH was performed using cells fixed with either 4% paraformaldehyde (PFA) or Carnoy's fixative as previously described (Sado et al., 2001). For PFA fixation, cells grown on a coverslip were fixed with 4% PFA and subsequently permeabilized with 0.5 % Triton X-100/0.5% BSA/PBS for 30 min and dehydrated through 70 and 100% ethanol. *Xist* and *Atrx* probes were prepared using pXist_SS12.9 and a BAC clone P23-260I15, respectively, as previously described (Sakata et al., 2017).

Whole-mount RNA-FISH was carried out according to Shiura et al. (2018). Briefly, embryos were first permeabilized in 0.1% Triton X-100/PBS for 10 s and fixed in 4% paraformaldehyde/PBS containing 0.1% Triton X-100 for 10 min. Following incubation in 2xSCC/0.05% Tween 20, 2xSSC/25% formamide/0.05% Tween 20, and 2xSSC/50% formamide/0.05% Tween 20 for 10 min each, the embryos were hybridized with *Xist* and *Atrx* probes overnight at 37°C, followed by washes for 5 min twice at 37°C in each of 2xSSC/50% formamide and 2xSSC/0.05% Tween 20 and subsequent counterstaining with Hoechst 33258.

For allele-specific RNA-FISH for *Xist*, probes were prepared according to Harris et al. (2019). Briefly, five B6- and JF1-specific oligonucleotides, the sequences of which were exactly the same as those designed by Harris et al. (2019) and contained an SNP between B6 and JF1, were labeled with Cy5 and Cy3, respectively, at their 3′-end. Five mask oligonucleotides complementary to the common 3′ part of the labeled oligonucleotides were also prepared. These Cy5-labeled B6-specific oligonucleotides, Cy3-labeled JF1-specific oligonucleotides, and mask oligonucleotides were included in hybridization buffer (2xSSC/10% dextran sulfate/2 mg/ml BSA/25% formamide) at a concentration of 5 μM for labeled and 10 μM for mask oligonucleotides. In the hybridization reaction, a Green-dUTP-labeled strand-specific probe for *Xist* prepared according to Shiura et al. (2018) was also included to validate the hybridization signal produced by Cy5- and Cy3-labeled oligonucleotides. Washing was carried out in 2xSSC/10% formamide for 30 min at 42°C twice and in 2xSSC for 5 min at room temperature.

RT-PCR

cDNA synthesis was carried out using random hexamer on 1 μg of total RNA treated with DNaseI in the presence or absence of SuperScript III (Invitrogen). One-fiftieth of the reaction was used as a template for PCR using a primer set, R700P2, and F1063AS (Sado et al., 2005). The PCR product was subsequently digested with *Xho*I, whose restriction site is present in the fragment derived from wild-type *Xist* RNA but destroyed in that derived from $Xist^{IVS}$ and $Xist^{16in}$ (Hoki et al., 2011).

RESULTS

The $Xist^{IVS}$ Allele Did Not Become Upregulated in the Epiblast

To investigate the effect of $Xist^{IVS}$ on XCI in the embryonic lineage, we established wild-type and $Xist^{IVS}/Xist^{IVS}$ ESCs from blastocysts obtained from a cross between $Xist^{IVS}$/+ females and $Xist^{IVS}$/Y males. They were induced to differentiate for 5 days and examined for *Xist* expression by RNA-FISH. While *Xist* was monoallelically upregulated in about 40% of wild-type cells, the majority of $Xist^{IVS}/Xist^{IVS}$ cells exhibited two pinpoint signals and none contained the *Xist* cloud (**Supplementary Figure 1**). Extension of differentiation did not change the situation. This raised a possibility that although the $Xist^{IVS}$ allele induced XCI by coating the X chromosome in the extraembryonic lineages of the embryo, it was not upregulated in the embryonic lineage. To make a more direct assessment of this issue, we set out to examine whether or not the $Xist^{IVS}$ allele became upregulated in the epiblast of the postimplantation embryo. We crossed females heterozygous for $Xist^{1lox}$ with males hemizygous for $Xist^{IVS}$, and embryos were recovered at embryonic day (E) 6.5 for whole-mount RNA-FISH. Since the $Xist^{1lox}$ allele is functionally null and does not form the *Xist* cloud (Sado et al., 2005), RNA-FISH allowed us to evaluate the ability of the $Xist^{IVS}$ allele to be upregulated in $Xist^{1lox}/Xist^{IVS}$ embryos. These embryos were morphologically reminiscent of previously reported $Xist^{IVS}/Xist^{IVS}$ embryos (Hoki et al., 2011), in which the epiblast is diminished, and therefore, they were differentiated from morphologically normal +/$Xist^{IVS}$ female embryos, in which wild-type *Xist* is uniformly upregulated, and male embryos ($Xist^{1lox}$/Y or +/Y) by visual inspection. **Figure 1** shows the distal part of +/$Xist^{IVS}$ and $Xist^{1lox}/Xist^{IVS}$ embryos examined at E6.5 for the expression of *Xist* and another X-linked *Atrx* gene by whole-mount RNA-FISH. In the +/$Xist^{IVS}$ embryo, the *Xist* cloud was detected in the epiblast as well as the visceral endoderm layer surrounding the epiblast. *Atrx* was monoallelically expressed in these tissues, suggesting that one *Atrx* allele was silenced by XCI. In contrast, the *Xist* cloud was detected in the visceral endoderm but not in the epiblast of $Xist^{1lox}/Xist^{IVS}$ embryos. A pinpoint *Xist* signal detected was, however, often juxtaposed to a pinpoint signal of *Atrx*. Since even if RNA was transcribed from the $Xist^{1lox}$ allele, it was truncated due to the insertion of an IRESEGFP-pA cassette in exon1 and would not be detected by the *Xist* probe used, and the pinpoint signals detected by the *Xist* probe should represent RNA transcribed from the $Xist^{IVS}$ allele, suggesting a defect of the $Xist^{IVS}$ allele to undergo upregulation. It was, therefore, likely that XCI was not induced in the epiblast of $Xist^{1lox}/Xist^{IVS}$ embryos, resulting in developmental failure of the epiblast. Taken together with the result in female ESCs homozygous for $Xist^{IVS}$, we concluded that the $Xist^{IVS}$ allele was defective in upregulating its transcription in the epiblast lineage. This contrasts with the fact that the $Xist^{IVS}$ allele is upregulated and capable of inducing XCI in the extraembryonic lineages and suggests some differences in the mechanism of *Xist* upregulation between the embryonic and extraembryonic lineages.

Forced Expression of $Xist^{IVS}$ RNA Compromised XCI in a Fraction of Differentiating Embryonic Stem Cells

To circumvent the defect of the $Xist^{IVS}$ allele to be upregulated, we attempted to replace the endogenous *Xist* promoter with the CAG promoter (Niwa et al., 1991) in F1 hybrid female ESCs between JF1 and B6 by homologous recombination with the aid of CRIPSR/Cas9 genome editing (**Supplementary Figure 2**). ESCs used here carried the neomycin resistance gene on the X chromosome derived from B6 (X^{B6}) and the Zeocin resistance gene on the X derived from JF1 (X^{JF1}) and, therefore, stably maintained both X chromosomes in the presence of G418 and Zeocin in culture medium (see *Materials and methods*). We isolated several lines harboring the $Xist^{CAGIVS2lox}$ allele on either X chromosome and chose two lines, one targeted on the X^{B6} (IVS-2L-B47) and the other on the X^{JF1} (IVS-2L-J19), for further analyses. A floxed PacECFPpA cassette was removed by transient expression of Cre recombinase in IVS-2L-B47 and IVS-2L-J19 lines to obtain IVS-B47#24 and IVS-J19#3, respectively, in which the $Xist^{CAGIVS2lox}$ allele was converted into the $Xist^{CAGIVS}$ allele.

These ESC lines thus generated carrying either $Xist^{CAGIVS2lox}$ (IVS-2L-B47 and IVS-2L-J19) or $Xist^{CAGIVS}$ (IVS-B47#24 and IVS-J19#3) were allowed to differentiate for up to 7 days in N2B27 medium without G418 and Zeocin, and expression of *Xist* and X-linked *Atrx* was examined by RNA-FISH. Given our previous studies of the targeted *Xist* alleles generated in a similar scheme, $Xist^{CAG2L}$ and $Xist^{CAG\Delta 5'\text{-}2L}$ (Amakawa et al., 2015; Sakata et al., 2017), it was reasonable to expect that the $Xist^{CAGIVS2L}$ allele would behave as a functionally null allele and the other wild-type allele would be upregulated upon differentiation. In IVS-2L-B47 and IVS-2L-J19 cells, the proportion of cells containing the *Xist* cloud increased over time to reach 65–70% of the population at day 7 (d7) (**Figures 2A,B**). Expression of *Atrx* was detected as a single pinpoint signal, which did not overlap with the *Xist* cloud, suggesting that *Atrx* on the *Xist* RNA-coated X was silenced. In IVS-B47#24 and IVS-J19#3 cells, although the $Xist^{CAGIVS}$ allele was driven by the CAG promoter, it was not upregulated prior to differentiation as previously reported (Amakawa et al., 2015). Following differentiation, although gradual upregulation of *Xist* was similarly observed in IVS-B47#24 and IVS-J19#3 cells over time, a large fraction of cells with a single *Xist* cloud expressed *Atrx* biallelically with expression of one allele overlapping with the *Xist* cloud at d5 and d7. In the remaining fraction of cells with a single *Xist* cloud, *Atrx* was monoallelically expressed, and its signal did not overlap with the *Xist* cloud. In addition, there

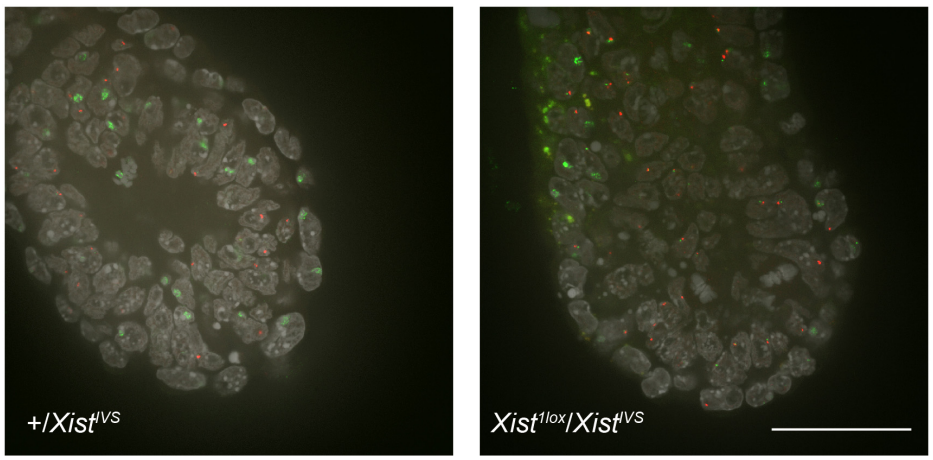


FIGURE 1 | Whole-mount RNA-FISH in the distal part of E6.5 embryos. Expression of *Xist* (green) and *Atrx* (red) was examined in +/$Xist^{IVS}$ and $Xist^{1lox}$/$Xist^{IVS}$ embryos recovered on E6.5. Six confocal sections of a 0.2-μm interval obtained by confocal microscopy were projected into a single image. epi, epiblast; ve, vesceral endoderm. Scale bar = 100 um.

was a gradual increase of cells that contained two *Xist* clouds, one of which overlapped with an *Atrx* signal. Such cells represented ∼30% of the population by d7. These results suggested that one of the two *Xist* clouds detected in a fraction of differentiating IVS-B47#24 and IVS-J19#3 cells allowed misexpression of *Atrx* and, therefore, was defective in silencing the X chromosome.

Kinetics of Allelic Expression of *Xist* in Differentiating Embryonic Stem Cells Heterozygous for $Xist^{CAGIVS}$

To identify the allelic origin of *Xist* RNA forming the *Xist* cloud in differentiating female ESCs, we employed allele-specific RNA-FISH using labeled B6- and JF1-specific oligo probes that differentially hybridized to *Xist* RNA derived from the B6 and JF allele, respectively (Harris et al., 2019). This confirmed that only the wild-type *Xist* allele expressed from the X^{JF1} and X^{B6} formed the cloud in IVS-2L-B47 and IVS-2L-J19 cells, respectively, at day 7 of differentiation.

Subsequently performed allele-specific *Xist* RNA-FISH revealed that *Xist* RNA forming the cloud in IVS-B47#24 cells was unexpectedly biased toward the RNA occurring from the wild-type allele on the X^{JF1} rather than the one from the $Xist^{CAGIVS}$ allele on the X^{B6} on d3 (**Figures 3A,B**). Although this bias was more pronounced on d5, two *Xist* clouds detected in a subset of cells were indeed originated from the wild-type and $Xist^{CAGIVS}$ alleles on the X^{B6} and X^{JF1}, respectively. The population of cells containing two *Xist* clouds increased from d5 to d7 with a decrease in the proportion of cells containing the single cloud of wild-type *Xist* on the X^{JF1}. This suggested that the $Xist^{CAGIVS}$ allele became upregulated in a subset of cells, which had undergone differentiation and initiated XCI *via* upregulation of wild-type *Xist* on the X^{JF1}, in IVS-B47#24 cells.

The kinetics of allelic expression of *Xist* in IVS-J19#3 cells was different. Cells that had initiated XCI by d3 seemed to have selected either the wild-type allele on the X^{B6} or the $Xist^{CAGIVS}$ allele on the X^{JF1} for upregulation in a random fashion (**Figures 3A,B**). Those that upregulated the $Xist^{CAGIVS}$ allele, however, had dominated in the population carrying the single *Xist* cloud by d5. As was the case in IVS-B47#24 cells, the proportion of these cells subsequently decreased from d5 to d7 with an increase of those containing two *Xist* clouds, which were derived from the wild-type and $Xist^{CAGIVS}$ allele on the X^{B6} and X^{JF1}, respectively. This suggested that the wild-type *Xist* became upregulated in a subset of cells, which had upregulated the $Xist^{CAGIVS}$ first during differentiation.

Although we expected that the $Xist^{CAGIVS}$ allele driven by the CAG promoter was preferentially upregulated upon differentiation, allele-specific RNA-FISH suggested that it was not necessarily the case and either allele could be chosen for upregulation when XCI initiated. Although two *Xist* clouds were subsequently formed in both IVS-B47#24 and IVS-J19#3 cells, the kinetics to become the two cloud states appeared different. Nonetheless, allele-specific RNA-FISH confirmed that the two *Xist* clouds were indeed formed by wild-type *Xist* RNA and $Xist^{IVS}$ RNA derived from the $Xist^{CAGIVS}$ allele. It was most likely that $Xist^{IVS}$ RNA was the one forming the cloud defective in *Atrx* silencing, and therefore, it was functionally compromised.

$Xist^{IVS}$ RNA Was Defective Because It Underwent Splicing

$Xist^{IVS}$ RNA inevitably contains an additional 16-base insertion derived from the targeting vector after splicing of the IVS2 sequence (Ohhata et al., 2008). To address the impact of the presence of the 16-base insertion on the function of *Xist* RNA, we generated another female ESCs harboring the $Xist^{CAG16in}$ allele, which produced $Xist^{16in}$ RNA containing exactly the same 16-base insertion at the same position without splicing under the control of the CAG promoter (**Supplementary Figure 3**). The ESC lines thus generated, 16in-B11#2 and 16in-B38#2, which

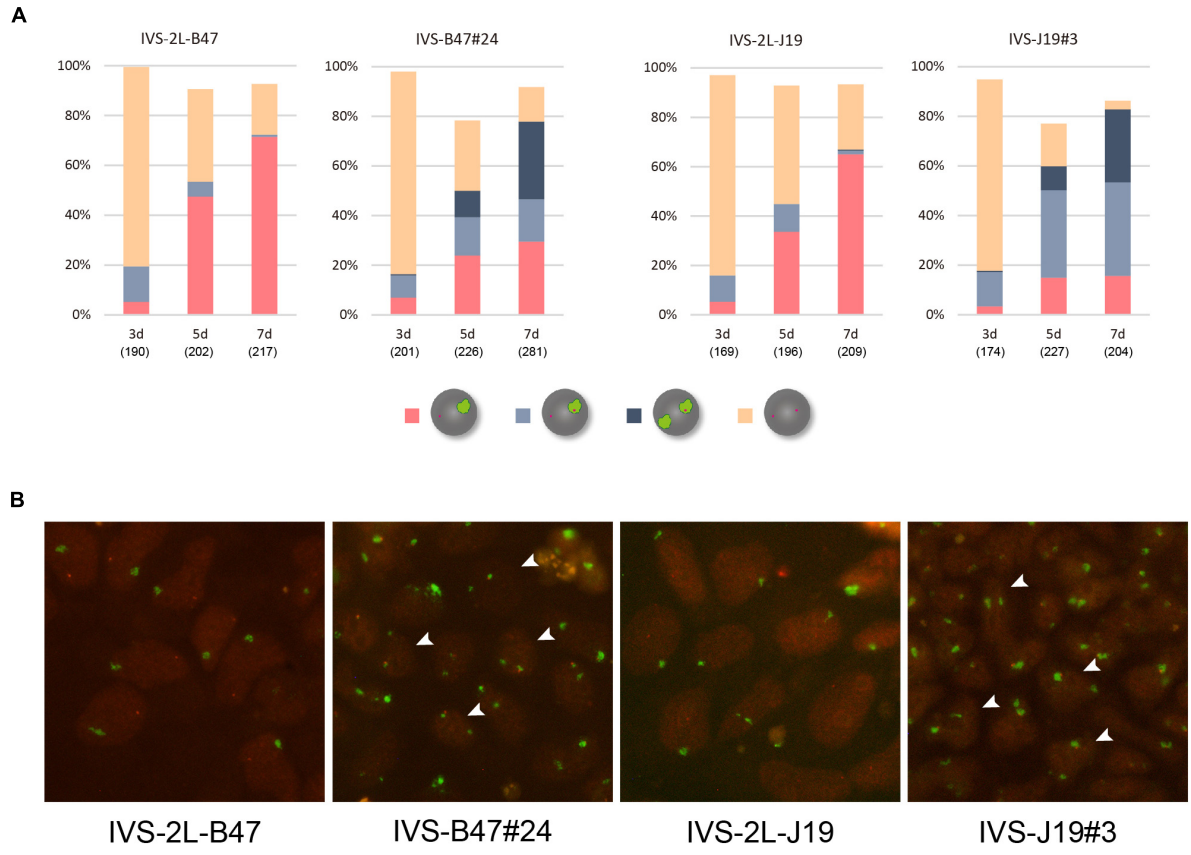


FIGURE 2 | RNA-FISH in differentiating ESCs heterozygous for either *Xist*^{CAGIVS2lox} or *Xist*^{CAGIVS}. **(A)** Expression of *Xist* and *Atrx* was examined in respective differentiating ESCs for day 3 (d3) up to day 7 (d7). IVS-2L-B47 and IVS-B47#24 carry the targeted *Xist* allele on the X^{B6}; IVS-2L-J19 and IVS-J19#3 on the X^{JF1}. The number of nuclei examined is indicated in parentheses under each day point. **(B)** Representative images of RNA-FISH in each ESC line on d7. *Xist* in green and *Atrx* in red. Arrowhead indicates some of the nuclei containing two *Xist* clouds, one of which juxtaposes an *Atrx* signal.

carried the targeted allele on the X^{B6} (those carrying the targeted allele on the X^{JF1} were not recovered), and their parental line containing a floxed PacECFPpA cassette, 16in-2L-B11 and 16in-2L-B38, respectively, were allowed to differentiate in N2B27 medium and examined for *Xist* and *Atrx* expression by RNA-FISH. In all cases, the proportion of cells containing the *Xist* cloud gradually increased over time (**Figure 4**) and *Xist* RNA forming a single cloud essentially silenced *Atrx* as few overlaps were observed. This suggested that *Xist*^{16in} RNA was capable of inducing and establishing the stably silenced state of the X chromosome. In contrast to IVS-B47#24 and IVS-J19#3 cells, only a minor population of the cells showed two *Xist* clouds in 16in-B11#2 and 16in-B38#2.

Allele-specific RNA-FISH for *Xist* revealed that the RNA forming the majority of the cloud derived from X^{B6}, on which the *Xist*^{16in} allele had been introduced, in 16in-B38#2 on d7 (**Figures 5A,B**). RT-PCR and subsequent restriction digestion of the amplified fragment further confirmed the expression of *Xist*^{16in} RNA in 16in-B11#2 and 16in-B38#2 (**Figure 5C**). The same analysis also demonstrated the production of not only the expected spliced product from the *Xist*^{CAGIVS} allele but also wild-type *Xist* RNA in IVS-B47#24 and IVS-J19#3 (**Figure 5C**), consistent with allele-specific RNA-FISH for *Xist*, shown in **Figure 3**. These results indicated that *Xist*^{16in} RNA was the one that formed the cloud stably silencing *Atrx* in 16in-B38#2 cells. *Xist*^{16in} RNA was therefore competent to induce stable XCI even though it contains exactly the same 16-base insertion at the same position as *Xist*^{IVS} RNA. It is, therefore, reasonable to conclude that the presence of the 16-base insertion per se does not compromise the function of *Xist* RNA, and it is splicing that deteriorates the function of the RNA expressed from the *Xist*^{CAGIVS} allele.

DISCUSSION

Xist^{IVS} Was Not Upregulated in the Embryonic Tissues

We previously reported that although *Xist*^{IVS} RNA expressed in the extraembryonic tissues could not establish a stable XCI state, it could support the early postimplantation development by inducing appreciable levels of XCI (Hoki et al., 2011). We also described that the *Xist* cloud was detected in 70% of the nuclei but absent in the remaining 30% in the distal part of

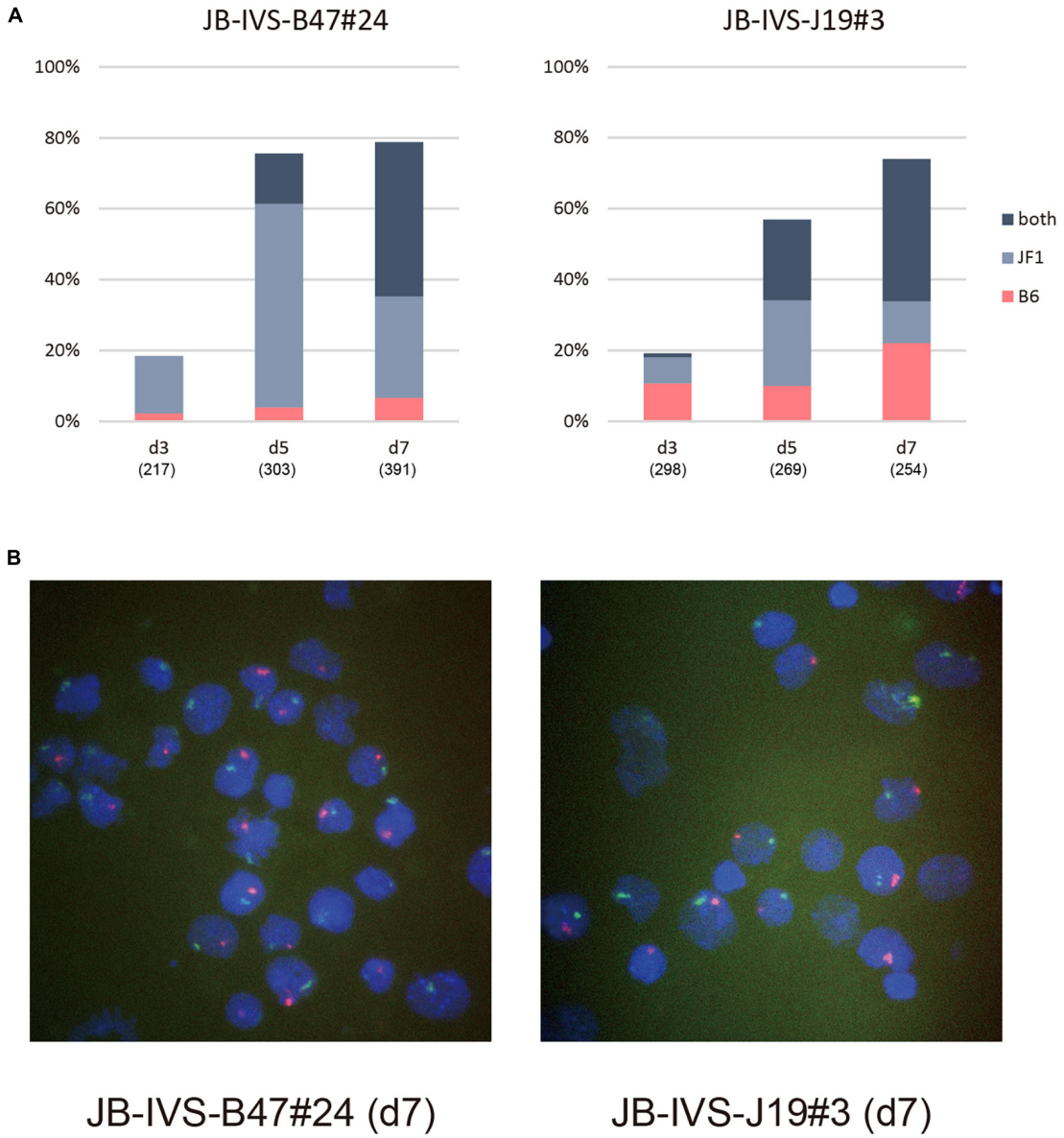


FIGURE 3 | Allele-specific RNA-FISH for *Xist* in IVS-B47#24 and IVS-J19#3. **(A)** Proportion of nuclei harboring the *Xist* cloud of the X^{B6} or X^{JF1} in origin or both in respective differentiating ESCs for d3 up to d7. IVS-B47#24 carries $Xist^{CAGIVS}$ on the X^{B6} and IVS-J19#3 on the X^{JF1}. **(B)** Representative images of allele-specific RNA-FISH for *Xist* in each ESC line on d7. *Xist* RNA originated from the X^{B6} and X^{JF1} is shown in red and green, respectively.

E7.5 $Xist^{IVS}/Xist^{IVS}$ embryos dissociated by treatment with lactic acid. This led us to speculate that although the $Xist^{IVS}$ allele was initially monoallelically upregulated, the transcripts coating one of the two Xs was lost over time in the embryonic tissue. However, the relative abundance of the epiblast cells in the distal part of an E7.5 embryo, which consists of the epiblast and visceral endoderm, would be much lower in $Xist^{IVS}/Xist^{IVS}$ embryos characterized by the small epiblast that fails to expand than in the morphologically normal wild-type embryos. Inevitable contamination of a relatively large proportion of the visceral endoderm in the mutant embryos might have led us to misinterpretation. In this study, however, whole-mount RNA-FISH using E6.5 compound heterozygotes, $Xist^{1lox}/Xist^{IVS}$, unequivocally demonstrated that the upregulation of the $Xist^{IVS}$ allele was confined only to the visceral endoderm and no *Xist* cloud was formed in the epiblast in the distal part of the embryo. This finding together with RNA-FISH analysis of ESCs homozygous for $Xist^{IVS}$ indicated that $Xist^{IVS}$ was not upregulated in the embryonic tissue. The differential behavior of the $Xist^{IVS}$ allele between the embryonic and extraembryonic tissues is most probably related to the mode of XCI, that is, imprinted or random. While in the tissues that undergo imprinted XCI, *Tsix*, an antisense RNA of *Xist*, is expressed on the maternal X but not on the paternal X, it is biallelically expressed

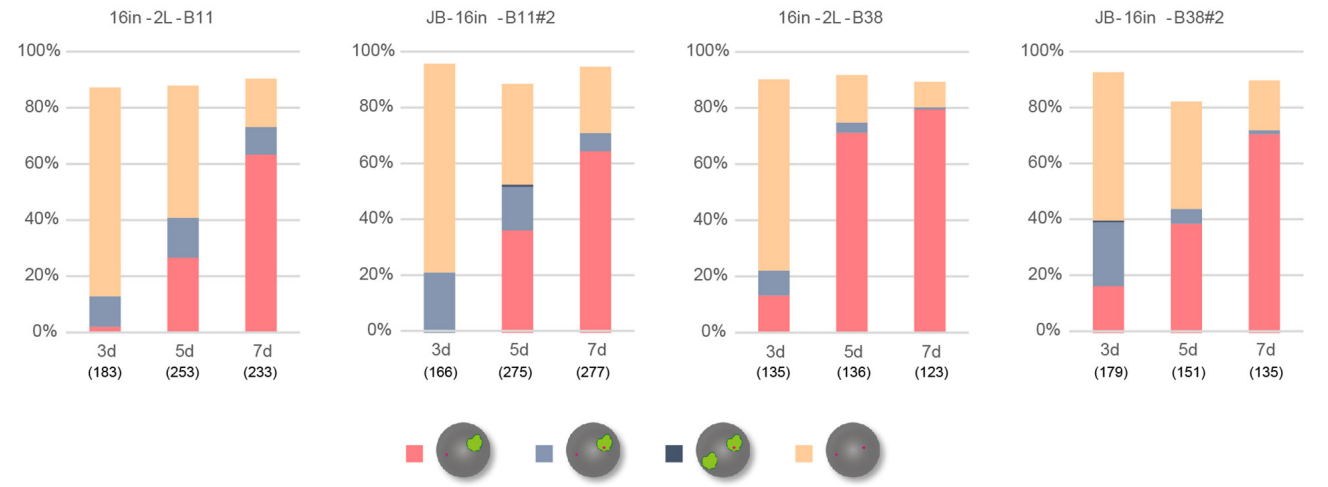


FIGURE 4 | RNA-FISH in differentiating ESCs heterozygous for either *Xist*^{CAG16in2lox} or *Xist*^{CAG16in}. Expression of *Xist* and *Atrx* was examined in respective differentiating ESCs for day 3 (d3) up to day 7 (d7). Each cell line examined (16in-2L-B11, 16in-B11#2, 16in-2L-B38, and 16in-B38#2) carried the targeted *Xist* allele on the X^{B6}. The number of nuclei examined is indicated in parentheses under each day point.

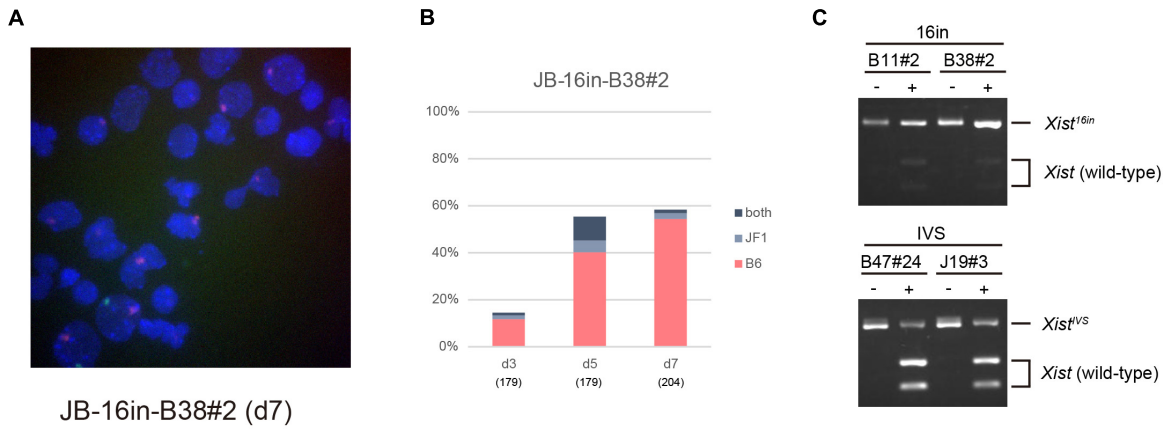


FIGURE 5 | *Xist*^{16in} RNA induces stable XCI. **(A)** Representative images of allele-specific RNA-FISH for *Xist* in 16in-B38#2 differentiated for 7 days. *Xist* RNA originated from the X^{B6} and X^{JF1} is shown in red and green, respectively. **(B)** Proportion of nuclei harboring the *Xist* cloud of the X^{B6} or X^{JF1} in origin or both in differentiating 16in-B38#2 ESCs for d3 up to d7. **(C)** Expression of *Xist*^{16in} and *Xist*^{IVS} RNA in differentiating ESCs heterozygous for the *Xist*^{16in} and *Xist*^{CAGIVS} allele, respectively. The unique *Xho*I site present in the amplified fragment derived from wild-type *Xist* RNA was destroyed by the presence of the 16-base insertion in the amplified fragment derived from *Xist*^{16in} and *Xist*^{IVS} RNA. The presence of an undigested band by *Xho*I indicates the expression of *Xist*^{16in} and *Xist*^{IVS} RNA. (–) and (+) indicate undigested and digested, respectively.

prior to upregulation of *Xist* in undifferentiated epiblast cells and ESCs, which undergo random XCI as cells differentiate (Lee, 2000; Sado et al., 2001). Given that *Tsix* negatively regulates *Xist* through modification of chromatin structure (Navarro et al., 2005; Sado et al., 2005), it is possible that the presence or absence of *Tsix* RNA or its transcription affects the potential of the *Xist*^{IVS} allele to be upregulated in response to cellular differentiation. *Tsix* expression is downregulated on one of the two Xs in the epiblast or ESCs during differentiation, and *Xist* becomes upregulated on that X to induce XCI (Lee et al., 1999). A series of these events may be compromised on the X chromosome carrying the *Xist*^{IVS} allele in the epiblast cells or ESCs. In the extraembryonic tissues, on the other hand, since *Tsix* is imprinted not to be expressed on the paternal X, there would not be any influence of *Tsix* on the paternal *Xist*^{IVS} allele, allowing its upregulation.

Splicing Deteriorates the Function of *Xist*^{IVS} RNA

We expected that when female ESCs heterozygous for *Xist*^{CAGIVS} were allowed to differentiate, the *Xist*^{CAGIVS} allele would become preferentially upregulated and its transcript, *Xist*^{IVS} RNA, would coat the X chromosome. Allele-specific RNA-FISH, however, revealed that this was not the case and either the wild-type or the *Xist*^{CAGIVS} allele could be monoallelically upregulated at the

onset of XCI. Although we had to admit that the $Xist^{CAGIVS}$ allele behaved a little different way from the one that we expected during the early phase of differentiation, still we could express the $Xist^{IVS}$ RNA in differentiating ESCs as a result of the formation of two *Xist* clouds in both cell lines, IVS-B47#24 and IVS-J19#3. We speculate that in IVS-B47#24, although the wild-type *Xist* allele on the X^{JF1} was initially preferentially upregulated upon initiation of XCI, the subsequent differentiated state of the cells allowed them to activate the CAG promoter, which barely drives transcription at the *Xist* locus in undifferentiated state, to express $Xist^{IVS}$ RNA from the $Xist^{CAGIVS}$ allele. In IVS-J19#3, on the other hand, one of either allele of *Xist* was selected for upregulation at the onset of XCI; the wild-type *Xist* allele appeared secondarily activated in those that upregulated the $Xist^{CAGIVS}$ allele first to compensate for the insufficient XCI initiated by $Xist^{IVS}$ RNA. It is likely that those that upregulated the wild-type *Xist* first in the population of IVS-J19#3 formed two *Xist* clouds in the same manner as IVS-B47#24. Whatever the reason for the formation of two *Xist* clouds is, upregulation of the $Xist^{CAGIVS}$ allele allowed us to evaluate the function of $Xist^{IVS}$ RNA in the embryonic lineage. Since one of the two *Xist* clouds seemed to be defective in silencing *Atrx*, it was most likely that $Xist^{IVS}$ RNA produced from the $Xist^{CAGIVS}$ allele was dysfunctional and failed to establish the proper XCI state. In contrast, $Xist^{16in}$ RNA containing the same 16-base insertion at the same position as $Xist^{IVS}$ RNA was indistinguishable from wild-type *Xist* RNA in terms of the kinetics of *Atrx* silencing. This strongly suggests that it is not the 16-base insertion per se that compromises the function of $Xist^{IVS}$ RNA. The difference between $Xist^{IVS}$ RNA and $Xist^{16in}$ RNA is whether or not the RNA undergoes splicing to remove the IVS2 sequence introduced 0.9 kb downstream of the major transcription start site of *Xist*. When an intron is removed and two exons are connected by splicing, many splicing-related proteins such as an exon junction complex (EJC) bind in the vicinity of the exon-exon junction. It is, therefore, reasonable to expect that such proteins would bind to the processed $Xist^{IVS}$ RNA but not to $Xist^{16in}$ RNA. The IVS2 sequence was located about 0.2 kb downstream of the A-repeat, which is essential for the silencing function of *Xist* RNA (Wutz et al., 2002) and mediates binding of some important proteins required for XCI such as SPEN (Chu et al., 2015; McHugh et al., 2015; Moindrot et al., 2015; Monfort et al., 2015) and RBM15 (Patil et al., 2016). We are tempted to speculate that binding of these factors to *Xist* RNA is disturbed by the presence of EJC and other factors brought on $Xist^{IVS}$ RNA as a result of splicing to remove the IVS2 sequence. It would be therefore particularly interesting to compare the proteins assembled on $Xist^{IVS}$ RNA with those on the wild-type *Xist* RNA. We cannot, however, exclude the possibilities that the inefficient silencing associated with $Xist^{IVS}$ RNA could result from the efficiency of the expected splicing event on the premature $Xist^{IVS}$ RNA or the difference of the overall levels of wild-type *Xist*, $Xist^{IVS}$, and $Xist^{16in}$ RNA. The fact that differentiating ESCs carrying the $Xist^{CAGIVS}$ allele give rise to the populations with biallelic expression of *Xist* (wild-type and $Xist^{IVS}$) as well as monoallelic expression of either allele makes it difficult to compare the quantity of the RNA produced from the respective wild-type and $Xist^{CAGIVS}$ allele. To circumvent this problem, we are currently attempting to derive a unique cell population with monoallelic expression of either allele or that with biallelic expression of both by inducing differentiation of ESCs into neural stem cells and subsequent cloning.

Since most of the *Xist* mutant allele so far generated compromise the initiation process of XCI, $Xist^{IVS}$ RNA is peculiar in that it can initiate XCI but fails to maintain the XCI state. This is most probably due to the failure of the establishment of robust heterochromatin. It is likely that proteins recruited by *Xist* RNA contribute to not only the initiation of XCI but also the establishment of a sustainable heterochromatin state. $Xist^{IVS}$ RNA would provide a unique opportunity to explore such factors involved in the latter process and how *Xist* participates in the establishment of the chromatin state required for the stable maintenance of the X-inactivated state.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by the committee for the Care and Use of Laboratory Animals of Kindai University.

AUTHOR CONTRIBUTIONS

TS and TN designed the study. RM, TN, SI, and TS performed the experiments and analyzed the data. TS wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported partly by Grants-in-Aid for Scientific Research (A) from the Japan Society for the Promotion of Science (JSPS) (17H01588 and 20H00550 to TS) and Takeda Science Foundation to TS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.751154/full#supplementary-material

REFERENCES

Amakawa, Y., Sakata, Y., Hoki, Y., Arata, S., Shioda, S., Fukagawa, T., et al. (2015). A new Xist allele driven by a constitutively active promoter is dominated by Xist locus environment and exhibits the parent-of-origin effects. *Development* 142, 4299–4308. doi: 10.1242/dev.128819

Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., et al. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* 351, 325–329. doi: 10.1038/351325a0

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., et al. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* 351:351329a0. doi: 10.1038/351329a0

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., et al. (2015). Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404–416. doi: 10.1016/j.cell.2015.03.025

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell. Biol.* 132, 259–275. doi: 10.1083/jcb.132.3.259

Harris, C., Cloutier, M., Trotter, M., Hinten, M., Gayen, S., Du, Z., et al. (2019). Conversion of random X-inactivation to imprinted X-inactivation by maternal PRC2. *Elife* 8:e44258. doi: 10.7554/eLife.44258

Hoki, Y., Ikeda, R., Mise, N., Sakata, Y., Ohhata, T., Sasaki, H., et al. (2011). Incomplete X-inactivation initiated by a hypomorphic Xist allele in the mouse. *Development* 138, 2649–2659. doi: 10.1242/dev.061226

Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* 72, 171–182. doi: 10.1016/0092-8674(93)90658-D

Lee, J., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* 21, 400–404. doi: 10.1038/7734

Lee, J. T. (2000). Disruption of imprinted X inactivation by parent-of-origin effects at tsix. *Cell* 103, 17–27. doi: 10.1016/S0092-8674(00)00101-X

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190, 372–373. doi: 10.1038/190372a0

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* 11, 156–166. doi: 10.1101/gad.11.2.156

McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521:232. doi: 10.1038/nature14443

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., et al. (2015). A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* 12, 562–572. doi: 10.1016/j.celrep.2015.06.053

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., et al. (2015). Identification of spen as a crucial factor for xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* 12, 554–561. doi: 10.1016/j.celrep.2015.06.067

Monk, M., and Harper, M. I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature* 281, 311–313. doi: 10.1038/281311a0

Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005). Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Gene Dev.* 19, 1474–1484. doi: 10.1101/gad.341105

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 108, 193–199. doi: 10.1016/0378-1119(91)90434-D

Ohhata, T., Hoki, Y., Sasaki, H., and Sado, T. (2008). Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. *Development* 135, 227–235. doi: 10.1242/dev.008490

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Leon, W. R. M., Damianov, A., et al. (2020). A protein assembly mediates Xist localization and gene silencing. *Nature* 587, 145–151. doi: 10.1038/s41586-020-2703-0

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* 10, 1991–2002. doi: 10.1101/gad.10.16.1991

Patil, D. P., Chen, C.-K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537, 369–373. doi: 10.1038/nature19342

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature* 379:379131a0. doi: 10.1038/379131a0

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B. A., Solcan, N., Cerase, A., et al. (2017). hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell* 68, 955–969.e10. doi: 10.1016/j.molcel.2017.11.013

Sado, T., Hoki, Y., and Sasaki, H. (2005). Tsix silences Xist through modification of chromatin structure. *Dev. Cell* 9, 159–165. doi: 10.1016/j.devcel.2005.05.015

Sado, T., Wang, Z., Sasaki, H., and Li, E. (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* 128, 1275–1286. doi: 10.1242/dev.128.8.1275

Sakata, Y., Nagao, K., Hoki, Y., Sasaki, H., Obuse, C., and Sado, T. (2017). Defects in dosage compensation impact global gene regulation in the mouse trophoblast. *Development* 144, 2784–2797. doi: 10.1242/dev.149138

Shiura, H., Sakata, Y., Abe, K., and Sado, T. (2018). RNA-FISH and immunofluorescence of mouse preimplantation and postimplantation embryos. *Methods Mol. Biol.* 1861, 161–176. doi: 10.1007/978-1-4939-8766-5_13

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* 256, 640–642. doi: 10.1038/256640a0

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* 30, 167–174. doi: 10.1038/ng820

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

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: 25 October 2021
doi: 10.3389/fcell.2021.730014

The Role of Long Non-coding RNAs in Human Imprinting Disorders: Prospective Therapeutic Targets

*Tingxuan Wang, Jianjian Li, Manyin Wu, Liuyi Yang and Qing Ma**

Shenzhen Key Laboratory of Synthetic Genomics, Guangdong Provincial Key Laboratory of Synthetic Genomics, CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Genomic imprinting is a term used for an intergenerational epigenetic inheritance and involves a subset of genes expressed in a parent-of-origin-dependent way. Imprinted genes are expressed preferentially from either the paternally or maternally inherited allele. Long non-coding RNAs play essential roles in regulating this allele-specific expression. In several well-studied imprinting clusters, long non-coding RNAs have been found to be essential in regulating temporal- and spatial-specific establishment and maintenance of imprinting patterns. Furthermore, recent insights into the epigenetic pathological mechanisms underlying human genomic imprinting disorders suggest that allele-specific expressed imprinted long non-coding RNAs serve as an upstream regulator of the expression of other protein-coding or non-coding imprinted genes in the same cluster. Aberrantly expressed long non-coding RNAs result in bi-allelic expression or silencing of neighboring imprinted genes. Here, we review the emerging roles of long non-coding RNAs in regulating the expression of imprinted genes, especially in human imprinting disorders, and discuss three strategies targeting the central long non-coding RNA *UBE3A-ATS* for the purpose of developing therapies for the imprinting disorders Prader–Willi syndrome and Angelman syndrome. In summary, a better understanding of long non-coding RNA-related mechanisms is key to the development of potential therapeutic targets for human imprinting disorders.

Keywords: genomic imprinting, lncRNA, epigenetic regulation, imprinting disorders, UBE3A-ATS, ASO, CRISPR-Cas9

OPEN ACCESS

Edited by:
Lan Jiang,
Key Laboratory of Genome Sciences & Information, Beijing Institute of Genomics (CAS), China

Reviewed by:
Joseph Mauro Calabrese,
University of North Carolina at Chapel Hill, United States
Janine M. LaSalle,
University of California, Davis, United States

***Correspondence:**
Qing Ma
qing.ma@siat.ac.cn

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

Received: *24 June 2021*
Accepted: *23 September 2021*
Published: *25 October 2021*

Citation:
Wang T, Li J, Wu M, Yang L and Ma Q (2021) The Role of Long Non-coding RNAs in Human Imprinting Disorders: Prospective Therapeutic Targets. Front. Cell Dev. Biol. 9:730014. doi: 10.3389/fcell.2021.730014

Abbreviations: ASE, allele-specific gene expression; MAE, monoallelic expression; PWS, Prader–Willi syndrome; AS, Angelman syndrome; BWS, Beckwith–Wiedemann syndrome; SRS, Silver–Russell syndrome; KOS14, Kagami–Ogata syndrome; TS14, Temple syndrome; ICRs, imprinting control centers or imprinting control regions; gDMRs, germline differentially methylated regions; lncRNAs, long non-coding RNAs; ASOs, antisense oligonucleotides; SNPs, single nucleotide polymorphisms; H4R3me2s, histone H4 arginine-3 symmetrical demethylation; H3K9me3, H3 lysine-9 trimethylation; Igf2, the insulin-like growth factor 2; Kcnq1/Kcnq1ot1, potassium voltage-gated channel subfamily Q member 1/Kcnq1 antisense transcript 1; KvDMR1, KvLQT1 differentially methylated region 1; PRC, polycomb repressive complex; EHMT2, histone methyltransferase euchromatic histone lysine *N*-methyltransferase-2; Airn, the antisense of Igf2r non-protein coding RNA; H2AK119ub1, lysine 119-monoubiquititinated histone H2A; Ube3a-ATS, Ube3a-antisense lncRNA; UPD, uniparental disomy; MLID, multi-locus imprinting disturbance; SNORD116, SnoRNA C/D box cluster 116; RBFOX2, RNA binding protein fox-1 homolog 2; DLK1/DIO3, delta-like homolog 1 gene/type III iodothyronine deiodinase gene; Rtl1as, the Rtl1-antisense; PI3K/Akt, phosphoinositide 3-kinase/protein kinase B; CRISPRi, CRISPR interference; CRISPRa, CRISPR activation; topoisomerase I (Top I) inhibitors.

INTRODUCTION

In diploid organisms, most genes are transcribed in an unbiased fashion from both alleles. However, in a small subset of genes, genetically identical alleles can be expressed differentially, a process referred to as ASE. In the mammalian genome, common epigenetic examples of ASE include random X-chromosome inactivation in females (Lee, 2011; Deng et al., 2014), genomic imprinting (Peters, 2014), random MAE (Reinius and Sandberg, 2015), allelic expression of antigen receptor (Bergman and Cedar, 2004; Vettermann and Schlissel, 2010), clustered protocadherin (Chen and Maniatis, 2013), and olfactory receptors (Monahan and Lomvardas, 2015). Imprinted genes are expressed strictly or preferentially from either paternally or maternally inherited alleles (referred to as parent-of-origin) (Barlow and Bartolomei, 2014; Huang et al., 2018; Chen and Zhang, 2020). The ASE of imprinted genes depends on differential epigenetic markings during gametogenesis in germline cells, as opposed to gene sequences. After imprinting patterns become established in mature germlines, genomic imprinting in an individual is maintained until genome-wide erasure of epigenetic modification occurs in gamete precursors.

Genomic imprinting has been described in diverse organisms, including marsupials, flowering plants, and insects (Macdonald, 2012). In the human and mouse genome, genomic imprinting has been extensively observed, indicating the conservation and evolutionary significance of this epigenetic regulatory mechanism. While the expression of 1% of human protein-coding genes is estimated to be regulated via genomic imprinting (Im et al., 2005; Patten et al., 2016; Elbracht et al., 2020), many of these imprinted genes are essential for metabolism, development, and the nervous system (Monk et al., 2019; Tucci et al., 2019). Not surprisingly, dysregulated imprinting is closely associated with a broad spectrum of human developmental defects and genetic disorders, including PWS, AS, BWS, SRS, KOS14, and TS14 (Bian and Sun, 2011; Wan et al., 2017). The association between imprinted genes and the clinical features of these human diseases has also been documented in mouse models through the identification of homologous imprinted gene regions corresponding to the imprinted gene regions implicated in human imprinting disorders (Peters, 2014; Tucci et al., 2019).

Long non-coding RNAs are a subgroup of non-coding RNAs defined as having a length longer than 200 nucleotides, and are extensively expressed among the genome (Derrien et al., 2012; Harrow et al., 2012; Knauss and Sun, 2013). The number of lncRNA genes in the human genome has been estimated at 20,000 to 100,000 (Zhao et al., 2016; Fang et al., 2018; Uszczynska-Ratajczak et al., 2018). This number is greater than the canonical protein-coding genes in the human genome (Southan, 2017). lncRNAs are primarily retained in the nucleus, having short half-lives and a rapid turn-over rate compared to mRNAs (Clark et al., 2012; Derrien et al., 2012; Yoon et al., 2015). lncRNAs can regulate gene expression in at least three ways: at the transcription level by modulating gene transcription and chromatin structure, at the post-transcription level by affecting splicing and stability of RNA, and at the translation level by modulating protein translation (referred to review Statello et al., 2021). In the human and mouse genome, imprinted genes often reside together within clusters (2–20 genes), called imprinted clusters or imprinted domains (Ferguson-Smith, 2011). In mammals, lncRNAs are generally located in imprinted clusters that contribute to the establishment and maintenance of monoallelic expression at a genome-scale and long time-range (Andergassen et al., 2019). Here, we summarize the roles of lncRNAs in the regulation of genomic imprinting using several well-established imprinted clusters as examples. We also discuss how the expression pattern of lncRNAs and their epigenetic regulatory functions are affected in imprinting disorders and some cancers. Three potential strategies have been developed to target the central long non-coding RNA *Ube3a-ATS* for the purpose of therapeutically correcting the PWS/AS locus imprinting disorders. We also discuss the functional mechanisms of imprinted lncRNAs in the regulation of mono-allelic imprinted gene expression and how it could help us understand ASE mechanisms and underlying pathological mechanisms of human imprinting disorders, hopefully inspiring additional efficient therapeutic strategies.

GENOMIC BASIS OF IMPRINTING

Along with more profound analysis of patient samples and well-established mouse reciprocal crossing models using high-throughput sequencing, the monoallelic expression of imprinted genes has been observed extensively in mice and humans (Tucci et al., 2019). Methylomes and transcriptomes derived from human peripheral blood and various adult tissue samples have been combined to identify imprinted methylation and the distribution of imprinted genes across the genome (Baran et al., 2015; Zink et al., 2018). In order to identify mouse imprinted genes, parents from strains with different genetic backgrounds were crossed to obtain heterozygotic individuals, permitting the discrimination of parent-of-origin-dependent transcriptional effects from sequence-dependent allelic expression (Babak et al., 2008; Wang et al., 2008). Imprinted genes in mice are identified based on SNPs specific to paternal or maternal genetic backgrounds, thus permitting the quantitation and comparison of expression levels from both alleles. To date, around 160 imprinted genes have been identified in the human genome, and 200 in the mouse genome (Tucci et al., 2019; Chen and Zhang, 2020). Sixty three of these imprinted genes are shared, suggesting that mouse models could be helpful for understanding imprinting regulation in humans.

In the human and mouse genome, imprinted genes often reside together within imprinted clusters (Ferguson-Smith, 2011). More than 80% of the known imprinted genes in the mouse genome are clustered together in multi-gene ranging in size from less than 100 kb to several megabases (Barlow, 2011). Imprinted lncRNAs located in one imprinted cluster are coordinately controlled by shared regulatory factors, including parent-of-origin-dependent differentially methylated regions (DMRs) and lncRNAs (Peters, 2014). In well-studied imprinted clusters, allele-specific DNA methylation occurs in

an independent ICR in the germline, referred to as germline-derived DMRs (gDMRs) or primary DMRs, and persists after fertilization. ICRs in imprinted clusters exhibit parent-of-origin-specific epigenetic modifications, including DNA methylation, governing different expression patterns of parentally inherited alleles (da Rocha and Gendrel, 2019). Around 35 imprinted gDMRs have been identified in the human genome (Monk et al., 2018) and 24 in the mouse genome to date (White et al., 2016). The establishment of gDMRs on paternal or maternal alleles (**Figure 1A**) is essential for regulating imprinted gene expression in embryonic development (Barlow, 2011; Kelsey and Feil, 2013; Elbracht et al., 2020). In early primordial germ cells, epigenetic marks are extensively erased genome-wide, including DNA methylation and histone modifications. In germline cells, DNA methylation of ICRs is re-established in gametes depending on the parent-of-origin. After fertilization, gDMRs escape secondary global epigenetic reprogramming. DNA methylation information at ICRs of the imprinted regions is retained. In this way, gDMRs of imprinted loci are established robustly during germline development and are resistant to genomic reprogramming after fertilization. Correspondingly, imprinting marks are inherited in a parent-specific manner (Chotalia et al., 2009; Henckel et al., 2012; da Rocha and Gendrel, 2019). gDMRs on the different parent-of-origin alleles are characterized by distinct chromatin configurations, marked with different histone modifications which are corresponding to 'open chromatin' and 'close chromatin' (Singh et al., 2011; Court et al., 2014; Sanli and Feil, 2015). The allele-specific methylation states of gDMRs are recognized by transcription factors with roles in maintaining parent-of-origin specific expression of the imprinted genes, such as ZFP57 protein (Riso et al., 2016). In total, differential methylation states of gDMRs on parental alleles are essential for the establishment of monoallelic gene expression.

Imprinting control regions govern DNA methylation and chromatin organization in early embryonic and adult lineages, resulting in the persistence of imprinting patterns across generations and their maintenance in adult tissues (Monk et al., 2019). After becoming established at early developmental stages in the germline, gDMRs are maintained in most somatic cells throughout life, resulting in the regulation of allelic expression of imprinted gene clusters. gDMRs also direct the rise of 'secondary' DMRs, normally corresponding to repressive chromatin modifications, condensed chromatin structure, and the gene-silencing function of imprinted lncRNAs (Sasaki et al., 1995; Nowak et al., 2011, p. 2; Rao et al., 2014; Tan et al., 2018; Zink et al., 2018). It has been shown that the imprinted expression of some genes is restricted to specific tissues or stages in developmental processes, along with additional allele-specific epigenetic marks further established in somatic cells. The expression patterns of these developmentally expressed imprinted genes are characterized by temporal- and spatial-specific biases (Perez et al., 2015; Andergassen et al., 2017). For example, *UBE3A* and *IGF2* show imprinted expression patterns in specific human brain cell types (Rougeulle et al., 1997; Vu and Hoffman, 1997; Pham et al., 1998; Yamasaki et al., 2003; Li J. et al., 2020). In a study of ASE in diverse tissues from 178 adult post-mortem donors, paternally silenced *IGF2* was reported in the human brain, different from the canonical paternal expression observed in other tissues (Baran et al., 2015). In the mouse E6.5 gastrulating epiblast, it has also been observed that *Igf2r* is expressed from both alleles and further becomes imprinted in the embryonic lineage at the gastrulation stage (Marcho et al., 2015). Besides, the placenta-specific imprinting has been observed, and the underlying mechanism has been well-understood, especially in the potassium voltage-gated channel subfamily Q member 1 (*Kcnq1*)/*Kcnq1* antisense transcript 1 (*Kcnq1ot1*) cluster and the antisense of *Igf2r* non-protein coding RNA (*Airn*)/*Igf2r* cluster (**Figure 1B**; Sleutels et al., 2002; Andergassen et al., 2019; Hanna, 2020). The establishment of the placenta-specific imprinting initiates by allelic DNA methylation in pre-implantation embryos. In the placenta, the genomic profile of DNA methylation in imprinted DMRs is different, likely the result of an overall different pattern of placenta compared to other tissues (Schroeder et al., 2013). After implantation, the silencing of imprinted genes on the paternal allele in the post-implantation placental trophoblast expands and tends to be larger than the post-implantation epiblast. This expansion of gene silencing is mediated by the spreading of H3K27me3 marks along the paternal chromosome (Calabrese et al., 2015; Andergassen et al., 2017).

LONG NON-CODING RNAs AND THEIR ROLES IN REGULATING THE EXPRESSION OF IMPRINTED GENES

Two major mechanisms have been described to explain the regulation of the gene expression within an imprinted cluster (Lee and Bartolomei, 2013; Barlow and Bartolomei, 2014; Chen and Zhang, 2020). The first model is the lncRNA model, which may be more common. In this model, imprinted lncRNAs regulate imprinted gene expression. In the lncRNA model, imprinted lncRNAs intimately associate with ICRs. Imprinted lncRNAs are characterized by their capacity to silence imprinted genes in the same cluster (Rao et al., 2014; Kanduri, 2016; Tan et al., 2018; Zink et al., 2018; Tucci et al., 2019). As illustrated by the *Kcnq1/Kcnq1ot1* imprinted cluster (**Figure 2A**), actively expressed imprinted lncRNA *Kcnq1ot1* on the paternal allele can silence multiple imprinted genes bidirectionally along their located gene region (Pauler et al., 2012). In contrast, a maternally methylated ICR on the paternal directly inhibits *Kcnq1ot1* and its silencing effects, leading to the released expression of imprinted genes from the silencing by *Kcnq1ot1*. Another model, the insulator model is identified in other imprinted regions, in which parental allele-specific epigenetic differences at ICRs contribute to topological alternations of imprinted gene regions, inducing gene silencing or activation of specific alleles. This model is mainly applied to explain how imprinted genes in the insulin-like growth factor 2 (*Igf2*)/*H19* locus are mechanistically regulated (**Figure 2B**; Kaffer et al., 2000). *H19* is a maternally expressed lncRNA (Bartolomei et al., 1991; DeChiara et al., 1991; Ferguson-Smith et al., 1991). The zinc-finger protein CTCF binds to the unmethylated maternal ICR and creates topologically associating

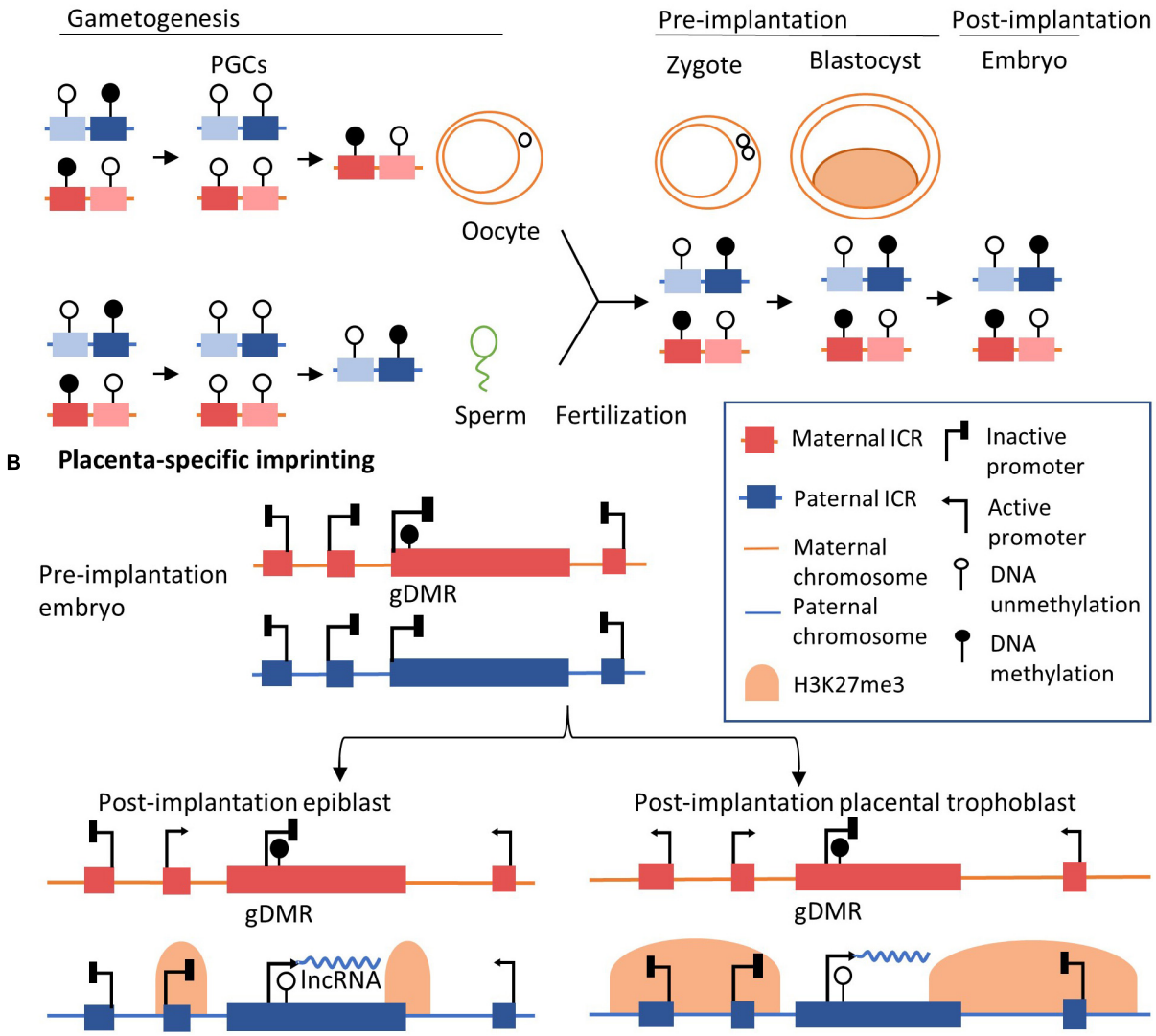


FIGURE 1 | Genomic basis of the regulation of imprinting clusters. **(A)** The inheritance of allele-specific imprinting epigenetic marks across generations. In the early primordial germ cells, epigenetic modifications are erased at a genomic scale before the formation of germline cells. In the germline, parent-of-origin DNA methylation is established, shown as gDMRs. After fertilization and the formation of the zygote, the gDMRs are further maintained. Established imprinting patterns are maintained in blastocyte and somatic cells in adult tissues. **(B)** Imprinting in epiblast and placenta in imprinting loci, such as the *Kcnq1ot1/Kcnq1* or *Airn/Igf2r* loci, is shown. In pre-implantation embryos, DNA methylation is inherited in the gDMR on the maternal allele, such as KvDMR1 of the *Kcnq1ot1/Kcnq1* imprinting cluster. After implantation, the expression of lncRNA on the maternal allele is repressed by DNA methylation in gDMR, allowing the expression of neighbor genes. On the contrary, lncRNA is expressed from the paternal allele, inducing the spreading of H3K27me3 modifications in adjacent regions in the embryonic lineage (epiblast). In extra-embryonic lineage (placental trophoblast), the extended scale of H3K27me3 marks is longer than that seen in embryonic cells. Adjacent genes further away are also silenced on the paternal allele, indicating placenta-specific imprinting, such as *Slc22a18* and *Tssc4* genes in the *Kcnq1ot1/Kcnq1* imprinting cluster. For simplicity, specific gene names are not shown.

domain boundaries, blocking *Igf2* access to the enhancer like an 'insulator' (Schoenherr et al., 2003; Gómez-Marín et al., 2015). On the paternal allele, methylated ICR prevents CTCF binding and leads to secondary methylation of the *H19* promoter and therefore silencing of lncRNA expression. The enhancers are then accessible to *Igf2*, permitting paternal-allele expression of *Igf2* (Thorvaldsen et al., 1998; Chen and Zhang, 2020). Different from the lncRNA model, imprinted lncRNAs *H19* in the insulator model are not the key regulation elements or whether imprinted lncRNAs affect other genes are not clear.

Here, we discuss on the role of imprinted lncRNAs in epigenetic regulation in the more common model, lncRNA model (Kopp and Mendell, 2018). lncRNA functions can be characterized based on their specific subcellular locations and interactions with DNA, RNA, and proteins, regulation of chromatin structure, expression of nearby and distal

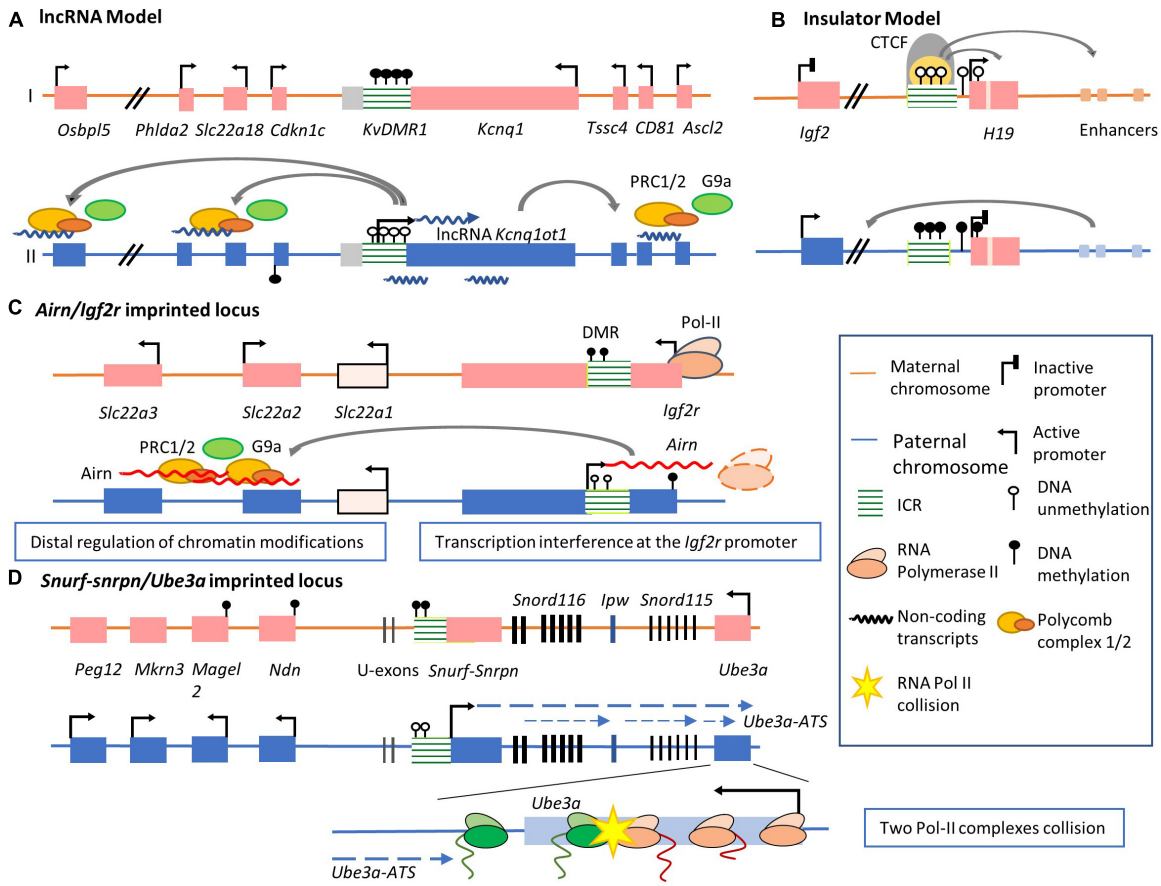


FIGURE 2 | Mechanisms by which imprinted lncRNA regulate allelic expression in imprinted clusters. **(A)** The lncRNA model of imprinted gene expression regulation. In *Kcnq1ot1* imprinted cluster, the ICR is unmethylated on the paternal allele, permitting lncRNA Kcnq1ot1 expression. The expression of this lncRNA recruits the PRC1/2 complex and histone methyltransferase G9a, leading to condensed chromatin and silencing of flanking protein-coding genes. The ICR is methylated on the maternal allele, inhibiting lncRNA expression. The expression of *Kcnq1* and several paternal silenced genes are activated. **(B)** The insulator model of imprinted gene expression regulation. The ICR on the maternal allele is unmethylated. CTCF binds to the maternal ICR and functions as an insulator to block *Igf2* access to its distal enhancers. In contrast, the ICR on the paternal allele is methylated, preventing binding of CTCF. The expression of *Igf2* is activated via enhancer regulation. **(C)** lncRNA *Airn* in *Airn/Igf2r* locus function in two distinctive mechanisms. On the one hand, methylated DMR on the maternal allele inhibits *Airn* expression, allowing access of transcription factors to the *Igf2r* promoter. The paternal DMR is unmethylated, permitting *Airn* transcription. *Airn* overlaps with the promoter of *Igf2r* and inhibits *Igf2r* expression. On the other hand, *Airn* transcripts recruit PRC2 complex to distal genes, such as *Slc22a3* and *Slc22a2*, where they silence expression. *Slc22a1* is a biallelic expressed protein-coding gene between distal regulated imprinted genes and Igf2r gene loci. **(D)** In Snurf-*Snrpn/Ube3a* imprinted cluster, the transcription of *Ube3a-ATS* starts from the exon upstream of the *Snurf-Snrpn* gene on the paternal allele. A group of non-coding RNAs are expressed, including *Snord116* and *Snord115* sno-lncRNAs and SnoRNAs. The elongation of this lncRNA overlaps with the *Ube3a* protein-coding region. A collision occurs between the converging elongation complexes of *Ube3a-ATS* and *Ube3a* resulting in the failure of *Ube3a* transcription elongation. By contrast, on the maternal allele, the ICR of *Snurf-Snrpn/Ube3a* cluster is methylated in the brain. G9a is recruited to the methylated DMR. This G9a accumulation leads to condensed chromatin and the silencing of flanking imprinted genes near the *Snurf/Snrpn* gene region. Consequently, the maternal *Ube3a* allele is expressed.

genes, RNA post-transcription modification, or mRNA translation (St Laurent et al., 2015; Kopp and Mendell, 2018; Statello et al., 2021). Imprinted lncRNAs range from 1.9 to 1,000 kb in length (Guenzl and Barlow, 2012) and regulate the expression of adjacent imprinted genes *in cis* through interacting with promoters and transcription factor binding sites, modifying chromatin status, or affecting higher-order structures (Barlow, 2011). Two major functional mechanisms of imprinted lncRNAs in the regulation of imprinted gene expression are hypothesized: interacting with promoters or enhancers of nearby target genes to affect transcription initiation, or overlapping imprinted gene regions, covering the gene body, and regulating the chromatin state of adjacent gene regions. We will also discuss the mechanisms underlying the regulation of imprinted gene expression by imprinted lncRNAs using well-characterized imprinted clusters as examples.

Transcriptional Interference

Inhibition of Transcriptional Initiation

Transcription of imprinted lncRNAs often overlaps with the promoters or enhancers of imprinted genes and influences their transcription (Lee and Bartolomei, 2013). These imprinted lncRNA transcripts often interfere with the transcription machinery of nearby imprinted genes, influencing the

recruitment of transcription factors at their promoters (Latos et al., 2012). Based on an analysis of lncRNA and DNA binding in imprinting clusters from multiple mammalian species, it was suggested that the binding of lncRNAs to promoters of imprinted genes may be common (Liu et al., 2017). The *Airn/Igf2r* imprinted cluster in the mouse genome is a well-studied example (**Figure 2C**; Latos et al., 2012). On the paternal allele, the transcription profile of *Airn* initiates from its promoter embedded within the ICR in a direction antisense to the transcription of the *Igf2r* gene (Latos and Barlow, 2009). It was noted that intragenic truncations of the endogenous lncRNA *Airn* in embryonic stem (ES) cells that do not include the overlapping region are unable to silence the *Igf2r* paternal allele, thus demonstrating that inhibition of RNA polymerase II recruitment to *Igf2r* promoter region does not depend on the overlap between *Airn* transcription and the promoter (Sleutels et al., 2002; Latos et al., 2012; Santoro and Pauler, 2013). Furthermore, during ES cells differentiation, *Airn* expression was also necessary and sufficient to silence *Igf2r* (Santoro et al., 2013). The overlapping regions between *Airn* transcription and *Igf2r* promoter and its gene body instead of *Airn* lncRNA products themselves lead to silencing of *Igf2r* expression.

The Disturbance of the Transcriptional Elongation

Another mechanism involves a collision between the converging elongation complexes of imprinted lncRNA and imprinted genes, leading to transcription stalling, premature termination, and subsequent degradation of the imprinted gene transcript (Hao et al., 2017). An example is the *UBE3A/UBE3A-ATS* imprinted domain on human chromosome 15q11-13, in which imprinted genes, including *MAGEL2, NDN, SNRPN, SNORD115,* and *SNORD116,* are silenced on the maternal allele (Horsthemke and Wagstaff, 2008). In contrast, *UBE3A*, which encodes an E3 ubiquitin ligase, is expressed from the maternal allele, especially in neurons in the brain. The homologous imprinted locus in mice has also been identified and studied, locating at a syntenic loci chromosome 7qC (Yang et al., 1998; **Figure 2D**). In this imprinted cluster, the ICR embedded within the *Snurf-Snrpn* gene is unmethylated on the paternal allele. In mouse neurons, *Ube3a-ATS* lncRNA is expressed specifically from its promoter embedded in the unmethylated ICR (Yin et al., 2012; Meng et al., 2013). Notably, the *Ube3a* promoter region is not methylated differently like *Ube3a-ATS*. This, combined with the observation that *Ube3a-ATS* transcription initiates from an exon region upstream of the *Snurf-Snrpn* gene and elongates approximately 1,000 kb as far as the intronic region of *Ube3a* between exons 4 and 5 (Landers et al., 2004; Lewis et al., 2019), it was hypothesized that the two opposing polymerases of *Ube3a* and *Ube3a-ATS* collide (**Figure 2D**). This transcriptional collision may lead to premature termination of *Ube3a* transcription inside its exon region on the paternal chromosome. In neurons from the monoallelic genetically engineered mouse model with the transcription of paternal *Ube3a-ATS* allele being terminated, *Ube3a* allele expression was activated on the paternal allele (Meng et al., 2013), resulting in increased expression comparable to maternal *Ube3a* (Meng et al., 2012). In cultured AS mouse neurons with biallelic silenced *Ube3a* expression, Antisense oligonucleotides (ASOs) targeting *Ube3a-ATS* rescued the expression of *Ube3a* efficiently (Meng et al., 2015). Consistently, in human induced pluripotent stem cells (iPSC)-derived neuron cells with biallelic silenced *UBE3A* expression, ASOs targeting *UBE3A-ATS* lncRNA transcripts lead to transcriptional termination by displacement of RNA Polymerase II, releasing the transcription of *UBE3A* (Germain et al., 2021). Recently, in human iPSCs, both sufficient expression of *UBE3A-ATS* lncRNA and two newly identified boundary elements were located inside the *IPW* gene and the *PWAR1* gene (Martins-Taylor et al., 2014; Hsiao et al., 2019). These two genes are located between *SNORD115* and *SNORD116.* In human iPSCs with the boundary elements deleted using gene editing technology, the expression of *UBE3A* was not silenced by up-regulated *UBE3A-ATS* expression (Hsiao et al., 2019). Mapping RNAPII density showed that reduced active RNAPII across the 3′ half of *UBE3A* corresponding to silenced UBE3A. These results together further support the hypothesized collision between *UBE3A-ATS* and *UBE3A* transcription complexes, leading to premature termination of the latter. In summary, the overlap between *Airn* and *Igf2r* promoter region disrupts the initiation of *Igf2r* transcription, while *Ube3a-ATS* silences the expression of *Ube3a* by disturbing its transcriptional elongation.

Chromatin Modification

Another lncRNA-related imprinting mechanism involves coating the bidirectionally flanking chromosomal region and recruiting repressive chromatin modification factors (Lee and Bartolomei, 2013; Sanli and Feil, 2015; Statello et al., 2021). The interactions between lncRNAs and these chromatin factors facilitate transcriptional silencing of target genes. The repressive chromatin-modification factors methylate DNA and produce histone modifications resulting in condensed chromatin structure and repressed gene expression. Among well-known repressive chromatin-modification factors, PRCs bind and spread across targeted chromatin facilitated by lncRNAs (Kotzin et al., 2016; Marín-Béjar et al., 2017). lncRNAs, genome structures, and CpG islands are essential factors in recruiting these PRCs, which have the capacity to catalyze lysine 119-mono-ubiquitinated histone H2A (H2AK119ub1) and H3K27me3 to repress gene expression through chromatin compaction and antagonization of transcriptional activators (Schwartz and Pirrotta, 2013; Simon and Kingston, 2013; Calabrese et al., 2015; Pintacuda et al., 2017; Colognori et al., 2019; Schertzer et al., 2019; Gil and Ulitsky, 2020; MacDonald and Mann, 2020). In genomic imprinting, some imprinted lncRNAs can bidirectionally direct repression of flanking neighbor imprinted gene region, such as *KCNQ1OT1* lncRNA. Some lncRNAs can target distal gene regions in the same imprinted clusters they locate, such as *Airn*.

Locally Recruiting Condensed Chromatin Structure to Neighbor Gene Region

The *Kcnq1/Kcnq1ot1* ICR, also known as KvDMR1 (*KvLQT1* differentially methylated region 1), with the embedded lncRNA *Kcnq1ot1* promoter, is unmethylated on the paternal allele (**Figure 2A**; Lee et al., 1999, p. 1; Smilinich et al., 1999; Beatty et al., 2006; Ager et al., 2008). lncRNA *Kcnq1ot1* transcripts from the promoter region recruit several epigenetic factors such as the Polycomb group proteins RING1B (Polycomb Repressive

Complex 1, PRC1), EZH2 (PRC2), and histone methyltransferase euchromatic histone lysine *N*-methyltransferase-2 (EHMT2 or G9a) to neighboring gene regions, forming repressive histone modifications such as H3K27me3 and H3K9me2 (**Figure 2A** II; Umlauf et al., 2004; Pandey et al., 2008). The chromatin state around the flanking regions of this lncRNA becomes condensed and results in silencing of flanking multi-protein coding genes such as *Cdkn1c*, *Slc22a18*, and *Tssc4*. On the maternal allele, DNA methylation of KvDMR1 silences the activation of the *Kcnq1ot1* promoter and represses the transcription, releasing the transcription of neighboring genes.

Recruiting Chromatin Modification Factors to Distal Imprinted Genes

A typical example of imprinted lncRNA regulating distal imprinted genes through epigenetic silencing is *Airn* and recruitment of PRCs in the placenta (**Figure 2C**; Latos et al., 2012; Lee and Bartolomei, 2013). As mentioned before, the transcription of *Airn* represses the expression of flanking imprinted gene *Igf2r* by transcriptional interference of the overlapping *Igf2r* promoter without repressive chromatin modification involved. In contrast, distal imprinted genes, such as *Slc22a2* (about 100 kb to *Airn* locus) and *Slc22a3* (about 300 kb to *Airn* locus), are also silenced by *Airn* in the extra-embryonic lineage, where *Airn* mediates the recruitment of PRC1 and PRC2 to distal targets on the paternal alleles (Terranova et al., 2008; Zhao et al., 2010; Schertzer et al., 2019). Recently, Airn was found to silence *Slc22a3* in mouse trophoblast stem cells (Andergassen et al., 2019). Allele-specific chromosome conformation capture studies have suggested that *Airn* transcription throughout the enhancer of *Slc22a3* may silence *Slc22a3* expression by disrupting its promoter-enhancer interactions. However, with monoallelic deletion of the entire *Airn* gene, no essential enhancers for the distal silenced genes were found in the *Airn* gene region. Nonetheless, it has also been shown that *Airn* lncRNA is enriched on the *Slc22a3* promoter together with an H3K9 dimethylase, G9a (Nagano et al., 2008). These results illustrate that *Airn* may target the promoters of distal imprinted genes by recruiting PRCs and G9a. The enrichment of these histone modification factors may lead to condensed chromatin in distal imprinted regions and silence imprinted genes.

THE ROLE OF IMPRINTED LONG NON-CODING RNAs IN HUMAN IMPRINTING DISORDERS AND CANCER

Long non-coding RNAs play essential roles in many biological processes and are related to various human diseases. Altered expression of imprinted loci has been linked to various neurodevelopmental disorders and cancers (Schaller et al., 2010; Huang et al., 2011; Riordan et al., 2013; Peters, 2014). Since imprinted regions are inherited in a parent-of-origin way, defects in one allele may be sufficient to lead to imprinting disorders (Kopp and Mendell, 2018). More specifically, silencing of parentally expressed imprinted genes can lead to the ultimate loss of its expression. Under abnormal conditions, DNA methylation status, allelic expression, and the biological functions of imprinted lncRNAs may be affected. These alterations may relate to human imprinting disorder-related disease phenotypes (Lee and Bartolomei, 2013). Here, we examine several well-studied imprinting disorders and emphasize the roles of imprinted lncRNAs in pathophysiological processes of imprinting-related diseases and cancers.

Common Molecular Mechanisms of Imprinting Disorders

Appropriate expression patterns of imprinted genes are important to growth and development. Correspondingly, imprinting disorder-related human diseases can be caused by genetic or epigenetic abnormalities on paternally or maternally inherited alleles (Lee and Bartolomei, 2013). Several common molecular mechanisms behind imprinting disorders have been defined, including molecular changes or genetic abnormalities, UPD, and epigenetic alterations (**Figures 3A–D**; Soellner et al., 2017; Carli et al., 2020). Firstly, genetic alterations, including SNPs and copy number variants on one imprinting allele, can affect imprinting (**Figure 3B**). Another mechanism is UPD, in which the inheritance of two copies of chromosomes or chromosomal regions are both from either the paternal or maternal allele, resulting in synchronous expression or silencing (**Figure 3C**; Robinson, 2000). Different from genetic alterations, epigenetic changes known as epimutations in DNA or histone modification without obvious genetic mutations have also been documented in imprinting disorders (**Figure 3D**; Horsthemke, 2010). Hypermethylation at imprinted DMRs can silence the active allele of the original monoallelic expressed imprinted genes. In contrast, hypomethylation can result in overexpression of the original silenced allele. Epimutations can arise randomly or be driven by their environment during the inheritance of germline epigenetic imprinting marks. DNA methylation in DMRs can thus be abnormally inherited in the absence of genetic sequence alterations (Robertson, 2005). Moreover, as with molecular or genetic alterations, epimutations can be permanently maintained in somatic tissues for life and cause developmental phenotypes (Ioannides et al., 2014; Gillessen-Kaesbach et al., 2018; Monk et al., 2019). Besides the imprinted disorder caused by variations in a single imprinted gene, imprinting disorders with epigenetic alterations at loci across the genome have also been observed in many imprinting diseases, referred to as MLID (Horsthemke, 2010; Fontana et al., 2018). Instead of changes at specific genetic loci, MLID may be caused by a globally disturbed imprinting inheritance process across the genome. However, since current research is mostly limited to a subset of imprinted genes and the mosaic character of MILD (Azzi et al., 2014; Eggermann et al., 2021), the role of MLID in imprinting disorders is still poorly understood.

Congenital Imprinting Disorders and Related Imprinted Long Non-coding RNAs

Molecular disturbances, like loss or gain of methylation at ICRs, and subsequent loss or gain of imprinted gene expression

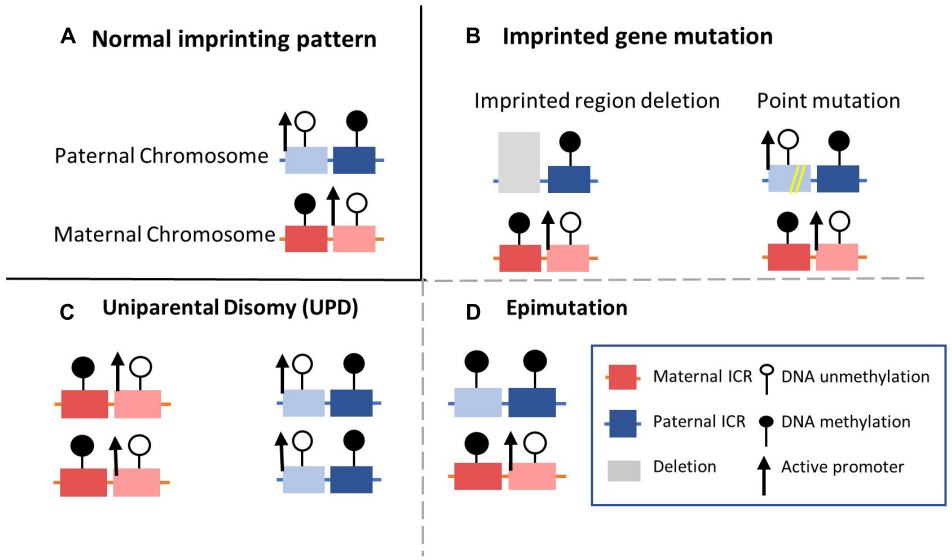


FIGURE 3 | Four common molecular mechanisms of imprinting disorders. **(A)** The normal state of the established DMR methylation pattern on the maternal and paternal alleles. **(B)** Imprinting disorder can be caused by copy number variations with imprinted cluster located such as duplication or deletion. Point mutations (single nucleotide polymorphisms) occurring in imprinted genes could also influence normal functions. **(C)** Both alleles are inherited from same parent-of-origin. In the case shown here, for example, maternally inherited alleles are duplicated without paternal allele participation. **(D)** Epimutations of DNA modification condition can disturb normal imprinting pattern without alterations in DNA sequences of the imprinted region. For example, DNA methylation of the imprinted gene on paternal allele are hypermethylated and silenced on both alleles.

have been described in various congenital human disorders. The frequencies of different molecular abnormalities vary among imprinting disorder-related diseases (Eggermann et al., 2015). More details about typical clinical syndromes and the pathological mechanisms are summarized in **Table 1**. However, details underlying imprinting disorder mechanisms and how they might impact adult neurobiology and developmental processes remain to be clarified. Fortunately, multiple mouse models of human imprinting disorders have been generated based on genomic conservation in most imprinted clusters. Strong correlates have been shown between these two genomes in imprinting loci, imprinting disorder phenotypes, and underlying molecular mechanisms (Perez et al., 2016). Spatial- and temporal-specific expression of allele-specific genes have been observed in several imprinting clusters in humans and mice. Here, imprinting disorders in three well-studied imprinting clusters are introduced as examples to demonstrate the roles of imprinted lncRNAs in imprinting-related congenital human disorders.

UBE3A-ATS in Prader–Willi Syndrome and Angelman Syndrome

Recent RNA-Seq data revealed strong allele-biased expression in the adult mouse brain, especially in imprinted regions (Perez et al., 2015, 2016), where many of these genes are expressed in cell type-specific manners. Importantly, mutations or disruptions in imprinted genes are linked with extensive neurobehavioral phenotypes, demonstrating that brain-specific imprinted genes may play important roles in neurodevelopmental disorders (Tucci et al., 2019). PWS and AS are two neurodevelopmental disorders caused by oppositely inherited deficiencies occurred in the same imprinted cluster (Peters, 2014; Kalsner and Chamberlain, 2015; Buiting et al., 2016). These two syndromes perform common phenotype characters, including hypotonia at the newborn stage, abnormal sleep patterns, and the deficiency in intellectual development (Buiting, 2010; Kalsner and Chamberlain, 2015). Children affected by PWS exhibit poor suck phenotypes with reduced muscle tone and mental abilities (Buiting et al., 1995; Buiting, 2010; Fontana et al., 2017), while AS is characterized by deficient motor function, intellectual development, and speech abilities (Buiting, 2010; Eggermann et al., 2015). These two disorders are caused by imprinting disorder in the imprinted PWS/AS locus (*UBE3A/UBE3A-ATS* imprinted cluster) on human chromosome 15q11-13 (**Figure 4A**). Similar to the mouse homologous locus mentioned previously in Section 3, the E3 ubiquitin ligase-encoding *UBE3A* gene is specifically imprinted in the brain (Vu and Hoffman, 1997). On the maternal allele, the methylated DMR encompasses the promoter of the *SNRPN* gene, silencing the *SNURF/SNRPN* gene and a series of downstream non-coding RNA genes (Rougeulle et al., 1997). In contrast, actively expressed *UBE3A-ATS* and the non-coding *SNORD* gene clusters are expressed from the paternal allele.

On the paternal allele of imprinted human PWS/AS locus, the unmethylated PWS-ICR is the region upstream to a protein-coding gene *SNRPN* and a lncRNA *SNHG14* (small nucleolar RNA host gene 14) (Sutcliffe et al., 1994; Buiting et al., 1995; Rougeulle et al., 1997; Runte et al., 2001; Vitali et al., 2010; Chamberlain, 2013; Stanurova et al., 2018; **Figure 4A**). The neuron-specific non-coding transcript *SNHG14* is processed to give rise to a series of non-coding RNA products, such as

TABLE 1 | Human imprinting disorder-related diseases.

Genomic location	Imprinted cluster/lncRNA	Imprinting disorder diseases	Clinical syndromes	Molecular mechanisms	Prevalence in population	References
Chromosome 15q11-13	*SNUPF-SNRPN/UBE3A* (imprinted lncRNA: *UBE3A-ATS*)	Prader–Willi syndrome (PWS) (OMIM #176270)	Obesity, reduced muscle tone, diminished swallowing and suckling, infantile hypotonia and hypogonadism, intellectual disability	Deletion the imprinted loci on the paternal allele (70–75%); Maternal UPD of chromosome 15 (20–25%); Epimutations of the DNA methylation at ICR 2%); Small deletions within the ICR (<0.5%)	1/25.000–1/10.000	Buiting et al. (1995), Buiting (2010), Fontana et al. (2017), Elbracht et al. (2020)
		Angelman syndrome (AS) (OMIM #105830)	Developmental delay, intellectual disability, absence of speech, microcephaly, seizures, specific excitable demeanor	Deletion of 15q.11–13 region on the maternal chromosome (70–75%); Point mutation in *UBE3A* gene (10%); Paternal UPD (3–7%); *SNURF* ICR loss of methylation (2–3%)	1/20.000–1/12.000	Buiting (2010), Eggermann et al. (2015), Elbracht et al. (2020)
Chromosome 11p-15.5	*H19/IGF2*; *KCNQ1OT1* (Imprinted lncRNA: *H19*) *H19/IGF2*	Beckwith–Wiedeman syndrome (BWS) (OMIM #130650)	Neonatal macrosomia, postnatal overgrowth, placental mesenchymal dysplasia, Tendency to embryonal tumors, cancer predisposition	Paternal UPD of chromosome 11p15.5 (20% to 25%); *KCNQ1OT1*-ICR loss of methylation (50%); H19/IGF2-ICR gain of methylation (5%); *CDKN1C* point mutations (5%); Cluster copy number variation (2–4%)	1/15.000	Eggermann et al. (2016), Mussa et al. (2016), Õunap (2016), Kalish et al. (2017)
		Silver–Russel syndrome (SRS) (OMIM #180860)	Severe intrauterine growth restriction (IUGR), postnatal growth failure with no catch-up, body hemihypoplasia, relative macrocephaly with triangular face, fifth finger clinodactyly and characteristic triangular face, lower birth weight	Loss of methylation at ICR on the paternal allele (40–60%); Maternal UPD of chromosome 7 (5–10%)	1/100.000–1/75.000	Cytrynbaum et al. (2016), Eggermann et al. (2016), Õunap (2016), Wakeling et al. (2017)
Chromosome 14q32.2	*MEG3/DLK1* (Imprinted lncRNA: *MEG3*)	Kagami–Ogata syndrome (KOS14) (OMIM #608149)	Polyhydramnios, placentomegaly, poor sucking and hypoventilation in the neonatal period, abdominal wall defects, a distinctive facial appearance, small bell-shaped thorax, coat-hanger ribs	Paternal UPD (65%); Microdeletion affecting the maternal 14q32.2 imprinted region (20%); Hypermethylation of the ICR (15%)	<1 in 1,000,000	Beygo et al. (2015), Kagami et al. (2015), Ogata and Kagami (2016), Prasasya et al. (2020)
		Temple syndrome (TS14) (OMIM #616222)	IUGR, PNGR (postnatal growth restriction), hypotonia and motor delay, feeding difficulties in infancy, truncal obesity, scoliosis, precocious puberty, small feet and hands	*MEG/DLK1* ICR loss of methylation (61%); Maternal UPD (29%); Deletion in imprinted region (10%)	<1 in 1,000,000	Ioannides et al. (2014), Gillessen-Kaesbach et al. (2018), Prasasya et al. (2020)

repeated C/D box small nucleolar RNAs (snoRNAs) and lncRNAs including *116HG*, *115HG*, and the antisense transcript to *UBE3A* (Mendiola and LaSalle, 2021; **Figure 4A**). The most studied RNA product from the host transcript *SNHG14* is *SNORD116* snoRNA, embedded within intronic regions of *SNORD116* gene locus (Cavaillé et al., 2000; de los Santos et al., 2000; Stanurova et al., 2018; Mendiola and LaSalle, 2021). *SNORD116* snoRNA present in ribonucleoprotein complexes (snoRNPs) and may participate in splicing, ribosomal RNA maturation, RNA modifications, and regulation of prohormone processing-related gene expression (Bazeley et al., 2008; Burnett et al., 2017). Meanwhile, *SNORD116* locus encoded *116HG* lncRNA was discovered recently (Vitali et al., 2010). *116HG* is stably retained in the nucleus 'RNA cloud' at its transcription site (Powell et al., 2013a). *116HG* potentially regulates transcript levels of circadian-related genes in the cortex and energy-related metabolism through in a time-of-day-dependent manner (Coulson et al., 2018b). Similarly, *SNORD115* locus encodes lncRNA *115HG* and *SNORD115* snoRNAs. While on the maternal allele, the methylated PWS-ICR occurs the upstream of the *SNRPN* gene. It silences the expression of the paternally expressed transcripts while allows the expression of *UBE3A* (Vu and Hoffman, 1997).

Prader–Willi syndrome is the first human disease identified to be caused by the abnormal expression of non-coding RNAs (Sahoo et al., 2008; de Smith et al., 2009; Duker et al., 2010). All cases of PWS in humans involve a deletion in the *SNORD116*

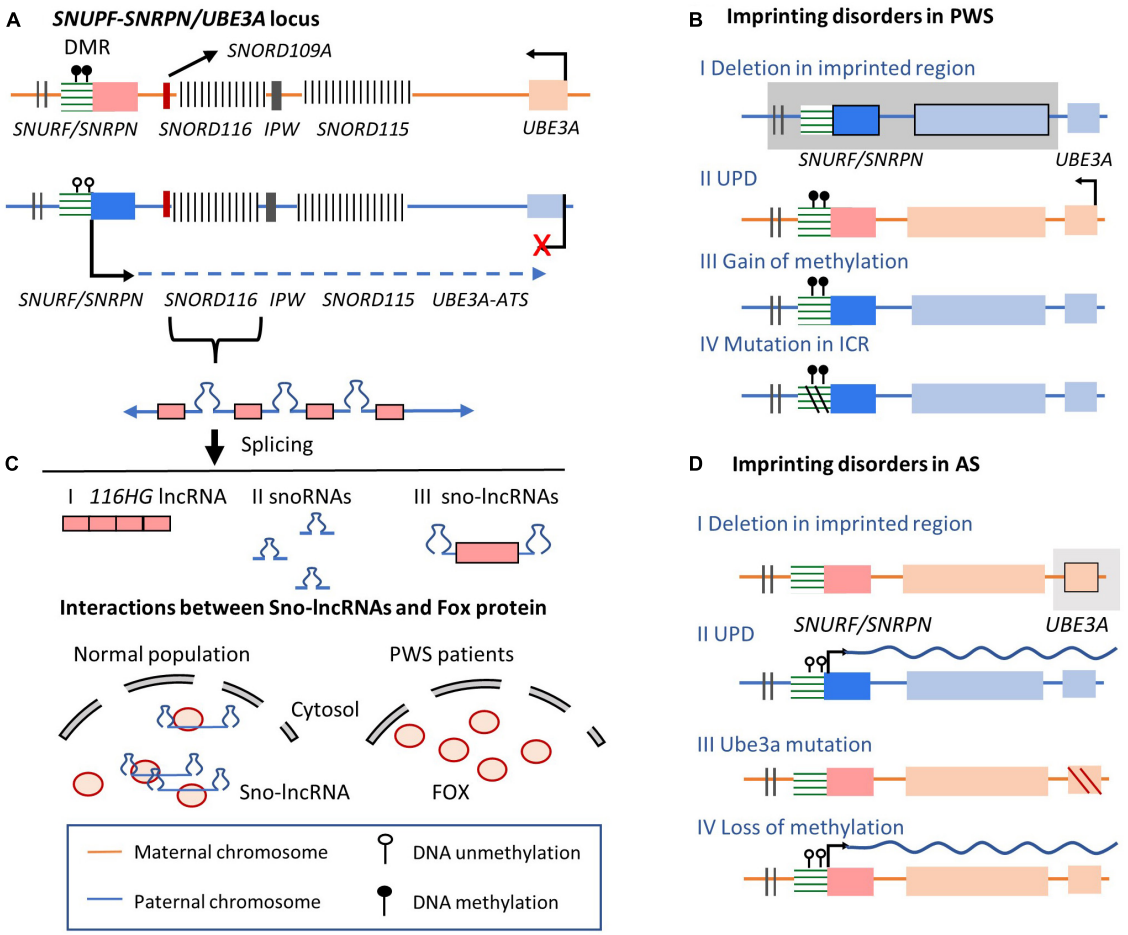


FIGURE 4 | *SNUPF-SNRPN/UBE3A* imprinted cluster on human chromosome 15 and related imprinting disorders. **(A)** Allelic expression pattern in *SNUPF-SNRPN/UBE3A* locus. On the maternal allele, the methylation of ICR silences the expression of *UBE3A-ATS*, permitting *UBE3A* expression. On the paternal allele, *UBE3A-ATS* is expressed from the *SNURF* gene region, overlapping the exon region of *UBE3A* gene. On the paternal allele, lncRNA host transcript are processed to give rise to snoRNAs (*SNORD115* and *SNORD116*), lncRNAs (*116HG*, *115HG*, and *UBE3A-ATS*). Three different spliced non-coding transcripts are produced from *SNORD116* gene locus, including *116HG* lncRNA, snoRNAs, and sno-lncRNAs. *SNORD116* sno-lncRNAs with snoRNAs on two ends are produced after splicing. **(B)** Imprinting disorders occur in PWS. PWS-related molecular alterations in *UBE3A* imprinted gene cluster. Line I indicates deletion in the imprinted region; Line II shows double maternal alleles are inherited, losing the paternal copy; Line III shows that the epigenetically mutated DNA methylation in DMR of the ICR leads to the silencing of lncRNA expression. Line IV: small deletion within the ICR. **(C)** Sno-lncRNAs transcribed from paternal allele can recruit Fox proteins and other related proteins, regulating Fox protein distribution and related alternative splicing functions. However, in PWS patients, loss of the *UBE3A-ATS* and other noncoding gene expression lead to the accumulation of Fox proteins in the nucleus and global abnormal splicing patterns. **(D)** Imprinting disorders occur in AS. Line I: deletions of the maternal imprinted regions containing the *UBE3A* and surrounding genes; Line II: both alleles are inherited from paternal chromosome; Line III: *UBE3A* mutations lead to transcript loss of function; Line IV: epimutations in the maternal allele lead to lncRNA expression from the maternal allele, preventing normal *UBE3A* expression.

non-coding gene locus, which regulates the maturation of the central nervous system. The overlap between the phenotype caused by *SNORD116* microdeletion and *MAGEL2* mutation suggests that transcripts from *SNORD116* locus may modify *MAGEL2* expression via long-range chromatin interactions (Meziane et al., 2015; Fountain and Schaaf, 2016; Langouët et al., 2018). The loss of the paternal expressed *SNORD116* in PWS can be caused by several factors, including large paternal deletions in the imprinted PWS/AS locus (60%), maternal UPD (36%), small microdeletion in *SNORD116* locus (<1%), and epigenetic alternations in DNA methylation of the PWS-ICR region (4%) (Sahoo et al., 2008; Duker et al., 2010; Bieth et al., 2015; Rozhdestvensky et al., 2016; Mendiola and LaSalle, 2021; **Figure 4B**). Rare microdeletions that encompass *SNORD116* and its adjacent genes, *SNRPN* or *SNORD115*, have been found in PWS patients (Sahoo et al., 2008; de Smith et al., 2009; Duker et al., 2010). Moreover, a small deletion that only covers *SNORD116* and its adjacent genes (*SNORD109A*, and *IPW*) was identified in a patient with typical PWS syndrome (Bieth et al., 2015; **Figure 4A**). Since there is no obvious involvement of *SNORD109A* and *IPW* genes in PWS, the observations in this PWS case further support that the *SNORD116* gene region play key roles in the PWS, independent with *SNORD115* or *SNRPN* deletion. Consistently,

SNORD116 is completely silenced in neuron cells derived from PWS patients (Cavaillé et al., 2000; Hsiao et al., 2019). Besides, *Snord116* deleted mouse model recapitulates major phenotypes of human PWS patients, including altered metabolism, growth deficiency, memory impairment, hyperphagia and increased anxiety (Skryabin et al., 2007; Ding et al., 2008; Zieba et al., 2015; Qi et al., 2016; Polex-Wolf et al., 2018; Adhikari et al., 2019).

Furthermore, an alternative RNA species (sno-lnRNAs) processed from *SNORD116* host non-coding transcript has been described in human (Yin et al., 2012; Powell et al., 2013a; **Figure 4A** III). The role of *SNORD116* sno-lncRNAs in RNA processing and decay of their target mRNAs is not well-understood but may facilitates our understanding of the connection between imprinting disorder and pathological mechanism of PWS (**Figure 4C**). *SNORD116* exon transcript is retained between two snoRNAs, forming sno-lncRNAs with two small nucleolar ribonucleoprotein ends (Yin et al., 2012). These sno-lncRNAs accumulate near the synthesis site together with a type of lncRNAs that are 5′ capped by snoRNAs and 3′ polyadenylated (SPAs) (Wu et al., 2016). These lncRNAs may interact with RNA binding proteins including TDP43 (TAR DNA-binding protein 43), RBFOX2 (RNA Binding Fox-1 Homolog 2), and hnRNP M (Heterogeneous nuclear ribonucleoprotein M). Especially, splicing regulator RBFOX2 are required for the neuron-specific splicing of *Snord116* transcript to produce *116HG* lncRNA and *Snord116* snoRNA (Yeo et al., 2009; Coulson et al., 2018a). Since immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) and RT-PCR assays confirmed that RBFOX2 directly binds to *Snord116* snoRNA, it is hypothesized that *Snord116* snoRNA may reduce the availability of these splicing-related proteins and regulate alternative splicing in the nucleus (Yin et al., 2012; Wu et al., 2016). Therefore, the disruption of *SNORD116* in PWS may lead to more uniform distribution of RBFOX2 protein and global changes in normal alternative splicing patterns, contributing to PWS phenotypes.

In contrast to the paternal-allelic imprinting disorder in PWS, AS, is mainly caused by the lack of maternal *UBE3A* gene expression (**Figure 4D**; Buiting et al., 2016). The brain-specific and maternally biased expression of *UBE3A* has been shown to function in regulating dendritic growth and influencing behavior and neurotransmitters (Avagliano Trezza et al., 2019). In AS patients, the expression of *UBE3A* or functional UBE3A protein is lost. These alternations can be caused by various imprinting disorder mechanisms including deletions of the maternally imprinted regions containing the *UBE3A* and surrounding genes. Besides pathological variants in the *UBE3A* gene, loss of *SNURF* DMR methylation has also been observed in AS cases (2–3%), in which the expression of *UBE3A* is silenced by *UBE3A-ATS* as discussed previously (Dagli et al., 1993).

KCNQ1OT1 and *H19/IGF* in Beckwith–Wiedemann Syndrome and Silver–Russell Syndrome

Beckwith–Wiedemann syndrome and SRS are clinically opposite growth-affecting disorders (Õunap, 2016). The underlying pathological mechanisms involve genetic and epigenetic perturbations of two imprinting clusters on human chromosome 11p15, the *KCNQ1/KCNQ1OT1* and *H19/IGF* loci (Carli et al., 2020; Chang and Bartolomei, 2020; **Figures 5A–C**). BWS is one of the most common congenital overgrowth conditions (Mussa et al., 2013), with common phenotypes including postnatal overgrowth, placenta mesenchymal dysplasia, and congenital and childhood cancer predisposition. In contrast, SRS patients exhibit postnatal growth failure with body hemihypoplasia, lower birth weight, fetal undergrowth and poor feeding predisposition (Wakeling et al., 2017).

Approximately 50% of BWS patients lose DNA methylation accompanied by loss of H3K9me2 on maternal KvDMR1 (**Figure 5A**; Robertson, 2005). This epigenetic disturbance results in biallelic expression of the *KCNQ1OT1* lncRNA. As a consequence, expression of this lncRNA silences adjacent imprinted genes on both alleles (Soejima and Higashimoto, 2013). Among these silenced genes, *CDKN1C* is linked to the development of BWS phenotypes (Yan et al., 1997; Zhang et al., 1997; Tunster et al., 2011). These epigenetic mutations in maternal KvDMR1 and biallelic expressed *KCNQ1OT1* lncRNAs lead to loss of CDKN1C expression and fetal overgrowth, thus contributing to BWS syndrome (Eggermann et al., 2016; Wakeling et al., 2017). Therefore, after the establishment of DMRs on imprinted alleles, monoallelic expression of *KCNQ1OT1* lncRNA is a crucial regulator of adjacent protein-coding genes, which have essential roles in maintaining normal growth processes during early development. Another major abnormal imprinted cluster identified in BWS patients is *H19/IGF2* (**Figure 5B**). Under normal conditions, *H19/IGF2* ICR is methylated on the paternal chromosome, controlling the expression of *H19*. In BWS patients, mutations or hypermethylation of the *H19/IGF2* ICR can lead to *H19* silencing and subsequent overexpression of *IGF2,* a circulating hormone and tissue growth factor. The upregulated expression of *IGF2* is linked to BWS overgrowth-related phenotypes (Pollak et al., 2004; Brioude et al., 2018a,b; Duffy et al., 2019). As for SRS, loss of *H19/IGF2* ICR methylation on the paternal chromosome 11p15 accounts for 40–60% of patients (Wakeling et al., 2017). ICR hypomethylation is bound by the insulator CTCF. The interaction of the *IGF2* promoter with its enhancer on both alleles is disrupted, resulting in decreased *IGF2* expression and subsequent growth and development delays (**Figure 5C**; Abi Habib et al., 2017).

Although some BWS and SRS patients can be identified based on clinical features alone, diagnosing imprinting disorders can be complicated by complex molecular alternations (Ibrahim et al., 2014; Wakeling et al., 2017). In addition to the two imprinted loci primarily relevant to BWS and SR phenotypes, MLID has also been observed in an increasingly growing fraction of patients with methylation abnormalities at other imprinted loci (Rossignol et al., 2006; Azzi et al., 2009; Eggermann et al., 2011; Fontana et al., 2018). In addition, symptoms vary widely in patients with imprinting disorders (Wakeling et al., 2017; Brioude et al., 2018b; Mantovani et al., 2018). Therefore, additional insights into the relationship between the epigenetic mechanisms of imprinting disorders and neurological diseases can help clarify more accurate diagnostic guidelines and appropriate clinical therapies.

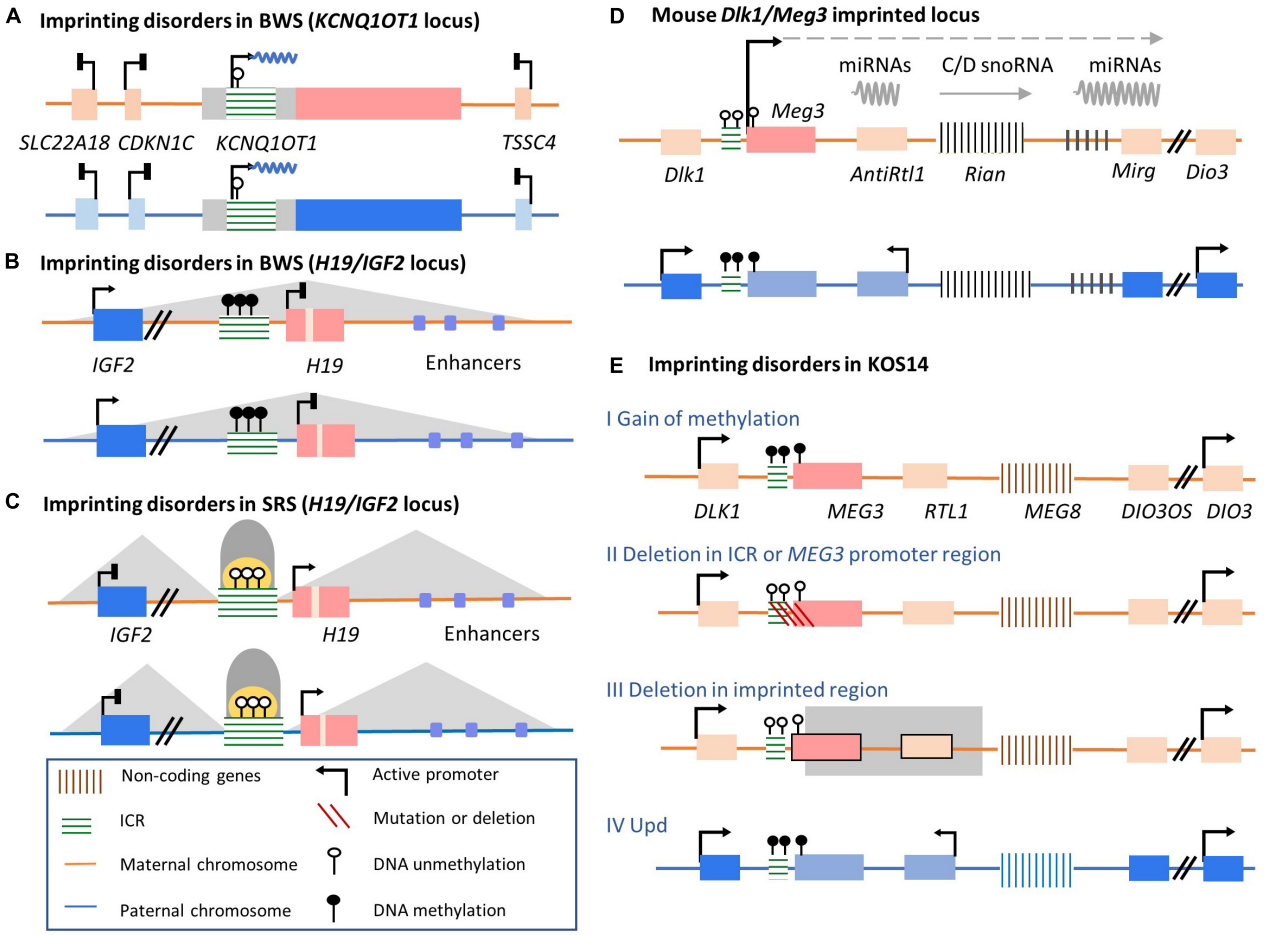


FIGURE 5 | Well-studied imprinting clusters and their conditions in disorder conditions involved in human imprinting disturbance-related diseases. **(A,**B) BWS and two related imprinted clusters, *H19/IGF2* and *KCNQ1OT1*. **(A)** The first situation is the hypomethylation of the maternal allele in *Kcnq1ot1* ICR leads to lncRNA *KCNQ1OT1* overexpression. The expression of neighboring imprinted genes, such as *SLC22A18*, *CDKN1C*, and *TSSC4*, is bi-allelically silenced. **(B)** The second major imprinting disorder responsible for BWS is hypermethylation of the maternal ICR, resulting in loss of *H19* expression and *IGF2* overexpression. **(C)** SRS and alterations in imprinting of *H19/IGF2* locus. Hypomethylation of the paternal *H19/IGF2* ICR resulting in *H19* overexpression and inhibited *Igf2* expression. **(D)** The regulation of mouse *Meg3* imprinted cluster. On the paternal allele, the gDMR of *Meg3* cluster ICR is methylated, repressing *Meg3* lncRNA expression. On the maternal allele, lncRNAs are transcribed from the promoter within the unmethylated ICR. **(E)** Four cases of *MEG3*-related imprinting disorders in KOS14 patients are shown. Line I, epimutations in normally activated maternal ICR of *MEG3* cluster result in loss of lncRNA transcription, releasing normally silenced adjacent imprinted genes; Line II: maternal deletion in ICR of *MEG3* regions; Line III: maternal deletion in the *MEG3* gene body; Line IV: both alleles are inherited by silenced paternal allele.

DLK1/DIO3 in Kagami–Ogata Syndrome and Temple Syndrome

Genetic and epigenetic alterations in delta-like homolog 1 gene/type III iodothyronine deiodinase gene (*DLK1/DIO3*) imprinted cluster on human chromosome 14q32 are associated with two human imprinting disorder-related diseases, KOS14 and TS14 (Temple et al., 1991; Wang et al., 1991; Ogata and Kagami, 2016). Common KOS14 phenotypes include neonatal respiratory difficulties, a distinctive facial appearance, variable developmental delay, and/or intellectual disability (Ogata and Kagami, 2016; Prasasya et al., 2020). Clinical syndromes observed in TS14 include severe intrauterine growth restriction, postnatal growth restriction, neonatal hypotonia, and feeding difficulties in infancy (Ioannides et al., 2014; Gillessen-Kaesbach et al., 2018; Prasasya et al., 2020).

The distribution of imprinted genes and regulatory mechanisms of *DLK1/DIO3* locus are highly conserved between humans and mice. The regulation of this imprinted locus has been revealed in mouse models established with genetic alterations in the *Dlk1/Dio3* locus on chromosome 12 (**Figure 5D**; Paulsen et al., 2001; da Rocha et al., 2008). Three paternally expressed imprinted protein-coding genes are *Dlk1*, *Rtl1*, and *Dio3*. lncRNA *Meg3* (also called *Gtl2*), the *Rtl1*-antisense *Rtl1as*, the C/D-box snoRNA cluster *Rian*, and the microRNA cluster *Mirg* are transcribed from the maternal allele (da Rocha et al., 2008; Kota et al., 2014). The regulation of imprinted gene expression in this locus relies on an intergenic

DMR (IG-DMR). On the maternal allele, AFF3 protein binds to an upstream enhancer of *Meg3*, activating lncRNA expression. In contrast, on the paternal allele, AFF3 binds instead to the methylated IG-DMR, leading to silencing of *Meg3* and other non-coding genes (Luo et al., 2016; Wang et al., 2017). It has also been suggested recently that maternally expressed lncRNA *Meg3* is involved in the regulation of the *Dlk1/Dio3* imprinted cluster (Sanli et al., 2018). The maternal expression of the *Meg3* lncRNA may play a role in preventing maternal *Dlk1* activation through interaction with the lysine methyltransferase (KMT) Ezh2 and PRC2 in the maternal *Dlk1* gene region (Kaneko et al., 2014; Sanli et al., 2018). Remarkably, *Meg3* lncRNA's regulation of imprinted protein-coding gene *Dlk1* is restricted to a developmental window as follows. In embryonic stem cells, the *Dlk1* gene is expressed biallelically at a low level. Upon neuronal differentiation, *Dlk1* expression is upregulated on the paternal allele. Conversely, the activation of the *Dlk1* gene on the maternal allele is prevented by the overlap of *Meg3* lncRNA *in cis* and the recruitment of Ezh2 to the *Dlk1* gene region (Sanli et al., 2018). Although the *Meg3* lncRNA is necessary for the silencing of *Dlk1* expression, the mechanisms underlying the connection between the *Meg3* lncRNA and repressed *Dlk1* expression on the maternal allele are unknown.

The *DLK1/DIO3* locus is predominantly imprinted in the human brain (Davis et al., 2005; Ferrón et al., 2011). Protein-coding genes *DLK1*, *RTL1,* and *DIO3* are expressed on the paternal allele; lncRNAs (*MEG3*, *MEG8*, *RTL1as*, *DIO3OS*), snoRNAs, and miRNAs are transcribed on the maternal allele. Importantly, the *DLK1* gene plays essential functions in regulating development and metabolism. In KOS14 patients, gain of DNA methylation on the maternal ICR leads to *MEG3* silencing (**Figure 5E** I; Sato et al., 2011). However, maternal micro-deletions of the *MEG3* promoter that don't affect ICR methylation are also observed in some cases (**Figure 5E** II; Kota et al., 2014). In another case, a maternal micro-deletion has been detected in the *MEG3* gene body instead of the IG-DMR or *MEG3* promoter (**Figure 5E** III; van der Werf et al., 2016).

In summary, in these conditions, imprinted lncRNAs play essential roles as upstream regulators of protein-coding genes in the same imprinted cluster. However, the detailed mechanisms are diverse and complicated in different imprinting disorders and remain to be further investigated.

Imprinted Long Non-coding RNAs and Human Cancers

Long non-coding RNAs play important roles in pathways implicated in many cancer types, including prostate (Hua et al., 2018, p. 19), breast (Zhang et al., 2017; Cho et al., 2018), and hepatocellular carcinoma (He et al., 2017; Lecerf et al., 2019; Ye et al., 2020). Long non-coding RNAs can serve as cancer enhancers or repressors in temporal- and spatial-specific manners (Calin et al., 2007; Kanduri, 2016; Quinn and Chang, 2016; Peng et al., 2017). Abnormal functions of lncRNAs have been observed in various tumors and cancer cell lines (Kitagawa et al., 2012; Bhan et al., 2017). Notably, abnormally regulated imprinted gene expression, altered ICR methylation conditions, and altered expression of cancer-related imprinted lncRNAs were observed in cancers such as breast cancer (Kim et al., 2015; Goovaerts et al., 2018). In addition, in imprinting disorders, abnormal silencing of imprinted lncRNAs contributes to congenital and childhood tumors. For instance, susceptibility to Wilm's tumor and adrenocortical carcinoma is increased in *H19*-silenced patients (Dao et al., 1999; DeBaun et al., 2000; Weksberg et al., 2010; Brioude et al., 2018a,b).

H19 is one of the most commonly implicated tumorigenesis-promoting lncRNAs (Zheng et al., 2020). The expression of *H19* occurs during embryonic development and decreases after birth in most tissues. However, *H19* is abnormally upregulated in various cancers, including breast, liver, lung, esophageal, pancreatic, ovarian, and bladder (Vennin et al., 2015; Zhang et al., 2016). *H19*'s tumor-promoting effects include the inhibition of cell death, promotion of proliferation, downregulation of growth suppressors, and promotion of invasion and metastasis (reviewed in Matouk et al., 2015; Lecerf et al., 2019). Moreover, high *H19* expression may be a molecular marker to predict cancers and prognoses after clinical treatment, including the rate of post-therapeutic relapse in hematological cancer patients (Liu et al., 2016). Increased risk of developing congenital and childhood tumors seen in BWS is also associated with aberrant *H19*. *H19* is also associated with growth suppression (Yoshimizu et al., 2008; Lecerf et al., 2019; Zhou et al., 2019). *H19*'s contribution to tumorigenesis varies by tissue and developmental windows and requires clarification in future investigations.

Another well-studied cancer-related imprinted lncRNA is *MEG3*, which acts as a cancer repressor. *MEG3* is downregulated in breast, neuroblastoma, meningioma, glioma, pituitary adenoma, and hematological malignancies (Benetatos et al., 2011; Cheunsuchon et al., 2011; Zhou et al., 2012; Lyu et al., 2017; Zhu et al., 2019). In pituitary neuroendocrine tumors, hypermethylation of the maternal *DLK1/MEG3* locus results in *MEG3* downregulation and impaired differentiation (Cheunsuchon et al., 2011; Chen et al., 2020). Hypermethylation of the *MEG3* promoter region has also been observed in AML patients (Lyu et al., 2017; Yao et al., 2017; Sellers et al., 2019), while recent studies have begun to reveal the underlying mechanisms in endometrial and breast cancers (Sun et al., 2017; Zhang et al., 2017; Zhu et al., 2019). One such mechanism involves *MEG3*'s inhibition of the phosphoinositide 3-kinase/protein kinase B (*PI3K/Akt*) signaling pathway, a well-known growth-related pathway. Therefore, unraveling the roles of imprinted lncRNAs in cancer may reveal novel biomarkers and therapeutic targets for cancer treatment.

MODULATION OF THE LONG NON-CODING RNA *UBE3A-ATS* TO RESCUE ABNORMAL IMPRINTING IN PRADER–WILLI SYNDROME/ANGELMAN SYNDROME IMPRINTED CLUSTER

Although our understanding of the mechanisms of imprinting disorders has grown, efficient molecular diagnosis and effective treatments are limited to nonexistent (Elbracht et al., 2020).

Modulation of imprinted lncRNAs has been proposed as a potential therapeutic strategy to target imprinted genes and rescue imprinting disorders (Peters, 2014; Statello et al., 2021). As the epigenetic regulatory mechanisms of the *Ube3a/Ube3a-ATS* imprinted cluster are understood best, attempts have been made to rescue *Ube3a* expression through modulating the collision between the transcriptional machinery of *Ube3a* and *Ube3a-ATS* in an allele-specific manner. Herein, three state-of-the art therapeutic strategies by targeting *Ube3a-ATS* lncRNA, editing *Ube3a-ATS* gene region, or modulating chromatin transcriptional state by small molecules are discussed along with recent preclinical studies of *UBE3A/UBE3A-ATS* imprinted cluster-related diseases.

Antisense Oligonucleotides for Imprinted Long Non-coding RNAs

Antisense oligonucleotides are single-stranded DNA oligos designed using sequence homology with their RNA targets that hybridize with the targeted RNA region based on complementary base pairs, and induce subsequent RNA degradation at the ASO-RNA heteroduplex part (Mishra et al., 2019; Li M. et al., 2020). ASOs can be used to alter splicing or gene expression. ASOs have been designed as potential therapies for various diseases, including AS, spinal muscular atrophy (SMA), Duchenne muscular dystrophy, Huntington disease, and hyperlipidemia (Beaudet and Meng, 2016; Dhuri et al., 2020). Several ASO-based therapies, such as Nusinersen (Spinraza) for SMA treatment, have received approval by the United States Food and Drug Administration (FDA) and other regional regulatory agencies (Karaki et al., 2019). Nusinersen is quite effective in rescuing protein deficiency by altering pre-mRNA splicing (Hoy, 2017; Groen et al., 2018; Claborn et al., 2019). The capacities of ASOs to access targeted RNAs through homology base pairing and in inducing RNase H-mediated cleavage at the pairing regions by exonucleases make them suitable to decrease lncRNA levels post-transcriptionally (Chan et al., 2006).

As mentioned before, on the paternal allele of *Ube3a/Ube3a-ATS* imprinted cluster, *Ube3a-ATS* represses *Ube3a* expression by prematurely terminating the elongation of *Ube3a* transcripts. Therefore, a potential strategy is to rescue the defective *Ube3a* transcription by targeting *Ube3a-ATS* transcripts using ASOs (**Figure 6A** I). To avoid influencing the transcription of sno-lncRNAs essential for neuronal development and PWS, ASOs were designed to be complementary to *Ube3a-ATS* transcripts downstream of the *Snord115* cluster. These ASOs were provided to cultured AS mouse neurons with deficient *Ube3a* expression (Meng et al., 2015). The treatment achieved sustained ectopic paternal expression of *Ube3a*, partially rescued *UBE3A* brain protein levels, and alleviated some cognitive deficits. Remarkably, other splicing products derived from *Ube3a-ATS* like *Snrpn* and *Snord116* were unaffected. Consistently, ASOs were designed to rescue the expression of *UBE3A* in AS iPSC-derived neuron cells with a large deletion of maternal 15q11-q13. ASOs targeting *UBE3A-ATS* transcripts at *SNORD115* and *SNORD109B*, or targeting the snoRNA located between *SNORD 115* locus and *UBE3A* gene region, cleave *UBE3A-ATS* and release the transcription of *UBE3A* on the paternal chromosome (Germain et al., 2021). *UBE3A-ATS* transcription is terminated by displacing RNA Polymerase II several kilobases downstream of the ASO targeting site. Therefore, targeting the lncRNA *UBE3A-ATS* by ASOs could be a potential strategy for rescuing *UBE3A* expression and related imprinting disorders. Besides, ASOs have several unique features in treating imprinted disorders, including high *in vivo* efficacy, broad tissue distribution, low adverse events, and long duration of action (Smith et al., 2006; Kordasiewicz et al., 2012). Considering that several mRNA-targeting ASOs have been approved (Dhuri et al., 2020), targeting lncRNAs using ASOs to treat imprinting diseases could achieve wide application. However, robust delivery systems devoid of associated toxicity should be carefully developed and evaluated.

Modulation of Imprinted Long Non-coding RNA Expression Using the CRISPR/Cas9 System

The CRISPR/Cas9 system permits *in vitro* and *in vivo* gene editing tool and is another novel strategy to modulate imprinted lncRNA expression (Deltcheva et al., 2011; Jinek et al., 2012; Konermann et al., 2015; Suzuki et al., 2016; Knott and Doudna, 2018). A series of CRISPR/Cas-engineered systems can be designed to manipulate lncRNAs, including deletion of the lncRNA encoding gene region (pre-transcription level), inhibition or activation of the expression of the lncRNA (transcription level), or direct degradation of the lncRNA transcripts (post-transcriptional level) (Perez-Pinera et al., 2013; Qi et al., 2013; Ran et al., 2013; Abudayyeh et al., 2017). For example, CRISPRi and CRISPRa can modulate lncRNA expression by recruiting transcriptional repressors or activators without inducing genetic mutations (Bester et al., 2018; Kampmann, 2018). At the same time, CRISPR/Cas9 is being studied as a strategy of *in vivo* genome editing therapy in neurological diseases like schizophrenia and Alzheimer's disease (Zhuo et al., 2017; Kuruvilla et al., 2018; Park et al., 2019; Sun et al., 2019). It is hoped that effective manipulation of the non-coding regions achieved in human cell lines and animal models could result in novel strategies to eliminate obstacles in developing therapies for lncRNA-related imprinting diseases (Cho et al., 2013; Cong et al., 2013; Jiang et al., 2013; Doudna and Charpentier, 2014, p. 9). Furthermore, when taking into account brain-specific expression of imprinted clusters, CRISPR/Cas9 could be designed to correct abnormal imprinting patterns (Han et al., 2014). Indeed, recently Cas9 gene therapy has shown promise in trapping *Ube3a-ATS* to activate paternal Ube3a expression (**Figure 6A** II; Wolter et al., 2020). In addition, a CRISPR/Cas9 system targeting the *Snord115* locus in cultured mouse cortical neurons and human neural progenitor-derived neurons was able to successfully increase total *Ube3a* protein expression while decreasing *Snord115* expression. Using a neuron-specific saCas9 and guide RNAs packaged in an adeno-associated virus delivering system and administered to an AS mouse brain during the embryonic and early postnatal stages led to silencing of paternal *Snord115* expression with long-lasting effects.

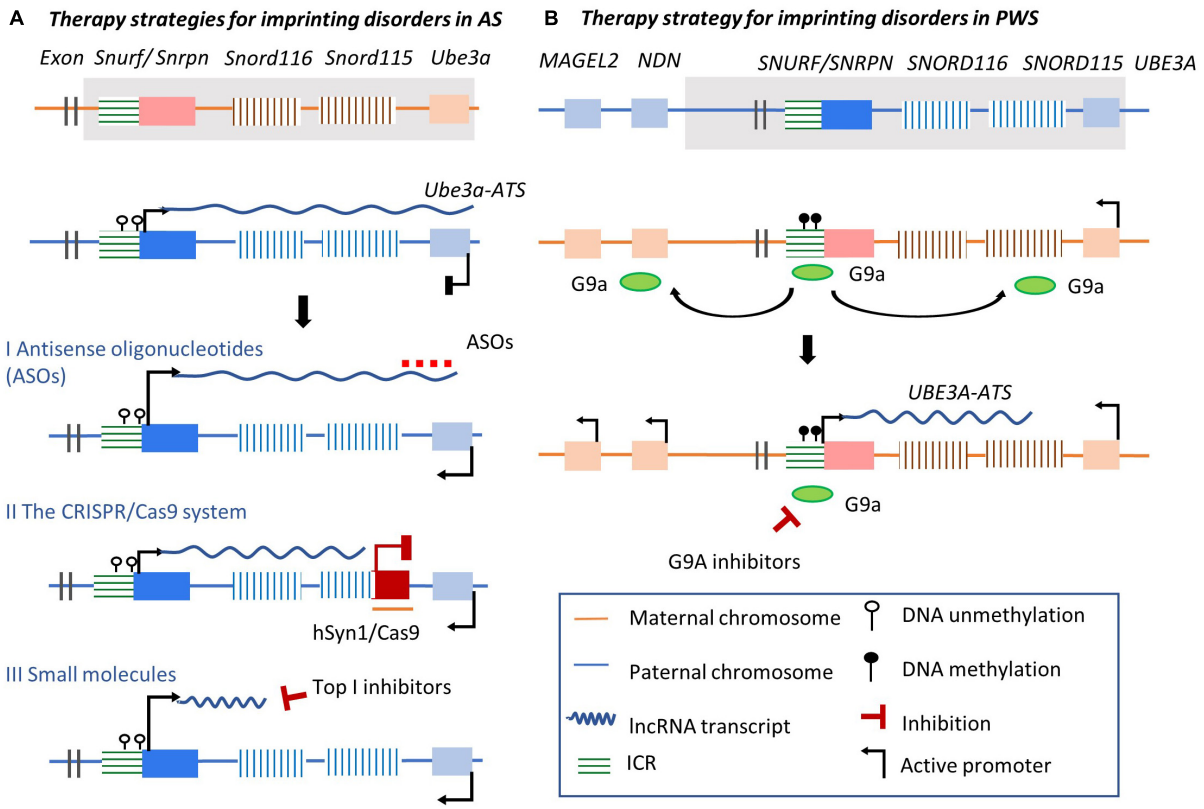


FIGURE 6 | Three state of the art strategies for imprinted disorders via targeting imprinted lncRNAs. **(A)** Therapeutic strategies for AS. Molecular alterations such as deletions in *Ube3a/Ube3a-ATS* imprinted cluster can cause the loss of effective *Ube3a* expression. Line I: ASOs are designed to target the overlapping regions of *Ube3a-ATS* transcripts and *Ube3a*, releasing paternal *Ube3a* expression; Line II: the human synapsin 1 (hSYN1) gene promoter is used drive neuron-specific expression and Cas9 packaged with adeno-associated virus delivering system is inserted into the gene region of the *Snord115*, leading to disrupted transcription of *Ube3a-ATS* before extending to the *Ube3a* gene encoding region. Line III: Top I inhibitors disrupt the elongation of the *Ube3a-ATS* at the *Snord116* region. *Ube3a* paternal expression is released from transcriptional collision. **(B)** Therapeutic strategy for PWS. G9A inhibitors prevent G9A recruitment to flanking regions near the ICR, releasing *Ube3a-ATS* lncRNA expression from its promoter in the ICR.

In summary, the CRISPR-Cas9 system offers promising therapeutic strategies with the potential to permanently alter imprinted gene expression with high specificity and low toxicity. Nevertheless, since lncRNAs lack open reading frames and functional protein products, the use of CRISPR-Cas9 system to achieve efficient lncRNA manipulation needs to be further improved (Statello et al., 2021). In addition, an optimal sgRNA design and an effective delivery mechanism to penetrate the blood-brain barrier need further investigation (Zhuo et al., 2017; Hana et al., 2021).

Small Molecules Targeting Histone Modifiers

Small molecules have been screened to target histone modification proteins involved in imprinted lncRNA regulation. As mentioned before, PWS and AS are two imprinting disorders related to the same imprinted cluster. In AS patients, *UBE3A* expression is decreased. Through high-content screening in mouse-derived primary cortical neurons, about 10 topoisomerase I (Top I) inhibitors have been identified with the capacity to downregulate *Ube3a-ATS* expression and induce reactivation of *UBE3A* expression from the paternal allele (Huang et al., 2011; Powell et al., 2013b). The Top I inhibitor topotecan blocks the elongation of the *Ube3a-ATS* transcription complex in cultured mouse neurons (Powell et al., 2013b). It inhibits sno-lncRNA transcription throughout the *Ube3a* encoding gene region by stabilizing the formation of R loops between RNA and DNA within paternal *Snord116*, leading to chromatin decondensation (Liu and Wang, 1987; Belotserkovskii et al., 2010; El Hage et al., 2010; Belotserkovskii and Hanawalt, 2011; French et al., 2011; Skourti-Stathaki et al., 2011; Aguilera and García-Muse, 2012; Ginno et al., 2012; **Figure 6A** III). The *Ube3a-ATS* transcription complex stalled before transcription of Sno-lncRNAs completed. Subsequent *Ube3a* expression was reactivated on the paternal allele. Additional candidates of other Top I inhibitors have also been assessed to identify inhibitors with better pharmacological profiles of *Ube3a* activation (Lee et al., 2018). Prospective therapeutic safety and central nervous system (CNS) bioavailability studies have also been performed recently in AS mouse neurons (Lee et al., 2018).

A therapeutic strategy for PWS based on the induction of *SNORD116* expression has been proposed. *SNORD116* is

normally silenced on the maternal allele, but its expression can be induced by modulating 'closed' chromatin condition into an 'open' state (Kim et al., 2017). The methylation of histone H3K9 performs allele-specific pattern in the ICR located upstream of *SNRPN* gene (PWS-ICR). On the maternal chromosome, histone methyltransferase euchromatic histone lysine N-methyltransferase-2 (G9a) locates at the methylated PWS-ICR and recruits repressive histone modifications (H3K9me2) along the PWS-ICR in a bidirectional manner. This leads to condensed chromatin structure and silencing of PWS-associated genes (**Figure 6B**). The inactivation of histone H3K9 methyltransferase G9a in mouse embryonic stem (ES) cells leads to reduced DNA methylation in PWS-ICR, and the expression of *Snrpn* was activated on both chromosome *in vitro* (Xin et al., 2003). However, in *in vivo* mouse model, two inhibitors of G9a selected lead to the activation of maternal copy of *Snord116* and improve survival of the PWS mouse without effect on the methylation state of the PWS-ICR or *Ube3a* expression on the maternal allele (Kim et al., 2017). Thus, further studies are needed to clarify the association between DNA methylation of PWS-ICR and allele-specific distribution of G9a. Meanwhile, the reactivation of *SNRPN* and *SNORD116* was recently achieved by preventing the recruitment of H3K9me3 repressive histone modification-related protein factor to *SNORD116* locus in PWS-derived iPSCs (Langouët et al., 2020). In summary, small molecules related epigenetic therapy for PWS through modulating the condition of specific chromatin regions could be a potential strategy to be translated in clinical relevance (Crunkhorn, 2017; Chung et al., 2020).

CONCLUDING REMARKS

Several imprinting gene clusters have been discovered and studied since the middle of the last century. These studies have shown that lncRNAs play crucial roles in regulating imprinted gene clusters and individual imprinted genes related to human health and diseases. However, from a genomic perspective, the characteristics of gene regulation among imprinting loci remain to be fully elucidated. This is despite the advancement in knowledge of the epigenetic regulatory mechanisms of a subset of genes in imprinted regions. In the three imprinted clusters (*Airn/Igfr2*, *Kcnq1/Kcnq1ot1*, and *Ube3a/Ube3a-ATS*), imprinted lncRNAs which play essential regulatory roles in silencing other imprinted genes are all expressed on the paternal allele. It has been reported that maternal expressed imprinted genes are prominent with protein-coding genes, while paternal expressed genes exhibit consistent distribution between non-coding and protein-coding sequences (Hutter et al., 2010). However, the difference between the establishment of maternal and paternal imprinted genes in lncRNA mechanisms remains unclear. Thus, comprehensive investigations are needed to understand further the mechanisms of imprinted lncRNAs in the epigenetic regulation of imprinted clusters. With technological advancements, studies on lncRNA-associated human imprinting disorders will lead to needed therapies.

Pharmacological treatments for congenital imprinting disorders are limited to symptomatic therapies, which are inefficient in promoting the patients' quality of life (Chung et al., 2020). Fortunately, the biological role of lncRNAs in the etiology of congenital imprinting disorders has been revealed thanks to the advancement in high-throughput genome-wide sequencing technologies. Therapeutic approaches based on disease-related lncRNAs have been investigated. In a recent study, lncRNA mimics were designed to restore the tissue-specific lncRNA *HULC* in mice, essential for phenylalanine metabolism (Li et al., 2021). In addition, three strategies mentioned above targeting Ube3a-ATS have efficiently rescued imprinting disorders in PWS/AS imprinted cluster in mouse models and human cell lines. Although the three strategies mentioned here targeting Ube3a-ATS have efficiently rescued imprinting disorders of PWS/AS imprinted cluster in mouse models and human cell lines, therapies for other disease-related clusters have not been investigated. Long non-coding RNA-based and lncRNA-targeting therapies have some unique advantages. For instance, in lncRNA-targeting methods like ASOs, synthesized RNA can be designed with organ-targeting peptides to achieve tissue-specific targeting of endogenous lncRNAs. Besides, synthesized RNA products could be modified to promote *in vivo* stability. Further translation of these strategies to real clinical tools will require further investigation to overcome related challenges. *In vivo* delivery of synthesized RNA molecules, cellular permeability, immunogenicity, and potential of organ toxicity also deserve further investigation (Perry and Ulitsky, 2021). Another challenge to extend the lessons learned in PWS and AS into other imprinting disorders is the epigenetic and molecular complexities in different imprinting disorders-related imprinted loci. Considering the complexity of the regulatory network of genomic imprinting, further efforts are needed to reveal underlying pathological mechanisms linked to imprinting disorder phenotypes and support continuous improvement of clinical management and therapeutic strategies.

AUTHOR CONTRIBUTIONS

TW and QM wrote and edited the manuscript and drew the pictures. JL, LY, and MW assisted in manuscript collation and review. QM provided critical inputs as the corresponding authors and obtained funds. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No: 32070870), Guangdong Province Natural Science Foundation of China (No: 2021A1515010758), Guangdong Provincial Key Laboratory of Synthetic Genomics (No: 2019B030301006), Shenzhen Key Laboratory of Synthetic Genomics (No: ZDSYS201802061806209) and Shenzhen Institute of Synthetic Biology Scientific Research Program (No: ZTXM20200008).

REFERENCES

Abi Habib, W., Brioude, F., Azzi, S., Salem, J., Das Neves, C., Personnier, C., et al. (2017). 11p15 ICR1 partial deletions associated with IGF2/H19 DMR hypomethylation and silver-russell syndrome. *Hum. Mutat.* 38, 105–111. doi: 10.1002/humu.23131

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR-Cas13. *Nature* 550, 280–284. doi: 10.1038/nature24049

Adhikari, A., Copping, N. A., Onaga, B., Pride, M. C., Coulson, R. L., Yang, M., et al. (2019). Cognitive deficits in the Snord116 deletion mouse model for Prader-Willi syndrome. *Neurobiol. Learn. Mem.* 165:106874. doi: 10.1016/j.nlm.2018.05.011

Ager, E. I., Pask, A. J., Gehring, H. M., Shaw, G., and Renfree, M. B. (2008). Evolution of the CDKN1C-KCNQ1 imprinted domain. *BMC Evol. Biol.* 8:163. doi: 10.1186/1471-2148-8-163

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol. Cell.* 46, 115–124. doi: 10.1016/j.molcel.2012.04.009

Andergassen, D., Dotter, C. P., Wenzel, D., Sigl, V., Bammer, P. C., Muckenhuber, M., et al. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *eLife* 6:e25125. doi: 10.7554/eLife.25125

Andergassen, D., Muckenhuber, M., Bammer, P. C., Kulinski, T. M., Theussl, H. C., Shimizu, T., et al. (2019). The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. *PLoS Genetics* 16:e1009151. doi: 10.1371/journal.pgen.1008268

Avagliano Trezza, R., Sonzogni, M., Bossuyt, S. N. V., Zampeta, F. I., Punt, A. M., van den Berg, M., et al. (2019). Loss of nuclear UBE3A causes electrophysiological and behavioral deficits in mice and is associated with Angelman syndrome. *Nat. Neurosci.* 22, 1235–1247. doi: 10.1038/s41593-019-0425-420

Azzi, S., Blaise, A., Steunou, V., Harbison, M. D., Salem, J., Brioude, F., et al. (2014). Complex tissue-specific epigenotypes in russell-silver syndrome associated with 11p15 ICR1 hypomethylation. *Hum. Mutat.* 35, 1211–1220. doi: 10.1002/humu.22623

Azzi, S., Rossignol, S., Steunou, V., Sas, T., Thibaud, N., Danton, F., et al. (2009). Multilocus methylation analysis in a large cohort of 11p15-related foetal growth disorders (Russell Silver and Beckwith Wiedemann syndromes) reveals simultaneous loss of methylation at paternal and maternal imprinted loci. *Hum. Mol. Genet.* 18, 4724–4733. doi: 10.1093/hmg/ddp435

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M. A., van der Kooy, D., et al. (2008). Global survey of genomic imprinting by transcriptome sequencing. *Curr. Biol.* 18, 1735–1741. doi: 10.1016/j.cub.2008.09.044

Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E. K., Rivas, M. A., et al. (2015). The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.* 25, 927–936. doi: 10.1101/gr.192278.115

Barlow, D. P. (2011). Genomic imprinting: a mammalian epigenetic discovery model. *Annu. Rev. Genet.* 45, 379–403. doi: 10.1146/annurev-genet-110410-132459

Barlow, D. P., and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6:a018382. doi: 10.1101/cshperspect.a018382

Bartolomei, M. S., Zemel, S., and Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153–155. doi: 10.1038/351153a0

Bazeley, P. S., Shepelev, V., Talebizadeh, Z., Butler, M. G., Fedorova, L., Filatov, V., et al. (2008). snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. *Gene* 408, 172–179. doi: 10.1016/j.gene.2007.10.037

Beatty, L., Weksberg, R., and Sadowski, P. D. (2006). Detailed analysis of the methylation patterns of the KvDMR1 imprinting control region of human chromosome 11. *Genomics* 87, 46–56. doi: 10.1016/j.ygeno.2005.05.015

Beaudet, A. L., and Meng, L. (2016). Gene-targeting pharmaceuticals for single-gene disorders. *Hum. Mol. Genet.* 25, R18–R26. doi: 10.1093/hmg/ddv476

Belotserkovskii, B. P., and Hanawalt, P. C. (2011). Anchoring nascent RNA to the DNA template could interfere with transcription. *Biophys. J.* 100, 675–684. doi: 10.1016/j.bpj.2010.12.3709

Belotserkovskii, B. P., Liu, R., Tornaletti, S., Krasilnikova, M. M., Mirkin, S. M., and Hanawalt, P. C. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. U S A.* 107, 12816–12821. doi: 10.1073/pnas.1007580107

Benetatos, L., Vartholomatos, G., and Hatzimichael, E. (2011). MEG3 imprinted gene contribution in tumorigenesis. *Int. J. Cancer* 129, 773–779. doi: 10.1002/ijc.26052

Bergman, Y., and Cedar, H. (2004). A stepwise epigenetic process controls immunoglobulin allelic exclusion. *Nat. Rev. Immunol.* 4, 753–761. doi: 10.1038/nri1458

Bester, A. C., Lee, J. D., Chavez, A., Lee, Y.-R., Nachmani, D., Vora, S., et al. (2018). An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance. *Cell* 173, 649–664.e20. doi: 10.1016/j.cell.2018.03.052

Beygo, J., Elbracht, M., de Groot, K., Begemann, M., Kanber, D., Platzer, K., et al. (2015). Novel deletions affecting the MEG3-DMR provide further evidence for a hierarchical regulation of imprinting in 14q32. *Eur. J. Hum. Genet.* 23, 180–188. doi: 10.1038/ejhg.2014.72

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 77, 3965–3981. doi: 10.1158/0008-5472.CAN-16-2634

Bian, S., and Sun, T. (2011). Functions of noncoding RNAs in neural development and neurological diseases. *Mol. Neurobiol.* 44, 359–373. doi: 10.1007/s12035-011-8211-8213

Bieth, E., Eddiry, S., Gaston, V., Lorenzini, F., Buffet, A., Conte Auriol, F., et al. (2015). Highly restricted deletion of the SNORD116 region is implicated in Prader-Willi syndrome. *Eur. J. Hum. Genet.* 23, 252–255. doi: 10.1038/ejhg.2014.103

Brioude, F., Hennekam, R., Bliek, J., Coze, C., Eggermann, T., Ferrero, G. B., et al. (2018a). Revisiting wilms tumour surveillance in beckwith-wiedemann syndrome with IC2 methylation loss, reply. *Eur. J. Hum. Genet.* 26, 471–472. doi: 10.1038/s41431-017-0074-72

Brioude, F., Kalish, J. M., Mussa, A., Foster, A. C., Bliek, J., Ferrero, G. B., et al. (2018b). Expert consensus document: clinical and molecular diagnosis, screening and management of Beckwith-Wiedemann syndrome: an international consensus statement. *Nat. Rev. Endocrinol.* 14, 229–249. doi: 10.1038/nrendo.2017.166

Buiting, K. (2010). Prader-Willi syndrome and Angelman syndrome. *Am. J. Med. Genet. C Semin. Med. Genet.* 154C, 365–376. doi: 10.1002/ajmg.c.30273

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D., et al. (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat. Genet.* 9, 395–400. doi: 10.1038/ng0495-395

Buiting, K., Williams, C., and Horsthemke, B. (2016). Angelman syndrome - insights into a rare neurogenetic disorder. *Nat. Rev. Neurol.* 12, 584–593. doi: 10.1038/nrneurol.2016.133

Burnett, L. C., Hubner, G., LeDuc, C. A., Morabito, M. V., Carli, J. F. M., and Leibel, R. L. (2017). Loss of the imprinted, non-coding Snord116 gene cluster in the interval deleted in the Prader Willi syndrome results in murine neuronal and endocrine pancreatic developmental phenotypes. *Hum. Mol. Genet.* 26, 4606–4616. doi: 10.1093/hmg/ddx342

Calabrese, J. M., Starmer, J., Schertzer, M. D., Yee, D., and Magnuson, T. (2015). A survey of imprinted gene expression in mouse trophoblast stem cells. *G3 (Bethesda)* 5, 751–759. doi: 10.1534/g3.114.016238

Calin, G. A., Liu, C., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., et al. (2007). Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12, 215–229. doi: 10.1016/j.ccr.2007.07.027

Carli, D., Riberi, E., Ferrero, G. B., and Mussa, A. (2020). Syndromic disorders caused by disturbed human imprinting. *J. Clin. Res. Pediatr. Endocrinol.* 12, 1–16. doi: 10.4274/jcrpe.galenos.2019.2018.0249

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., et al. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. U S A.* 97, 14311–14316. doi: 10.1073/pnas.250426397

Chamberlain, S. J. (2013). RNAs of the human chromosome 15q11-q13 imprinted region. *Wiley Interdiscip. Rev. RNA* 4, 155–166. doi: 10.1002/wrna.1150

Chan, J. H. P., Lim, S., and Wong, W. S. F. (2006). Antisense oligonucleotides: from design to therapeutic application. *Clin. Exp. Pharmacol. Physiol.* 33, 533–540. doi: 10.1111/j.1440-1681.2006.04403.x

Chang, S., and Bartolomei, M. S. (2020). Modeling human epigenetic disorders in mice: beckwith-Wiedemann syndrome and Silver-Russell syndrome. *Dis. Model Mech.* 13:dmm044123. doi: 10.1242/dmm.044123

Chen, W. V., and Maniatis, T. (2013). Clustered protocadherins. *Development* 140, 3297–3302. doi: 10.1242/dev.090621

Chen, Y., Gao, H., Liu, Q., Xie, W., Li, B., Cheng, S., et al. (2020). Functional characterization of DLK1/MEG3 locus on chromosome 14q32.2 reveals the differentiation of pituitary neuroendocrine tumors. *Aging (Albany NY)* 13, 1422–1439. doi: 10.18632/aging.202376

Chen, Z., and Zhang, Y. (2020). Maternal H3K27me3-dependent autosomal and X chromosome imprinting. *Nat. Rev. Genet.* 21, 555–571. doi: 10.1038/s41576-020-0245-249

Cheunsuchon, P., Zhou, Y., Zhang, X., Lee, H., Chen, W., Nakayama, Y., et al. (2011). Silencing of the imprinted DLK1-MEG3 locus in human clinically nonfunctioning pituitary adenomas. *Am. J. Pathol.* 179, 2120–2130. doi: 10.1016/j.ajpath.2011.07.002

Cho, S. W., Kim, S., Kim, J. M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232. doi: 10.1038/nbt.2507

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., et al. (2018). Promoter of lncRNA Gene PVT1 is a tumor-suppressor DNA boundary element. *Cell* 173, 1398–1412.e22. doi: 10.1016/j.cell.2018.03.068

Chotalia, M., Smallwood, S. A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., et al. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev.* 23, 105–117. doi: 10.1101/gad.495809

Chung, M. S., Langouët, M., Chamberlain, S. J., and Carmichael, G. G. (2020). Prader-Willi syndrome: reflections on seminal studies and future therapies. *Open Biol.* 10:200195. doi: 10.1098/rsob.200195

Claborn, M. K., Stevens, D. L., Walker, C. K., and Gildon, B. L. (2019). Nusinersen: a treatment for spinal muscular atrophy. *Ann. Pharmacother.* 53, 61–69. doi: 10.1177/1060028018789956

Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., et al. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22, 885–898. doi: 10.1101/gr.131037.111

Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C.-Y., and Lee, J. T. (2019). Xist deletional analysis reveals an interdependency between xist RNA and polycomb complexes for spreading along the inactive X. *Mol. Cell* 74, 101–117.e10. doi: 10.1016/j.molcel.2019.01.015

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Coulson, R. L., Powell, W. T., Yasui, D. H., Dileep, G., Resnick, J., and LaSalle, J. M. (2018a). Prader-Willi locus Snord116 RNA processing requires an active endogenous allele and neuron-specific splicing by Rbfox3/NeuN. *Hum. Mol. Genet.* 27, 4051–4060. doi: 10.1093/hmg/ddy296

Coulson, R. L., Yasui, D. H., Dunaway, K. W., Laufer, B. I., Vogel Ciernia, A., Zhu, Y., et al. (2018b). Snord116-dependent diurnal rhythm of DNA methylation in mouse cortex. *Nat. Commun.* 9:1616. doi: 10.1038/s41467-018-03676-3670

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., et al. (2014). Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res.* 24, 554–569. doi: 10.1101/gr.164913.113

Crunkhorn, S. (2017). Genetic disorders: steps towards epigenetic therapy for PWS. *Nat. Rev. Drug Discov.* 16:85. doi: 10.1038/nrd.2017.3

Cytrynbaum, C., Chong, K., Hannig, V., Choufani, S., Shuman, C., Steele, L., et al. (2016). Genomic imbalance in the centromeric 11p15 imprinting center in three families: further evidence of a role for IC2 as a cause of russell-silver syndrome. *Am. J. Med. Genet. A* 170, 2731–2739. doi: 10.1002/ajmg.a.37819

da Rocha, S. T., Edwards, C. A., Ito, M., Ogata, T., and Ferguson-Smith, A. C. (2008). Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends Genet.* 24, 306–316. doi: 10.1016/j.tig.2008.03.011

da Rocha, S. T., and Gendrel, A.-V. (2019). The influence of DNA methylation on monoallelic expression. *Essays Biochem.* 63, 663–676. doi: 10.1042/EBC20190034

Dagli, A. I., Mathews, J., and Williams, C. A. (1993). "Angelman Syndrome," in *GeneReviews*, eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. Bean, G. Mirzaa, et al. (Seattle, WA: University of Washington, Seattle).

Dao, D., Walsh, C. P., Yuan, L., Gorelov, D., Feng, L., Hensle, T., et al. (1999). Multipoint analysis of human chromosome 11p15/mouse distal chromosome 7: inclusion of H19/IGF2 in the minimal WT2 region, gene specificity of H19 silencing in Wilms' tumorigenesis and methylation hyper-dependence of H19 imprinting. *Hum. Mol. Genet.* 8, 1337–1352. doi: 10.1093/hmg/8.7.1337

Davis, E., Caiment, F., Tordoir, X., Cavaillé, J., Ferguson-Smith, A., Cockett, N., et al. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr. Biol.* 15, 743–749. doi: 10.1016/j.cub.2005.02.060

de los Santos, T., Schweizer, J., Rees, C. A., and Francke, U. (2000). Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader-Willi deletion region, which Is highly expressed in brain. *Am. J. Hum. Genet.* 67, 1067–1082. doi: 10.1086/303106

de Smith, A. J., Purmann, C., Walters, R. G., Ellis, R. J., Holder, S. E., Van Haelst, M. M., et al. (2009). A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum. Mol. Genet.* 18, 3257–3265. doi: 10.1093/hmg/ddp263

DeBaun, M. R., King, A. A., and White, N. (2000). Hypoglycemia in Beckwith-Wiedemann syndrome. *Semin. Perinatol.* 24, 164–171. doi: 10.1053/sp.2000.6366

DeChiara, T. M., Robertson, E. J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859. doi: 10.1016/0092-8674(91)90513-x

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., et al. (2011). CRISPR RNA maturation by trans -encoded small RNA and host factor RNase III. *Nature* 471, 602–607. doi: 10.1038/nature09886

Deng, X., Berletch, J. B., Nguyen, D. K., and Disteche, C. M. (2014). X chromosome regulation: diverse patterns in development, tissues and disease. *Nat. Rev. Genet.* 15, 367–378. doi: 10.1038/nrg3687

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111

Dhuri, K., Bechtold, C., Quijano, E., Pham, H., Gupta, A., Vikram, A., et al. (2020). Antisense oligonucleotides: an emerging area in drug discovery and development. *J. Clin. Med.* 9:2004. doi: 10.3390/jcm9062004

Ding, F., Li, H. H., Zhang, S., Solomon, N. M., Camper, S. A., Cohen, P., et al. (2008). SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3:e1709. doi: 10.1371/journal.pone.0001709

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096

Duffy, K. A., Cielo, C. M., Cohen, J. L., Gonzalez-Gandolfi, C. X., Griff, J. R., Hathaway, E. R., et al. (2019). Characterization of the beckwith-wiedemann spectrum: diagnosis and management. *Am. J. Med. Genet. C Semin. Med. Genet.* 181, 693–708. doi: 10.1002/ajmg.c.31740

Duker, A. L., Ballif, B. C., Bawle, E. V., Person, R. E., Mahadevan, S., Alliman, S., et al. (2010). Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur. J. Hum. Genet.* 18, 1196–1201. doi: 10.1038/ejhg.2010.102

Eggermann, T., Brioude, F., Russo, S., Lombardi, M. P., Bliek, J., Maher, E. R., et al. (2016). Prenatal molecular testing for Beckwith-Wiedemann and Silver-Russell syndromes: a challenge for molecular analysis and genetic counseling. *Eur. J. Hum. Genet.* 24, 784–793. doi: 10.1038/ejhg.2015.224

Eggermann, T., Kadgien, G., Begemann, M., and Elbracht, M. (2021). Biallelic PADI6 variants cause multilocus imprinting disturbances and miscarriages in the same family. *Eur. J. Hum. Genet.* 29, 575–580. doi: 10.1038/s41431-020-00762-760

Eggermann, T., Perez, de Nanclares, G., Maher, E. R., Temple, I. K., Tümer, Z., et al. (2015). Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clin. Epigenetics* 7:123. doi: 10.1186/s13148-015-0143-148

Eggermann, T., Buiting, K., and Temple, I. K. (2011). Clinical utility gene card for: silver-russell syndrome. *Eur. J. Hum. Genet.* 19. doi: 10.1038/ejhg.2010.202

El Hage, A., French, S. L., Beyer, A. L., and Tollervey, D. (2010). Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* 24, 1546–1558. doi: 10.1101/gad.573310

Elbracht, M., Mackay, D., Begemann, M., Kagan, K. O., and Eggermann, T. (2020). Disturbed genomic imprinting and its relevance for human reproduction: causes and clinical consequences. *Hum. Reprod. Update* 26, 197–213. doi: 10.1093/humupd/dmz045

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 46, D308–D314. doi: 10.1093/nar/gkx1107

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* 12, 565–575. doi: 10.1038/nrg3032

Ferguson-Smith, A. C., Cattanach, B. M., Barton, S. C., Beechey, C. V., and Surani, M. A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature* 351, 667–670. doi: 10.1038/351667a0

Ferrón, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., et al. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* 475, 381–385. doi: 10.1038/nature10229

Fontana, L., Bedeschi, M. F., Maitz, S., Cereda, A., Faré, C., Motta, S., et al. (2018). Characterization of multi-locus imprinting disturbances and underlying genetic defects in patients with chromosome 11p15.5 related imprinting disorders. *Epigenetics* 13, 897–909. doi: 10.1080/15592294.2018.1514230

Fontana, P., Grasso, M., Acquaviva, F., Gennaro, E., Galli, M. L., Falco, M., et al. (2017). SNORD116 deletions cause Prader-Willi syndrome with a mild phenotype and macrocephaly. *Clin. Genet.* 92, 440–443. doi: 10.1111/cge.13005

Fountain, M. D., and Schaaf, C. P. (2016). Prader-Willi syndrome and schaaf-yang syndrome: neurodevelopmental diseases intersecting at the MAGEL2 gene. *Diseases* 4:E2. doi: 10.3390/diseases4010002

French, S. L., Sikes, M. L., Hontz, R. D., Osheim, Y. N., Lambert, T. E., El Hage, A., et al. (2011). Distinguishing the roles of topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. *Mol. Cell. Biol.* 31, 482–494. doi: 10.1128/MCB.00589-510

Germain, N. D., Gorka, D., Drennan, R., Jafar-nejad, P., Whipple, A., Core, L., et al. (2021). Antisense oligonucleotides targeting UBE3A-ATS restore expression of UBE3A by relieving transcriptional interference. *bioRxiv [preprint]* doi: 10.1101/2021.07.09.451826

Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* 21, 102–117. doi: 10.1038/s41576-019-0184-185

Gillessen-Kaesbach, G., Albrecht, B., Eggermann, T., Elbracht, M., Mitter, D., Morlot, S., et al. (2018). Molecular and clinical studies in 8 patients with temple syndrome. *Clin. Genet.* 93, 1179–1188. doi: 10.1111/cge.13244

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell.* 45, 814–825. doi: 10.1016/j.molcel.2012.01.017

Gómez-Marín, C., Tena, J. J., Acemel, R. D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. U S A.* 112, 7542–7547. doi: 10.1073/pnas.1505463112

Goovaerts, T., Steyaert, S., Vandenbussche, C. A., Galle, J., Thas, O., Van Criekinge, W., et al. (2018). A comprehensive overview of genomic imprinting in breast and its deregulation in cancer. *Nat. Commun.* 9:4120. doi: 10.1038/s41467-018-06566-6567

Groen, E. J. N., Talbot, K., and Gillingwater, T. H. (2018). Advances in therapy for spinal muscular atrophy: promises and challenges. *Nat. Rev. Neurol.* 14, 214–224. doi: 10.1038/nrneurol.2018.4

Guenzl, P. M., and Barlow, D. P. (2012). Macro lncRNAs: a new layer of cis-regulatory information in the mammalian genome. *RNA Biol.* 9, 731–741. doi: 10.4161/rna.19985

Han, J., Zhang, J., Chen, L., Shen, B., Zhou, J., Hu, B., et al. (2014). Efficient *in vivo* deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol.* 11, 829–835. doi: 10.4161/rna.29624

Hana, S., Peterson, M., McLaughlin, H., Marshall, E., Fabian, A. J., McKissick, O., et al. (2021). Highly efficient neuronal gene knockout *in vivo* by CRISPR-Cas9 via neonatal intracerebroventricular injection of AAV in mice. *Gene Ther.* doi: 10.1038/s41434-021-00224-222 [Epub ahead of print].

Hanna, C. W. (2020). Placental imprinting: emerging mechanisms and functions. *PLoS Genet* 16:e1008709. doi: 10.1371/journal.pgen.1008709

Hao, N., Palmer, A. C., Dodd, I. B., and Shearwin, K. E. (2017). Directing traffic on DNA-How transcription factors relieve or induce transcriptional interference. *Transcription* 8, 120–125. doi: 10.1080/21541264.2017.1285851

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111

He, J.-H., Han, Z.-P., Liu, J.-M., Zhou, J.-B., Zou, M.-X., Lv, Y.-B., et al. (2017). Overexpression of long non-coding RNA MEG3 inhibits proliferation of hepatocellular carcinoma Huh7 cells via negative modulation of miRNA-664. *J. Cell. Biochem.* 118, 3713–3721. doi: 10.1002/jcb.26018

Henckel, A., Chebli, K., Kota, S. K., Arnaud, P., and Feil, R. (2012). Transcription and histone methylation changes correlate with imprint acquisition in male germ cells. *EMBO J.* 31, 606–615. doi: 10.1038/emboj.2011.425

Horsthemke, B. (2010). Mechanisms of imprint dysregulation. *Am. J. Med. Genet. C Semin. Med. Genet.* 154C, 321–328. doi: 10.1002/ajmg.c.30269

Horsthemke, B., and Wagstaff, J. (2008). Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am. J. Med. Genet. A* 146A, 2041–2052. doi: 10.1002/ajmg.a.32364

Hoy, S. M. (2017). Nusinersen: first global approval. *Drugs* 77, 473–479. doi: 10.1007/s40265-017-0711-717

Hsiao, J. S., Germain, N. D., Wilderman, A., Stoddard, C., Wojenski, L. A., Villafano, G. J., et al. (2019). A bipartite boundary element restricts UBE3A imprinting to mature neurons. *Proc. Natl. Acad. Sci. U S A.* 116, 2181–2186. doi: 10.1073/pnas.1815279116

Hua, J. T., Ahmed, M., Guo, H., Zhang, Y., Chen, S., Soares, F., et al. (2018). Risk SNP-Mediated promoter-enhancer switching drives prostate cancer through lncRNA PCAT19. *Cell* 174, 564–575.e18. doi: 10.1016/j.cell.2018.06.014

Huang, H.-S., Allen, J. A., Mabb, A. M., King, I. F., Miriyala, J., Taylor-Blake, B., et al. (2011). Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* 481, 185–189. doi: 10.1038/nature10726

Huang, W. C., Bennett, K., and Gregg, C. (2018). Epigenetic and cellular diversity in the brain through allele-specific effects. *Trends Neurosci.* 41, 925–937. doi: 10.1016/j.tins.2018.07.005

Hutter, B., Bieg, M., Helms, V., and Paulsen, M. (2010). Imprinted genes show unique patterns of sequence conservation. *BMC Genomics* 11:649. doi: 10.1186/1471-2164-11-649

Ibrahim, A., Kirby, G., Hardy, C., Dias, R. P., Tee, L., Lim, D., et al. (2014). Methylation analysis and diagnostics of Beckwith-Wiedemann syndrome in 1,000 subjects. *Clin. Epigenetics* 6:11. doi: 10.1186/1868-7083-6-11

Im, M., Jp, R., and Hg, S. (2005). A census of mammalian imprinting. *Trends Genet.* 21, 457–465. doi: 10.1016/j.tig.2005.06.008

Ioannides, Y., Lokulo-Sodipe, K., Mackay, D. J. G., Davies, J. H., and Temple, I. K. (2014). Temple syndrome: improving the recognition of an underdiagnosed chromosome 14 imprinting disorder: an analysis of 51 published cases. *J. Med. Genet.* 51, 495–501. doi: 10.1136/jmedgenet-2014-102396

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233–239. doi: 10.1038/nbt.2508

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

Kaffer, C. R., Srivastava, M., Park, K. Y., Ives, E., Hsieh, S., Batlle, J., et al. (2000). A transcriptional insulator at the imprinted H19/Igf2 locus. *Genes Dev.* 14, 1908–1919.

Kagami, M., Mizuno, S., Matsubara, K., Nakabayashi, K., Sano, S., Fuke, T., et al. (2015). Epimutations of the IG-DMR and the MEG3-DMR at the 14q32.2 imprinted region in two patients with Silver-Russell Syndrome-compatible phenotype. *Eur. J. Hum. Genet.* 23, 1062–1067. doi: 10.1038/ejhg.2014.234

Kalish, J. M., Biesecker, L. G., Brioude, F., Deardorff, M. A., Di Cesare-Merlone, A., Druley, T., et al. (2017). Nomenclature and definition in asymmetric regional body overgrowth. *Am. J. Med. Genet. A* 173, 1735–1738. doi: 10.1002/ajmg.a.38266

Kalsner, L., and Chamberlain, S. J. (2015). Prader-Willi, angelman, and 15q11-q13 duplication syndromes. *Pediatr. Clin. North Am.* 62, 587–606. doi: 10.1016/j.pcl.2015.03.004

Kampmann, M. (2018). CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chem. Biol.* 13, 406–416. doi: 10.1021/acschembio.7b00657

Kanduri, C. (2016). Long noncoding RNAs: lessons from genomic imprinting. *Biochim. Biophys. Acta (BBA) - Gene Regulatory Mechan.* 1859, 102–111. doi: 10.1016/j.bbagrm.2015.05.006

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., et al. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell.* 53, 290–300. doi: 10.1016/j.molcel.2013.11.012

Karaki, S., Paris, C., and Rocchi, P. (2019). "Antisense oligonucleotides, a novel developing targeting therapy," in *Antisense Therapy*, eds S. Sharad and S. Kapur (London: IntechOpen), doi: 10.5772/intechopen.82105

Kelsey, G., and Feil, R. (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20110336. doi: 10.1098/rstb.2011.0336

Kim, J., Bretz, C. L., and Lee, S. (2015). Epigenetic instability of imprinted genes in human cancers. *Nucleic Acids Res.* 43, 10689–10699. doi: 10.1093/nar/gkv867

Kim, Y., Lee, H.-M., Xiong, Y., Sciaky, N., Hulbert, S. W., Cao, X., et al. (2017). Targeting the histone methyltransferase G9a activates imprinted genes and improves survival of a mouse model of Prader-Willi syndrome. *Nat. Med.* 23, 213–222. doi: 10.1038/nm.4257

Kitagawa, M., Kotake, Y., and Ohhata, T. (2012). Long non-coding RNAs involved in cancer development and cell fate determination. *Curr. Drug Targets* 13, 1616–1621. doi: 10.2174/138945012803530026

Knauss, J. L., and Sun, T. (2013). Regulatory mechanisms of long noncoding RNAs in vertebrate central nervous system development and function. *Neuroscience* 235, 200–214. doi: 10.1016/j.neuroscience.2013.01.022

Knott, G. J., and Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science* 361, 866–869. doi: 10.1126/science.aat5011

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588. doi: 10.1038/nature14136

Kopp, F., and Mendell, J. T. (2018). Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172, 393–407. doi: 10.1016/j.cell.2018.01.011

Kordasiewicz, H. B., Stanek, L. M., Wancewicz, E. V., Mazur, C., McAlonis, M. M., Pytel, K. A., et al. (2012). Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* 74, 1031–1044. doi: 10.1016/j.neuron.2012.05.009

Kota, S. K., Llères, D., Bouschet, T., Hirasawa, R., Marchand, A., Begon-Pescia, C., et al. (2014). ICR noncoding RNA expression controls imprinting and DNA replication at the Dlk1-Dio3 domain. *Dev. Cell* 31, 19–33. doi: 10.1016/j.devcel.2014.08.009

Kotzin, J. J., Spencer, S. P., McCright, S. J., Kumar, D. B. U., Collet, M. A., Mowel, W. K., et al. (2016). The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature* 537, 239–243. doi: 10.1038/nature19346

Kuruvilla, J., Sasmita, A. O., and Ling, A. P. K. (2018). Therapeutic potential of combined viral transduction and CRISPR/Cas9 gene editing in treating neurodegenerative diseases. *Neurol. Sci.* 39, 1827–1835. doi: 10.1007/s10072-018-3521-3520

Landers, M., Bancescu, D. L., Le Meur, E., Rougeulle, C., Glatt-Deeley, H., Brannan, C., et al. (2004). Regulation of the large (approximately 1000 kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. *Nucleic Acids Res.* 32, 3480–3492. doi: 10.1093/nar/gkh670

Langouët, M., Glatt-Deeley, H. R., Chung, M. S., Dupont-Thibert, C. M., Mathieux, E., Banda, E. C., et al. (2018). Zinc finger protein 274 regulates imprinted expression of transcripts in Prader-Willi syndrome neurons. *Hum. Mol. Genet.* 27, 505–515. doi: 10.1093/hmg/ddx420

Langouët, M., Gorka, D., Orniacki, C., Dupont-Thibert, C. M., Chung, M. S., Glatt-Deeley, H. R., et al. (2020). Specific ZNF274 binding interference at SNORD116 activates the maternal transcripts in Prader-Willi syndrome neurons. *Hum. Mol. Genet.* 29, 3285–3295. doi: 10.1093/hmg/ddaa210

Latos, P. A., and Barlow, D. P. (2009). Regulation of imprinted expression by macro non-coding RNAs. *RNA Biol.* 6, 100–106. doi: 10.4161/rna.6.2.7854

Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472. doi: 10.1126/science.1228110

Lecerf, C., Le Bourhis, X., and Adriaenssens, E. (2019). The long non-coding RNA H19: an active player with multiple facets to sustain the hallmarks of cancer. *Cell Mol. Life. Sci.* 76, 4673–4687. doi: 10.1007/s00018-019-03240-z

Lee, H.-M., Clark, E. P., Kuijer, M. B., Cushman, M., Pommier, Y., and Philpot, B. D. (2018). Characterization and structure-activity relationships of indenoisoquinoline-derived topoisomerase I inhibitors in unsilencing the dormant Ube3a gene associated with Angelman syndrome. *Mol. Autism* 9:45. doi: 10.1186/s13229-018-0228-222

Lee, J. T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat. Rev. Mol. Cell Biol.* 12, 815–826. doi: 10.1038/nrm3231

Lee, J. T., and Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308–1323. doi: 10.1016/j.cell.2013.02.016

Lee, M. P., DeBaun, M. R., Mitsuya, K., Galonek, H. L., Brandenburg, S., Oshimura, M., et al. (1999). Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc. Natl. Acad. Sci. U S A.* 96, 5203–5208. doi: 10.1073/pnas.96.9.5203

Lewis, M. W., Vargas-Franco, D., Morse, D. A., and Resnick, J. L. (2019). A mouse model of Angelman syndrome imprinting defects. *Hum. Mol. Genet.* 28, 220–229. doi: 10.1093/hmg/ddy345

Li, J., Lin, X., Wang, M., Hu, Y., Xue, K., Gu, S., et al. (2020). Potential role of genomic imprinted genes and brain developmental related genes in autism. *BMC Med. Genomics* 13:54. doi: 10.1186/s12920-020-0693-692

Li, M., Ding, X., Zhang, Y., Li, X., Zhou, H., Yang, L., et al. (2020). Antisense oligonucleotides targeting lncRNA AC104041.1 induces antitumor activity through Wnt2B/β-catenin pathway in head and neck squamous cell carcinomas. *Cell Death Dis.* 11:672. doi: 10.1038/s41419-020-02820-2823

Li, Y., Tan, Z., Zhang, Y., Zhang, Z., Hu, Q., Liang, K., et al. (2021). A noncoding RNA modulator potentiates phenylalanine metabolism in mice. *Science* 373, 662–673. doi: 10.1126/science.aba4991

Liu, F., Pan, H., Xia, G., Qiu, C., and Zhu, Z. (2016). Prognostic and clinicopathological significance of long noncoding RNA H19 overexpression in human solid tumors: evidence from a meta-analysis. *Oncotarget* 7, 83177–83186. doi: 10.18632/oncotarget.13076

Liu, H., Shang, X., and Zhu, H. (2017). LncRNA/DNA binding analysis reveals losses and gains and lineage specificity of genomic imprinting in mammals. *Bioinformatics.* 33, 1431–1436. doi: 10.1093/bioinformatics/btw818

Liu, L. F., and Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U S A.* 84, 7024–7027. doi: 10.1073/pnas.84.20.7024

Luo, Z., Lin, C., Woodfin, A. R., Bartom, E. T., Gao, X., Smith, E. R., et al. (2016). Regulation of the imprinted Dlk1-Dio3 locus by allele-specific enhancer activity. *Genes Dev.* 30, 92–101. doi: 10.1101/gad.270413.115

Lyu, Y., Lou, J., Yang, Y., Feng, J., Hao, Y., Huang, S., et al. (2017). Dysfunction of the WT1-MEG3 signaling promotes AML leukemogenesis via p53-dependent and -independent pathways. *Leukemia* 31, 2543–2551. doi: 10.1038/leu.2017.116

Macdonald, W. A. (2012). Epigenetic mechanisms of genomic imprinting: common themes in the regulation of imprinted regions in mammals, plants, and insects. *Genet Res. Int.* 2012:585024. doi: 10.1155/2012/585024

MacDonald, W. A., and Mann, M. R. W. (2020). Long noncoding RNA functionality in imprinted domain regulation. *PLoS Genet* 16:e1008930. doi: 10.1371/journal.pgen.1008930

Mantovani, G., Bastepe, M., Monk, D., de Sanctis, L., Thiele, S., Usardi, A., et al. (2018). Diagnosis and management of pseudohypoparathyroidism and related disorders: first international consensus statement. *Nat. Rev. Endocrinol.* 14, 476–500. doi: 10.1038/s41574-018-0042-40

Marcho, C., Bevilacqua, A., Tremblay, K. D., and Mager, J. (2015). Tissue-specific regulation of Igf2r/Airn imprinting during gastrulation. *Epigenet. Chromatin* 8:10. doi: 10.1186/s13072-015-0003-y

Marín-Béjar, O., Mas, A. M., González, J., Martinez, D., Athie, A., Morales, X., et al. (2017). The human lncRNA LINC-PINT inhibits tumor cell invasion through a

highly conserved sequence element. *Genome Biol.* 18:202. doi: 10.1186/s13059-017-1331-y

Martins-Taylor, K., Hsiao, J. S., Chen, P.-F., Glatt-Deeley, H., De Smith, A. J., Blakemore, A. I. F., et al. (2014). Imprinted expression of UBE3A in non-neuronal cells from a Prader-Willi syndrome patient with an atypical deletion. *Hum. Mol. Genet.* 23, 2364–2373. doi: 10.1093/hmg/ddt628

Matouk, I. J., Halle, D., Gilon, M., and Hochberg, A. (2015). The non-coding RNAs of the H19-IGF2 imprinted loci: a focus on biological roles and therapeutic potential in Lung Cancer. *J. Trans. Med.* 13:113. doi: 10.1186/s12967-015-0467-463

Mendiola, A. J. P., and LaSalle, J. M. (2021). Epigenetics in Prader-Willi syndrome. *Front. Genet.* 12:624581. doi: 10.3389/fgene.2021.624581

Meng, L., Person, R. E., and Beaudet, A. L. (2012). Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. *Hum. Mol. Genet.* 21, 3001–3012. doi: 10.1093/hmg/dds130

Meng, L., Person, R. E., Huang, W., Zhu, P. J., Costa-Mattioli, M., and Beaudet, A. L. (2013). Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. *PLoS Genet.* 9:e1004039. doi: 10.1371/journal.pgen.1004039

Meng, L., Ward, A. J., Chun, S., Bennett, C. F., Beaudet, A. L., and Rigo, F. (2015). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518, 409–412. doi: 10.1038/nature13975

Meziane, H., Schaller, F., Bauer, S., Villard, C., Matarazzo, V., Riet, F., et al. (2015). An early postnatal oxytocin treatment prevents social and learning deficits in adult mice deficient for Magel2, a gene involved in Prader-willi syndrome and autism. *Biol. Psychiatry* 78, 85–94. doi: 10.1016/j.biopsych.2014.11.010

Mishra, S., Verma, S. S., Rai, V., Awasthee, N., Chava, S., Hui, K. M., et al. (2019). Long non-coding RNAs are emerging targets of phytochemicals for cancer and other chronic diseases. *Cell Mol. Life. Sci.* 76, 1947–1966. doi: 10.1007/s00018-019-03053-3050

Monahan, K., and Lomvardas, S. (2015). Monoallelic expression of olfactory receptors. *Annu. Rev. Cell Dev. Biol.* 31, 721–740. doi: 10.1146/annurev-cellbio-100814-125308

Monk, D., Mackay, D. J. G., Eggermann, T., Maher, E. R., and Riccio, A. (2019). Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* 20, 235–248. doi: 10.1038/s41576-018-0092-90

Monk, D., Morales, J., den Dunnen, J. T., Russo, S., Court, F., Prawitt, D., et al. (2018). Recommendations for a nomenclature system for reporting methylation aberrations in imprinted domains. *Epigenetics* 13, 117–121. doi: 10.1080/15592294.2016.1264561

Mussa, A., Russo, S., De Crescenzo, A., Chiesa, N., Molinatto, C., Selicorni, A., et al. (2013). Prevalence of beckwith-wiedemann syndrome in north West of Italy. *Am. J. Med. Genet. A* 161A, 2481–2486. doi: 10.1002/ajmg.a.36080

Mussa, A., Russo, S., de Crescenzo, A., Freschi, A., Calzari, L., Maitz, S., et al. (2016). Fetal growth patterns in Beckwith-Wiedemann syndrome. *Clin. Genet.* 90, 21–27. doi: 10.1111/cge.12759

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., et al. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717–1720. doi: 10.1126/science.1163802

Nowak, K., Stein, G., Powell, E., He, L. M., Naik, S., Morris, J., et al. (2011). Establishment of paternal allele-specific DNA methylation at the imprinted mouse Gtl2 locus. *Epigenetics* 6, 1012–1020. doi: 10.4161/epi.6.8.16075

Ogata, T., and Kagami, M. (2016). Kagami-Ogata syndrome: a clinically recognizable upd(14)pat and related disorder affecting the chromosome 14q32.2 imprinted region. *J. Hum. Genet.* 61, 87–94. doi: 10.1038/jhg.2015.113

Õunap, K. (2016). Silver-Russell syndrome and beckwith-wiedemann syndrome: opposite phenotypes with heterogeneous molecular etiology. *Mol. Syndromol.* 7, 110–121. doi: 10.1159/000447413

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., et al. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell.* 32, 232–246. doi: 10.1016/j.molcel.2008.08.022

Park, H., Oh, J., Shim, G., Cho, B., Chang, Y., Kim, S., et al. (2019). *In vivo* neuronal gene editing via CRISPR-Cas9 amphiphilic nanocomplexes alleviates deficits in mouse models of Alzheimer's disease. *Nat. Neurosci.* 22, 524–528. doi: 10.1038/s41593-019-0352-0

Patten, M. M., Cowley, M., Oakey, R. J., and Feil, R. (2016). Regulatory links between imprinted genes: evolutionary predictions and consequences. *Proc. Biol. Sci.* 283:20152760. doi: 10.1098/rspb.2015.2760

Pauler, F. M., Barlow, D. P., and Hudson, Q. J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Curr. Opin. Genet. Dev* 22, 283–289. doi: 10.1016/j.gde.2012.02.005

Paulsen, M., Takada, S., Youngson, N. A., Benchaib, M., Charlier, C., Segers, K., et al. (2001). Comparative sequence analysis of the imprinted Dlk1-Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2-H19 region. *Genome Res.* 11, 2085–2094. doi: 10.1101/gr.206901

Peng, W.-X., Koirala, P., and Mo, Y.-Y. (2017). LncRNA-mediated regulation of cell signaling in cancer. *Oncogene* 36, 5661–5667. doi: 10.1038/onc.2017.184

Perez, J. D., Rubinstein, N. D., and Dulac, C. (2016). New perspectives on genomic imprinting, an essential and multifaceted mode of epigenetic control in the developing and adult brain. *Annu. Rev. Neurosci.* 39, 347–384. doi: 10.1146/annurev-neuro-061010-113708

Perez, J. D., Rubinstein, N. D., Fernandez, D. E., Santoro, S. W., Needleman, L. A., Ho-Shing, O., et al. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *eLife* 4:e07860. doi: 10.7554/eLife.07860

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* 10, 973–976. doi: 10.1038/nmeth.2600

Perry, R. B.-T., and Ulitsky, I. (2021). Therapy based on functional RNA elements. *Science* 373, 623–624. doi: 10.1126/science.abj7969

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. *Nat. Rev. Genet.* 15, 517–530. doi: 10.1038/nrg3766

Pham, N. V., Nguyen, M. T., Hu, J. F., Vu, T. H., and Hoffman, A. R. (1998). Dissociation of IGF2 and H19 imprinting in human brain. *Brain Res.* 810, 1–8. doi: 10.1016/s0006-8993(98)00783-785

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B. A., Solcan, N., Cerase, A., et al. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell.* 68, 955–969.e10. doi: 10.1016/j.molcel.2017.11.013

Polex-Wolf, J., Lam, B. Y., Larder, R., Tadross, J., Rimmington, D., Bosch, F., et al. (2018). Hypothalamic loss of Snord116 recapitulates the hyperphagia of Prader-Willi syndrome. *J. Clin. Invest.* 128, 960–969. doi: 10.1172/JCI97007

Pollak, M. N., Schernhammer, E. S., and Hankinson, S. E. (2004). Insulin-like growth factors and neoplasia. *Nat. Rev. Cancer* 4, 505–518. doi: 10.1038/nrc1387

Powell, W. T., Coulson, R. L., Crary, F. K., Wong, S. S., Ach, R. A., Tsang, P., et al. (2013a). A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. *Hum. Mol. Genet.* 22, 4318–4328. doi: 10.1093/hmg/ddt281

Powell, W. T., Coulson, R. L., Gonzales, M. L., Crary, F. K., Wong, S. S., Adams, S., et al. (2013b). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc. Natl. Acad. Sci. U S A.* 110, 13938–13943. doi: 10.1073/pnas.1305426110

Prasasya, R., Grotheer, K. V., Siracusa, L. D., and Bartolomei, M. S. (2020). Temple syndrome and Kagami-Ogata syndrome: clinical presentations, genotypes, models and mechanisms. *Hum. Mol. Genet.* 29, R107–R116. doi: 10.1093/hmg/ddaa133

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Qi, Y., Purtell, L., Fu, M., Lee, N. J., Aepler, J., Zhang, L., et al. (2016). Snord116 is critical in the regulation of food intake and body weight. *Sci. Rep.* 6:18614. doi: 10.1038/srep18614

Quinn, J. J., and Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* 17, 47–62. doi: 10.1038/nrg.2015.10

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

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

Reinius, B., and Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat. Rev. Genet.* 16, 653–664. doi: 10.1038/nrg3888

Riordan, J. D., Keng, V. W., Tschida, B. R., Scheetz, T. E., Bell, J. B., Podetz-Pedersen, K. M., et al. (2013). Identification of rtl1, a retrotransposon-derived imprinted gene, as a novel driver of hepatocarcinogenesis. *PLoS Genet* 9:e1003441. doi: 10.1371/journal.pgen.1003441

Riso, V., Cammisa, M., Kukreja, H., Anvar, Z., Verde, G., Sparago, A., et al. (2016). ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. *Nucleic Acids Res.* 44, 8165–8178. doi: 10.1093/nar/gkw505

Robertson, K. D. (2005). DNA methylation and human disease. *Nat. Rev. Genet.* 6, 597–610. doi: 10.1038/nrg1655

Robinson, W. P. (2000). Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* 22, 452–459.

Rossignol, S., Steunou, V., Chalas, C., Kerjean, A., Rigolet, M., Viegas-Pequignot, E., et al. (2006). The epigenetic imprinting defect of patients with Beckwith-Wiedemann syndrome born after assisted reproductive technology is not restricted to the 11p15 region. *J. Med. Genet.* 43, 902–907. doi: 10.1136/jmg.2006.042135

Rougeulle, C., Glatt, H., and Lalande, M. (1997). The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain. *Nat. Genet.* 17, 14–15. doi: 10.1038/ng0997-14

Rozhdestvensky, T. S., Robeck, T., Galiveti, C. R., Raabe, C. A., Seeger, B., Wolters, A., et al. (2016). Maternal transcription of non-protein coding RNAs from the PWS-critical region rescues growth retardation in mice. *Sci. Rep.* 6:20398. doi: 10.1038/srep20398

Runte, M., Hüttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B., and Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* 10, 2687–2700. doi: 10.1093/hmg/10.23.2687

Sahoo, T., del Gaudio, D., German, J. R., Shinawi, M., Peters, S. U., Person, R. E., et al. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.* 40, 719–721. doi: 10.1038/ng.158

Sanli, I., and Feil, R. (2015). Chromatin mechanisms in the developmental control of imprinted gene expression. *Int. J. Biochem. Cell Biol.* 67, 139–147. doi: 10.1016/j.biocel.2015.04.004

Sanli, I., Lalevée, S., Cammisa, M., Perrin, A., Rage, F., Llères, D., et al. (2018). Meg3 Non-coding RNA expression controls imprinting by preventing transcriptional upregulation in cis. *Cell Rep.* 23, 337–348. doi: 10.1016/j.celrep.2018.03.044

Santoro, F., Mayer, D., Klement, R. M., Warczok, K. E., Stukalov, A., Barlow, D. P., et al. (2013). Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. *Development* 140, 1184–1195. doi: 10.1242/dev.088849

Santoro, F., and Pauler, F. M. (2013). Silencing by the imprinted Airn macro lncRNA: transcription is the answer. *Cell Cycle* 12, 711–712. doi: 10.4161/cc.23860

Sasaki, H., Ferguson-Smith, A. C., Shum, A. S., Barton, S. C., and Surani, M. A. (1995). Temporal and spatial regulation of H19 imprinting in normal and uniparental mouse embryos. *Development* 121, 4195–4202.

Sato, S., Yoshida, W., Soejima, H., Nakabayashi, K., and Hata, K. (2011). Methylation dynamics of IG-DMR and Gtl2-DMR during murine embryonic and placental development. *Genomics* 98, 120–127. doi: 10.1016/j.ygeno.2011.05.003

Schaller, F., Watrin, F., Sturny, R., Massacrier, A., Szepetowski, P., and Muscatelli, F. (2010). A single postnatal injection of oxytocin rescues the lethal feeding behaviour in mouse newborns deficient for the imprinted Magel2 gene. *Hum. Mol. Genet.* 19, 4895–4905. doi: 10.1093/hmg/ddq424

Schertzer, M. D., Braceros, K. C. A., Starmer, J., Cherney, R. E., Lee, D. M., Salazar, G., et al. (2019). lncRNA-Induced spread of polycomb controlled by genome architecture, RNA abundance, and CpG Island DNA. *Mol. Cell.* 75, 523–537.e10. doi: 10.1016/j.molcel.2019.05.028

Schoenherr, C. J., Levorse, J. M., and Tilghman, S. M. (2003). CTCF maintains differential methylation at the Igf2/H19 locus. *Nat. Genet.* 33, 66–69. doi: 10.1038/ng1057

Schroeder, D. I., Blair, J. D., Lott, P., Yu, H. O. K., Hong, D., Crary, F., et al. (2013). The human placenta methylome. *Proc. Natl. Acad. Sci. U S A.* 110, 6037–6042. doi: 10.1073/pnas.1215145110

Schwartz, Y. B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–864. doi: 10.1038/nrg3603

Sellers, Z. P., Bolkun, L., Kloczko, J., Wojtaszewska, M. L., Lewandowski, K., Moniuszko, M., et al. (2019). Increased methylation upstream of the MEG3 promotor is observed in acute myeloid leukemia patients with better overall survival. *Clin. Epigenetics* 11:50. doi: 10.1186/s13148-019-0643-z

Simon, J. A., and Kingston, R. E. (2013). Occupying chromatin: polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell.* 49, 808–824. doi: 10.1016/j.molcel.2013.02.013

Singh, P., Wu, X., Lee, D.-H., Li, A. X., Rauch, T. A., Pfeifer, G. P., et al. (2011). Chromosome-wide analysis of parental allele-specific chromatin and DNA methylation. *Mol. Cell. Biol.* 31, 1757–1770. doi: 10.1128/MCB.00961-910

Skourti-Stathaki, K., Proudfoot, N. J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell.* 42, 794–805. doi: 10.1016/j.molcel.2011.04.026

Skryabin, B. V., Gubar, L. V., Seeger, B., Pfeiffer, J., Handel, S., Robeck, T., et al. (2007). Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet.* 3:e235. doi: 10.1371/journal.pgen.0030235

Sleutels, F., Zwart, R., and Barlow, D. P. (2002). The non-coding air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813. doi: 10.1038/415810a

Smilinich, N. J., Day, C. D., Fitzpatrick, G. V., Caldwell, G. M., Lossie, A. C., Cooper, P. R., et al. (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl. Acad. Sci. U S A.* 96, 8064–8069. doi: 10.1073/pnas.96.14.8064

Smith, R. A., Miller, T. M., Yamanaka, K., Monia, B. P., Condon, T. P., Hung, G., et al. (2006). Antisense oligonucleotide therapy for neurodegenerative disease. *J. Clin. Invest.* 116, 2290–2296. doi: 10.1172/JCI25424

Soejima, H., and Higashimoto, K. (2013). Epigenetic and genetic alterations of the imprinting disorder Beckwith-Wiedemann syndrome and related disorders. *J. Hum. Genet.* 58, 402–409. doi: 10.1038/jhg.2013.51

Soellner, L., Begemann, M., Mackay, D. J. G., Grønskov, K., Tümer, Z., Maher, E. R., et al. (2017). Recent advances in imprinting disorders. *Clin. Genet.* 91, 3–13. doi: 10.1111/cge.12827

Southan, C. (2017). Last rolls of the yoyo: assessing the human canonical protein count. *F1000Res* 6:448. doi: 10.12688/f1000research.11119.1

St Laurent, G., Wahlestedt, C., and Kapranov, P. (2015). The landscape of long noncoding RNA classification. *Trends Genet.* 31, 239–251. doi: 10.1016/j.tig.2015.03.007

Stanurova, J., Neureiter, A., Hiber, M., de Oliveira Kessler, H., Stolp, K., Goetzke, R., et al. (2018). Corrigendum: angelman syndrome-derived neurons display late onset of paternal UBE3A silencing. *Sci. Rep.* 8:46952. doi: 10.1038/srep46952

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi: 10.1038/s41580-020-00315-319

Sun, J., Carlson-Stevermer, J., Das, U., Shen, M., Delenclos, M., Snead, A. M., et al. (2019). CRISPR/Cas9 editing of APP C-terminus attenuates β-cleavage and promotes α-cleavage. *Nat. Commun.* 10:53. doi: 10.1038/s41467-018-07971-7978

Sun, K.-X., Wu, D.-D., Chen, S., Zhao, Y., and Zong, Z.-H. (2017). LncRNA MEG3 inhibit endometrial carcinoma tumorigenesis and progression through PI3K pathway. *Apoptosis* 22, 1543–1552. doi: 10.1007/s10495-017-1426-1427

Sutcliffe, J. S., Nakao, M., Christian, S., Orstavik, K. H., Tommerup, N., Ledbetter, D. H., et al. (1994). Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nat. Genet.* 8, 52–58. doi: 10.1038/ng0994-52

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E. J., et al. (2016). *In vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 540, 144–149. doi: 10.1038/nature20565

Tan, L., Xing, D., Chang, C.-H., Li, H., and Xie, X. S. (2018). Three-dimensional genome structures of single diploid human cells. *Science* 361, 924–928. doi: 10.1126/science.aat5641

Temple, I. K., Cockwell, A., Hassold, T., Pettay, D., and Jacobs, P. (1991). Maternal uniparental disomy for chromosome 14. *J. Med. Genet.* 28, 511–514. doi: 10.1136/jmg.28.8.511

Terranova, R., Yokobayashi, S., Stadler, M. B., Otte, A. P., van Lohuizen, M., Orkin, S. H., et al. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* 15, 668–679. doi: 10.1016/j.devcel.2008.08.015

Thorvaldsen, J. L., Duran, K. L., and Bartolomei, M. S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.* 12, 3693–3702. doi: 10.1101/gad.12.23.3693

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., Erice Imprinting, and Group. (2019). Genomic imprinting and physiological processes in mammals. *Cell* 176, 952–965. doi: 10.1016/j.cell.2019.01.043

Tunster, S. J., Van de Pette, M., and John, R. M. (2011). Fetal overgrowth in the Cdkn1c mouse model of Beckwith-Wiedemann syndrome. *Dis. Model. Mech.* 4, 814–821. doi: 10.1242/dmm.007328

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., et al. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat. Genet.* 36, 1296–1300. doi: 10.1038/ng1467

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* 19, 535–548. doi: 10.1038/s41576-018-0017-y

van der Werf, I. M., Buiting, K., Czeschik, C., Reyniers, E., Vandeweyer, G., Vanhaesebrouck, P., et al. (2016). Novel microdeletions on chromosome 14q32.2 suggest a potential role for non-coding RNAs in Kagami-Ogata syndrome. *Eur. J. Hum. Genet.* 24, 1724–1729. doi: 10.1038/ejhg.2016.82

Vennin, C., Spruyt, N., Dahmani, F., Julien, S., Bertucci, F., Finetti, P., et al. (2015). H19 non coding RNA-derived miR-675 enhances tumorigenesis and metastasis of breast cancer cells by downregulating c-Cbl and Cbl-b. *Oncotarget* 6, 29209–29223. doi: 10.18632/oncotarget.4976

Vettermann, C., and Schlissel, M. S. (2010). Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol. Rev.* 237, 22–42. doi: 10.1111/j.1600-065X.2010.00935.x

Vitali, P., Royo, H., Marty, V., Bortolin-Cavaillé, M.-L., and Cavaillé, J. (2010). Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J. Cell Sci.* 123, 70–83. doi: 10.1242/jcs.054957

Vu, T. H., and Hoffman, A. R. (1997). Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain. *Nat. Genet.* 17, 12–13. doi: 10.1038/ng0997-12

Wakeling, E. L., Brioude, F., Lokulo-Sodipe, O., O'Connell, S. M., Salem, J., Bliek, J., et al. (2017). Diagnosis and management of Silver-Russell syndrome: first international consensus statement. *Nat. Rev. Endocrinol.* 13, 105–124. doi: 10.1038/nrendo.2016.138

Wan, P., Su, W., and Zhuo, Y. (2017). The role of long noncoding RNAs in neurodegenerative diseases. *Mol. Neurobiol.* 54, 2012–2021. doi: 10.1007/s12035-016-9793-9796

Wang, J. C., Passage, M. B., Yen, P. H., Shapiro, L. J., and Mohandas, T. K. (1991). Uniparental heterodisomy for chromosome 14 in a phenotypically abnormal familial balanced 13/14 Robertsonian translocation carrier. *Am. J. Hum. Genet.* 48, 1069–1074.

Wang, X., Sun, Q., McGrath, S. D., Mardis, E. R., Soloway, P. D., and Clark, A. G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One* 3:e3839. doi: 10.1371/journal.pone.0003839

Wang, Y., Shen, Y., Dai, Q., Yang, Q., Zhang, Y., Wang, X., et al. (2017). A permissive chromatin state regulated by ZFP281-AFF3 in controlling the imprinted Meg3 polycistron. *Nucleic Acids Res.* 45, 1177–1185. doi: 10.1093/nar/gkw1051

Weksberg, R., Shuman, C., and Beckwith, J. B. (2010). Beckwith-Wiedemann syndrome. *Eur. J. Hum. Genet.* 18, 8–14. doi: 10.1038/ejhg.2009.106

White, C. R., MacDonald, W. A., and Mann, M. R. W. (2016). Conservation of DNA methylation programming between mouse and human gametes and preimplantation embryos. *Biol. Reprod.* 95:61. doi: 10.1095/biolreprod.116.140319

Wolter, J. M., Mao, H., Fragola, G., Simon, J. M., Krantz, J. L., Bazick, H. O., et al. (2020). Cas9 gene therapy for Angelman syndrome traps Ube3a-ATS long non-coding RNA. *Nature* 587, 281–284. doi: 10.1038/s41586-020-2835-2832

Wu, H., Yin, Q.-F., Luo, Z., Yao, R.-W., Zheng, C.-C., Zhang, J., et al. (2016). Unusual processing generates SPA LncRNAs that sequester multiple RNA binding proteins. *Mol. Cell.* 64, 534–548. doi: 10.1016/j.molcel.2016.10.007

Xin, Z., Tachibana, M., Guggiari, M., Heard, E., Shinkai, Y., and Wagstaff, J. (2003). Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. *J. Biol. Chem.* 278, 14996–15000. doi: 10.1074/jbc.M211753200

Yamasaki, K., Joh, K., Ohta, T., Masuzaki, H., Ishimaru, T., Mukai, T., et al. (2003). Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Hum. Mol. Genet.* 12, 837–847. doi: 10.1093/hmg/ddg106

Yan, Y., Frisén, J., Lee, M. H., Massagué, J., and Barbacid, M. (1997). Ablation of the CDK inhibitor p57Kip2 results in increased apoptosis and delayed differentiation during mouse development. *Genes Dev.* 11, 973–983. doi: 10.1101/gad.11.8.973

Yang, T., Adamson, T. E., Resnick, J. L., Leff, S., Wevrick, R., Francke, U., et al. (1998). A mouse model for Prader-Willi syndrome imprinting-centre mutations. *Nat. Genet.* 19, 25–31. doi: 10.1038/ng0598-25

Yao, H., Duan, M., Lin, L., Wu, C., Fu, X., Wang, H., et al. (2017). TET2 and MEG3 promoter methylation is associated with acute myeloid leukemia in a Hainan population. *Oncotarget* 8, 18337–18347. doi: 10.18632/oncotarget.15440

Ye, Y., Guo, J., Xiao, P., Ning, J., Zhang, R., Liu, P., et al. (2020). Macrophages-induced long noncoding RNA H19 up-regulation triggers and activates the miR-193b/MAPK1 axis and promotes cell aggressiveness in hepatocellular carcinoma. *Cancer Lett.* 469, 310–322. doi: 10.1016/j.canlet.2019.11.001

Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X.-D., and Gage, F. H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* 16, 130–137. doi: 10.1038/nsmb.1545

Yin, Q.-F., Yang, L., Zhang, Y., Xiang, J.-F., Wu, Y.-W., Carmichael, G. G., et al. (2012). Long noncoding RNAs with snoRNA ends. *Mol. Cell.* 48, 219–230. doi: 10.1016/j.molcel.2012.07.033

Yoon, J.-H., Kim, J., and Gorospe, M. (2015). Long noncoding RNA turnover. *Biochimie* 117, 15–21. doi: 10.1016/j.biochi.2015.03.001

Yoshimizu, T., Miroglio, A., Ripoche, M.-A., Gabory, A., Vernucci, M., Riccio, A., et al. (2008). The H19 locus acts *in vivo* as a tumor suppressor. *Proc. Natl. Acad. Sci. U S A.* 105, 12417–12422. doi: 10.1073/pnas.0801540105

Zhang, C.-Y., Yu, M.-S., Li, X., Zhang, Z., Han, C.-R., and Yan, B. (2017). Overexpression of long non-coding RNA MEG3 suppresses breast cancer cell proliferation, invasion, and angiogenesis through AKT pathway. *Tumour Biol,* 39:1010428317701311. doi: 10.1177/1010428317701311

Zhang, K., Luo, Z., Zhang, Y., Zhang, L., Wu, L., Liu, L., et al. (2016). Circulating lncRNA H19 in plasma as a novel biomarker for breast cancer. *Cancer Biomark.* 17, 187–194. doi: 10.3233/CBM-160630

Zhang, P., Liégeois, N. J., Wong, C., Finegold, M., Hou, H., Thompson, J. C., et al. (1997). Altered cell differentiation and proliferation in mice lacking p57KIP2 indicates a role in Beckwith-Wiedemann syndrome. *Nature* 387, 151–158. doi: 10.1038/387151a0

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., et al. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell.* 40, 939–953. doi: 10.1016/j.molcel.2010.12.011

Zhao, Y., Li, H., Fang, S., Kang, Y., Wu, W., Hao, Y., et al. (2016). NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* 44, D203–D208. doi: 10.1093/nar/gkv1252

Zheng, J.-F., Guo, N.-H., Zi, F.-M., and Cheng, J. (2020). Long noncoding RNA H19 promotes tumorigenesis of multiple myeloma by activating BRD4 signaling by targeting MicroRNA 152-3p. *Mol. Cell. Biol.* 40:e00382-19. doi: 10.1128/MCB.00382-319

Zhou, H., Wang, B., Yang, Y.-X., Jia, Q.-J., Zhang, A., Qi, Z.-W., et al. (2019). Long Noncoding RNAs in pathological cardiac remodeling: a review of the update literature. *Biomed. Res. Int.* 2019:7159592. doi: 10.1155/2019/7159592

Zhou, Y., Zhang, X., and Klibanski, A. (2012). MEG3 noncoding RNA: a tumor suppressor. *J. Mol. Endocrinol.* 48, R45–R53. doi: 10.1530/JME-12-18

Zhu, M., Wang, X., Gu, Y., Wang, F., Li, L., and Qiu, X. (2019). MEG3 overexpression inhibits the tumorigenesis of breast cancer by downregulating miR-21 through the PI3K/Akt pathway. *Arch. Biochem. Biophys.* 661, 22–30. doi: 10.1016/j.abb.2018.10.021

Zhuo, C., Hou, W., Hu, L., Lin, C., Chen, C., and Lin, X. (2017). Genomic editing of Non-Coding RNA genes with CRISPR/Cas9 ushers in a potential novel approach to study and treat schizophrenia. *Front. Mol. Neurosci.* 10:28. doi: 10.3389/fnmol.2017.00028

Zieba, J., Low, J. K., Purtell, L., Qi, Y., Campbell, L., Herzog, H., et al. (2015). Behavioural characteristics of the Prader-Willi syndrome related biallelic Snord116 mouse model. *Neuropeptides* 53, 71–77. doi: 10.1016/j.npep.2015.06.009

Zink, F., Magnusdottir, D. N., Magnusson, O. T., Walker, N. J., Morris, T. J., Sigurdsson, A., et al. (2018). Insights into imprinting from parent-of-origin phased methylomes and transcriptomes. *Nat. Genet.* 50, 1542–1552. doi: 10.1038/s41588-018-0232-237

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

REVIEW
published: 11 January 2022
doi: 10.3389/fcell.2021.720798

May the Odds Be Ever in Your Favor: Non-deterministic Mechanisms Diversifying Cell Surface Molecule Expression

*Donnell L. Williams[1,2†], Veronica Maria Sikora[1†], Max A. Hammer[1†], Sayali Amin[1], Taema Brinjikji[1], Emily K. Brumley[1], Connor J. Burrows[1], Paola Michelle Carrillo[1], Kirin Cromer[1], Summer J. Edwards[1], Olivia Emri[1], Daniel Fergle[1], M. Jamal Jenkins[1,2], Krishangi Kaushik[1], Daniella D. Maydan[1], Wrenn Woodard[1] and E. Josephine Clowney[2]**

[1]MCDB 464 – Cellular Diversity in the Immune and Nervous Systems, University of Michigan, Ann Arbor, MI, United States, [2]Department of Molecular, Cellular and Developmental Biology, The University of Michigan, Ann Arbor, MI, United States

How does the information in the genome program the functions of the wide variety of cells in the body? While the development of biological organisms appears to follow an explicit set of genomic instructions to generate the same outcome each time, many biological mechanisms harness molecular noise to produce variable outcomes. Non-deterministic variation is frequently observed in the diversification of cell surface molecules that give cells their functional properties, and is observed across eukaryotic clades, from single-celled protozoans to mammals. This is particularly evident in immune systems, where random recombination produces millions of antibodies from only a few genes; in nervous systems, where stochastic mechanisms vary the sensory receptors and synaptic matching molecules produced by different neurons; and in microbial antigenic variation. These systems employ overlapping molecular strategies including allelic exclusion, gene silencing by constitutive heterochromatin, targeted double-strand breaks, and competition for limiting enhancers. Here, we describe and compare five stochastic molecular mechanisms that produce variety in pathogen coat proteins and in the cell surface receptors of animal immune and neuronal cells, with an emphasis on the utility of non-deterministic variation.

Keywords: monogenic, monoallelic, stochastic gene choice, V(D)J recombination, Dscam, protocadherin, olfactory receptor, antigenic variation

OPEN ACCESS

Edited by:
Jin Xu,
Sun Yat-Sen University, China

Reviewed by:
Alexander Gimelbrant,
Harvard Medical School, United States
Daniele Canzio,
University of California, San Francisco, United States

****Correspondence:***
E. Josephine Clowney
jclowney@umich.edu

[†]These authors have contributed equally to this work and share first authorship

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

Received: *05 June 2021*
Accepted: *24 November 2021*
Published: *11 January 2022*

Citation:
Williams DL, Sikora VM, Hammer MA, Amin S, Brinjikji T, Brumley EK, Burrows CJ, Carrillo PM, Cromer K, Edwards SJ, Emri O, Fergle D, Jenkins MJ, Kaushik K, Maydan DD, Woodard W and Clowney EJ (2022) May the Odds Be Ever in Your Favor: Non-deterministic Mechanisms Diversifying Cell Surface Molecule Expression. Front. Cell Dev. Biol. 9:720798. doi: 10.3389/fcell.2021.720798

INTRODUCTION

Despite the shocking complexity of eukaryotic life, eukaryotic genomes often contain less than 20,000 protein-coding genes. While most genes are expressed in a deterministic manner, a variety of molecular mechanisms have been discovered that expand the coding capacity of the genome by expressing cell surface molecules in a quasi-random manner. Expression systems that accomplish cell surface molecule diversification make use of genomic rearrangement, RNA splicing, and epigenetic restriction to create a vast array of molecular variants from a limited amount of DNA. In this way, the static information within genomes can generate a wider diversity of cells throughout the body or across unicellular populations.

TABLE 1 | Comparison of non-deterministic systems of cell surface molecule expression. For brevity, references are not included; they are provided throughout the main text description of each system.

	VSG	IgG	Dscam	Pcdh	ORs
Combinatorial diversification	Yes—construction of mosaic VSGs increases repertoire	Yes—V, D, and J exons are variably combined	Yes—exons 4, 6, and 9 are variably combined; multiple isoforms per cell	Yes—cells can express isoforms from A, B, and G clusters; some cells express multiple genes from one cluster	No
Monoallelic	N/A—active VSG copied from a "genomic archive; " expression sites can be hemizygous	Yes	No	Sometimes	Yes
Exclusive (i.e. exactly one isoform/ cell)	Yes	Yes	No	Cells generally express isoforms from A, B, and G clusters. The choice within cluster is not necessarily exclusive	Yes
Dependence on limiting enhancer	Active VSG associates with genomic locus encoding splice-leader RNA	Yes, for proximal V promoter selection	At the RNA level, dependent on unique RNA "chooser" elements	Yes, e.g. HS5-1 for PCDHA	Yes, Greek Islands
Mechanism of choice/variation	Recombination into active site, active site switching, construction of mosaic VSGs	Recombination and AID-induced point mutation	Alternative splicing	Promoter choice via limiting enhancer(s)	Promoter choice via limiting enhancers
Expression choice in each cell is initially more promiscuous, and then refines	Yes	No	N/A, not exclusive	Yes	Yes
Choice is stable once refined	Choice is heritable. Switching is critical, but unclear how it is induced	Yes	No	Unknown	Yes
Feedback	Selection by immune system clearance	Unfolded protein response	Unknown	Unknown	Unfolded protein response
Function of non-deterministic choice	Immune system can't predict what antigen will be expressed next, mosaic VSGs expand repertoire	Pathogen can't predict what antibodies will be present, allows defense against novel pathogens that were not predicted by evolution	Allows neurons of the same ontogenetic type to have distinct barcodes, and allows neurons to respond differently to themselves than to ontogenetically identical sisters	Allows neurons of the same ontogenetic type to have distinct barcodes, and allows neurons to respond differently to themselves than to ontogenetically identical sisters	Provides a concise mechanism for activating OR expression; new ORs can be expressed without evolution of new transcriptional mechanism; only need one regulatory system instead of 1,000
Drawbacks of non-deterministic choice	Not obvious	Since antibodies are produced randomly, many arise that bind self-antigens. These must be selected against	Not obvious	Not obvious	Requires receptor-dependent mechanisms to wire OSNs to olfactory bulb glomeruli
Function of restricted expression and diverse cell surface phenotypes	Prolongs infection by allowing host immune system to "see" only one VSG at once	Allows binding of diverse and novel antigens; compartmentalization allows cellular somatic selection of effective receptors	Neuronal self-identification and self-avoidance	Neuronal self-identification and self-avoidance	Olfactory perception—each cell senses limited and distinct odorants

This diversity is crucial for the proper functioning of many different biological systems. The immune system, for example, relies on diverse antigen receptors to bind to and recognize an incredible range of potential pathogens and harmful molecules. Without the stochastic mechanisms driving this variation in expression, the proper functioning of the immune system would be severely compromised. The nervous system is similar in that it also relies on diversity in gene expression for proper functioning. Like the immune system, neurons in chemosensory systems express diverse receptors to bind a wide array of environmental molecules. In addition, neurons, even of the same class, must be sufficiently different from one another in order to properly identify self vs. non-self. Finally, pathogens also diversify their surface molecules in an arms race with the adaptive

immune system. Thus, understanding the mechanisms that produce non-deterministic cellular heterogeneity is an important area of study.

In this review, we will focus specifically on non-deterministic processes that select one or a few surface molecules to be expressed on a particular cell from among many copies of similar sequences encoded in the genome. We highlight 5 such systems: the expression of variable surface glycoproteins (VSGs) by the parasite *Trypanosoma brucei*, pathogen identification by B cell and T cell receptors, neuronal self-avoidance through expression of Dscams and protocadherins, and the perception of stimuli through the olfactory system. While other reviews have compared subsets of these systems, here we broaden the scope of the comparison by considering both single-celled organisms and animals and by considering both neuronal and barrier functions (Magklara and Lomvardas 2013; Khamlichi and Feil 2018; Aresta-Branco et al., 2019a). In addition to comparing molecular mechanisms, we highlight the distinct types of utility gained by non-deterministic expression in different systems. Often, procedural or algorithmic mechanisms are simply more concise than deterministic mechanisms. In other cases, unpredictability in molecular outcomes is itself crucial for cellular function.

There are many similarities across these five examples (summarized in **Table 1**). First, they all have some type of restriction mechanism, often heterochromatin-based, that ensures that all of the coding sequences that could possibly be expressed aren't expressed at the same time. Each system also involves stochastic selection of a single (or a few) isoform(s) that will be expressed. For antigen receptors, Dscams, protocadherins, and olfactory receptors, stochastic selection involves a unique enhancer or locus control region. Such a region or enhancer has not yet been identified for *VSG*s. Lastly, in three of these systems, there are feedback mechanisms downstream of selection that can act to help correct any flaws that were made during selection. In antigenic (VSG) variation, this feedback is whether or not the cell survives the host immune system. In V(D)J recombination, feedback takes place within the germinal center when higher affinity B cells win the competition for antigen. In olfactory receptor choice, the feedback mechanism allows the cell to choose a different olfactory receptor gene if it initially chose a flawed one—or stops the cell from choosing another gene if the one it already chose is functional. Although similar feedback processes may take place in Dscam and protocadherin expression, they have not yet been discovered.

While we restrict our analysis here to mechanisms that diversify cell surface molecule repertoires by choosing among genetically encoded paralogues, we note that all biological diversification ultimately relies on noise in genome replication that produces mutations, and that noise is often harnessed and regulated to do biological work. For example, HIV immune evasion has been suggested to result from the virus's retention of an unusually error-prone replication enzyme, and switches between lytic and latent phases are thought to occur stochastically (Roberts et al., 1988; Weinberger and Weinberger 2013; Cuevas et al., 2015). Behavioral switches are also likely governed by probabilistic rather than deterministic mechanisms. The degree of variation in gene expression between cells is itself subject to selection, and such variation can alter the penetrance of mutant alleles (Raj et al., 2010; Metzger et al., 2015; Duveau et al., 2018) Stochastic processes can also reduce the fitness costs of mutations, as in the case of X inactivation in female mammals. While the processes of life contravene entropy, in many cases the otherwise robust and predictable mechanisms of cellular development allow molecular noise to peek through in a regulated manner to influence phenotype.

As we discuss throughout, the monogenic and/or monoallelic expression of cell surface molecules allow each of these systems to appropriately interact with the outside or extracellular world. The functional purpose of selecting cell surface molecules in a non-deterministic rather than a predictable manner varies across them. In some cases, non-deterministic processes may be the only way for cells in otherwise almost identical environments and with identical differentiation regimens to become distinct from one another. In the nervous system, for example, groups of neurons that are developmentally equivalent and located in the same location can produce different cell-surface proteins by randomly selecting and expressing certain gene segments or genes—as is the case with olfactory receptors in olfactory sensory neurons. Non-deterministic expression systems likely also allow for a larger array of different proteins to be made than can otherwise be deterministically encoded by the genome, as is likely the case for the immune system. Because it is inherently unpredictable, non-deterministic expression may also increase fitness for hosts and pathogens locked in battle with one another.

Definitions

In these fields, the terms "stochastic" and "random" are used to refer to processes in which knowing the ontogenetic identity of a cell predicts a distribution of possible gene expression choice but is insufficient to deterministically predict cell surface molecule expression. We note that in most of these systems, molecular choices follow biased distributions—for example, olfactory receptor choice is biased by position in the olfactory epithelium, VSG choice by the time course of infection, and Dscam choice by the neuronal cell type. Biased distributions are consistent with the mathematical definitions of stochastic or random, and we continue to use those terms here. We use the terms "non-deterministic," "probabilistic," and "unpredictable" as additional descriptors. Further, we use "monoallelic" to refer to molecular choice between two copies of the same gene, and "monogenic" to refer to selection among paralogues. We include gene families here, for example the Dscams, in which surface molecule choice is not exclusively monogenic, i.e. where multiple choices are made in each cell but most of the available choices are still repressed. We note that in the VSG field, expression of a single VSG paralogue per trypanosome is typically referred to as "monoallelic" expression. For consistency with the other topics covered here, we use the term "monogenic."

VARIANT SURFACE GLYCOPROTEINS

Many pathogens have learned to survive in host environments that are hostile to their growth. One such method that pathogens have evolved is antigenic variation. Here, we will discuss coat protein switching in trypanosome infection as a model. *Trypanosoma brucei* is a single-celled eukaryotic pathogen that

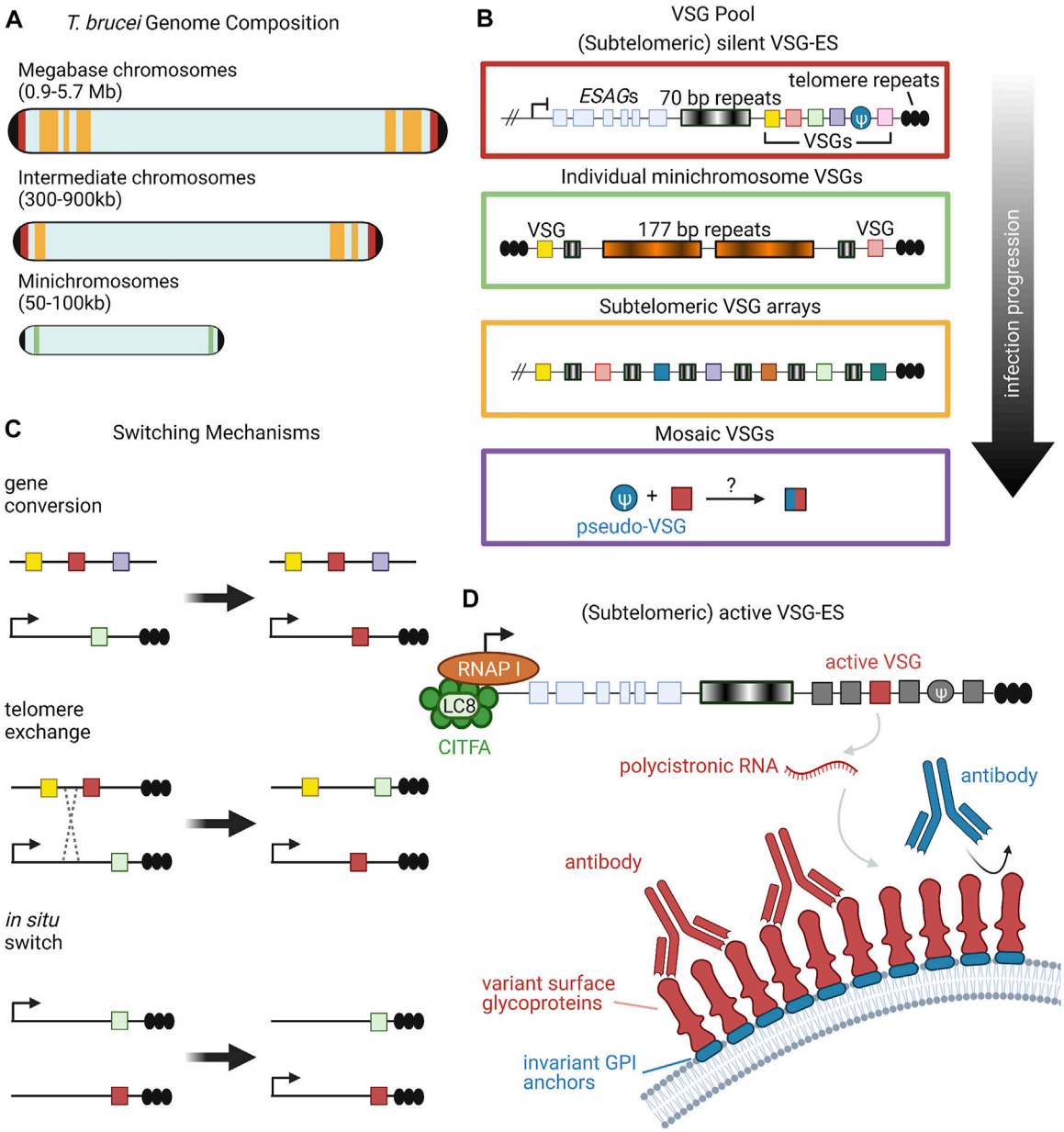


FIGURE 1 | Organization and Expression of VSGs. **(A)** *T. brucei* has an unusual karyotype consisting of large, megabase-sized chromosomes, intermediate chromosomes, and minichromosomes. Size ranges of each chromosome type are listed in parentheses (Berriman et al., 2005). Black caps on the ends of the chromosomes represent telomeres. Red, yellow, and green bands denote the typical locations of different VSG repertoires, corresponding to the colored-coded insets in panel **(B)**. **(B)** Candidate VSG genes are located on megabase and intermediate chromosomes within silent subtelomeric bloodstream expression sites (VSG-ES; red box) or subtelomeric arrays (yellow box). Individual VSG genes can also be found in subtelomeric regions of minichromosomes (green box) or can be generated from recombination of intact and/or VSG pseudogenes from various sources (purple box). Throughout the course of infection, *T. brucei* will draw upon the VSG pool in a semi-predictable manner according to the location of candidate genes; for example, VSG gene arrays from silent expression sites are typically used before minichromosome VSGs (Sima et al., 2019). **(C)** VSG expression proceeds from a single active expression site. To shift expression to a new VSG gene, *T. brucei* can employ one of three switching mechanisms: gene conversion, telomere exchange, or *in situ* switching (Liu et al., 1983; Rudenko et al., 1996; Horn and Cross, 1997; Robinson et al., 1999; Li, 2015). **(D)** RNA polymerase I transcribes polycistronic RNAs from active VSG expression sites. The CITFA transcription factor complex, which in *T. brucei* consists of CITFA subunits 1–7 (green circles) complexed with LC8/DYNLL1 (light green oval), is a basal transcription factor required for RNA pol I initiation (Kirkham et al., 2016). The active VSG gene is typically transcribed last, preceded by expression-site associated genes (ESAGs) (Pays et al., 2001; Hertz-Fowler et al., 2008). VSG RNA is translated into variant surface glycoproteins, which form a densely-packed coat that prevents recognition of underlying invariant cell-surface molecules, such as GPI anchors (Hertz-Fowler et al., 2008; Horn and McCulloch, 2010). Figure inspiration was drawn from various sources (Rudenko et al., 1996; Berriman et al., 2005; Li, 2015). All figures in this review were created using BioRender.

has dedicated a large amount of its genome to this process. This parasite is found mostly in sub-Saharan Africa and is the cause of a vector-borne disease known as sleeping sickness. *T. brucei* is coated by a dense layer of variant surface glycoproteins (VSGs) and is able to switch which *VSG* is expressed in order to evade host immune systems (Boothroyd et al., 2009; Hoeijmakers et al., 1980; Sima et al., 2019). The high density of surface VSG molecules shields the pathogen's other non-variable surface proteins (**Figure 1**), making the pathogen's immunological identity tied to the particular *VSG* it expresses (Hertz-Fowler et al., 2008; Horn and McCulloch, 2010).

Research suggests that there are about 2000 genes that constitute the *VSG* repertoire of *T. brucei,* clustered into subtelomeric arrays as well as on several minichromosomes (**Figure 1A**) (Cross et al., 2014). Though each individual *T. brucei* organism expresses only a single *VSG* gene at a time, VSG switching has been shown to occur at a rate as high as 10^{-3} switches per cell per generation (Mugnier et al., 2015; Turner and Barry, 1989). This high switching rate, along with the large number and diversity of cells present, leads to a sinusoidal pattern of infection where the immune system eliminates cells expressing a given VSG, but not before new variants arise in the population. These variants then grow in number, only to be wiped out again by the immune system, followed by the emergence of new variants (Mugnier et al., 2016). This constant back-and-forth between new VSG variants and the host immune system allows for *T. brucei* to remain inside of a host for long periods of time, creating chronic infections. Interestingly, studies of *T. brucei* population dynamics have begun to reveal semi-predictable patterns in *VSG* expression based on gene location and other gene family characteristics (**Figure 1B**), but much is still unknown about the level of determinism in the system (Morrison et al., 2005; Mugnier et al., 2015).

Multiple overlapping mechanisms accomplish this dynamic VSG switching (**Figure 1C**). The predominant mechanism is duplicative *VSG* gene conversion, in which a silent *VSG* gene is copied into an active expression site (Liu et al., 1983; Robinson et al., 1999; Li, 2015). *VSG* expression can also swap via *in situ* switching where a previously silenced expression site is activated, while the previously active site is silenced (Horn and Cross, 1997). A third mechanism is telomere exchange, where telomeric regions undergo crossing over that swaps which *VSG* is downstream of the active promoter (Rudenko et al., 1996). The field has primarily focused on these first two mechanisms, with telomere exchange still relatively underexplored, so we will focus more on gene conversion and *in situ* switching in this review.

As in all the monogenic expression systems described in this review, repression of the majority of possible loci is a necessary condition for restricted use of the chosen locus. The sub-telomeric location of *VSG* expression sites plays a part in their repression (Ersfeld et al., 1999; Berriman et al., 2005). Telomere proximity is inversely related to transcriptional activity of genes generally (Robin et al., 2014), and this trend holds true for DNA Pol I transcribed genes such as *VSG* genes (Glover and Horn, 2006). The telomere binding protein RAP1 is an essential component of the telomere complex and has been associated with *VSG* repression (Yang et al., 2009). The protein phosphatidylinositol 5-phosphatase binds to RAP1, and, together with phosphatidylinositol 5-kinase, helps to control *VSG* gene repression near the telomere by phosphorylating and dephosphorylating key regulatory molecules (Cestari and Stuart, 2015).

VSG Expression Sites

The trypanosome genome has 20–40 polycistronic, sub-telomeric expression sites (ES) that promote transcription of *VSG*s as well as ES-associated genes (*ESAG*s) (Hertz-Fowler et al., 2008; Pays et al., 2001). They contain a Pol I promoter and are typically around 45 kb in length, with the *VSG* gene the most distal gene transcribed (**Figure 1D**) (Pays et al., 2001; Hertz-Fowler et al., 2008). We will focus our attention on expression sites active during the bloodstream stage of expression (bloodstream expression site, BES). Interestingly, though VSG proteins are monogenically expressed, it has been observed that multiple BES can be transcriptionally active at a time (Kassem et al., 2014). The additional BES transcripts do not fully elongate, are transcribed at lower levels, and are not translated, indicating additional regulation at the transcriptional and post-transcriptional levels to maintain monogenic expression (Kassem et al., 2014). Due to differing recombination into BESs, the two alleles of a particular BES could contain different contents; regardless, expression is from only one BES per cell and is therefore monoallelic.

VSG Induction, Inheritance, and Switching

Monogenic expression and switching of a single VSG gene is what allows the parasite to successfully evade the host immune system. In fact, parasites that express multiple VSG proteins at once are quickly cleared by the immune system (Aresta-Branco et al., 2019b). VSG expression initiates in parasites that reside in the salivary gland of the tsetse fly, prior to bloodstream infection. Recent data has shown that multiple VSG genes are initially transcribed within pre-metacyclic cells, with a single gene being expressed within mature metacyclic cells (Hutchinson et al., 2021). A "race" model has been proposed to explain this phenomenon in which different VSG expression sites race to hit a certain threshold level of transcription. Once a particular gene hits this threshold, the other transcribed expression sites become downregulated, possibly due to the limited expression machinery being used up at this single site, or by the actively transcribed RNA transcripts silencing expression at the other sites (Hutchinson et al., 2021). The particular transcribed VSG and its localization to the nuclear expression site can be inherited following cell division and this inheritance depends on the chromatin assembly factor CAF1 (Faria et al., 2019). Remarkably, simply loosening chromatin structure through ectopic overexpression of the high-mobility group box protein TDP1 is sufficient to allow expression of multiple VSGs per cell (Aresta-Branco et al., 2019b).

While the choice of active VSG can be stable within the life of a cell and through cell division, occasional VSG switching is critical for immune evasion and long-term infection. How VSG switching is regulated—whether this is a probabilistic event or induced by parasite or host factors—remains unknown. The molecular mechanisms that induce VSG exchange are also mysterious. Some possible explanations include collapse of the replication fork due to continuous VSG transcription (Glover et al., 2013), or translocations triggered by frequent DNA damage,

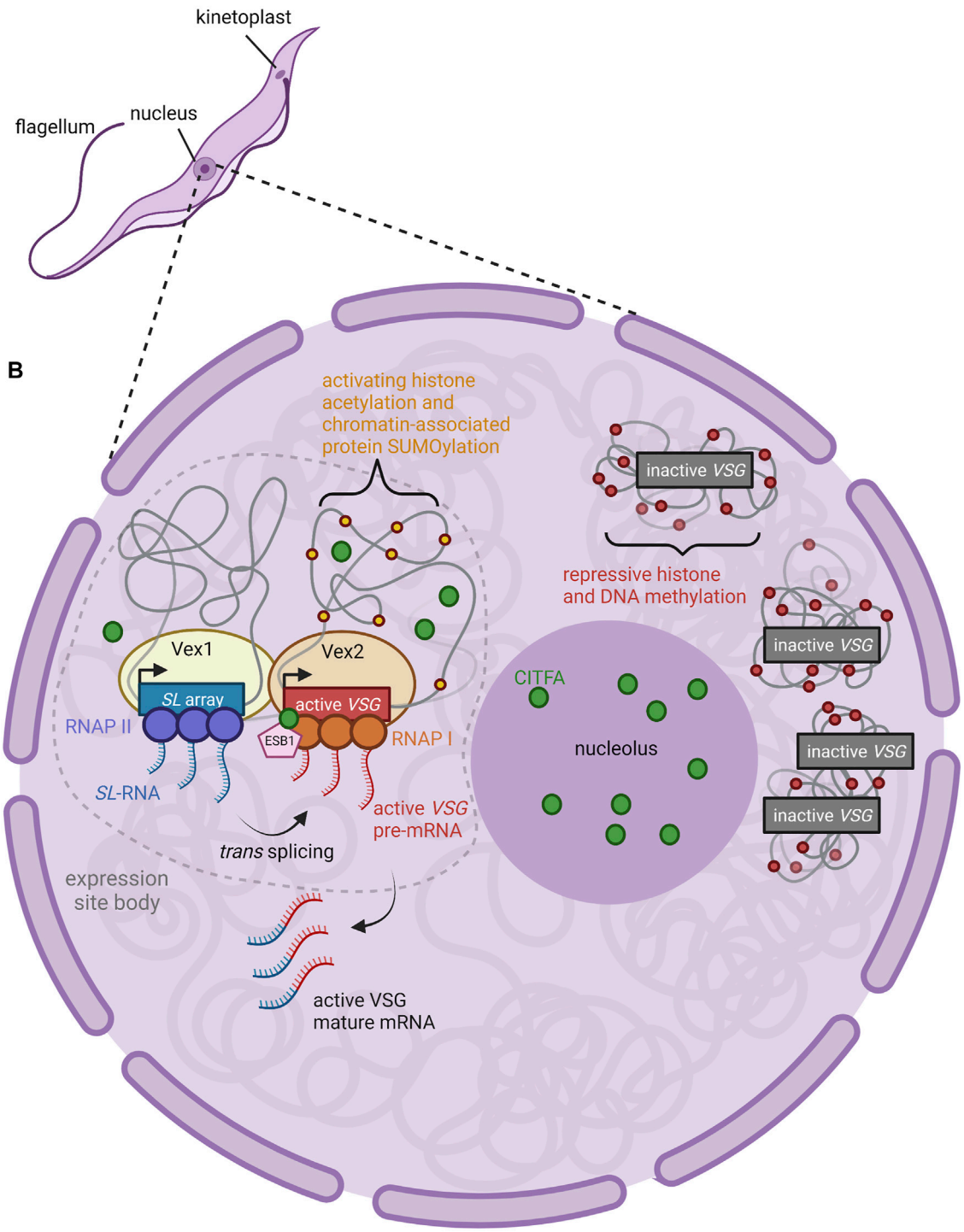


FIGURE 2 | Nuclear Organization and Control of VSG Expression. **(A)** Anatomy of a *T. brucei* cell. The long slender shape is characteristic of the bloodstream form of the parasite, an actively proliferating stage which causes chronic parasitemia in infected hosts (Matthews, 2005). **(B)** The active BES VSG is expressed within an extranucleolar expression site body (ESB) (Navarro and Gull, 2001). ESBs are enriched for RNA Pol I, CITFA transcription factor complexes (green circles), and the ESB-specific transcription factor ESB1 (pink pentagon) (Brandenburg et al., 2007; Nguyen et al., 2012; Nguyen et al., 2014; Escobar et al., 2021). Both CITFA and ESB1 are required for transcription of the active VSG gene; the absence of these factors from inactive VSGs (relegated to heterochromatic regions outside of the ESB) contributes to the repression of inactive VSGs (Brandenburg et al., 2007; Nguyen et al., 2012; Nguyen et al., 2014; Escobar et al., 2021). Within the ESB, the proteins

(*Continued*)

FIGURE 2 | VEX1 (yellow oval) and VEX2 (light orange oval) complex together, associating an mRNA splicing locus (SL array, blue rectangle; RNAP II, purple circles) with the active VSG expression locus (active VSG gene, red rectangle; RNAP I, orange circles) (Glover et al., 2016; Faria et al., 2021). Thus, transcription of the splice leader (SL) RNA by RNAP II proceeds adjacent to transcription of the active VSG pre-mRNA by RNAP I. The 5′ end of the SL RNA is then spliced in trans to the VSG pre-mRNA to form mature VSG mRNA (Faria et al., 2021). Figure inspiration was drawn from various sources (López-Farfán et al., 2014; Martínez-Calvillo et al., 2019; Nguyen et al., 2014; Escobar et al., 2021; Faria et al., 2021).

such as double stranded breaks, within unstable regions surrounding expression sites (Boothroyd et al., 2009). However, loss of RECQ2, a helicase which repairs DNA breaks within the telomere, leads to an increase in DNA recombination, indicating a possibility that double stranded breaks are not responsible for inducing VSG switching (Devlin et al., 2016). In contrast, VSG transcription and DNA replication have been shown to be associated with one another (Devlin et al., 2016). Thus, VSG switching could be induced by DNA fragility brought about by DNA replication (Devlin et al., 2016).

Telomere length, telomere stability, and the regulation of the chromatin structure surrounding VSG expression sites has also been shown to be important for VSG switching (Hovel-Miner et al., 2012; Aresta-Branco et al., 2016). The degree to which this VSG expression choice is stochastic versus deterministic has also come into question with studies revealing a degree of predictability in VSG emergence, which could result from either ordering of VSG choice or from differential selection (Morrison et al., 2005; Mugnier et al., 2015).

Expression Site Activation and In Situ Switching

While only one BES at a time produces an actively translated product, the active BES can switch between the repertoire of available BES through *in situ* switching (**Figure 1C**). Several factors have been identified as characteristic features of the active BES that must be altered in order for *in situ* switching to occur (Cestari and Stuart, 2018). Reminiscent of the importance of nuclear organization in *OR* gene selection (described below), active BES are localized to an extranucleolar region termed the expression site body (**Figure 2**) (Navarro and Gull, 2001). The expression site body and active BES promoter are enriched for Pol I along with the *basal class I transcription factor A* (CITFA) complex (Brandenburg et al., 2007; Nguyen et al., 2012; Nguyen et al., 2014). The novel transcription regulator NLP similarly associates selectively with the active BES (Narayanan et al., 2011).

What mechanisms ensure that only a single BES can produce functional VSG in each cell? Work in the last 5 years has focused on two factors identified in genetic screens that are required for VSG expression: *VSG exclusion 1* (*VEX1*) and *VSG exclusion 2* (*VEX2*). VEX1 has been shown to positively regulate the active VSG site in *cis* while also negatively regulating all other VSGs in *trans* (Glover et al., 2016). VEX1 binds to VEX2 independently of transcription, and together, they are responsible for VSG exclusion (Faria et al., 2019). Recent work suggests that the single VSG chosen for expression in a particular cell is physically associated with the genomic locus that encodes mRNA splice-leader sequences (Faria et al., 2021). In *T. brucei*, mature mRNA is produced by *trans*-splicing to leader sequences produced from this locus. The active BES can associate with the splice-leader locus in *trans*, across chromosomes (Faria et al., 2021). VEX1 associates with the splice leader locus, and VEX2 with the actively expressed VSG in the BES.

Thus, VSG transcription and mRNA splicing take place within a specific compartment of the nucleus and are associated closely with VEX1 and VEX2 proteins (**Figure 2**) (Faria et al., 2021). The coalescence of the VEX1-bound splice leader locus and the VEX2-bound BES into the expression site body may serve to activate Pol I transcription and subsequently induce repression of the remaining BESs (Glover et al., 2016; Schulz and Papavasiliou, 2016; Cestari and Stuart, 2018; Faria et al., 2021). Transcription-mediated gene silencing is a hallmark of heterochromatinization in other stochastic systems such as yeast mating-type switching, and could be involved in olfactory receptor heterochromatinization as well (Allshire and Madhani, 2018). Another possibility is that failure to concentrate access to the splice leader cassette and transcription factors on one VSG gene allows multiple VSGs to be expressed at lower levels.

Chromatin modifications also appear to play an important role in BES activation. The active BES is significantly depleted of histones, especially H3, in comparison to other, silent BESs (Stanne and Rudenko, 2010). In support of this, the knockdown of H1, H3, H3.V, and H4.V increased the accessibility and transcription at previously inactive BES promoters and *VSG* genes (Povelones et al., 2012; Reynolds et al., 2016; Schulz et al., 2016; Müller et al., 2018). This effect has been shown to be mediated by chromatin remodeling proteins such as ASF1A, CAF-1b, and SIR2rp1, alongside a handful of histone methyltransferases and acetylases/deacetylases (Alsford et al., 2007; Figueiredo et al., 2008; Kawahara et al., 2008; Wang et al., 2010; Alsford and Horn, 2012). SUMOylation also plays a crucial role, with SUMOylated chromatin-associated proteins serving as a distinct marker of the active BES in the expression site body (López-Farfán et al., 2014). The *VSG* transcriptional activator SNF2PH is recruited to SUMOylation-rich BES where it is itself SUMOylated to subsequently facilitate Pol I transcription (Saura et al., 2019). Pol I is then further regulated by activating SUMOylation via TbSIZ1/PIAS1 (López-Farfán et al., 2014). All of these SUMOylation events appear to be localized to the active BES.

While nuclear localization, transcription factor recruitment, and chromatin modification have all been shown to be relevant in *VSG* expression at the selected BES, the order and dominance of these activating events is still unclear. It is also still uncertain

what induces these activation signals to switch between BESs during *in situ* switching.

Duplicative Gene Conversion

Early genetic experiments revealed that VSG switching can involve the overwriting of genomic loci (Horn, 2014). A form of recombination, duplicative gene conversion involves the removal of the active *VSG* gene from the expression site, which is replaced by a duplicated form of a previously silent *VSG* gene (**Figure 1C**) (Liu et al., 1983; Robinson et al., 1999). Boothroyd et al. found that gene conversion switches are initiated by DNA double-strand breaks which are subsequently repaired by homologous recombination (Boothroyd et al., 2009). Each *VSG* gene possesses an upstream region of 70 bp repeats. Double-strand breaks adjacent to these repeats were both necessary and sufficient to induce VSG switching, suggesting that the repeats serve as a guide for homologous recombination that allows for the active *VSG* site to be overwritten (Boothroyd et al., 2009). The BESs, along with many of the silent *VSG* arrays, are located in sub-telomeric regions of the genome (Cross et al., 2014). These regions are inherently unstable portions of the genome where recombination and double-strand breaks frequently occur (Glover et al., 2013; Horn, 2014). As such, it is suspected that *T. brucei* takes advantage of this natural instability to induce *VSG* gene conversion; however, alternative mechanisms for DNA lesion production have been proposed (reviewed in da Silva et al., 2018). *VSG* recombination requires RAD51 and BRCA2, while TOPO3α has been shown to suppress recombination and restrict it to the 70bp repeats in partnership with RMI1 (McCulloch and Barry, 1999; Hartley and McCulloch, 2008; Kim and Cross, 2010; Kim and Cross, 2011).

Mosaic VSGs

While trypanosomes predominantly switch between existing intact *VSG* genes at the beginning of an infection, long read sequencing has confirmed that there is a significant increase in the number of novel mosaic *VSG* genes as infection time increases (Jayaraman et al., 2019; Mugnier et al., 2015). Because 80% of *T. brucei's* ~ 2000 *VSG* genes are incomplete or pseudogenes, the repertoire of complete genes is eventually exhausted during chronic infections (Berriman et al., 2005; Cross et al., 2014). Once the majority of complete genes have been expressed and recognized by the immune system, *T. brucei* utilizes segmental gene conversion to merge fragments of different *VSG* genes in what is termed mosaic recombination (**Figure 1B**) (Mugnier at el., 2015). This process is not uncommon, as other pathogens have been known to utilize segmental gene conversion to further diversify their pool of variant genes (Zhuang et al., 2007). Trypanosomes are able to construct functional mosaic *VSG* genes from pseudogenes and gene fragments, suggesting that *T. brucei's* large abundance of partial genes are important for continued diversification (Hall et al., 2013). It is still unclear what cellular process is used to merge the *VSG* segments together. One possibility is that mosaics are generated by homologous recombination within the *VSG* gene similar to duplicative recombination or by crossover events, as in telomeric exchange. It is also unknown whether mosaic formation occurs within expression sites, or if instead they are formed somewhere else in the genome before being moved into the expression site.

Sleeping sickness remains a deadly and difficult to treat disease, so increasing our understanding of the mechanisms that allow this parasite to evade host immune systems will provide advances in our ability to fight *T. brucei* infections. A more detailed analysis of remaining questions in the field is reviewed by McCulloch and colleagues (McCulloch et al., 2017). Similar methods of variation utilized by trypanosomes can also be found in the systems they are meant to evade. Just as antigen diversity aids pathogens in evading the immune system, antigen receptor diversity allows for greater detection. B and T cells in the immune system create this diversity through stochastic genome editing. This process is often initiated by the introduction and subsequent repair of DNA double-stranded breaks, similar to gene conversion in *VSG*'s (Papavasiliou and Schatz, 2000). It has also been suggested that B and T cells expand their receptor diversity via segmental gene conversion, similar to mosaic *VSG*s (Barbet and Kamper, 1993).

ANTIGEN RECEPTOR DIVERSITY

The coexistence of host and pathogen has largely driven the diversification of both the host's immune surveillance and the pathogen's antigenic determinants (Chang et al., 2011). Mammalian genomes contain approximately 20,000 protein-coding genes, and yet the B and T cells of the adaptive immune system produce receptors that can bind to a vast array of arbitrary antigens regardless of evolutionary experience. Receptor-level diversity was ultimately shown to be produced via two stochastic processes: V(D)J recombination, which alone can generate 10^{11} possible binding domains, and somatic hypermutation, which can introduce mutations in any of these recombination products to further expand receptor possibilities (Janeway et al., 2001). These processes allow for a truly outstanding level of diversity to emerge from just a few germline-level genes, preparing the immune system for any antigen it might face without taking up very much genomic space. Much as learning mechanisms in the nervous system allow animals to relate arbitrary sensory stimuli to the contexts in which they are experienced, selective processes during B and T cell development in the context of an immune response shape cellular immune responses according to the "meaning" of self, benign, or pernicious antigens.

Antibodies, which are immunoglobulin proteins, possess variable binding surfaces that can recognize diverse antigens. These antibodies can be secreted in the serum or bound to the surface of B lymphocytes to form B cell receptors (BCRs). T lymphocytes also have surface receptors (TCRs) that recognize antigens in combination with antigen presenting

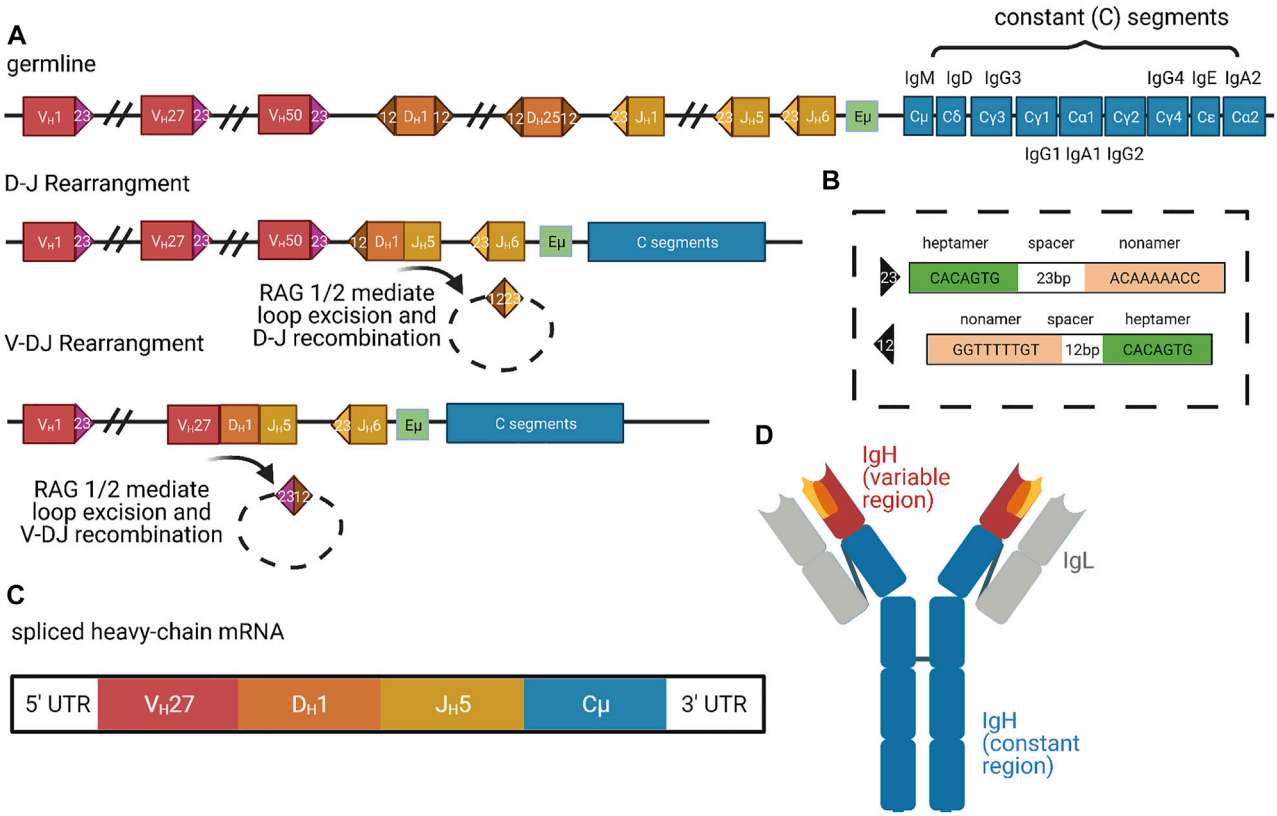


FIGURE 3 | Organization and Expression of Immunoglobulins. **(A)** Successive organization of the human heavy-chain locus in the germline, after D-J rearrangement, and after V-DJ rearrangement. The germline locus contains roughly 50 V segments (red), 25 D segments (orange), 6 J segments (yellow), and 9 constant segments (blue) (Rodriguez et al., 2020). In the first step of D-J and V-DJ rearrangement, RAG1 and RAG2 complexes bind RSS motifs (colored triangles) (Schatz et al., 1989; Oettinger et al., 1990). **(B)** The RSS motifs consist of consensus heptamer and nonamer sequences, which are separated by 23 or 12 bp spacers (Ferrier, 2009). **(C)** In the rearranged DNA locus (last line of panel **(A)**), transcription only proceeds from the promoter of the most-proximal V segment because it competes most effectively for access to the limiting enhancer Eμ and other enhancers (green in **(A)**) (Roy et al., 2011). Additionally, intervening J segments are spliced out and alternative splicing selects either Cμ or Cδ constant segments for inclusion, generating IgM or IgD class BCRs, respectively. Upon activation, B cells can switch expression to different downstream C_H genes via additional recombination, called class switching. Thus, mature heavy-chain mRNA consists of just 1 V, D, J, and C exon (Charles A Janeway et al., 2001). **(D)** Mature IgM protein, color-coded according to contributing heavy-chain exons. Gray regions correspond to light-chain segments, which are combined with heavy chains post-translationally. Figure inspiration was drawn from various sources (Little et al., 2015; Feederle and Schepers, 2017; Backhaus, 2018).

major histocompatibility complex (MHC) proteins. Like TCRs and BCRs, MHC proteins are present in the population as diverse alleles. While TCR and BCR diversity is generated via somatic mechanisms, population-level MHC diversity is maintained at the germline level via balancing selection.

Immunoglobulin Gene Loci and V(D)J Recombination

Immunoglobulins are composed of covalently-linked heavy and light chains, both of which possess a variable N-terminus that recognizes antigens and a constant C-terminus that can recruit effectors (Schroeder and Cavacini, 2010). Here, we will focus on the generation of BCRs and antibodies as a model for immunoglobulin diversification. The germline-encoded heavy chain locus produces IgM and IgD constant regions via alternative splicing; both IgM and IgD can be membrane-bound or secreted as antibodies. DNA rearrangements of the heavy chain locus during the course of the lymphatic germinal center reaction can also produce secreted IgG, IgA, and IgE antibodies by joining the variable N-terminus to different constant regions; we will focus on variable region diversification. V(D)J recombination during lymphocyte development produces the initial diversity of mature IgM and IgD. Once B cells bind their given antigen, somatic hypermutation in the germinal centers allows for further diversification of the variable region alongside the production of IgG, IgA, and IgE through class switching.

In V(D)J recombination, the N termini of the heavy and light chains are rearranged to bring distinct V (variable) segments, followed by the D (diversity) segments, and then the J (joining) segments into proximity with the constant region (**Figure 3**) (Schroeder and Cavacini, 2010). The human heavy chain locus on

chromosome 14 possesses roughly 50 functional V segments, 27 functional D segments, and six functional J segments (Rodriguez et al., 2020). The two major classes of light chains are kappa and lambda, both of which do not possess D segments but still undergo VJ recombination. The kappa locus is on chromosome 2 with roughly 44 functional V segments and 5 J segments, whereas the lambda locus is on chromosome 22 with roughly 37 functional V segments and only 1 J segment (Collins and Watson, 2018; Watson et al., 2015). The ability to create combinatorial V(D)J regions allows for an incredible diversity of heavy and light chains, which are both combined to further expand the possibilities for mature immunoglobulin proteins. There are roughly 3.5×10^6 potential combinatorial products that can arise from these V(D)J and heavy-light chain pairings, and final protein products are additionally varied by junctional diversification that occurs during recombination. Recombination is induced by RAG1/2 (recombination-activating gene) which target discrete locations within the immunoglobulin loci through conserved and repeated DNA sequence elements (**Figure 3A**) (Schatz et al., 1989; Oettinger et al., 1990). As the V(D)J recombination process is "settled science," we refer readers to other reviews or textbooks for more detailed description.

Remarkably, diversification of antigen binding repertoires through alteration of germline DNA has evolved more than once. In the lamprey, leucine rich repeat (LRR) proteins are diversified during lymphocyte maturation via insertion of LRR modules from flanking regions of the locus (reviewed in Boehm et al., 2012). Random combinatorial usage of immunoglobulin modules has also arisen in other systems: neuronal self-recognition in insects is mediated by randomized alternative splicing of the *Dscam* immunoglobulin locus.

Enhancer-Mediated Restriction

While V(D)J recombination removes V regions proximal to D or J segments, distal V options remain intact, and each V has its own upstream promoter. Nevertheless, transcription always begins at the V region most proximal to D/J and therefore ensures that only a single V—the most proximal—is included in the transcript (Roy et al., 2011). This selection was initially thought to be performed by a limiting enhancer element located between the V and D/J domains, called Eμ (**Figure 3A**) (Serwe and Sablitzky, 1993; Li and Eckhardt, 2009). More recent work has suggested that several additional enhancers, including 3′RR and DICE, participate in a complex promoter selection process (Bébin et al., 2010; Roy et al., 2011). The most proximal V region promoter that remains after recombination is thought to compete most effectively for looping interactions with the limiting enhancers, thus conferring deterministic use of the most proximal V in the context of stochastic removal of alternate distal Vs. As we will describe below, the process of clustered protocadherin transcription in mammalian neuronal self-recognition also involves competition among nearly identical promoters for access to a single enhancer, and expression variability is produced by suppression of spatial bias for the proximal promoter, rather than recombination of different segment choices into proximity with the enhancer.

Monoallelic Expression

Similar to the selective expression of a single allele in the OR, PCDH, and VSG systems, there is extensive evidence that each B cell expresses only a single BCR, which makes each B cell specific for one particular antigen (Weiler, 1965; Vettermann and Schlissel, 2010). This specificity is important for subsequent clonal selection of antibody-producing cells and proper immune response. Interestingly, the *Ig* alleles are transcribed biallelically early in B cell development, indicating that transcriptional activation alone does not govern the allelic exclusion of these loci (Singh et al., 2003). At the level of translation, transcripts from *Ig* genes that have not undergone proper V(D)J recombination possess premature stop codons that prevent production of functional protein (Bühler et al., 2004; Eberle et al., 2009). Moreover, B cells co-opt the unfolded protein response to trigger differentiation in response to BCR translation (Hetz et al., 2020). A similar process links olfactory receptor choice, stabilization of singular olfactory receptor translation, and olfactory neuron differentiation (Dalton et al., 2013).

Studies have suggested that complete V(D)J recombination of one allele induces the suppression of the second non-recombined allele, preventing subsequent recombination and productive transcription (Vettermann and Schlissel, 2010). This is supported by the observation that D-J recombination occurs in both IgH alleles, yet only one productive V-DJ recombination proceeds (Jung et al., 2006). To accomplish this, the active recombination of the locus appears to induce RAG- and ATM-mediated repositioning of the inactive allele to repressive heterochromatin alongside inducing locus decontraction that has been associated with recombination inhibition (Goldmit et al., 2005; Hewitt et al., 2009). Furthermore, the production of a complete immunoglobulin protein chain then induces progression of B cell development that subsequently downregulates RAG proteins to prevent further recombination (Grawunder et al., 1995; Galler et al., 2004). This model allows for developing B cells to make multiple attempts at performing proper recombination, as complete suppression of the alternative allele does not occur until one of the alleles has produced protein. However, in order for this process to produce a monoallelic product the induction of recombination must be asynchronous.

Early models suggested that the low rate of recombination allowed for a probabilistic first-come, first-serve mechanism where allelic selection was purely based on which allele happened to recombine first (Perry et al., 1980; Liang et al., 2004), but continued studies have revealed that the process is likely more controlled than this. It has been shown that the selected allele is replicated first and localized to the euchromatic nuclear center whereas the non-selected allele is found in the repressive heterochromatin of the nuclear periphery (Mostoslavsky et al., 2001; Skok et al., 2001). The active allele is subsequently found to have activation signatures: hypomethylation of CpG dinucleotides; hypersensitivity to DNA nucleases and restriction enzymes; and increased activating histone marks, including histone H3/H4 acetylation (Outters et al., 2015). These differences lead the two alleles to be

differentially available for RAG binding and recombination (Vettermann and Schlissel, 2010). The order and significance of these influences is still unclear, and the initial mechanism that dictates the selected allele remains debated. A detailed discussion of competing models can be found here (Vettermann and Schlissel, 2010; Outters et al., 2015).

Positive Selection of B Lymphocytes

The vast pool of antigen receptors allows for modest binding and subsequent detection of most antigens, but once lymphocytes are activated they undergo numerous rounds of selection to increase their affinity for the antigen. We will focus on the positive selection of antigen-selective B lymphocytes in the germinal center of lymph nodes. Of course, stochastic production of TCRs and BCRs can also lead to dangerous autoimmune reactions; these are minimized due to distinct processes of negative selection that occur during lymphocyte development (reviewed in Klein et al., 2014; Nemazee, 2017; Rose, 2017).

Somatic hypermutation (SHM), a key process in affinity maturation, functions to diversify BCRs and promote the adaptive immune response. During SHM, the *BCR* locus undergoes a significant increase in the rate of point mutations compared to the rest of the genome (Forrest and Oprea, 1999). These mutation "hotspots" usually encode the complementarity-determining regions in the variable N-terminus of the antibody that interact with and recognize the antigen. SHM occurs when the enzyme activation induced deaminase (AID) targets mature rearranged V(D)J and switch regions of *Ig* genes (Pilzecker and Jacobs, 2019). AID functions by binding to single-stranded DNA and removing the amino group from cytosine, which produces highly mutagenic deoxy-uracil in the DNA of both *Ig* strands at a high rate. DNA damage response processes then generate base substitutions at and around the lesion created by the deoxy-uracil (Pilzecker and Jacobs, 2019).

Lymph node germinal centers (GCs) are the site of B lymphocyte clonal selection that drives affinity maturation to produce memory B cells and antibody-secreting plasma cells (Victora and Mesin, 2014). The GC is separated into a dark and light zone. B cells undergo SHM while proliferating in the dark zone (McKean et al., 1984; Victora and Mesin, 2014). This generates a diverse clonal pool that migrates to the light zone for selection. In the light zone, B cells use their antigen receptors to retrieve antigen from the surface of follicular dendritic cells (FDCs) and then present this antigen to T follicular helper (Tfh) cells to receive survival signals. Tfh cells were found to be the limiting factor in GC selection, as they can only interact with a small portion of the B cells (Victora et al., 2010). This creates competition between B cell clones to retrieve antigen from FDCs and present it to Tfh cells, with higher affinity BCRs being able to present more antigen and receive the limited Tfh support (Victora et al., 2010). Tfh cells send support signals in the form of cytokines and cell surface receptors like CD40L, IL-21, and IL-4 to allow B cell survival and migration into the dark zone for further proliferation and SHM (Crotty, 2014). Post-transcriptional regulation of the chemokine CXCL12 receptor CXCR4, along with differential expression of polycomb proteins, have been shown to mediate zonal migration and polarization (Allen et al., 2004; Okada et al., 2005; Allen et al., 2007a; Allen et al., 2007b; Victora et al., 2010).

Multiple rounds of this selective process produces a robust pool of antibodies that have significantly improved affinity for the antigen. Though selection of B cells in the GC starts out from mostly *inter*clonal competition, competition eventually progresses to *intra*clonal competition between variants generated by SHM (Jacob et al., 1993). Some GCs will show clonal dominance of high affinity lineages, but this dominance is not required for high affinity clones to emerge and is not present in all GCs (Tas et al., 2016).

DOWN SYNDROME CELL ADHESION MOLECULES

Neural circuit wiring is an extremely important process that is highly dependent on the proper patterning of neurons within the developing nervous system. While neurons positively select their partners through recognition of deterministically expressed cell surface molecules, neurons also have to avoid synapsing with themselves in order to establish their typical anatomies and heterologous partners. This process, called neuronal self-avoidance, requires neurons to distinguish "self" from "nonself."

In both vertebrates and insects, neuronal self identity is determined by randomized expression of subsets of possible cell surface molecules. These expression patterns are distinct across individual neurons, even neurons of the same type, and can be thought of as a unique barcode displayed on the surface of each individual cell. In *Drosophila*, the protein family used for this purpose is the Dscam (Down syndrome cell adhesion molecule) family of immunoglobulins. Via alternative splicing, the *Drosophila Dscam1* locus encodes up to 38,016 distinct Dscam isoforms, all of which contain the same basic structure: an ectodomain comprised of 10 immunoglobulin (Ig) domains and six fibronectin type III repeats, a transmembrane domain, and a C-terminal cytoplasmic tail (**Figure 4A**) (Schmucker et al., 2000). Four variable domains are encoded by clusters of exon variants, which are spliced independently of each other: *Ig2* (12 variants), *Ig3* (48 variants), *Ig7* (33 variants), and the transmembrane domains (2 variants). This means that for 38,016 distinct isoforms there are potentially ($12 \times 48 \times 33 = 19{,}008$) distinct ectodomains; at least 18,048 of these ectodomains are confirmed to support "homophilic" binding, or binding between identical isoforms (Wojtowicz et al., 2007).

Homophilic binding of two Dscams generates a repulsive response. When coupled with the immense diversity of Dscam isoforms, which makes it unlikely that neighboring neurons will express identical sets of isoforms, it becomes clear how Dscams mediate neuronal self-avoidance: neurites within the same neurons will express the same Dscams and repel each other, while neurites between neighboring neurons will express different Dscams and allow synapsing. The power of this "barcoding system" is evident from mutation and ablation studies: Where *Dscam1* function is disrupted, dramatic defects in self-recognition are observed, including increases in intraneuronal dendritic

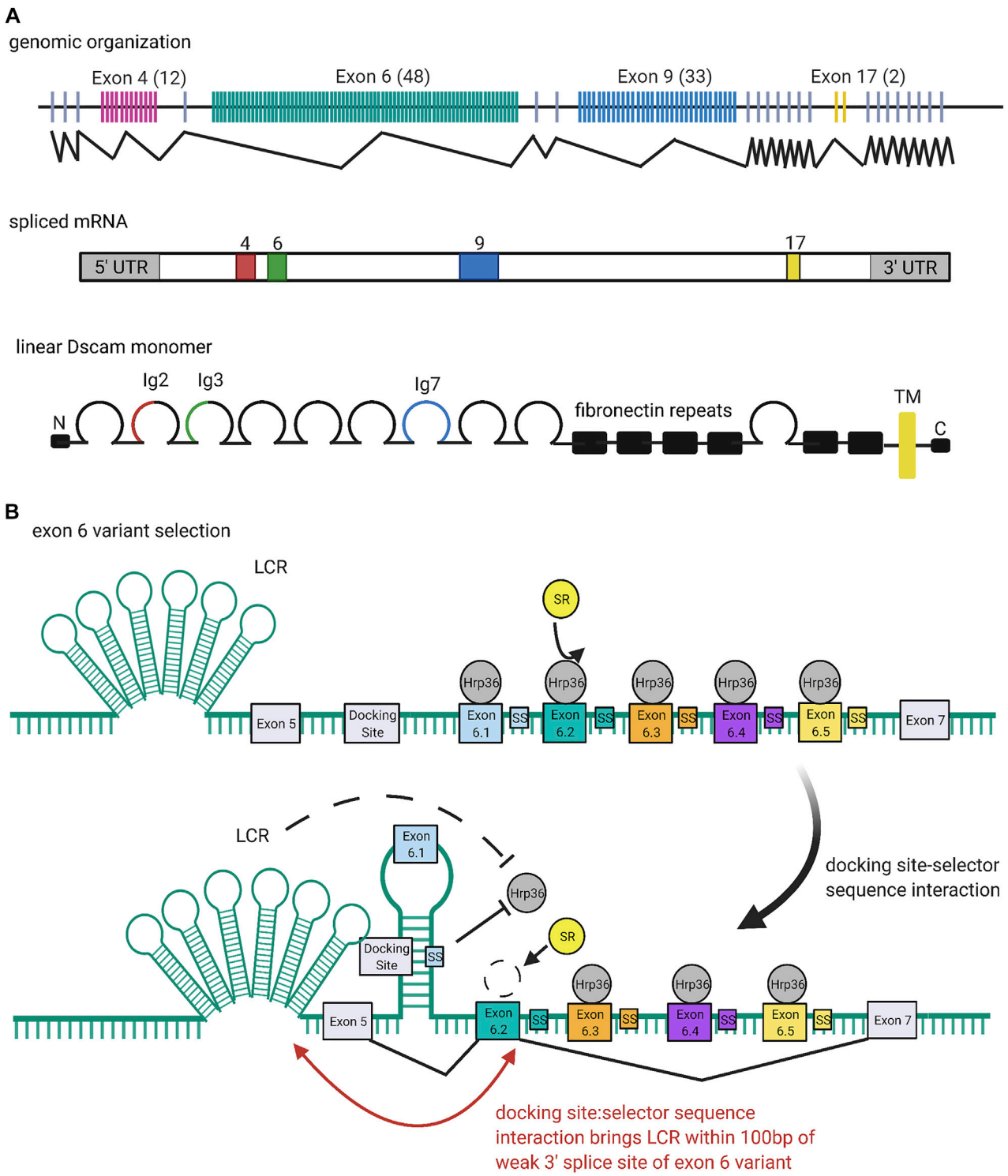


FIGURE 4 | Organization and Expression of Dscams. **(A)** Genomic organization of the *Dscam1* locus in *D. melanogaster*. Numbers of exon variants are listed in parentheses next to the exon cluster number. Alternative splicing at exon clusters 4, 6, 9, and 17 (black zigzagging lines) proceeds independently, such that the *Dscam1* locus can combinatorially encode up to 38,016 unique isoforms (Schmucker et al., 2000). Exon 4 (red) and exon 6 (green) variants encode the N-terminal halves of the Ig2 and Ig3 domains, respectively, while exon 9 (blue) variants encode the entire Ig7 domain (Schmucker et al., 2000; Zhan et al., 2004). Exon 17 (yellow) codes for one of two alternative transmembrane domains, which appear to play a role in the subcellular localization of Dscams (Wang J. et al., 2004; Yang et al., 2012; Liu et al., 2020). **(B)** Mechanism of exon six variant exon selection; for simplicity, only the first 5 exon six variants and selector sequences (colored boxes) are shown, flanked by exons 5 and 7 and the common upstream docking site (grey rectangles). The exon 6 cluster of the *Dscam1* locus is maintained in a globally repressed state by binding of Hrp36 (*Continued*)

FIGURE 4 | proteins (dark grey circles) to each variant (Olson et al., 2007). Binding of a variant selector sequence to the docking site forms an RNA hairpin loop which prevents inclusion of variants contained within the loop but promotes inclusion of the variant immediately downstream (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011; Hemani and Soller, 2012; Xu et al., 2019). This interaction also brings the locus control region (LCR) within 100 bp of the weak 3′ splice site of the downstream variant, which may also promote variant inclusion by facilitating recognition of the splice site (Wang X. et al., 2012). Both the docking-site selector sequence interaction and LCR are thought to promote variant inclusion by antagonizing binding of the repressive Hrp36 proteins, which allows binding of inclusion-promoting SR proteins (yellow circles) at the selected variant (Olson et al., 2007; Wang S.-Z. et al., 2012). Figure inspiration was drawn from various sources (Schmucker et al., 2000; Wang X. et al., 2012).

crossing in dendritic arborization (da) neurons (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007) and failure of sister branch segregation in the axons of mushroom body (MB) neurons (Wang et al., 2002; Hattori et al., 2007). In addition to self-recognition, Dscams have also been suggested to mediate synaptic target selection and axon guidance in several kinds of neurons (Wang et al., 2002; Hummel et al., 2003; Zhan et al., 2004; Zhu et al., 2006; Millard et al., 2010).

Structure and Function of Dscam Homophilic Binding

As discussed, binding specificity is critical to Dscams function in neuronal self-avoidance (Neves et al., 2004; Zipursky et al., 2006). Indeed, both *in vitro* and *in vivo* studies have demonstrated highly specific homophilic binding, to the point that isoforms differing in just a few residues exhibit very weak or no binding (Wojtowicz et al., 2004; Zipursky and Sanes, 2010). How is this exquisite specificity determined? Furthermore, how does attractive homophilic binding generate a repulsive response? Briefly, ELISA binding assays have determined that the 8 N-terminal domains (Ig1-Ig8) of Dscam proteins are sufficient to support normal binding (**Figure 5A**) (Wojtowicz et al., 2004). Contained in this sequence are the Ig2, Ig3, and Ig7 variant domains, which determine the binding specificity of isoforms by selectively "matching" with their identical counterparts (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Sawaya et al., 2008). While binding of individual variable domains is modular, binding of whole Dscams is all-or-nothing (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Sawaya et al., 2008). That is, the particular identities of the variable domains do not matter as long as they are the *same* between isoforms, as even a minor mismatch between one pair of variable domains is sufficient to totally disrupt binding (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Sawaya et al., 2008).

Upon homophilic ectodomain binding, the C-terminal cytoplasmic tail of Dscam initiates repulsive signaling in the cell, which eventually leads to repulsion between cells expressing identical isoforms (Matthews et al., 2007). Although the mechanism that promotes this repulsion is still poorly understood, studies have identified a few conspicuous binding partners involved in cytoskeletal rearrangement (Hing et al., 1999; Schmucker et al., 2000; Worby et al., 2001). Among these is the adaptor protein Dock, which binds the Dscam cytoplasmic tail and recruits the effector kinase Pak1, which is implicated in several pathways underlying neurite repulsion (**Figure 5A**) (Hing et al., 1999). However, while the Dock-Pak axis seems to be necessary for Dscam-mediated axon guidance (Schmucker et al., 2000), loss of Dock or Pak1 does not produce dendrite self-crossing phenotypes (Hughes et al., 2007). This indicates that repulsive signaling required for Dscam-mediated self-avoidance can proceed independently of Dock and Pak.

A recent study investigating DSCAMs, the mammalian homologs of fly Dscams, found that both DSCAMs and Dscams share a predicted nuclear localization signal (NLS) and can be cleaved by proteolysis *in vivo* (Sachse et al., 2019). In mammalian DSCAMs, the NLS was found to promote translocation of cleaved DSCAM cytoplasmic fragments into the nucleus, where they affect expression of genes associated with synapse formation (**Figure 5A**) (Sachse et al., 2019). Future research should assess whether this signaling mode occurs in *Drosophila*.

Mutually Exclusive Splicing Generates Isoform Diversity

Similar to their cousins in the immunoglobulin superfamily, Dscams rely on large arrays of diverse variants to function (Hattori et al., 2007). But unlike TCRs and BCRs, Dscam isoform diversity is generated at the RNA level. Schmucker and colleagues were the first to note the wide variety of Dscam isoforms in fruit flies (Schmucker et al., 2000). cDNA and genomic analyses of Bolwig's nerves in *D. melanogaster* embryos revealed alternative sequences for the extracellular *Ig* domains 2, 3, and 7, as well as the transmembrane domain (Schmucker et al., 2000). The N-terminal half of Ig2 is encoded by variants of exon 4, the N-terminal half of Ig3 is encoded by variants of exon 6, the entire Ig7 domain is encoded by variants of exon 9, and the entire transmembrane domain is encoded by variants of exon 17 (Schmucker et al., 2000; Zhan et al., 2004). Exon clusters 4, 6, 9, and 17 were found to have 12, 48, 33, and 2 exon variants, respectively. Each variant is spliced in a mutually exclusive manner, such that each *Dscam* cDNA sequence only contains one of each variable exon 4, 6, 9, and 17 (**Figure 4A**) (Schmucker et al., 2000). Further, splicing of different exon clusters proceeds independently, which is why the *Dscam1* locus can combinatorially encode up to 38,016 unique isoforms (Schmucker et al., 2000). Combined with the fact that individual neurons express several isoforms simultaneously, *Dscam1* turns out to be a powerful system for uniquely "barcoding" cells (Celotto and Graveley, 2001; Neves et al., 2004; Zhan et al., 2004).

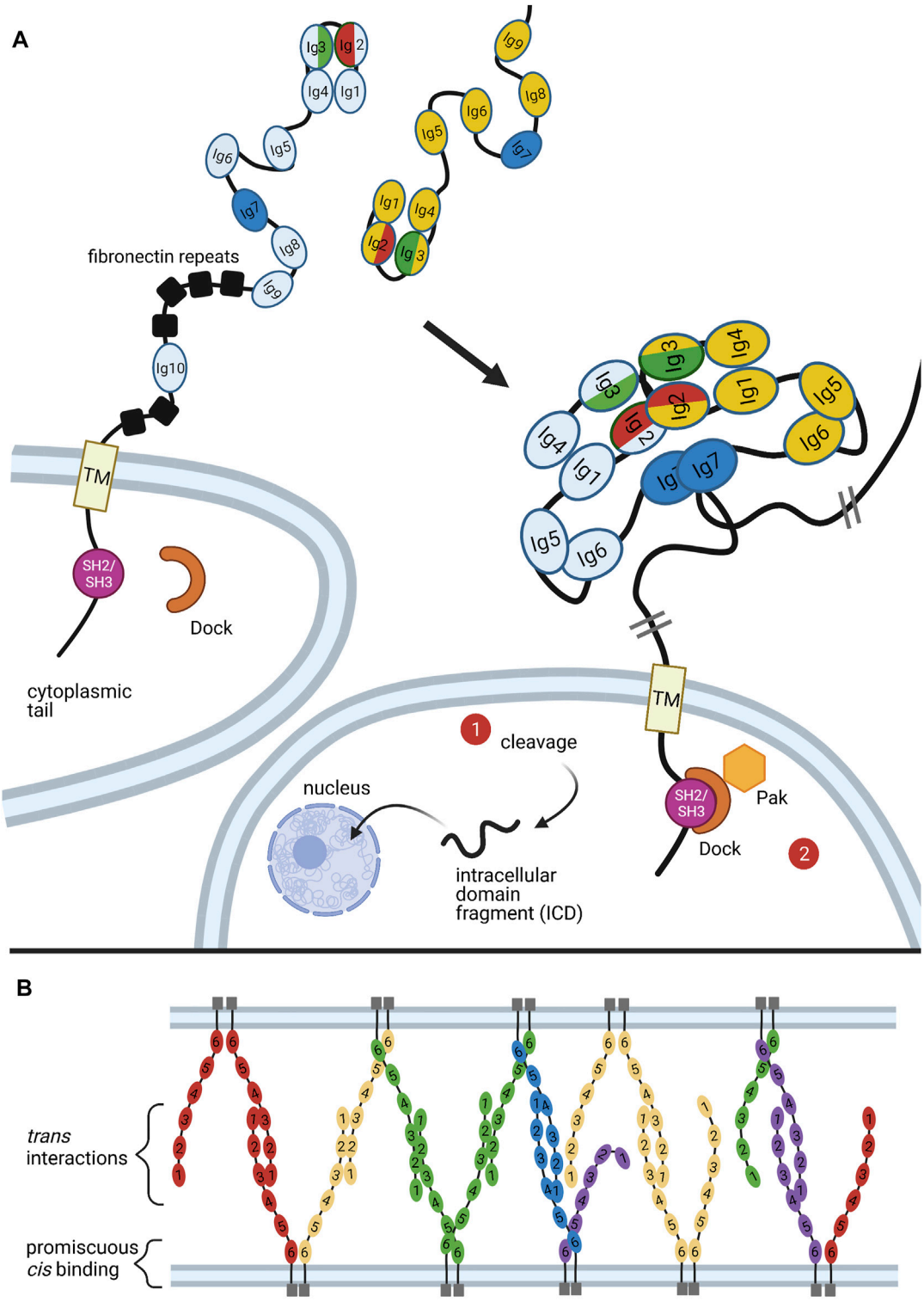


FIGURE 5 | Binding of Dscams and Protocadherins. **(A)** Homophilic binding of Dscam domains Ig1-Ig8 produces an S-shaped dimer (Meijers et al., 2007; Sawaya et al., 2008). The variant domains Ig2, Ig3, and Ig7 confer binding specificity, with the variant halves of Ig2 (red) and Ig3 (green) forming a composite binding interface and Ig7 (blue) binding independently (Sawaya et al., 2008). Upon homophilic binding, a repulsive signal is generated (Matthews et al., 2007). While it is still unclear what mediates this signal, two possible pathways are illustrated. First, studies in mammalian DSCAMs revealed a nuclear localization signal (NLS) in the cytoplasmic tail. Upon cleavage of the tail, the NLS promotes translocation of the tail fragment into the nucleus, where it affects expression of synapsing genes. Although similar cleavage and nuclear translocation has not yet been demonstrated in flies, fly Dscams do have a conserved NLS in the cytoplasmic tail (Sachse et al., 2019). Second, the adaptor

(*Continued*)

FIGURE 5 | protein Dock has been shown to bind the SH2/SH3 domains of the Dscam cytoplasmic tail and recruit the effector kinase Pak1 (Hing et al., 1999). But while Pak1 has been shown to mediate axon guidance, it does not seem to be necessary for neuronal self-avoidance (Hughes et al., 2007). **(B)** Similar to Dscams, protocadherins bind homophilically. A mixture of *trans* and *cis* interactions forms a zipper-like lattice spanning neighboring cell membranes (Brasch et al., 2019). Figure inspiration was drawn from various sources (Sawaya et al., 2008; Schmucker and Chen, 2009; Guo et al., 2012; Chen and Maniatis, 2013; Goodman et al., 2017).

Variant Exon Inclusion in Alternative Splicing Is Probabilistic

Based on data from *D. melanogaster* exon 4 splicing reporter lines, Miura and colleagues proposed that this mutually exclusive alternative splicing proceeds in a probabilistic fashion (Miura et al., 2013). Within class IV da neurons, they observed that all 12 variants of exon 4 had different yet stable probabilities of inclusion (Miura et al., 2013). In contrast, the probabilities of variant inclusion differed between different classes of neurons. For example, exon 4.2 was expressed in more than half of class IV da neurons, but negligibly expressed in Kenyon cells (Miura et al., 2013). Further, in comparing class IV da neurons in late second and wandering third instar larval stage, Miura and colleagues found that the inclusion frequencies of exon 4 variants changed over time (Miura et al., 2013). Their findings coincide with other studies which demonstrate that *Dscam1* exon selection is biased by developmental stage, tissue type, and even by neuronal subtype (Celotto and Graveley, 2001; Neves et al., 2004; Zhan et al., 2004).

Aside from the general observation that alternative splicing of exon clusters is probabilistic, the mechanism underlying exon variant selection remains to be elucidated; that is, it is still unclear how exon variants are specifically, exclusively, and stochastically selected for inclusion during splicing (Hemani and Soller, 2012). The picture is complicated by the fact that different exon clusters in the *Dscam1* locus seem to employ different methods for mutually exclusive alternative splicing (Hemani and Soller, 2012). To consider one proposed mechanism in detail, this review will focus on mutually exclusive splicing in the exon 6 cluster of the *D. melanogaster Dscam1* locus. More information on regulation of this or other clusters in *Drosophila* or other organisms can be found elsewhere (Graveley, 2005; Anastassiou et al., 2006; Yang et al., 2011; Wang S.-Z. et al., 2012; Hemani and Soller, 2012; Yue et al., 2016a; Yue et al., 2016b; Haussmann et al., 2019; Ustaoglu et al., 2019; Xu et al., 2019).

RNA Secondary Structures Mediate Inclusion of Single Exon Variants

Similar to the selection mechanisms for *VSG*s and protocadherin exon variants (discussed below), the exon six variants of the *Dscam1* locus are maintained in a repressed state until a selection event specifically activates a variant for expression (**Figure 4B**). In the case of the *Dscam1* locus, the selection event is the formation of RNA secondary structures which antagonize binding of repressive heterogeneous nuclear ribonucleoproteins (hnRNPs) and promote binding of serine-arginine rich (SR) proteins (Olson et al., 2007).

Graveley (2005) first reported conserved sequences within the exon 6 cluster that seem to be required for mutually exclusive selection of exon six variants: a "docking site," located in an intron upstream of the first exon six variant, and a "selector sequence," one of which is located directly upstream of each of the 48 exon six variants (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011). The docking site and each selector sequence are predicted to form an RNA stem-loop structure by base-pairing (Graveley, 2005; May et al., 2011). This stem-loop prevents splicing inclusion of the exon variants contained within the loop but promotes specific inclusion of the exon directly downstream of it (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011; Hemani and Soller, 2012; Xu et al., 2019). Because the selector sequences bind to offset, overlapping portions of the docking site, only one selector sequence is predicted to bind, ensuring that there is only one such stem-loop structure at any given time (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011). In addition to competing docking site-selector sequence interactions, it appears that a locus control region (LCR) is necessary for the activation of exon six variants (Wang X. et al., 2012). The LCR is a large tandem stem-loop RNA structure. In *Drosophila* species it forms a highly conserved "hexaleaf" consisting of ~700 bp of scattered upstream intronic sequences (Wang S.-Z. et al., 2012).

RNAi screens by Graveley and colleagues identified Hrp36 (Hrb87F) as the hnRNP responsible for global repression of the exon 6 cluster (Olson et al., 2007). Hrp36 was shown to bind to the exon 6 cluster and is required to repress the inclusion of extra exon six variants. Further, Hrp36 was shown to inhibit binding of SR proteins, which are known to regulate alternative splicing and promote exon inclusion (Olson et al., 2007). Thus, the current model is that an Hrp36 protein binds at each exon six variant within the cluster, maintaining it in a repressed state until an upstream docking site-selector sequence stem-loop somehow dislodges Hrp36 from the proximal variant. The LCR may also help destabilize Hrp36 binding (Wang X. et al., 2012). Dissociation of the Hrp36 protein allows SR proteins to bind the proximal variant and promote its inclusion in splicing (Graveley, 2005; Olson et al., 2007; Hemani and Soller, 2012; Xu et al., 2019). Separately, it has also been suggested that the LCR promotes recognition of weak splice sites in exon variants. Specifically, upon formation of a docking site-selector sequence stem-loop, the LCR is brought within 100 bp of both splice sites of the proximal variant, allowing it to activate inclusion in a proximity-dependent manner (Wang S.-Z. et al., 2012).

To sum, exon inclusion in the *Dscam1* exon 6 cluster appears to be determined by the binding ability of different selector sequences, which may be modulated by splicing factors and RNA-binding proteins (RBPs) such as SR proteins and hnRNPs (or other, noncanonical RBPs, as in the case of the exon 9 cluster, reported elsewhere (Ustaoglu et al., 2019)). It is possible that *deterministic* regulation of these protein factors, which themselves mediate probabilistic events in splicing, underlies the "stochastic yet biased" expression of different exon variants among different cell types and at different times

(Neves et al., 2004; Zhan et al., 2004; Miura et al., 2013). These protein factors may stably associate with splicing machinery in a complex with chromatin, allowing them to sterically exclude exon variants; this may also explain the fact that individual cells only express a finite number of Dscam isoforms (Miura et al., 2013). Future work should investigate the possibility of active negative feedback mechanisms regulating the number of expressed isoforms.

Experimental evidence and comparative genomic analyses indicate that selection of exons 4, 9, and 17 also relies on competing RNA secondary structures, which may be recognized by distinct but overlapping sets of RBPs. This is discussed further elsewhere (Park et al., 2004; Anastassiou et al., 2006; Olson et al., 2007; Yang et al., 2011; Wang X. et al., 2012; Hemani and Soller, 2012; Yue et al., 2016a; Yue et al., 2016b; Haussmann et al., 2019; Ustaoglu et al., 2019; Xu et al., 2019). There is still much to be understood about the mechanisms regulating exon choice within each cluster. It may also prove fruitful to investigate possible crosstalk between the different exon clusters.

Dscam Diversity Is Required for Proper Neural Patterning

Studies that reduced the number of possible Dscams underline the importance of great isoform variety. Regarding self-avoidance, studies found that reducing the Dscam repertoire to just one isoform produced marked neural circuit defects in MB and olfactory receptor (OR) neurons (Hattori et al., 2007; Matthews et al., 2007; Soba et al., 2007). In another study that reduced the isoform repertoire, it was found that flies with at least 4,752 Dscam isoforms were indistinguishable from wild-type controls, while flies with 1,152 isoforms or less demonstrated substantial self-branching defects in da neurons (Hattori et al., 2009). These branching defects improved as the number of potential isoforms increased, indicating that self-avoidance requires several thousand different isoforms (Hattori et al., 2009). The finding that neurons require diverse Dscams not only to avoid synapsing with themselves but also to perform anatomic work such as axonal branching suggest that the repulsive force of self-avoidance is used to generate neuronal shapes. How the strength of this force is regulated or differentially harnessed in the production of distinct neuronal shapes is of interest in future work.

While a large variety of isoforms is clearly required, it is unclear whether any one isoform is necessary for normal patterning. In particular, studies that serially deleted different exon 4 variants (thereby eliminating particular Ig2 domains) did not produce observable phenotypes in MB or da neurons, indicating that self-avoidance does not require any specific isoform (Wang J. et al., 2004; Hattori et al., 2009). On the other hand, another study reducing diversity to 22,176 isoforms in mechanosensory neurons found defects in axonal branch extension and branching patterns that correlated with particular deletion alleles, suggesting that some connectivity patterns may be mediated by specific isoforms (Chen et al., 2006). It may be that specific isoforms are needed for some types of neural patterning processes, such as axonal targeting and branching, but not for dendritic self-avoidance. If so, this may also reconcile the bias for certain exon variants at certain developmental stages and within specific cell types: different Dscam isoforms may be required for different developmental and patterning processes (Celotto and Graveley, 2001; Zipursky and Sanes, 2010).

Protocadherins as Analogs for Dscams

Vertebrate protocadherins function analogously to invertebrate Dscams in that both systems mediate neuronal self-avoidance (Garrett et al., 2018). As with invertebrate Dscams, diverse sets of protocadherin isoforms are generated and go on to mediate processes required for proper neural circuit wiring in vertebrates (Schreiner and Weiner, 2010). A notable difference between invertebrate Dscams and vertebrate protocadherins is how variation is produced in each system: alternative splicing generates diverse Dscam isoforms, while utilization of alternative promoters generates diverse protocadherin isoforms (Schmucker et al., 2000; Wojtowicz et al., 2004; Jin et al., 2018). Specifically, Dscams rely on splicing and associated proteins, while PCDHs utilize a CTCF/cohesin-mediated DNA looping mechanism to select proper promoters (Schmucker et al., 2000, p. 200; Wojtowicz et al., 2004). Both mechanisms, however, are prime examples of the ability of non-deterministic events to generate great protein diversity. Such diversity is especially helpful in patterning the nervous system, given that each neuron is likely to have several neighbors, each of which needs to have a different "barcode" if it is to form unique and overlapping networks of connections. Further, the binding of variable domains in either type of protein is exquisitely specific, much like the binding of immunoglobulins and T-cell receptors (**Figure 5B**) (Zipursky and Sanes, 2010). It is remarkable that Dscams and protocadherins have such convergent functions, given the significant differences in their phylogeny, morphology, and mechanistic origins.

While the *DSCAM* gene is conserved in mammals, it does not encode diverse isoforms. In some tissues, including the vertebrate eye, DSCAM has been shown to play a deterministic role in synaptic matching (Yamagata and Sanes, 2008). It has also been shown to increase stringency of synaptic partnerships in other areas of the brain (Garrett et al., 2018). However, a recent analysis found that mammalian DSCAM and DSCAML1, along with other members of the basigin-related family, are sometimes expressed in a mutually exclusive pattern (Iakovlev et al., 2021). It is therefore possible that *DSCAM* genes also contribute to non-deterministic aspects of cell identity in mammals.

PROTOCADHERINS

The lack of Dscam diversity in mammals prompted a hunt for protein families performing barcoding or self-avoidance roles in mammals. Work over the last 20 years has shown that a subset of cadherins, the clustered protocadherins, mediates neuronal self-identification in mammals. Just like Dscams, protocadherins generate unique signatures, or "barcodes," in individual neurons

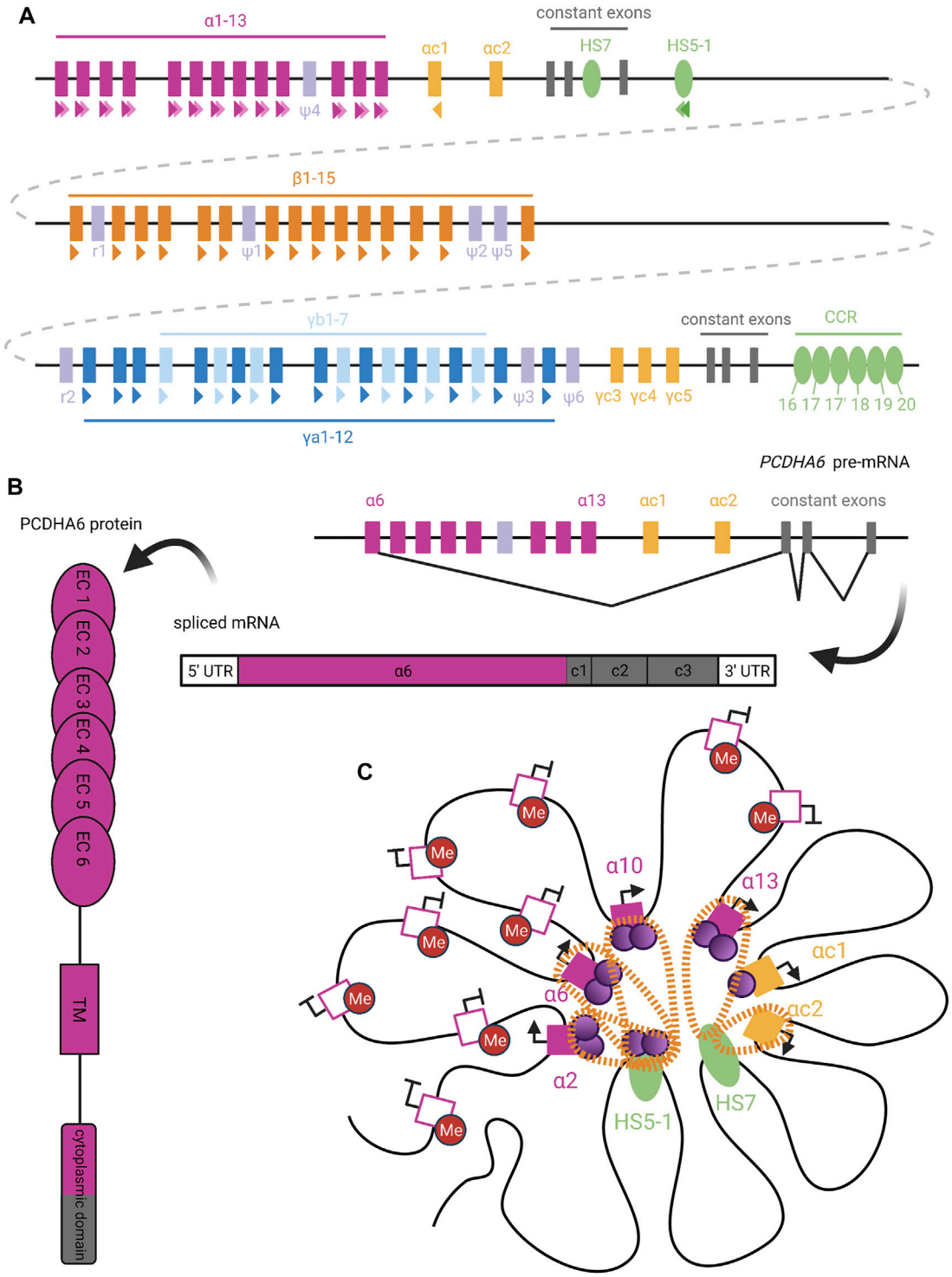


FIGURE 6 | Organization and Expression of Protocadherins. **(A)** Genomic organization of the human clustered protocadherins. In humans, the *PCDHA*, *PCDHB*, and *PCDHG* genes are arranged in tandem clusters, with a few intervening pseudogenes (lavender), in the 5q31 region of chromosome 5 (Wu and Maniatis, 1999). The *PCDHA* and *PCDHG* clusters also have deterministically-expressed terminal V exons (yellow) and constant exons (grey), while the *PCDHB* cluster only contains stochastically-expressed V exons (Tasic et al., 2002). Colored triangles under exons or enhancers denote CSEs, with staggered triangles indicating two CSEs for a given element. **(B)** Upon *PCDHA* promoter selection, transcription proceeds through all downstream V and C exons. Intervening V exons are spliced out such that mature *PCDHA* mRNA contains just one variable exon spliced to three constant exons (Tasic et al., 2002). The variable exon encodes six ectodomains, the transmembrane domain, and the N-terminal half of the cytoplasmic domain in the PCDHA protein; the rest is encoded by the constant exons (Mah and Weiner, 2016). **(C)**

(*Continued*)

FIGURE 6 | Extensive DNA looping between active *PCDHA* promoters (filled pink or yellow squares) and long-range enhancers (HS5-1 and HS7; green ovals) is mediated by complexes of CTCF (purple circles) and cohesin (orange perforated bands) (Kehayova et al., 2011; Guo et al., 2012). In certain cases, binding of cohesin or CTCF alone (as with the *PCDHAC1* and *PCDHAC2* exons, respectively) is able to mediate promoter-enhancer interactions (Guo et al., 2012). To allow CTCF binding, candidate *PCDHA* genes must escape repressive DNA methylation (red circles), perhaps with the help of long non-coding RNA (Canzio et al., 2019). We diagram the potential for co-expression of a few isoforms in the same cell, as has been observed in single-cell analyses (Esumi et al., 2005; Kaneko et al., 2006; Mountoufaris et al., 2017). Figure inspiration was drawn from various sources (Sawaya et al., 2008; Weiner and Jontes, 2013; Massah et al., 2015; Mancini et al., 2020; Wu and Jia, 2021).

and function via homophilic repulsion (Weiner and Jontes, 2013). The regulation of clustered *PCDH* expression shares much in common with other systems discussed in this review: protocadherins utilize enhancers to stochastically select a promoter, similar to olfactory receptors; they contain variable and constant regions analogous to those of immunoglobulins; and their expression is sometimes monoallelic and restricted by heterochromatinization.

Genomic Organization of Clustered Protocadherins

Clustered *PCDH*s are organized into alpha (*PCDHA*), beta (*PCDHB*), and gamma (*PCDHG*) clusters, arranged in tandem along chromosome 5 in humans. The *PCDHA* genes are encoded by a set of 15 large variable (V) exons that precede 3 constant (C) exons (**Figure 6A**) (Wu and Maniatis, 1999). Each V exon is preceded by its own promoter, and transcription can initiate from any of these 15 promoters. The first transcribed V exon, determined by which promoter is selected, is then spliced to the C exons, removing the intervening V exons. The *PCDHG* cluster is arranged similarly to the *PCDHA* cluster, with 22 V exons preceding 3 C exons (Morishita and Yagi, 2007). The *PCDHB* cluster differs from the other two in that it does not contain any C exons to complement its 22 V exons and is thus a set of distinct genes (Tasic et al., 2002).

The final 2 V exons within the *PCDHA* cluster, *PCDHAC1* and *PCDHAC2*, and the *PCDHGC3*, *PCDHGC4*, and *PCDHGC5* from the *PCDHG* cluster are very similar to each other but not to other V exons (Wu and Maniatis, 1999). These five exons are expressed deterministically and will not be discussed further here.

Stochastic Selection of V Exon Promoters Dictates Isoform Expression

The promoters preceding *PCDH* V exons contain a conserved sequence element (CSE) (Wu et al., 2001), which has been found to serve as an essential promoter binding motif (Tasic et al., 2002). With the exception of *PCDHAC1* and *PCDHAC2*, a second CSE has been observed within the exonic sequence of each *PCDHA* V exon (**Figure 6A**) (Chen and Maniatis, 2013). Similar CSE regions are also present in transcriptional *cis*-enhancer elements that are specific to each *PCDH* gene cluster (Hirayama and Yagi, 2017): The *PCDHA* cluster enhancer element HS5-1 is found downstream of the cluster's third C exon and contains two CSE sites (Ribich et al., 2006). A long-range *PCDHB* enhancer, known as the clustered control region (CCR), has also been identified downstream of the *PCDHG* cluster (Yokota et al., 2011). Deletion of these enhancers affected the expression of their respective clusters, with little effect on *PCDHG* cluster expression. This suggests the existence of distinct *PCDHG* enhancer element(s) (Xγ elements) whose exact genomic location is unknown (Yokota et al., 2011).

In order to initiate transcription at a given V exon, the CSE sites within its promoter region, its respective cluster's enhancer element(s), and exon sequence (for *PCDHA* only), must escape repressive methylation established by DNMT3B during embryogenesis (Garrett et al., 2019). A recent study has implicated antisense long non-coding RNA (as-lncRNA) in the demethylation of promoter and exonic CSE sites within the *PCDHA* cluster (Canzio et al., 2019). This process may facilitate variation in *PCDHA* promoter choice by equalizing interaction of the enhancer with distal and proximal promoters. That is, in the absence of methylation, proximal promoters are preferentially chosen (Canzio et al., 2019). Thus, global methylation of the alpha cluster, alleviated by stochastic demethylation by as-lncRNA, prevents inclusion bias caused by proximity of certain promoters to enhancer elements. The same study did not detect the presence of any as-lncRNA corresponding to the *PCDHB* and *PCDHG* clusters. As such, it remains unclear what mechanisms are responsible for the demethylation of CSE sites within these two clusters.

Once the appropriate CSE sites have been demethylated, the next step of transcription initiation can proceed. The CSEs serve as binding sites for the insulator protein CTCF, which can interact with cohesin to form a CTCF/cohesin complex (Guo et al., 2012). Chromosome conformation capture (3C) assays have demonstrated the ability of CTCF/cohesin complexes to mediate interactions between the V exon promoters and enhancers through DNA-looping (**Figure 6C**) (Guo et al., 2012). Interestingly, a second *PCDHA* enhancer element (HS7), which is located within the intron between the second and third C exons, lacks a CSE site but is still able to mediate DNA-looping through interaction with cohesin alone (Guo et al., 2012). Nonetheless, several studies have pointed towards CTCF/cohesin complex interactions between V exon promoters and enhancers as a necessary step for expression of most PCDH isoforms from all three clusters (Monahan et al., 2012; Hirayama and Yagi, 2017). This conclusion is further supported by the fact that in non-neuronal cell types, competitive binding of the REST/NRSF repressor complex to the HS5-1 enhancer, rather than the CTCF/cohesin complex, led to significantly decreased *PCDHA* expression (Kehayova et al., 2011). Unlike the proposed model for an "enhancer hub" similar to that involved in olfactory receptor choice, it remains unclear how transcription proceeds following the formation of these promoter/enhancer interactions (Guo et al., 2012).

Incredibly, *PCDH* genes are expressed both monoallelically and biallelically. All three *PCDH* clusters show monoallelic, combinatorial expression of the V exons (Esumi et al., 2005; Kaneko et al., 2006; Hirano et al., 2012). However, the 5 C-type variable exons of the alpha and gamma clusters, *PCDHAC1*, *PCDHAC2*, *PCDHGC3*, *PCDHGC4*, and *PCDHGC5*, are all expressed biallelically (Kaneko et al., 2006). Therefore, both the *PCDHA* and *PCDHG* clusters are regulated under different allelic gene regulation mechanisms, which may help to increase neuronal diversity (Kaneko et al., 2006).

Protocadherin expression is not fully monogenic—in at least some neuronal types, expression of >2 non-C-type isoforms together is common (Kehayova et al., 2011). This could result from unstable promoter choice, such that different isoforms are expressed sequentially and still present contemporaneously, or from incomplete dependence on exclusive enhancers. In support of the second model, PCDHA promoters display differential dependence on the HS5-1 enhancer; PCDHA1-5 promoters are only moderately affected by loss of HS5-1 and could be chosen by an additional enhancer (Kehayova et al., 2011). Remarkably, Dscam choice in insects is also non-exclusive. Mixing of isoforms, or indeed variation in chosen isoforms over time, should not impede the function of these gene families in self-recognition as long as all parts of the same neuron have the same isoform mix as one another.

Roles in Neural Circuit Development

PCDHG genes are commonly studied and highly expressed in the dendrites of hippocampal neurons where homophilic cell-cell interactions between isoforms facilitate circuit complexity (Molumby et al., 2016). In these neurons, PCDHG acts locally to promote arborization via homophilic matching. This was shown by increasing the likelihood of PCDHG homophilic interactions using mutations, which subsequently increased dendritic and circuit complexity (Molumby et al., 2016). This is opposite of what would be expected in other cell types, like retinal starburst amacrine cells and Purkinje cells, in which homophilic PCDH interactions lead to a clear self avoidance pattern (Lefebvre et al., 2012). Thus, depending on the cell type, PCDHG interactions may lead to attraction, repulsion, or other dendritic arborization signaling events. It has also been shown that the *PCDHA* cluster is involved in these expression patterns and that PCDHA and PCDHG isoforms work synergistically to facilitate dendritic arborization pattern formation (Molumby et al., 2016; Ing-Esteves et al., 2018). Comparisons of an allelic series mutant support the conclusion that PCDHA and PCDHG function together in a dose-dependent and cell-type-specific manner to provide a critical threshold for PCDH activity. Although this does create a type of redundancy in *PCDH* stochastic expression, having both *PCDHA* and *PCDHG* expressed is critical for neuronal development (Ing-Esteves et al., 2018). PCDHB isoforms, similar to PCDHA and PCDHG, form *trans*-homophilic interactions (interactions with identical molecules on other cells), but expression patterns are not well classified (Yokota et al., 2011; Hasegawa et al., 2016).

Structural Characterizations

Like Dscams, clustered protocadherins bind homophilically, and all the isoforms expressed in a cell must find a match in *trans* in order to induce strong enough binding to initiate downstream signaling cascades and cellular responses. Recent structural characterizations of the gene family show that the structure of the PCDH ectodomain is a zipper-like lattice formed by alternating *cis*- and *trans*-interactions (**Figure 5B**). As the protein extends out of the cell and interacts with the proteins of adjacent cells, a larger two-dimensional structure is created between the cell membranes (Brasch et al., 2019). This structure is the basis for the initial self-recognition step in neuronal self-avoidance (Brasch et al., 2019). Once these structures are formed, members of all three PCDH clusters can mediate highly specific homophilic recognition, maximizing the most favorable protein interactions. This favorable homophilic interaction is maximized when identical isoforms of clustered PCDHs are present. For example, neurons expressing five isoforms prefer to form homophilic aggregates with neurons expressing a full identical set of five isoforms rather than those expressing just three or four out of the five. Thus, self-recognition between different neurons is avoided by the expression of a single mismatched isoform (Thu et al., 2014). If there is a perfect match between isoforms (i.e. branches of the same Soma), repulsion of the neurite branches will occur to avoid overlapping. However, the downstream signaling mechanisms mediating repulsion are not yet understood.

OLFACTORY RECEPTORS

Similar to the protocadherin family, mammalian olfactory receptor genes are also expressed monoallelically and monogenically in neurons. Olfactory receptor (OR) genes are organized in large clusters in the genome and, like *VSG* genes or the *PCDHB* cluster, are each independent transcriptional units, not sets of overlapping possibilities like the *BCR*/*TCR*, *Dscam1*, and *PCDHA* and *PCDHG* loci. ORs are transmembrane chemoreceptors, found on the cell membranes of olfactory sensory neurons (OSNs), which recognize odorant molecules to detect smells. Buck and Axel correctly estimated that there are roughly 1,000 *OR* genes in the mammalian genome (Buck and Axel, 1991). Stochastic expression of individual *OR* genes across individual olfactory sensory neurons allows each neuron to serve as a sensor for a limited repertoire of odors, simplifying higher-order interpretation of olfactory information.

The mammalian olfactory system is composed of the main and accessory olfactory systems (Spehr and Munger, 2009). The main olfactory epithelium (MOE), located in the main olfactory system in the nasal cavity, contains the OSNs, which synapse onto the main olfactory bulb. The accessory olfactory system, which is located in the vomeronasal organ in the nasal cavity, contains OSNs that express vomeronasal receptors and is absent in humans and other primates. A further look into the *OR*s of mice and humans reveals that they share many subfamilies of *OR* genes. However, homology relationships are difficult to discern at the gene level, and mouse subfamilies typically include more *OR* genes than human subfamilies (Godfrey et al., 2004). This variation in the olfactory receptor repertoire occurs through a

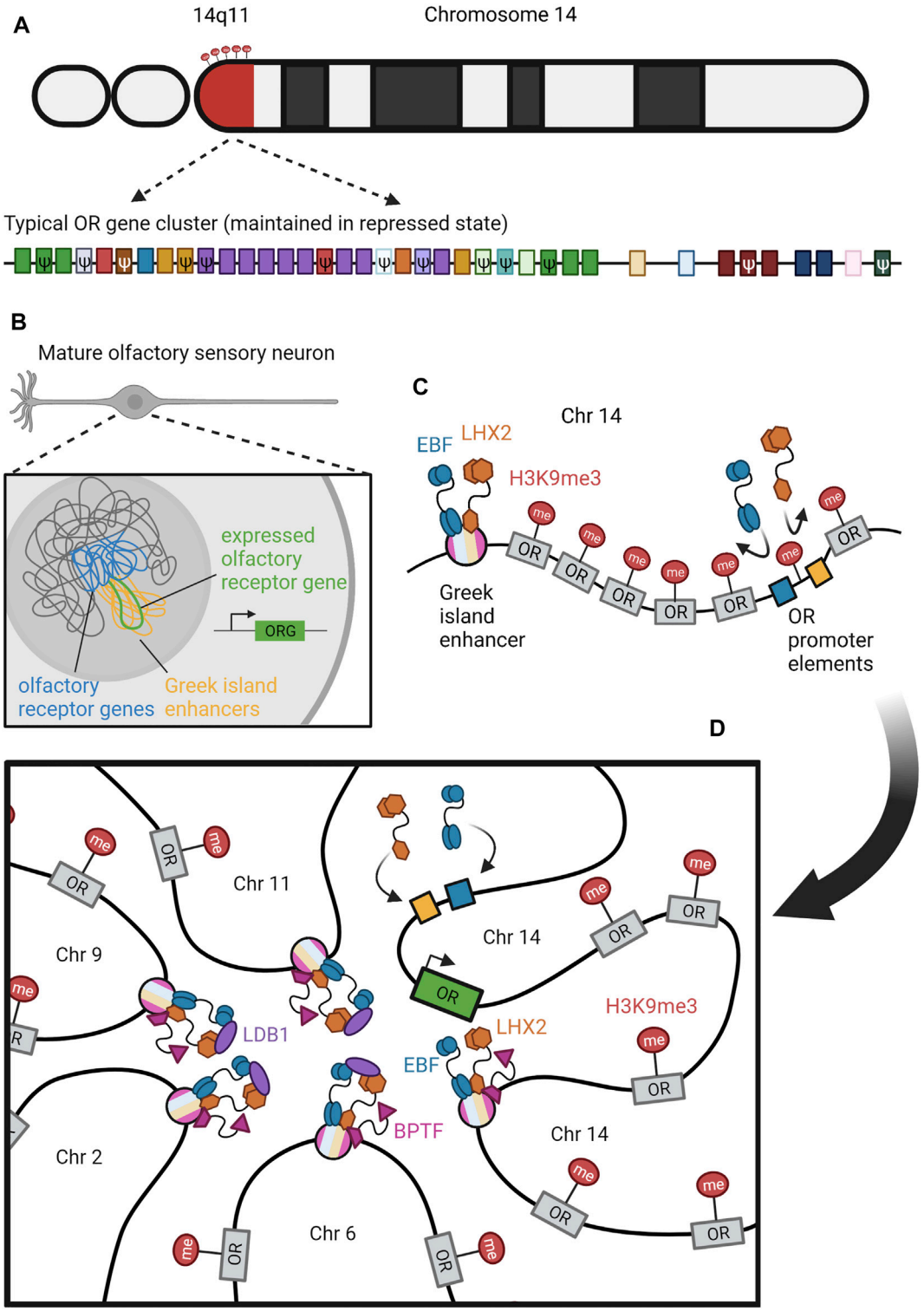


FIGURE 7 | Organization and Expression of ORs. **(A)** Olfactory receptor (*OR*) genes are distributed throughout all chromosomes in the human genome except for chromosome 20. An example of a typical *OR* gene cluster, maintained in a repressed state by histone methylation, is shown on region 14q11 of chromosome 14. H3K9me3 is provided as an example of a heterochromatic modification (red lollipops). A close-up of this cluster reveals an array of functional and pseudo (ψ) *OR* genes belonging to various (color-coded) subfamilies (Olender et al., 2008). **(B)** In mature olfactory sensory neurons, *OR* genes are clustered into constitutive chromatin in the center of the nucleus (Clowney et al., 2012; Armelin-Correa et al., 2014; Armelin-Correa et al., 2015; Tan et al., 2019). Greek island enhancers (yellow strands) that are *cis* to inactive *OR* genes (blue strands) will form *trans* interactions to cluster around one active *OR* gene (green strand). Panel **(C,D)** detail the temporal progression

(*Continued*)

FIGURE 7 | leading to this arrangement. **(C)** Although Greek island enhancers can make *cis OR* genes competent for activation, they cannot completely relieve heterochromatic repression (red lollipops) (Serizawa et al., 2003; Nishizumi et al., 2007; Khan et al., 2011). Transcription factors BPTF (pink), EBF (blue), and LHX2 (orange) can bind to juxtaposed motifs within enhancers (striped circles), but are not strong enough to overcome heterochromatic repression of the same motifs in *OR* promoters (yellow and blue boxes) (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2017). **(D)** *Trans* interactions between Greek island enhancers on different chromosomes select one competent *OR* gene for activation (green box) (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2017; Monahan et al., 2019; Tan et al., 2019). Long-range, interchromosomal interactions are mediated by LDB1 (purple) in concert with EBF and LHX2 (Monahan et al., 2019). These interactions relieve repression on the promoter *cis* to the selected *OR*, allowing EBF and LHX2 to bind and promote transcription (Monahan et al., 2017). Figure inspiration was drawn from various sources (Malnic et al., 2004; Olender et al., 2008; Monahan et al., 2017; Bashkirova and Lomvardas, 2019; Monahan et al., 2019; Spitz, 2019) (Malnic et al., 2004; Olender et al., 2008; Monahan et al., 2017; Bashkirova and Lomvardas, 2019; Monahan et al., 2019; Spitz, 2019).

process of birth-and-death evolution, and species-level differences are thought to arise through a combination of adaptation to different niches and genetic drift (Nozawa and Nei, 2009). Selection on any one member of such a large gene family is expected to be weak.

Like Dscam isoforms, ORs are expressed through a biased stochastic process: While expression of each OR is restricted to a particular zone of the olfactory epithelium, within the zone each olfactory sensory neuron randomly expresses just one allele of one of the ~1000 *OR* genes. Individual olfactory sensory neurons don't "know" in advance which receptor they will express. This stochastic expression system may facilitate evolutionary turnover at the receptor level as newly born ORs can be expressed privately in existing cell types.

The axons of OSNs expressing the same receptor converge onto glomeruli in the olfactory bulb, which act as a functional unit of odor coding (Chesler et al., 2007). Since a given odorant activates multiple OR species and a given OR responds to multiple odorants, this results in received odorants producing unique combinations of activated glomeruli with varying magnitudes of activity in the olfactory bulb (Malnic et al., 1999). This ensures precise odorant perception and allows the mouse olfactory system, for example, to detect orders-of-magnitude more odors than the number of receptors encoded in the genome (Brochtrup and Hummel, 2011).

Monoallelic and Monogenic Expression of ORs

The expression of just a single olfactory receptor gene per OSN plays two critical roles in odor perception: First, it ensures that each OSN senses just a small set of odorant molecules; second, it allows olfactory sensory neurons to target specific glomeruli in the olfactory bulb. Multiple studies of differentiated OSNs have demonstrated that *OR* genomic regions remain intact throughout development (Eggan et al., 2004; Li et al., 2004). This is unlike the genetic recombination mechanisms we see in VSG and immunoglobulin diversity, which rearrange genomic loci. Further sequencing data confirmed that ORs are each encoded by individual genes rather than variable exons like *Dscam* and *Pcdh* genes (Li et al., 2004). Not only that, but a single *OR* gene is selected via the interaction of multiple enhancers (**Figure 7**), followed by feedback inhibition of remaining ORs (Monahan and Lomvardas, 2015). Investigations into *OR* selection have therefore focused on epigenetic and transcriptional regulators in an attempt to understand this stochastic system (Shykind, 2005).

In addition to being expressed monogenically (i.e. one OR per neuron), *OR*s are also expressed monoallelically. It is not clear if this is functionally important or if it is a side effect of the monogenic expression system. If the neuron treats all *OR* genes as identical and makes a single choice, then this would result in monoallelic as well as monogenic expression. However, previous studies have suggested that monoallelic silencing is a separate process from monogenic *OR* choice, and that it occurs earlier in development via DNA methylation (Chess et al., 1994; Armelin-Correa et al., 2014).

Transcription of multiple *OR* genes per OSN precedes the eventual translation of just a single OR per neuron (Hanchate et al., 2015; Tan et al., 2015). Immature neurons may express multiple receptors at low concentrations in overlapping regions of the nasal epithelium (Hanchate et al., 2015). This is reminiscent of the VSG system, in which cells may transcribe multiple *VSG*s before choice stabilization, and may implicate RNA-based mechanisms of gene silencing and/or feedback mechanisms informing the cell that successful translation has occurred.

Heterochromatic Silencing of Non-chosen ORs

Despite coexpression of multiple *OR* genes in some immature OSNs, only one functional OR protein is translated in mature OSNs, and the rest of the gene family is transcriptionally silenced. Similar to many other surface molecules previously described, epigenetic regulation is essential for *OR* choice. Formation of heterochromatin throughout *OR* gene clusters likely occurs in an early stage preceding *OR* gene expression, although the extent of heterochromatic silencing in an OSN's ground state is still undetermined (Magklara et al., 2011; Nagai et al., 2016). *OR* genes are marked by the constitutive heterochromatin marks H3K9me3 and H4K20me3, which is a common characteristic amongst other monoallelic expression patterns (**Figures 7C,D**) (Magklara et al., 2011). Histone methyltransferases G9a and GLP assist in H3K9 trimethylation and have been shown to be essential for promoting OR diversity, but their significance in the timeline of OSN development has not been explicitly explained (Ferreira et al., 2014; Lyons et al., 2014). Since *OR* genes are widely repressed by trimethylation, histone demethylases are required to remove constitutive repressive marks from heterochromatin in order to permit expression of the chosen *OR* (Magklara et al., 2011). The enzymes that demethylate each heterochromatin mark to its dimethylated state have yet to be discovered, but one H3K9me2 demethylase, LSD1, is required for *OR* transcription (Lyons et al., 2013).

Heterochromatinization of olfactory receptor loci likely allows their reorganization into dense nuclear aggregates and therefore prevents most alleles from contacting the transcriptional machinery (Magklara et al., 2011; Clowney et al., 2012; Le Gros et al., 2016). While heterochromatin is found at the periphery of most mammalian cell nuclei, OSNs and some other types of neurons display a peculiar architecture with constitutive heterochromatin concentrated at the center of the nucleus (**Figure 7B**) (Solovei et al., 2009; Clowney et al., 2012; Armelin-Correa et al., 2015; Tan et al., 2015). Moreover, *OR* genes from all of the chromosomal arrays coalesce into interchromosomal aggregates during OSN development (Clowney et al., 2012; Monahan et al., 2019; Tan et al., 2019). Dismantling these aggregates allows expression of multiple *OR*s per OSN, even when heterochromatin marks remain intact (Clowney et al., 2012). These aggregates are also known to be disrupted in COVID-induced anosmia (Zazhytska et al., 2021).

Enhancer Activation

Reminiscent of *PCDH* promoter selection, enhancer activation is essential for *OR* gene expression. There is a large family of enhancers, called Greek islands, each serving a necessary function in activating their cluster of *cis OR*s (Serizawa et al., 2003; Nishizumi et al., 2007; Khan et al., 2011; Markenscoff-Papadimitriou et al., 2014). Additionally, these enhancers interact with one another in *trans* to form a large enhancer cluster that associates with a single *OR*, selecting it for activation (**Figure 7D**) (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2017; Monahan et al., 2019; Tan et al., 2019). These enhancer-*OR* clusters are specific to OSNs, and are not found in other olfactory epithelial cells (Tan et al., 2019). The enhancer regions contain homeodomain binding motifs that are essential for their regulatory function (Nishizumi et al., 2007; Monahan et al., 2019). In transgenic mice, the accumulation of these homeodomain binding motifs upstream of an *OR* promoter led to an increasing probability of reporter transcription (Vassalli et al., 2011). These observations suggest that the accumulation of multiple enhancers provides this homeodomain enrichment to promote *OR* transcription.

The dynamics of these clusters follow a pattern in which the *cis* interactions form first, and then aggregate with other *cis* clusters on separate chromosomes to form larger *trans* complexes (**Figures 7C–D**) (Monahan et al., 2019). The formation of the *OR* foci also plays an important role in bringing *trans* enhancers and *OR* genes together. As such, inhibition of *OR* foci leads to a decrease in enhancer cluster formation and association with *OR* genes, causing a decrease in *OR* expression (Clowney et al., 2012; Markenscoff-Papadimitriou et al., 2014). Recent single cell Hi-C analysis has revealed that the number of enhancers needed to activate a given *OR* is rather low compared to the total number available, with active *OR* genes being associated with an average of six enhancers (Tan et al., 2019).

Several integral proteins have been characterized as key facilitators of cluster formation and *OR* activation. BPTF and LDB1 bind the enhancers and facilitate the aggregation of large enhancer-*OR* complexes to promote *OR* expression (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2019). Additionally, LHX2 binds homeodomains (Hirota and Mombaerts, 2004; Rothman et al., 2005; Monahan et al., 2017; Monahan et al., 2019) and the Olf/EBF family bind O/E binding sites (**Figures 7C–D**) (Wang et al., 1997; Wang S. S. et al., 2004), both of which are found on the *OR* promoters and enhancers. The deletion of these transcription factors reduces *OR* expression, and the Olf/EBF family is tightly regulated by Ffp433/OAZ throughout OSN development (Cheng and Reed, 2007; Roby et al., 2012).

Taken together, these discoveries support a model in which each *OR* enhancer first associates in *cis* with a single *OR* gene to promote transcriptional competence, after which the nuclear localization of *OR* foci facilitates the aggregation of these *cis* enhancer-*OR* pairs into larger enhancer complexes (Bashkirova and Lomvardas, 2019). Stochastic association of this large enhancer complex with one of the *OR* genes selected by its *cis* enhancer then promotes its transcription via the accumulated recruitment of transcription factors. Several of these activating clusters emerge in a given OSN nucleus, necessitating a feedback mechanism to silence non-selected *OR* genes and stabilize the selected gene (Tan et al., 2019).

Negative Feedback Regulation

In order to achieve proper OSN identity, *OR* gene silencing is necessary to ensure only one OR will function once the neuron has fully matured (Lewcock and Reed, 2004; Fleischmann et al., 2008; Zhou and Belluscio, 2012). Once a functional OR is translated and expressed at robust levels, a negative feedback mechanism called the unfolded protein response (UPR) is triggered (Dalton et al., 2013). During this response, the successful and abundant translation of an OR leads to downstream signaling that turns off important translation-initiation factors, ceasing translation of most other ORs while also inducing expression of adenylyl cyclase type 3 (ADCY3), a molecule that is essential for odor signaling (Ron and Walter, 2007; Wang S.-Z. et al., 2012; Dalton et al., 2013). This feedback mechanism is very similar to the process that takes place within B cells that confirms successful translation of a functional BCR (Hetz et al., 2020). Many chaperones are also essential for OR singularity by coordinating the UPR between organelles and the cell membrane (Ron and Walter, 2007; Sharma et al., 2017). Evidence from transgene studies suggests there are also *cis*-regulatory elements in the *OR* coding region that contribute to monogenic expression, but these elements have yet to be identified (Nguyen et al., 2007).

G protein signaling also plays a role in stabilizing *OR* choice following expression of the OR protein in the cell membrane. However, it is undetermined whether the chosen *OR* is the first to be expressed or the most robustly expressed. Nevertheless, the selected OR protein acts as a GPCR to activate G_{olf} to signal its presence as a functional OR. Specifically, the beta-gamma subunit of G_{olf} has been shown to contribute to *OR* gene silencing by heterochromatin regulation in zebrafish, while other model systems have pointed to the importance of the alpha subunit of G_{olf} and its mediation by RIC8B (Von Dannecker et al., 2006;

Ferreira et al., 2014; Machado et al., 2017). Regardless of each subunit's role, G_{olf} then activates ADCY3, which in turn downregulates the aforementioned histone demethylase LSD to prevent activation of other *OR* genes and stabilize *OR* choice (Imai et al., 2006; Dalton et al., 2013; Lyons et al., 2013). The opposing roles of ADCY3 and LSD1 are further reviewed in (Monahan and Lomvardas, 2015).

Altogether, there are multiple known mechanisms of *OR* gene expression and silencing that are essential for monoallelic and monogenic expression. Achievement of singularity has been shown to involve heterochromatin formation, nuclear compartmentalization, enhancer interactions, and a feedback inhibition pathway. However, the temporal and molecular overlaps between these pathways are not yet understood. More research is needed to outline the developmental timeline of *OR* expression and identify causal relationships between these mechanisms that ensure *OR* singularity. This diversity and singularity of *OR* expression facilitates the effective axon guidance and complex circuitry that allows us to detect such a wide spectrum of odors.

SIMILARITIES AND DIFFERENCES ACROSS SYSTEMS

The major shared feature of the five systems covered in this review is their selection of a specific genetic option among a collection of many alternatives in order to express a surface protein that is integral to the cell's identity. While each system has evolved unique mechanisms of monogenic and/or monoallelic expression, there are many parallels in the strategies adopted to accomplish selection, expression, and stabilization of the selected sequence (**Table 1**).

First, generation of combinatorial products is a key feature in many of these systems that expands the number of potential surface molecules relative to the number of coding segments encoded in the genome. The antigen receptors utilize V(D)J recombination, clustered *PCDH* genes select variable segments, and *Dscam1* splicing combines variable exons to create a massive repertoire of possible combinations from a small number of gene segments. *VSG* genes similarly combine pre-existing *VSG* gene segments to create new mosaic *VSG*s, and, unlike the other systems, these repertoires are constantly evolving and growing as new *VSG*s are made and stored.

These systems also exhibit context-driven restrictions and biases regarding which genes or gene segments are chosen. Within the olfactory epithelium, OSNs are spatially segregated into different zones, and each OR is expressed only within a single zone. Dscams and PCDHs show a similar restriction, but based on cell type rather than spatial location, as the pattern of variable domains expressed appears to vary between different populations of neurons. *VSG* expression instead follows a temporal bias for which the probability of certain VSGs appears to change throughout the course of infection, with mosaic VSGs emerging in the later stages of infection. Within the confines of these spatial, cell type, and temporal restrictions the available genes and gene segments still seem to be stochastically distributed. The mechanisms that restrict stochasticity are largely unknown.

The process determining which *VSG* expression site is active relies heavily on nuclear localization and chromatin remodeling, a feature shared by the other four systems. During VSG, OR, and antigen receptor selection the inactive alleles are repressed in heterochromatic regions, with *VSG* and antigen receptor loci localized to the nuclear periphery and *OR* gene arrays compacted into repressive foci. The selected allele in these systems is found in accessible euchromatic regions, which for antigen receptors is in the nuclear center, for *OR*s entails the active gene escaping the repressive foci, and for *VSG* genes requires the single active expression site to localize to the expression body. These nuclear localization events are intimately tied to chromatin modifications and DNA methylation. *VSG* expression sites localized to the expression body show significant nucleosome depletion. Interestingly, the active *VSG* expression site is enriched with H3K10 acetylation as well as H3K4 trimethylation. H3K4me3 is an activating signature that is also found in the active regions of the OR, *PCDH*, and antigen receptor loci. Histone modifications are prominent features of these systems. The antigen receptor loci are further activated by H3K9 acetylation and H3K36 trimethylation, whereas the inactive locus is enriched for repressive H3K9 di- and trimethylation. H3K9 methylation is also utilized to repress *OR* loci alongside H4K20 di- and trimethylation.

Singular *cis*-acting elements are also a common feature activating the selected sequences throughout many of these systems. Antigen receptors and *PCDH* loci share mechanistically similar downstream enhancer elements that loop back to the loci and activate the selected promoter, in both cases scanning for the first available promoter sequence either at the most proximal V segment or hypomethylated *PCDH* segment. Enhancers play a role not only in inducing expression of the selected *OR* gene, but also in carrying out the selection process, as enhancers coalesce first in *cis*, and then in *trans* onto a given *OR* gene for selective expession. Even though Dscam selection occurs at the RNA level, an enhancer-like element (the docking site-selector sequence stem-loop) occurs in the secondary structure of the transcript. Reminiscent of the previous systems, the stochastic binding of variant selector sequences to the docking site (which forms the stem-loop) dictates the pattern of exon splicing. While *VSG* expression has not been shown to require enhancer interactions, the selected *VSG* still must be located within the active expression site. Further research into whether a limiting enhancer selects a *VSG* for selection at a particular time will be of great interest.

CONCLUSION

In this review, we have emphasized the importance of stochastic processes in achieving variability in barrier and nervous systems. By analyzing immune system evasion by the parasite *Trypanosoma brucei*, immune system pathogen identification by B cell and T cell receptors, neuronal self-avoidance through diverse expression and self-binding processes of Dscams and protocadherins, and the perception of stimuli through the olfactory system, we have shown that the role of diversity is crucial to the survival and function of organisms through both internal and environmental

interactions. Common processes such as monoallelic expression, epigenetic regulation, and specificity of binding of variable domains were discussed. It is important to recognize that all five systems studied are examples of surface molecules that utilize the stochastic selection of genetic material to create cellular diversity. Like aleatoric music and the world-generating algorithms of sandbox games, these systems maximize limited genetic information to construct complex and unpredictable outcomes.

AUTHOR CONTRIBUTIONS

Writing—Original Draft Preparation: all authors. Writing—Review & Editing: DW, VS, MH, CB, and EJC. Preparation of figures: VS. Funding acquisition: EJC. DW, VS, and MH share first authorship. This review was written in the Fall 2020 semester by students enrolled in a course at the University of Michigan, MCDB464: Cellular Diversity in the Immune and Nervous Systerms.

FUNDING

The preparation of this review was supported by the University of Michigan Program in Neuroscience, by the Rita Allen Foundation Milton Cassell Scholar Award (to EJC), and by NIDCD (Grant number 1R01DC018032, to EJC).

REFERENCES

Allen, C. D. C., Ansel, K. M., Low, C., Lesley, R., Tamamura, H., Fujii, N., et al. (2004). Germinal center Dark and Light Zone Organization Is Mediated by CXCR4 and CXCR5. *Nat. Immunol.* 5, 943–952. doi:10.1038/ni1100

Allen, C. D. C., Okada, T., and Cyster, J. G. (2007a). Germinal-center Organization and Cellular Dynamics. *Immunity* 27, 190–202. doi:10.1016/j.immuni.2007.07.009

Allen, C. D. C., Okada, T., Tang, H. L., and Cyster, J. G. (2007b). Imaging of Germinal center Selection Events during Affinity Maturation. *Science* 315, 528–531. doi:10.1126/science.1136736

Allshire, R. C., and Madhani, H. D. (2018). Ten Principles of Heterochromatin Formation and Function. *Nat. Rev. Mol. Cel. Biol.* 19, 229–244. doi:10.1038/nrm.2017.119

Alsford, S., and Horn, D. (2012). Cell-cycle-regulated Control of VSG Expression Site Silencing by Histones and Histone Chaperones ASF1A and CAF-1b in Trypanosoma Brucei. *Nucleic Acids Res.* 40, 10150–10160. doi:10.1093/nar/gks813

Alsford, S., Kawahara, T., Isamah, C., and Horn, D. (2007). A Sirtuin in the African Trypanosome Is Involved in Both DNA Repair and Telomeric Gene Silencing but Is Not Required for Antigenic Variation. *Mol. Microbiol.* 63, 724–736. doi:10.1111/j.1365-2958.2006.05553.x

Anastassiou, D., Liu, H., and Varadan, V. (2006). Variable Window Binding for Mutually Exclusive Alternative Splicing. *Genome Biol.* 7, R2. doi:10.1186/gb-2006-7-1-r2

Aresta-Branco, F., Erben, E., Papavasiliou, F. N., and Stebbins, C. E. (2019a). Mechanistic Similarities between Antigenic Variation and Antibody Diversification during Trypanosoma Brucei Infection. *Trends Parasitology* 35 (4), 302–315. doi:10.1016/j.pt.2019.01.011

Aresta-Branco, F., Pimenta, S., and Figueiredo, L. M. (2016). A Transcription-independent Epigenetic Mechanism Is Associated with Antigenic Switching inTrypanosoma Brucei. *Nucleic Acids Res.* 44 (7), 3131–3146. doi:10.1093/nar/gkv1459

Aresta-Branco, F., Sanches-Vaz, M., Bento, F., Rodrigues, J. A., and Figueiredo, L. M. (2019b). African Trypanosomes Expressing Multiple VSGs Are Rapidly Eliminated by the Host Immune System. *Proc. Natl. Acad. Sci. USA* 116 (41), 20725–20735. doi:10.1073/pnas.1905120116

Armelin-Correa, L. M., Gutiyama, L. M., Brandt, D. Y. C., and Malnic, B. (2014). Nuclear Compartmentalization of Odorant Receptor Genes. *Proc. Natl. Acad. Sci.* 111, 2782–2787. doi:10.1073/pnas.1317036111

Armelin-Correa, L. M., Nagai, M. H., Silva, A. G. L., and Malnic, B. (2015). Nuclear Architecture and Gene Silencing in Olfactory Sensory Neurons. *Bioarchitecture* 4, 160–163. doi:10.4161/19490992.2014.982934

Backhaus, O. (2018). "Generation of Antibody Diversity," in *Book: Antibody Engineering*. doi:10.5772/intechopen.72818

Barbet, A. F., and Kamper, S. M. (1993). The Importance of Mosaic Genes to Trypanosome Survival. *Parasitol. Today* 9, 63–66. doi:10.1016/0169-4758(93)90039-I

Bashkirova, E., and Lomvardas, S. (2019). Olfactory Receptor Genes Make the Case for Inter-chromosomal Interactions. *Curr. Opin. Genet. Development* 55, 106–113. doi:10.1016/j.gde.2019.07.004

Bébin, A.-G., Carrion, C., Marquet, M., Cogné, N., Lecardeur, S., Cogné, M., et al. (20101950). *In Vivo* Redundant Function of the 3′ IgH Regulatory Element HS3b in the Mouse. *J. Immunol.* 184, 3710–3717. doi:10.4049/jimmunol.0901978

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., et al. (2005). The Genome of the African Trypanosome Trypanosoma Brucei. *Science* 309, 416–422. doi:10.1126/science.1112642

Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., and Cooper, M. D. (2012). VLR-based Adaptive Immunity. *Annu. Rev. Immunol.* 30, 203–220. doi:10.1146/annurev-immunol-020711-075038

Boothroyd, C. E., Dreesen, O., Leonova, T., Ly, K. I., Figueiredo, L. M., Cross, G. A. M., et al. (2009). A Yeast-Endonuclease-Generated DNA Break Induces Antigenic Switching in Trypanosoma Brucei. *Nature* 459, 278–281. doi:10.1038/nature07982

Brandenburg, J., Schimanski, B., Nogoceke, E., Nguyen, T. N., Padovan, J. C., Chait, B. T., et al. (2007). Multifunctional Class I Transcription in Trypanosoma Brucei Depends on a Novel Protein Complex. *EMBO J.* 26, 4856–4866. doi:10.1038/sj.emboj.7601905

Brasch, J., Goodman, K. M., Noble, A. J., Rapp, M., Mannepalli, S., Bahna, F., et al. (2019). Visualization of Clustered Protocadherin Neuronal Self-Recognition Complexes. *Nature* 569, 280–283. doi:10.1038/s41586-019-1089-3

Brochtrup, A., and Hummel, T. (2011). Olfactory Map Formation in the Drosophila Brain: Genetic Specificity and Neuronal Variability. *Curr. Opin. Neurobiol.* 21, 85–92. doi:10.1016/j.conb.2010.11.001

Buck, L., and Axel, R. (1991). A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition. *Cell* 65, 175–187. doi:10.1016/0092-8674(91)90418-X

Bühler, M., Paillusson, A., and Mühlemann, O. (2004). Efficient Downregulation of Immunoglobulin mRNA with Premature Translation-Termination Codons Requires the 5'-half of the VDJ Exon. *Nucleic Acids Res.* 32, 3304–3315. doi:10.1093/nar/gkh651

Canzio, D., Nwakeze, C. L., Horta, A., Rajkumar, S. M., Coffey, E. L., Duffy, E. E., et al. (2019). Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice. *Cell* 177, 639–653. e15. doi:10.1016/j.cell.2019.03.008

Celotto, A. M., and Graveley, B. R. (2001). Alternative Splicing of the Drosophila Dscam Pre-mRNA Is Both Temporally and Spatially Regulated. *Genetics* 159, 599–608. doi:10.1093/genetics/159.2.599

Cestari, I., and Stuart, K. (2015). Inositol Phosphate Pathway Controls Transcription of Telomeric Expression Sites in Trypanosomes. *Proc. Natl. Acad. Sci. USA* 112, E2803–E2812. doi:10.1073/pnas.1501206112

Cestari, I., and Stuart, K. (2018). Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes. *Cg* 19, 119–132. doi:10.2174/1389202918666170911161831

Chang, C. X. L., Dai, L., Tan, Z. W., Choo, J. A. L., Bertoletti, A., and Grotenbreg, G. M. (20113014). Sources of Diversity in T Cell Epitope Discovery. *Front. Biosci.* 16, 3014–3035. doi:10.2741/3895

Charles A Janeway, J., Travers, P., Walport, M., and Shlomchik, M. J. (2001). *The Generation of Diversity in Immunoglobulins*. 5th Ed. New York: Immunobiol. Immune Syst. Health Dis.

Chen, B. E., Kondo, M., Garnier, A., Watson, F. L., Püettmann-Holgado, R., Lamar, D. R., et al. (2006). The Molecular Diversity of Dscam Is Functionally Required for Neuronal Wiring Specificity in Drosophila. *Cell* 125, 607–620. doi:10.1016/j.cell.2006.03.034

Chen, W. V., and Maniatis, T. (2013). Clustered Protocadherins. *Development* 140, 3297–3302. doi:10.1242/dev.090621

Cheng, L. E., and Reed, R. R. (2007). Zfp423/OAZ Participates in a Developmental Switch during Olfactory Neurogenesis. *Neuron* 54, 547–557. doi:10.1016/j.neuron.2007.04.029

Chesler, A. T., Zou, D.-J., Le Pichon, C. E., Peterlin, Z. A., Matthews, G. A., Pei, X., et al. (2007). A G protein/cAMP Signal cascade Is Required for Axonal Convergence into Olfactory Glomeruli. *Pnas* 104, 1039–1044. doi:10.1073/pnas.0609215104

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic Inactivation Regulates Olfactory Receptor Gene Expression. *Cell* 78, 823–834. doi:10.1016/s0092-8674(94)90562-2

Clowney, E. J., LeGros, M. A., Mosley, C. P., Clowney, F. G., Markenskoff-Papadimitriou, E. C., Myllys, M., et al. (2012). Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. *Cell* 151, 724–737. doi:10.1016/j.cell.2012.09.043

Collins, A. M., and Watson, C. T. (2018). Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire. *Front. Immunol.* 9, 2249. doi:10.3389/fimmu.2018.02249

Cross, G. A. M., Kim, H.-S., and Wickstead, B. (2014). Capturing the Variant Surface Glycoprotein Repertoire (The VSGnome) of Trypanosoma Brucei Lister 427. *Mol. Biochem. Parasitol.* 195, 59–73. doi:10.1016/j.molbiopara.2014.06.004

Crotty, S. (2014). T Follicular Helper Cell Differentiation, Function, and Roles in Disease. *Immunity* 41, 529–542. doi:10.1016/j.immuni.2014.10.004

Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J., and Sanjuán, R. (2015). Extremely High Mutation Rate of HIV-1 *In Vivo*. *Plos Biol.* 13 (9), e1002251. doi:10.1371/journal.pbio.1002251

da Silva, M. S., Hovel-Miner, G. A., Briggs, E. M., Elias, M. C., and McCulloch, R. (2018). Evaluation of Mechanisms that May Generate DNA Lesions Triggering Antigenic Variation in African Trypanosomes. *Plos Pathog.* 14, e1007321. doi:10.1371/journal.ppat.1007321

Dalton, R. P., Lyons, D. B., and Lomvardas, S. (2013). Co-Opting the Unfolded Protein Response to Elicit Olfactory Receptor Feedback. *Cell* 155, 321–332. doi:10.1016/j.cell.2013.09.033

Devlin, R., Marques, C. A., Paape, D., Prorocic, M., Zurita-Leal, A. C., Campbell, S. J., et al. (2016). Mapping Replication Dynamics in Trypanosoma Brucei Reveals a Link with Telomere Transcription and Antigenic Variation. *Elife* 5, e12765. doi:10.7554/eLife.12765

Duveau, F., Hodgins-Davis, A., Metzger, B. P., Yang, B., Tryban, S., Walker, E. A., et al. (2018). Fitness Effects of Altering Gene Expression Noise in *Saccharomyces cerevisiae*. *Elife* 7, e37272. doi:10.7554/eLife.37272

Eberle, A. B., Herrmann, K., Jäck, H.-M., and Mühlemann, O. (2009). Equal Transcription Rates of Productively and Nonproductively Rearranged Immunoglobulin μ Heavy Chain Alleles in a Pro-B Cell Line. *RNA* 15, 1021–1028. doi:10.1261/rna.1516409

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., et al. (2004). Mice Cloned from Olfactory Sensory Neurons. *Nature* 428, 44–49. doi:10.1038/nature02375

Ersfeld, K., Melville, S. E., and Gull, K. (1999). Nuclear and Genome Organization of Trypanosoma Brucei. *Parasitol. Today* 15, 58–63. doi:10.1016/S0169-4758(98)01378-7

Escobar, L. L., Hänisch, B., Halliday, C., Dean, S., Sunter, J. D., Wheeler, R. J., et al. (2021). *Monoallelic Antigen Expression in Trypanosomes Requires a Stage-specific Transcription Activator*. Headington, Oxford: Cold Spring Harbor Laboratory. doi:10.1101/2021.05.06.442931

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., et al. (2005). Monoallelic yet Combinatorial Expression of Variable Exons of the Protocadherin-α Gene Cluster in Single Neurons. *Nat. Genet.* 37, 171–176. doi:10.1038/ng1500

Faria, J., Glover, L., Hutchinson, S., Boehm, C., Field, M. C., and Horn, D. (2019). Monoallelic Expression and Epigenetic Inheritance Sustained by a Trypanosoma Brucei Variant Surface Glycoprotein Exclusion Complex. *Nat. Commun.* 10 (1), 1–14. doi:10.1038/s41467-019-10823-8

Faria, J., Luzak, V., Müller, L. S. M., Brink, B. G., Hutchinson, S., Glover, L., et al. (2021). Spatial Integration of Transcription and Splicing in a Dedicated Compartment Sustains Monogenic Antigen Expression in African Trypanosomes. *Nat. Microbiol.* 6 (3), 289–300. doi:10.1038/s41564-020-00833-4

Feederle, R., and Schepers, A. (2017). Antibodies Specific for Nucleic Acid Modifications. *RNA Biol.* 14, 1089–1098. doi:10.1080/15476286.2017.1295905

Ferreira, T., Wilson, S. R., Choi, Y. G., Risso, D., Dudoit, S., Speed, T. P., et al. (2014). Silencing of Odorant Receptor Genes by G Protein βγ Signaling Ensures the Expression of One Odorant Receptor Per Olfactory Sensory Neuron. *Neuron* 81, 847–859. doi:10.1016/j.neuron.2014.01.001

Figueiredo, L. M., Janzen, C. J., and Cross, G. A. M. (2008). A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes. *Plos Biol.* 6, e161. doi:10.1371/journal.pbio.0060161

Fleischmann, A., Shykind, B. M., Sosulski, D. L., Franks, K. M., Glinka, M. E., Mei, D. F., et al. (2008). Mice with a "Monoclonal Nose": Perturbations in an Olfactory Map Impair Odor Discrimination. *Neuron* 60, 1068–1081. doi:10.1016/j.neuron.2008.10.046

Forrest, S., and Oprea, M. (1999). *Antibody Repertoires and Pathogen Recognition: The Role of Germline Diversity and Somatic Hypermutation [WWW Document]*. Available at: https://www.semanticscholar.org/paper/Antibody-repertoires-and-pathogen-recognition%3A-the-Forrest-Oprea/7be2f1a22f928d3a0f3ca8ca5129eeaf34987997 (accessed 1 12, 20).

Galler, G. R., Mundt, C., Parker, M., Pelanda, R., Mårtensson, I.-L., and Winkler, T. H. (2004). Surface μ Heavy Chain Signals Down-Regulation of the V(D)J-Recombinase Machinery in the Absence of Surrogate Light Chain Components. *J. Exp. Med.* 199, 1523–1532. doi:10.1084/jem.20031523

Garrett, A. M., Bosch, P. J., Steffen, D. M., Fuller, L. C., Marcucci, C. G., Koch, A. A., et al. (2019). CRISPR/Cas9 Interrogation of the Mouse Pcdhg Gene Cluster Reveals a Crucial Isoform-specific Role for Pcdhgc4. *PLOS Genet.* 15, e1008554–32. doi:10.1371/journal.pgen.1008554

Garrett, A. M., Khalil, A., Walton, D. O., and Burgess, R. W. (2018). DSCAM Promotes Self-Avoidance in the Developing Mouse Retina by Masking the Functions of Cadherin Superfamily Members. *Proc. Natl. Acad. Sci. USA* 115, E10216–E10224. doi:10.1073/pnas.1809430115

Glover, L., Alsford, S., and Horn, D. (2013). DNA Break Site at Fragile Subtelomeres Determines Probability and Mechanism of Antigenic Variation in African Trypanosomes. *Plos Pathog.* 9, e1003260. doi:10.1371/journal.ppat.1003260

Glover, L., and Horn, D. (2006). Repression of Polymerase I-mediated Gene Expression at Trypanosoma Brucei Telomeres. *EMBO Rep.* 7, 93–99. doi:10.1038/sj.embor.7400575

Glover, L., Hutchinson, S., Alsford, S., and Horn, D. (2016). VEX1 Controls the Allelic Exclusion Required for Antigenic Variation in Trypanosomes. *Proc. Natl. Acad. Sci. USA* 113, 7225–7230. doi:10.1073/pnas.1600344113

Godfrey, P. A., Malnic, B., and Buck, L. B. (2004). The Mouse Olfactory Receptor Gene Family. *Pnas* 101, 2156–2161. doi:10.1073/pnas.0308051100

Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., et al. (2005). Epigenetic Ontogeny of the Igk Locus during B Cell Development. *Nat. Immunol.* 6, 198–203. doi:10.1038/ni1154

Goodman, K. M., Rubinstein, R., Dan, H., Bahna, F., Mannepalli, S., Ahlsén, G., et al. (2017). Protocadherin Cis-Dimer Architecture and Recognition Unit Diversity. *Proc. Natl. Acad. Sci. USA* 114, E9829–E9837. doi:10.1073/pnas.1713449114

Graveley, B. R. (2005). Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell* 123, 65–73. doi:10.1016/j.cell.2005.07.028

Grawunder, U., Leu, T. M. J., Schatz, D. G., Werner, A., Rolink, A. G., Melchers, F., et al. (1995). Down-regulation of RAG1 and RAG2 Gene Expression in preB Cells after Functional Immunoglobulin Heavy Chain Rearrangement. *Immunity* 3, 601–608. doi:10.1016/1074-7613(95)90131-0

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K. E., Li, W., et al. (2012). CTCF/cohesin-mediated DNA Looping Is Required for Protocadherin Promoter Choice. *Proc. Natl. Acad. Sci.* 109, 21081–21086. doi:10.1073/pnas.1219280110

Hall, J. P. J., Wang, H., and Barry, J. D. (2013). Mosaic VSGs and the Scale of Trypanosoma Brucei Antigenic Variation. *Plos Pathog.* 9, e1003502. doi:10.1371/journal.ppat.1003502

Hanchate, N. K., Kondoh, K., Lu, Z., Kuang, D., Ye, X., Qiu, X., et al. (2015). Single-cell Transcriptomics Reveals Receptor Transformations during Olfactory Neurogenesis. *Science* 350, 1251–1255. doi:10.1126/science.aad2456

Hartley, C. L., and McCulloch, R. (2008). Trypanosoma Brucei BRCA2 Acts in Antigenic Variation and Has Undergone a Recent Expansion in BRC Repeat Number that Is Important during Homologous Recombination. *Mol. Microbiol.* 68, 1237–1251. doi:10.1111/j.1365-2958.2008.06230.x

Hasegawa, S., Kumagai, M., Hagihara, M., Nishimaru, H., Hirano, K., Kaneko, R., et al. (2016). Distinct and Cooperative Functions for the Protocadherin-α, -β and -γ Clusters in Neuronal Survival and Axon Targeting. *Front. Mol. Neurosci.* 9, 155. doi:10.3389/fnmol.2016.00155

Hattori, D., Chen, Y., Matthews, B. J., Salwinski, L., Sabatti, C., Grueber, W. B., et al. (2009). Robust Discrimination between Self and Non-self Neurites Requires Thousands of Dscam1 Isoforms. *Nature* 461, 644–648. doi:10.1038/nature08431

Hattori, D., Demir, E., Kim, H. W., Viragh, E., Zipursky, S. L., and Dickson, B. J. (2007). Dscam Diversity Is Essential for Neuronal Wiring and Self-Recognition. *Nature* 449, 223–227. doi:10.1038/nature06099

Haussmann, I. U., Ustaoglu, P., Brauer, U., Hemani, Y., Dix, T. C., and Soller, M. (2019). Plasmid-based gap-repair Recombineered Transgenes Reveal a central Role for Introns in Mutually Exclusive Alternative Splicing in Down Syndrome Cell Adhesion Molecule Exon 4. *Nucleic Acids Res.* 47, 1389–1403. doi:10.1093/nar/gky1254

Hemani, Y., and Soller, M. (2012). Mechanisms of Drosophila Dscam Mutually Exclusive Splicing Regulation. *Biochem. Soc. Trans.* 40, 804–809. doi:10.1042/BST20120060

Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., et al. (2008). Telomeric Expression Sites Are Highly Conserved in Trypanosoma Brucei. *PLOS ONE* 3, e3527. doi:10.1371/journal.pone.0003527

Hetz, C., Zhang, K., and Kaufman, R. J. (2020). Mechanisms, Regulation and Functions of the Unfolded Protein Response. *Nat. Rev. Mol. Cel. Biol.* 21, 421–438. doi:10.1038/s41580-020-0250-z

Hewitt, S. L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., et al. (2009). RAG-1 and ATM Coordinate Monoallelic Recombination and Nuclear Positioning of Immunoglobulin Loci. *Nat. Immunol.* 10, 655–664. doi:10.1038/ni.1735

Hing, H., Xiao, J., Harden, N., Lim, L., and Zipursky, S. L. (1999). Pak Functions Downstream of Dock to Regulate Photoreceptor Axon Guidance in Drosophila. *Cell* 97, 853–863. doi:10.1016/S0092-8674(00)80798-9

Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., and Yagi, T. (2012). Single-neuron Diversity Generated by Protocadherin-β Cluster in Mouse central and Peripheral Nervous Systems. *Front. Mol. Neurosci.* 5, 90. doi:10.3389/fnmol.2012.00090

Hirayama, T., and Yagi, T. (2017). Regulation of Clustered Protocadherin Genes in Individual Neurons. *Semin. Cel Developmental Biol.* 69, 122–130. doi:10.1016/j.semcdb.2017.05.026

Hirota, J., and Mombaerts, P. (2004). The LIM-Homeodomain Protein Lhx2 Is Required for Complete Development of Mouse Olfactory Sensory Neurons. *Proc. Natl. Acad. Sci.* 101, 8751–8755. doi:10.1073/pnas.0400940101

Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P., and Cross, G. a. M. (1980). Novel Expression-Linked Copies of the Genes for Variant Surface Antigens in Trypanosomes. *Nature* 284, 78–80. doi:10.1038/284078a0

Horn, D. (2014). Antigenic Variation in African Trypanosomes. *Mol. Biochem. Parasitol.* 195, 123–129. doi:10.1016/j.molbiopara.2014.05.001

Horn, D., and Cross, G. A. M. (1997). Analysis of Trypanosoma Brucei Vsg Expression Site Switching *In Vitro*. *Mol. Biochem. Parasitol.* 84, 189–201. doi:10.1016/s0166-6851(96)02794-6

Horn, D., and McCulloch, R. (2010). Molecular Mechanisms Underlying the Control of Antigenic Variation in African Trypanosomes. *Curr. Opin. Microbiol.* 13, 700–705. doi:10.1016/j.mib.2010.08.009

Hovel-Miner, G. A., Boothroyd, C. E., Mugnier, M., Dreesen, O., Cross, G. A. M., and Papavasiliou, F. N. (2012). Telomere Length Affects the Frequency and Mechanism of Antigenic Variation in Trypanosoma Brucei. *Plos Pathog.* 8, e1002900. doi:10.1371/journal.ppat.1002900

Hughes, M. E., Bortnick, R., Tsubouchi, A., Bäumer, P., Kondo, M., Uemura, T., et al. (2007). Homophilic Dscam Interactions Control Complex Dendrite Morphogenesis. *Neuron* 54, 417–427. doi:10.1016/j.neuron.2007.04.013

Hummel, T., Vasconcelos, M. L., Clemens, J. C., Fishilevich, Y., Vosshall, L. B., and Zipursky, S. L. (2003). Axonal Targeting of Olfactory Receptor Neurons in Drosophila Is Controlled by Dscam. *Neuron* 37, 221–231. doi:10.1016/s0896-6273(02)01183-2

Hutchinson, S., Foulon, S., Crouzols, A., Menafra, R., Rotureau, B., Griffiths, A. D., et al. (2021). The Establishment of Variant Surface Glycoprotein Monoallelic Expression Revealed by Single-Cell RNA-Seq ofTrypanosoma Bruceiin the Tsetse Fly Salivary Glands. *BioRxiv* 1, 1. doi:10.1101/2021.03.01.433049

Iakovlev, M., Faravelli, S., and Becskei, A. (2021). Gene Families with Stochastic Exclusive Gene Choice Underlie Cell Adhesion in Mammalian Cells. *Front. Cel. Dev. Biol.* 9. doi:10.3389/fcell.2021.642212

Imai, T., Suzuki, M., and Sakano, H. (2006). Odorant Receptor–Derived cAMP Signals Direct Axonal Targeting. *Science* 314, 657–661. doi:10.1126/science.1131794

Ing-Esteves, S., Kostadinov, D., Marocha, J., Sing, A. D., Joseph, K. S., Laboulaye, M. A., et al. (2018). Combinatorial Effects of Alpha- and Gamma-Protocadherins on Neuronal Survival and Dendritic Self-Avoidance. *J. Neurosci.* 38, 2713–2729. doi:10.1523/JNEUROSCI.3035-17.2018

Jacob, J., Przylepa, J., Miller, C., and Kelsoe, G. (1993). *In Situ* studies of the Primary Immune Response to (4-Hydroxy-3-Nitrophenyl)acetyl. III. The Kinetics of V Region Mutation and Selection in Germinal center B Cells. *J. Exp. Med.* 178, 1293–1307. doi:10.1084/jem.178.4.1293

Jayaraman, S., Harris, C., Paxton, E., Donachie, A.-M., Vaikkinen, H., McCulloch, R., et al. (2019). Application of Long Read Sequencing to Determine Expressed Antigen Diversity in Trypanosoma Brucei Infections. *Plos Negl. Trop. Dis.* 13, e0007262. doi:10.1371/journal.pntd.0007262

Jin, Y., Dong, H., Shi, Y., and Bian, L. (2018). Mutually Exclusive Alternative Splicing of Pre-mRNAs. *WIREs RNA* 9, e1468. doi:10.1002/wrna.1468

Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F. W. (2006). Mechanism and Control of V(d)j Recombination at the Immunoglobulin Heavy Chain Locus. *Annu. Rev. Immunol.* 24, 541–570. doi:10.1146/annurev.immunol.23.021704.115830

Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., et al. (2006). Allelic Gene Regulation of Pcdh-α and Pcdh-γ Clusters Involving Both Monoallelic and Biallelic Expression in Single Purkinje Cells. *J. Biol. Chem.* 281, 30551–30560. doi:10.1074/jbc.M605677200

Kassem, A., Pays, E., and Vanhamme, L. (2014). Transcription Is Initiated on Silent Variant Surface Glycoprotein Expression Sites Despite Monoallelic Expression in Trypanosoma Brucei. *Proc. Natl. Acad. Sci.* 111, 8943–8948. doi:10.1073/pnas.1404873111

Kawahara, T., Siegel, T. N., Ingram, A. K., Alsford, S., Cross, G. A. M., and Horn, D. (2008). Two Essential MYST-Family Proteins Display Distinct Roles in Histone H4K10 Acetylation and Telomeric Silencing in Trypanosomes. *Mol. Microbiol.* 69, 1054–1068. doi:10.1111/j.1365-2958.2008.06346.x

Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory Elements Required for the Activation and Repression of the Protocadherin-Gene Cluster. *Proc. Natl. Acad. Sci.* 108, 17195–17200. doi:10.1073/pnas.1114357108

Khamlichi, A. A., and Feil, R. (2018). Parallels between Mammalian Mechanisms of Monoallelic Gene Expression. *Trends Genet.* 34 (12), 954–971. doi:10.1016/j.tig.2018.08.005

Khan, M., Vaes, E., and Mombaerts, P. (2011). Regulation of the Probability of Mouse Odorant Receptor Gene Choice. *Cell* 147, 907–921. doi:10.1016/j.cell.2011.09.049

Kim, H.-S., and Cross, G. A. M. (2011). Identification of Trypanosoma Brucei RMI1/BLAP75 Homologue and its Roles in Antigenic Variation. *PLoS One* 6, e25313. doi:10.1371/journal.pone.0025313

Kim, H.-S., and Cross, G. A. M. (2010). TOPO3α Influences Antigenic Variation by Monitoring Expression-Site-Associated VSG Switching in Trypanosoma Brucei. *Plos Pathog.* 6, e1000992. doi:10.1371/journal.ppat.1000992

Kirkham, J. K., Park, S. H., Nguyen, T. N., Lee, J. H., and Günzl, A. (2016). Dynein Light Chain LC8 Is Required for RNA Polymerase I-Mediated Transcription in Trypanosoma Brucei, Facilitating Assembly and Promoter Binding of Class I Transcription Factor A. *Mol. Cel. Biol.* 36, 95–107. doi:10.1128/MCB.00705-15

Klein, L., Kyewski, B., Allen, P. M., and Hogquist, K. A. (2014). Positive and Negative Selection of the T Cell Repertoire: what Thymocytes See (And Don't See). *Nat. Rev. Immunol.* 14, 377–391. doi:10.1038/nri3667

Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T., and Sanes, J. R. (2012). Protocadherins Mediate Dendritic Self-Avoidance in the Mammalian Nervous System. *Nature* 488, 517–521. doi:10.1038/nature11305

Lewcock, J. W., and Reed, R. R. (2004). A Feedback Mechanism Regulates Monoallelic Odorant Receptor Expression. *Proc. Natl. Acad. Sci.* 101, 1069–1074. doi:10.1073/pnas.0307986100

Le Gros, M. A., Clowney, E. J., Magklara, A., Yen, A., Markenscoff-Papadimitriou, E., Colquitt, B., et al. (2016). Soft X-Ray Tomography Reveals Gradual Chromatin Compaction and Reorganization during Neurogenesis *In Vivo*. *Cel. Rep.* 17, 2125–2136. doi:10.1016/j.celrep.2016.10.060

Li, B. (2015). DNA Double-Strand Breaks and Telomeres Play Important Roles in Trypanosoma Brucei Antigenic Variation. *Eukaryot. Cel.* 14, 196–205. doi:10.1128/EC.00207-14

Li, F., and Eckhardt, L. A. (2009). A Role for the IgH Intronic Enhancer Eμ in Enforcing Allelic Exclusion. *J. Exp. Med.* 206, 153–167. doi:10.1084/jem.20081202

Li, J., Ishii, T., Feinstein, P., and Mombaerts, P. (2004). Odorant Receptor Gene Choice Is Reset by Nuclear Transfer from Mouse Olfactory Sensory Neurons. *Nature* 428, 393–399. doi:10.1038/nature02433

Liang, H.-E., Hsu, L.-Y., Cado, D., and Schlissel, M. S. (2004). Variegated Transcriptional Activation of the Immunoglobulin κ Locus in Pre-B Cells Contributes to the Allelic Exclusion of Light-Chain Expression. *Cell* 118, 19–29. doi:10.1016/j.cell.2004.06.019

Little, A. J., Matthews, A., Oettinger, M., Roth, D. B., and Schatz, D. G. (2015). "The Mechanism of V(D)J Recombination," in *Molecular Biology of B Cells*. Editors F. W. Alt, T. Honjo, A. Radbruch, and M. Reth. Second Edition (London: Academic Press), 13–34. doi:10.1016/B978-0-12-397933-9.00002-3

Liu, A. Y. C., Van der Ploeg, L. H. T., Rijsewijk, F. A. M., Borst, P., and Chambon, P. (1983). The Transposition Unit of Variant Surface Glycoprotein Gene 118 of Trypanosoma Brucei. *J. Mol. Biol.* 167, 57–75. doi:10.1016/s0022-2836(83)80034-5

Liu, H., Pizzano, S., Li, R., Zhao, W., Veling, M. W., Hu, Y., et al. (2020). isoTarget: A Genetic Method for Analyzing the Functional Diversity of Splicing Isoforms *In Vivo*. *Cel. Rep.* 33, 108361. doi:10.1016/j.celrep.2020.108361

López-Farfán, D., Bart, J.-M., Rojas-Barros, D. I., and Navarro, M. (2014). SUMOylation by the E3 Ligase TbSIZ1/PIAS1 Positively Regulates VSG Expression in Trypanosoma Brucei. *Plos Pathog.* 10, e1004545. doi:10.1371/journal.ppat.1004545

Lyons, D. B., Allen, W. E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An Epigenetic Trap Stabilizes Singular Olfactory Receptor Expression. *Cell* 154, 325–336. doi:10.1016/j.cell.2013.06.039

Lyons, D. B., Magklara, A., Goh, T., Sampath, S. C., Schaefer, A., Schotta, G., et al. (2014). Heterochromatin-mediated Gene Silencing Facilitates the Diversification of Olfactory Neurons. *Cel. Rep.* 9, 884–892. doi:10.1016/j.celrep.2014.10.001

Machado, C. F., Nagai, M. H., Lyra, C. S., Reis-Silva, T. M., Xavier, A. M., Glezer, I., et al. (2017). Conditional Deletion of Ric-8b in Olfactory Sensory Neurons Leads to Olfactory Impairment. *J. Neurosci.* 37, 12202–12213. doi:10.1523/JNEUROSCI.0943-17.2017

Magklara, A., and Lomvardas, S. (2013). Stochastic Gene Expression in Mammals: Lessons from Olfaction. *Trends Cell Biology* 23 (9), 449–456. doi:10.1016/j.tcb.2013.04.005

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., et al. (2011). An Epigenetic Signature for Monoallelic Olfactory Receptor Expression. *Cell* 145, 555–570. doi:10.1016/j.cell.2011.03.040

Mah, K. M., and Weiner, J. A. (2016). "Clustered Protocadherins," in *The Cadherin Superfamily: Key Regulators of Animal Development and Physiology*. Editors S. T. Suzuki and S. Hirano (Tokyo: Springer Japan), 195–221. doi:10.1007/978-4-431-56033-3_8

Malnic, B., Godfrey, P. A., and Buck, L. B. (2004). The Human Olfactory Receptor Gene Family. *Proc. Natl. Acad. Sci.* 101, 2584–2589. doi:10.1073/pnas.0307882100

Malnic, B., Hirono, J., Sato, T., and Buck, L. B. (1999). Combinatorial Receptor Codes for Odors. *Cell* 96, 713–723. doi:10.1016/S0092-8674(00)80581-4

Mancini, M., Bassani, S., and Passafaro, M. (2020). Right Place at the Right Time: How Changes in Protocadherins Affect Synaptic Connections Contributing to the Etiology of Neurodevelopmental Disorders. *Cells* 9, 2711. doi:10.3390/cells9122711

Markenscoff-Papadimitriou, E., Allen, W. E., Colquitt, B. M., Goh, T., Murphy, K. K., Monahan, K., et al. (2014). Enhancer Interaction Networks as a Means for Singular Olfactory Receptor Expression. *Cell* 159, 543–557. doi:10.1016/j.cell.2014.09.033

Massah, S., Beischlag, T. V., and Prefontaine, G. G. (2015). Epigenetic Events Regulating Monoallelic Gene Expression. *Crit. Rev. Biochem. Mol. Biol.* 50, 337–358. doi:10.3109/10409238.2015.1064350

Matthews, B. J., Kim, M. E., Flanagan, J. J., Hattori, D., Clemens, J. C., Zipursky, S. L., et al. (2007). Dendrite Self-Avoidance Is Controlled by Dscam. *Cell* 129, 593–604. doi:10.1016/j.cell.2007.04.013

Matthews, K. R. (2005). The Developmental Cell Biology of Trypanosoma Brucei. *J. Cel. Sci.* 118, 283–290. doi:10.1242/jcs.01649

May, G. E., Olson, S., McManus, C. J., and Graveley, B. R. (2011). Competing RNA Secondary Structures Are Required for Mutually Exclusive Splicing of the Dscam Exon 6 Cluster. *Rna* 17, 222–229. doi:10.1261/rna.2521311

McCulloch, R., and Barry, J. D. (1999). A Role for RAD51 and Homologous Recombination in Trypanosoma Brucei Antigenic Variation. *Genes Development* 13, 2875–2888. doi:10.1101/gad.13.21.2875

McCulloch, R., Cobbold, C. A., Figueiredo, L., Jackson, A., Morrison, L. J., Mugnier, M. R., et al. (2017). Emerging Challenges in Understanding Trypanosome Antigenic Variation. *Emerg. Top. Life Sci.* 1, 585–592. doi:10.1042/ETLS20170104

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W., and Weigert, M. (1984). Generation of Antibody Diversity in the Immune Response of BALB/c Mice to Influenza Virus Hemagglutinin. *Proc. Natl. Acad. Sci.* 81, 3180–3184. doi:10.1073/pnas.81.10.3180

Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J.-h., Walz, T., Wang, J.-h., et al. (2007). Structural Basis of Dscam Isoform Specificity. *Nature* 449, 487–491. doi:10.1038/nature06147

Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duveau, F., and Wittkopp, P. J. (2015). Selection on Noise Constrains Variation in a Eukaryotic Promoter. *Nature* 521 (7552), 344–347. doi:10.1038/nature14244

Millard, S. S., Lu, Z., Zipursky, S. L., and Meinertzhagen, I. A. (2010). Drosophila Dscam Proteins Regulate Postsynaptic Specificity at Multiple-Contact Synapses. *Neuron* 67, 761–768. doi:10.1016/j.neuron.2010.08.030

Miura, S. K., Martins, A., Zhang, K. X., Graveley, B. R., and Zipursky, S. L. (2013). Probabilistic Splicing of Dscam1 Establishes Identity at the Level of Single Neurons. *Cell* 155, 1166–1177. doi:10.1016/j.cell.2013.10.018

Molumby, M. J., Keeler, A. B., and Weiner, J. A. (2016). Homophilic Protocadherin Cell-Cell Interactions Promote Dendrite Complexity. *Cel. Rep.* 15, 1037–1050. doi:10.1016/j.celrep.2016.03.093

Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-Mediated Trans Interactions Regulate Olfactory Receptor Choice. *Nature* 565, 448–453. doi:10.1038/s41586-018-0845-0

Monahan, K., and Lomvardas, S. (2015). Monoallelic Expression of Olfactory Receptors. *Annu. Rev. Cel. Dev. Biol.* 31, 721–740. doi:10.1146/annurev-cellbio-100814-125308

Monahan, K., Rudnick, N. D., Kehayova, P. D., Pauli, F., Newberry, K. M., Myers, R. M., et al. (2012). Role of CCCTC Binding Factor (CTCF) and Cohesin in the Generation of Single-Cell Diversity of Protocadherin- Gene Expression. *Proc. Natl. Acad. Sci.* 109, 9125–9130. doi:10.1073/pnas.1205074109

Monahan, K., Schieren, I., Cheung, J., Mumbey-Wafula, A., Monuki, E. S., and Lomvardas, S. (2017). Cooperative Interactions Enable Singular Olfactory Receptor Expression in Mouse Olfactory Neurons. *eLife* 6. doi:10.7554/eLife.28620

Morishita, H., and Yagi, T. (2007). Protocadherin Family: Diversity, Structure, and Function. *Curr. Opin. Cel. Biol.* 19, 584–592. doi:10.1016/j.ceb.2007.09.006

Morrison, L. J., Majiwa, P., Read, A. F., and Barry, J. D. (2005). Probabilistic Order in Antigenic Variation of Trypanosoma Brucei. *Int. J. Parasitol.* 35, 961–972. doi:10.1016/j.ijpara.2005.05.004

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., et al. (2001). Asynchronous Replication and Allelic Exclusion in the Immune System. *Nature* 414, 221–225. doi:10.1038/35102606

Mountoufaris, G., Chen, W. V., Hirabayashi, Y., O'Keeffe, S., Chevee, M., Nwakeze, C. L., et al. (2017). Multicluster Pcdh Diversity Is Required for Mouse Olfactory Neural Circuit Assembly. *Science* 356, 411–414. doi:10.1126/science.aai8801

Mugnier, M. R., Cross, G. A. M., and Papavasiliou, F. N. (2015). The *In Vivo* Dynamics of Antigenic Variation in Trypanosoma Brucei. *Science* 347, 1470–1473. doi:10.1126/science.aaa4502

Mugnier, M. R., Stebbins, C. E., and Papavasiliou, F. N. (2016). Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma Brucei. *Plos Pathog.* 12, e1005784. doi:10.1371/journal.ppat.1005784

Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., et al. (2018). Genome Organization and DNA Accessibility Control Antigenic Variation in Trypanosomes. *Nature* 563, 121–125. doi:10.1038/s41586-018-0619-8

Nagai, M. H., Armelin-Correa, L. M., and Malnic, B. (2016). Monogenic and Monoallelic Expression of Odorant Receptors. *Mol. Pharmacol.* 90, 633–639. doi:10.1124/mol.116.104745

Narayanan, M. S., Kushwaha, M., Ersfeld, K., Fullbrook, A., Stanne, T. M., and Rudenko, G. (2011). NLP Is a Novel Transcription Regulator Involved in VSG Expression Site Control in Trypanosoma Brucei. *Nucleic Acids Res.* 39, 2018–2031. doi:10.1093/nar/gkq950

Navarro, M., and Gull, K. (2001). A Pol I Transcriptional Body Associated with VSG Mono-Allelic Expression in Trypanosoma Brucei. *Nature* 414, 759–763. doi:10.1038/414759a

Nemazee, D. (2017). Mechanisms of central Tolerance for B Cells. *Nat. Rev. Immunol.* 17, 281–294. doi:10.1038/nri.2017.19

Neves, G., Zucker, J., Daly, M., and Chess, A. (2004). Stochastic yet Biased Expression of Multiple Dscam Splice Variants by Individual Cells. *Nat. Genet.* 36, 240–246. doi:10.1038/ng1299

Nguyen, M. Q., Zhou, Z., Marks, C. A., Ryba, N. J. P., and Belluscio, L. (2007). Prominent Roles for Odorant Receptor Coding Sequences in Allelic Exclusion. *Cell* 131, 1009–1017. doi:10.1016/j.cell.2007.10.050

Nguyen, T. N., Müller, L. S. M., Park, S. H., Siegel, T. N., and Günzl, A. (2014). Promoter Occupancy of the Basal Class I Transcription Factor A Differs Strongly between Active and Silent VSG Expression Sites in Trypanosoma Brucei. *Nucleic Acids Res.* 42, 3164–3176. doi:10.1093/nar/gkt1301

Nguyen, T. N., Nguyen, B. N., Lee, J. H., Panigrahi, A. K., and Günzl, A. (2012). Characterization of a Novel Class I Transcription Factor A (CITFA) Subunit that Is Indispensable for Transcription by the Multifunctional RNA Polymerase I of Trypanosoma Brucei. *Eukaryot. Cel.* 11, 1573–1581. doi:10.1128/EC.00250-12

Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A., and Sakano, H. (2007). Deletion of the Core-H Region in Mice Abolishes the Expression of Three Proximal Odorant Receptor Genes in Cis. *Proc. Natl. Acad. Sci.* 104, 20067–20072. doi:10.1073/pnas.0706544105

Nozawa, M., and Nei, M. (2009). Genomic Drift and Copy Number Variation of Chemosensory Receptor Genes in Humans and Mice. *Cytogenet. Genome Res.* 123, 263–269. doi:10.1159/000184716

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, Adjacent Genes that Synergistically Activate V(D)J Recombination. *Science* 248, 1517–1523. doi:10.1126/science.2360047

Okada, T., Miller, M. J., Parker, I., Krummel, M. F., Neighbors, M., Hartley, S. B., et al. (2005). Antigen-engaged B Cells Undergo Chemotaxis toward the T Zone and Form Motile Conjugates with Helper T Cells. *Plos Biol.* 3, e150. doi:10.1371/journal.pbio.0030150

Olender, T., Lancet, D., and Nebert, D. W. (2008). Update on the Olfactory Receptor (OR) Gene Superfamily. *Hum. Genomics* 3, 87. doi:10.1186/1479-7364-3-1-87

Olson, S., Blanchette, M., Park, J., Savva, Y., Yeo, G. W., Yeakley, J. M., et al. (2007). A Regulator of Dscam Mutually Exclusive Splicing Fidelity. *Nat. Struct. Mol. Biol.* 14, 1134–1140. doi:10.1038/nsmb1339

Outters, P., Jaeger, S., Zaarour, N., and Ferrier, P. (2015). "Long-Range Control of V(D)J Recombination & Allelic Exclusion," in *Advances in Immunology, Molecular Mechanisms that Orchestrate the Assembly of Antigen Receptor Loci*. Editor C. Murre (Academic Press), 363–413. doi:10.1016/bs.ai.2015.08.002

Papavasiliou, F. N., and Schatz, D. G. (2000). Cell-cycle-regulated DNA Double-Strand Breaks in Somatic Hypermutation of Immunoglobulin Genes. *Nature* 408, 216–221. doi:10.1038/35041599

Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A., and Graveley, B. R. (2004). Identification of Alternative Splicing Regulators by RNA Interference in Drosophila. *Proc. Natl. Acad. Sci.* 101, 15974–15979. doi:10.1073/pnas.0407004101

Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Pérez-Morga, D. (2001). The VSG Expression Sites of Trypanosoma Brucei: Multipurpose Tools for the Adaptation of the Parasite to Mammalian Hosts. *Mol. Biochem. Parasitol.* 114, 1–16. doi:10.1016/S0166-6851(01)00242-0

Perry, R. P., Kelley, D. E., Coleclough, C., Seidman, J. G., Leder, P., Tonegawa, S., et al. (1980). Transcription of Mouse Kappa Chain Genes: Implications for Allelic Exclusion. *Proc. Natl. Acad. Sci.* 77, 1937–1941. doi:10.1073/pnas.77.4.1937

P. Ferrier (Editor) (2009). *V(D)J Recombination, Advances in Experimental Medicine and Biology* (New York: Springer-Verlag). doi:10.1007/978-1-4419-0296-2

Pilzecker, B., and Jacobs, H. (2019). Mutating for Good: DNA Damage Responses during Somatic Hypermutation. *Front. Immunol.* 10. doi:10.3389/fimmu.2019.00438

Povelones, M. L., Gluenz, E., Dembek, M., Gull, K., and Rudenko, G. (2012). Histone H1 Plays a Role in Heterochromatin Formation and VSG Expression Site Silencing in Trypanosoma Brucei. *Plos Pathog.* 8, e1003010. doi:10.1371/journal.ppat.1003010

Raj, A., Rifkin, S. A., Andersen, E., and Van Oudenaarden, A. (2010). Variability in Gene Expression Underlies Incomplete Penetrance. *Nature* 463 (7283), 913–918. doi:10.1038/nature08781

Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., et al. (2016). Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma Brucei. *Plos Genet.* 12, e1005758. doi:10.1371/journal.pgen.1005758

Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of Long-Range Regulatory Elements in the Protocadherin- Gene Cluster. *Proc. Natl. Acad. Sci.* 103, 19719–19724. doi:10.1073/pnas.0609445104

Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988). The Accuracy of Reverse Transcriptase from HIV-1. *Science* 242 (4882), 1171–1173. doi:10.1126/science.2460925

Robin, J. D., Ludlow, A. T., Batten, K., Magdinier, F., Stadler, G., Wagner, K. R., et al. (2014). Telomere Position Effect: Regulation of Gene Expression with Progressive Telomere Shortening over Long Distances. *Genes Dev.* 28, 2464–2476. doi:10.1101/gad.251041.114

Robinson, N. P., Burman, N., Melville, S. E., and Barry, J. D. (1999). Predominance of Duplicative VSG Gene Conversion in Antigenic Variation in African Trypanosomes. *Mol. Cel. Biol.* 19, 5839–5846. doi:10.1128/MCB.19.9.5839

Roby, Y. A., Bushey, M. A., Cheng, L. E., Kulaga, H. M., Lee, S.-J., and Reed, R. R. (2012). Zfp423/OAZ Mutation Reveals the Importance of Olf/EBF Transcription Activity in Olfactory Neuronal Maturation. *J. Neurosci.* 32, 13679–13688a. doi:10.1523/JNEUROSCI.6190-11.2012

Rodriguez, O. L., Gibson, W. S., Parks, T., Emery, M., Powell, J., Strahl, M., et al. (2020). A Novel Framework for Characterizing Genomic Haplotype Diversity in the Human Immunoglobulin Heavy Chain Locus. *Front. Immunol.* 11. doi:10.3389/fimmu.2020.02136

Ron, D., and Walter, P. (2007). Signal Integration in the Endoplasmic Reticulum Unfolded Protein Response. *Nat. Rev. Mol. Cel. Biol.* 8, 519–529. doi:10.1038/nrm2199

Rose, N. R. (2017). Negative Selection, Epitope Mimicry and Autoimmunity. *Curr. Opin. Immunol.* 49, 51–55. doi:10.1016/j.coi.2017.08.014

Rothman, A., Feinstein, P., Hirota, J., and Mombaerts, P. (2005). The Promoter of the Mouse Odorant Receptor Gene M71. *Mol. Cell Neurosci.* 28, 535–546. doi:10.1016/j.mcn.2004.11.006

Roy, A. L., Sen, R., and Roeder, R. G. (2011). Enhancer-promoter Communication and Transcriptional Regulation of Igh. *Trends Immunol.* 32, 532–539. doi:10.1016/j.it.2011.06.012

Rudenko, G., McCulloch, R., Dirks-Mulder, A., and Borst, P. (1996). Telomere Exchange Can Be an Important Mechanism of Variant Surface Glycoprotein Gene Switching in Trypanosoma Brucei. *Mol. Biochem. Parasitol.* 80, 65–75. doi:10.1016/0166-6851(96)02669-2

Sachse, S. M., Lievens, S., Ribeiro, L. F., Dascenco, D., Masschaele, D., Horré, K., et al. (2019). Nuclear Import of the DSCAM -cytoplasmic Domain Drives Signaling Capable of Inhibiting Synapse Formation. *EMBO J.* 38, e99669. doi:10.15252/embj.201899669

Saura, A., Iribarren, P. A., Rojas-Barros, D., Bart, J. M., López-Farfán, D., Andrés-León, E., et al. (2019). SUMOylated SNF2PH Promotes Variant Surface Glycoprotein Expression in Bloodstream Trypanosomes. *EMBO Rep.* 20, e48029. doi:10.15252/embr.201948029

Sawaya, M. R., Wojtowicz, W. M., Andre, I., Qian, B., Wu, W., Baker, D., et al. (2008). A Double S Shape Provides the Structural Basis for the Extraordinary Binding Specificity of Dscam Isoforms. *Cell* 134, 1007–1018. doi:10.1016/j.cell.2008.07.042

Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). The V(D)J Recombination Activating Gene, RAG-1. *Cell* 59, 1035–1048. doi:10.1016/0092-8674(89)90760-5

Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: Complex Genes in Simple Animals, Complex Animals yet Simple Genes. *Genes Dev.* 23, 147–156. doi:10.1101/gad.1752909

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., et al. (2000). Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell* 101, 671–684. doi:10.1016/S0092-8674(00)80878-8

Schreiner, D., and Weiner, J. A. (2010). Combinatorial Homophilic Interaction between -protocadherin Multimers Greatly Expands the Molecular Diversity of Cell Adhesion. *Proc. Natl. Acad. Sci.* 107, 14893–14898. doi:10.1073/pnas.1004526107

Schroeder, H. W., and Cavacini, L. (2010). Structure and Function of Immunoglobulins. *J. Allergy Clin. Immunol.* 125, S41–S52. doi:10.1016/j.jaci.2009.09.046

Schulz, D., and Papavasiliou, F. N. (2016). The VEXing Problem of Monoallelic Expression in the African Trypanosome. *Proc. Natl. Acad. Sci. USA* 113, 7017–7019. doi:10.1073/pnas.1608546113

Schulz, D., Zaringhalam, M., Papavasiliou, F. N., and Kim, H.-S. (2016). Base J and H3.V Regulate Transcriptional Termination in Trypanosoma Brucei. *Plos Genet.* 12, e1005762. doi:10.1371/journal.pgen.1005762

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., et al. (2003). Negative Feedback Regulation Ensures the One Receptor-One Olfactory Neuron Rule in Mouse. *Science* 302, 2088–2094. doi:10.1126/science.1089122

Serwe, M., and Sablitzky, F. (1993). V(D)J Recombination in B Cells Is Impaired but Not Blocked by Targeted Deletion of the Immunoglobulin Heavy Chain Intron Enhancer. *EMBO J.* 12, 2321–2327. doi:10.1002/j.1460-2075.1993.tb05886.x

Sharma, R., Ishimaru, Y., Davison, I., Ikegami, K., Chien, M.-S., You, H., et al. (2017). Olfactory Receptor Accessory Proteins Play Crucial Roles in Receptor Function and Gene Choice. *eLife* 6, e21895. doi:10.7554/eLife.21895

Shykind, B. M. (2005). Regulation of Odorant Receptors: One Allele at a Time. *Hum. Mol. Genet.* 14, R33–R39. doi:10.1093/hmg/ddi105

Sima, N., McLaughlin, E. J., Hutchinson, S., and Glover, L. (2019). Escaping the Immune System by DNA Repair and Recombination in African Trypanosomes. *Open Biol.* 9, 190182. doi:10.1098/rsob.190182

Singh, N., Bergman, Y., Cedar, H., and Chess, A. (2003). Biallelic Germline Transcription at the κ Immunoglobulin Locus. *J. Exp. Med.* 197, 743–750. doi:10.1084/jem.20021392

Skok, J. A., Brown, K. E., Azuara, V., Caparros, M.-L., Baxter, J., Takacs, K., et al. (2001). Nonequivalent Nuclear Location of Immunoglobulin Alleles in B Lymphocytes. *Nat. Immunol.* 2, 848–854. doi:10.1038/ni0901-848

Soba, P., Zhu, S., Emoto, K., Younger, S., Yang, S.-J., Yu, H.-H., et al. (2007). Drosophila Sensory Neurons Require Dscam for Dendritic Self-Avoidance and Proper Dendritic Field Organization. *Neuron* 54, 403–416. doi:10.1016/j.neuron.2007.03.029

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., et al. (2009). Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* 137, 356–368. doi:10.1016/j.cell.2009.01.052

Spehr, M., and Munger, S. D. (2009). Olfactory Receptors: G Protein-Coupled Receptors and beyond. *J. Neurochem.* 109, 1570–1583. doi:10.1111/j.1471-4159.2009.06085.x

Spitz, F. (2019). Chromosomes Come Together to Help Mice Distinguish Odours. *Nature* 565, 439–440. doi:10.1038/d41586-019-00010-6

Stanne, T. M., and Rudenko, G. (2010). Active VSG Expression Sites in Trypanosoma Brucei Are Depleted of Nucleosomes. *Eukaryot. Cel.* 9, 136–147. doi:10.1128/EC.00281-09

Tan, L., Li, Q., and Xie, X. S. (2015). Olfactory Sensory Neurons Transiently Express Multiple Olfactory Receptors during Development. *Mol. Syst. Biol.* 11, 844. doi:10.15252/msb.20156639

Tan, L., Xing, D., Daley, N., and Xie, X. S. (2019). Three-dimensional Genome Structures of Single Sensory Neurons in Mouse Visual and Olfactory Systems. *Nat. Struct. Mol. Biol.* 26, 297–307. doi:10.1038/s41594-019-0205-2

Tas, J. M. J., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J. T., Mano, Y. M., et al. (2016). Visualizing Antibody Affinity Maturation in Germinal Centers. *Science* 351, 1048–1054. doi:10.1126/science.aad3439

Tasic, B., Nabholz, C. E., Baldwin, K. K., Kim, Y., Rueckert, E. H., Ribich, S. A., et al. (2002). Promoter Choice Determines Splice Site Selection in Protocadherin α and γ Pre-mRNA Splicing. *Mol. Cel.* 10, 21–33. doi:10.1016/S1097-2765(02)00578-6

Thu, C. A., Chen, W. V., Rubinstein, R., Chevee, M., Wolcott, H. N., Felsovalyi, K. O., et al. (2014). Single-Cell Identity Generated by Combinatorial Homophilic Interactions between α, β, and γ Protocadherins. *Cell* 158, 1045–1059. doi:10.1016/j.cell.2014.07.012

Turner, C. M. R., and Barry, J. D. (1989). High Frequency of Antigenic Variation inTrypanosoma Brucei Rhodesienseinfections. *Parasitology* 99, 67–75. doi:10.1017/S0031182000061035

Ustaoglu, P., Haussmann, I. U., Liao, H., Torres-Mendez, A., Arnold, R., Irimia, M., et al. (2019). Srrm234, but Not Canonical SR and hnRNP Proteins, Drive Inclusion of Dscam Exon 9 Variable Exons. *RNA* 25, 1353–1365. doi:10.1261/rna.071316.119

Vassalli, A., Feinstein, P., and Mombaerts, P. (2011). Homeodomain Binding Motifs Modulate the Probability of Odorant Receptor Gene Choice in Transgenic Mice. *Mol. Cell Neurosci.* 46, 381–396. doi:10.1016/j.mcn.2010.11.001

Vettermann, C., and Schlissel, M. S. (2010). Allelic Exclusion of Immunoglobulin Genes: Models and Mechanisms. *Immunol. Rev.* 237, 22–42. doi:10.1111/j.1600-065X.2010.00935.x

Victora, G. D., and Mesin, L. (2014). Clonal and Cellular Dynamics in Germinal Centers. *Curr. Opin. Immunol.* 28, 90–96. doi:10.1016/j.coi.2014.02.010

Victora, G. D., Schwickert, T. A., Fooksman, D. R., Kamphorst, A. O., Meyer-Hermann, M., Dustin, M. L., et al. (2010). Germinal Center Dynamics Revealed by Multiphoton Microscopy with a Photoactivatable Fluorescent Reporter. *Cell* 143, 592–605. doi:10.1016/j.cell.2010.10.032

Von Dannecker, L. E. C., Mercadante, A. F., and Malnic, B. (2006). Ric-8B Promotes Functional Expression of Odorant Receptors. *Proc. Natl. Acad. Sci.* 103, 9310–9314. doi:10.1073/pnas.0600697103

Wang, J., Ma, X., Yang, J. S., Zheng, X., Zugates, C. T., Lee, C.-H. J., et al. (2004a). Transmembrane/Juxtamembrane Domain-dependent Dscam Distribution and Function during Mushroom Body Neuronal Morphogenesis. *Neuron* 43, 663–672. doi:10.1016/j.neuron.2004.06.033

Wang, J., Zugates, C. T., Liang, I. H., Lee, C.-H. J., and Lee, T. (2002). Drosophila Dscam Is Required for Divergent Segregation of Sister Branches and Suppresses Ectopic Bifurcation of Axons. *Neuron* 33, 559–571. doi:10.1016/s0896-6273(02)00570-6

Wang, Q.-P., Kawahara, T., and Horn, D. (2010). Histone Deacetylases Play Distinct Roles in Telomeric VSG Expression Site Silencing in African Trypanosomes. *Mol. Microbiol.* 77, 1237–1245. doi:10.1111/j.1365-2958.2010.07284.x

Wang, S.-Z., Ou, J., Zhu, L. J., and Green, M. R. (2012a). Transcription Factor ATF5 Is Required for Terminal Differentiation and Survival of Olfactory Sensory Neurons. *Proc. Natl. Acad. Sci.* 109, 18589–18594. doi:10.1073/pnas.1210479109

Wang, S. S., Lewcock, J. W., Feinstein, P., Mombaerts, P., and Reed, R. R. (2004b). Genetic Disruptions of O/E2 and O/E3 Genes Reveal Involvement in Olfactory Receptor Neuron Projection. *Development* 131, 1377–1388. doi:10.1242/dev.01009

Wang, S. S., Tsai, R. Y. L., and Reed, R. R. (1997). The Characterization of the Olf-1/ebf-like HLH Transcription Factor Family: Implications in Olfactory Gene Regulation and Neuronal Development. *J. Neurosci.* 17, 4149–4158. doi:10.1523/JNEUROSCI.17-11-04149.1997

Wang, X., Li, G., Yang, Y., Wang, W., Zhang, W., Pan, H., et al. (2012b). An RNA Architectural Locus Control Region Involved in Dscam Mutually Exclusive Splicing. *Nat. Commun.* 3, 1255. doi:10.1038/ncomms2269

Watson, C. T., Steinberg, K. M., Graves, T. A., Warren, R. L., Malig, M., Schein, J., et al. (2015). Sequencing of the Human IG Light Chain Loci from a Hydatidiform Mole BAC Library Reveals Locus-specific Signatures of Genetic Diversity. *Genes Immun.* 16, 24–34. doi:10.1038/gene.2014.56

Weiler, E. (1965). Differential Activity of Allelic Gamma-Globulin Genes in Antibody-Producing Cells. *Proc. Natl. Acad. Sci.* 54, 1765–1772. doi:10.1073/pnas.54.6.1765

Weinberger, A. D., and Weinberger, L. S. (2013). Stochastic Fate Selection in HIV-Infected Patients. *Cell* 155 (3), 497–499. doi:10.1016/j.cell.2013.09.039

Weiner, J. A., and Jontes, J. D. (2013). Protocadherins, Not Prototypical: a Complex Tale of Their Interactions, Expression, and Functions. *Front. Mol. Neurosci.* 6, 4. doi:10.3389/fnmol.2013.00004

Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L., and Clemens, J. C. (2004). Alternative Splicing of Drosophila Dscam Generates Axon Guidance Receptors that Exhibit Isoform-specific Homophilic Binding. *Cell* 118, 619–633. doi:10.1016/j.cell.2004.08.021

Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S. L. (2007). A Vast Repertoire of Dscam Binding Specificities Arises from Modular Interactions of Variable Ig Domains. *Cell* 130, 1134–1145. doi:10.1016/j.cell.2007.08.026

Worby, C. A., Simonson-Leff, N., Clemens, J. C., Kruger, R. P., Muda, M., and Dixon, J. E. (2001). The Sorting Nexin, DSH3PX1, Connects the Axonal Guidance Receptor, Dscam, to the Actin Cytoskeleton. *J. Biol. Chem.* 276, 41782–41789. doi:10.1074/jbc.M107080200

Wu, Q., and Jia, Z. (2021). Wiring the Brain by Clustered Protocadherin Neural Codes. *Neurosci. Bull.* 37, 117–131. doi:10.1007/s12264-020-00578-4

Wu, Q., and Maniatis, T. (1999). A Striking Organization of a Large Family of Human Neural Cadherin-like Cell Adhesion Genes. *Cell* 97, 779–790. doi:10.1016/S0092-8674(00)80789-8

Wu, Q., Zhang, T., Cheng, J.-F., Kim, Y., Grimwood, J., Schmutz, J., et al. (2001). Comparative DNA Sequence Analysis of Mouse and Human Protocadherin Gene Clusters. *Genome Res.* 11, 389–404. doi:10.1101/gr.167301

Xu, B., Shi, Y., Wu, Y., Meng, Y., and Jin, Y. (2019). Role of RNA Secondary Structures in Regulating Dscam Alternative Splicing. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1862, 194381. doi:10.1016/j.bbagrm.2019.04.008

Yamagata, M., and Sanes, J. R. (2008). Dscam and Sidekick Proteins Direct Lamina-specific Synaptic Connections in Vertebrate Retina. *Nature* 451, 465–469. doi:10.1038/nature06469

Yang, X., Figueiredo, L. M., Espinal, A., Okubo, E., and Li, B. (2009). RAP1 Is Essential for Silencing Telomeric Variant Surface Glycoprotein Genes in Trypanosoma Brucei. *Cell* 137, 99–109. doi:10.1016/j.cell.2009.01.037

Yang, Y., Zhan, L., Zhang, W., Sun, F., Wang, W., Tian, N., et al. (2011). RNA Secondary Structure in Mutually Exclusive Splicing. *Nat. Struct. Mol. Biol.* 18, 159–168. doi:10.1038/nsmb.1959

Yang, Z., Huh, S. U., Drennan, J. M., Kathuria, H., Martinez, J. S., Tsuda, H., et al. (2012). Drosophila Vap-33 Is Required for Axonal Localization of Dscam Isoforms. *J. Neurosci.* 32, 17241–17250. doi:10.1523/JNEUROSCI.2834-12.2012

Yokota, S., Hirayama, T., Hirano, K., Kaneko, R., Toyoda, S., Kawamura, Y., et al. (2011). Identification of the Cluster Control Region for the Protocadherin-β Genes Located beyond the Protocadherin-γ Cluster. *J. Biol. Chem.* 286, 31885–31895. doi:10.1074/jbc.M111.245605

Yue, Y., Meng, Y., Ma, H., Hou, S., Cao, G., Hong, W., et al. (2016a). A Large Family of Dscam Genes with Tandemly Arrayed 5′ Cassettes in Chelicerata. *Nat. Commun.* 7, 11252. doi:10.1038/ncomms11252

Yue, Y., Yang, Y., Dai, L., Cao, G., Chen, R., Hong, W., et al. (2016b). Long-range RNA Pairings Contribute to Mutually Exclusive Splicing. *RNA* 22, 96–110. doi:10.1261/rna.053314.115

Zazhytska, M., Kodra, A., Hoagland, D. A., Fullard, J. F., Shayya, H., Omer, A., et al. (20212021). Disruption of Nuclear Architecture as a Cause of COVID-19 Induced Anosmia. *BioRxiv* 1, 430314. doi:10.1101/2021.02.09.430314

Zhan, X.-L., Clemens, J. C., Neves, G., Hattori, D., Flanagan, J. J., Hummel, T., et al. (2004). Analysis of Dscam Diversity in Regulating Axon Guidance in Drosophila Mushroom Bodies. *Neuron* 43, 673–686. doi:10.1016/j.neuron.2004.07.020

Zhou, Z., and Belluscio, L. (2012). Coding Odorant Concentration through Activation Timing between the Medial and Lateral Olfactory Bulb. *Cel. Rep.* 2, 1143–1150. doi:10.1016/j.celrep.2012.09.035

Zhu, H., Hummel, T., Clemens, J. C., Berdnik, D., Zipursky, S. L., and Luo, L. (2006). Dendritic Patterning by Dscam and Synaptic Partner Matching in the Drosophila Antennal Lobe. *Nat. Neurosci.* 9, 349–355. doi:10.1038/nn1652

Zhuang, Y., Futse, J. E., Brown, W. C., Brayton, K. A., and Palmer, G. H. (2007). Maintenance of Antibody to Pathogen Epitopes Generated by Segmental Gene Conversion Is Highly Dynamic during Long-Term Persistent Infection. *Infect. Immun.* 75, 5185–5190. doi:10.1128/IAI.00913-07

Zipursky, S. L., and Sanes, J. R. (2010). Chemoaffinity Revisited: Dscams, Protocadherins, and Neural Circuit Assembly. *Cell* 143, 343–353. doi:10.1016/j.cell.2010.10.009

Zipursky, S. L., Wojtowicz, W. M., and Hattori, D. (2006). Got Diversity? Wiring the Fly Brain with Dscam. *Trends Biochem. Sci.* 31, 581–588. doi:10.1016/j.tibs.2006.08.003

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