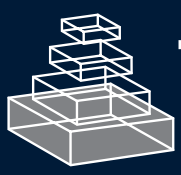
frontiers

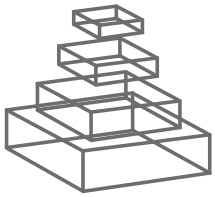
RESEARCH TOPICS

ADMINISTRATION AND COORDINATION OF GENETIC EXPRESSION BY PROTEINS STRUCTURING THE GENOME

Topic Editor
Michèle Amouyal



frontiers in GENETICS



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-135-2
DOI 10.3389/978-2-88919-135-2

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: researchtopics@frontiersin.org

ADMINISTRATION AND COORDINATION OF GENETIC EXPRESSION BY PROTEINS STRUCTURING THE GENOME

Topic Editor:
Michèle Amouyal, CNRS, France

These last years, a new class of proteins involved in the regulation of gene expression has been unraveled besides such classical elements as polymerases, transcription factors or enhancers, directly responsible for gene expression. This class introduces a second order level of regulation which is superimposed on that of the standard regulators and is revealed by epigenetic modifications.

One end of this group is occupied by the so-called gene insulators such as CTCF that organize the action of the standard regulators and of invasive chromatin in order to limit and target their action to a specific gene or genetic entity.

At the other end, one would find the proteins operating at the level of the whole genome and of cellular programming, such as SATB1.

They are specific proteins nearly totally dedicated to organization and coordination of gene expression, like CTCF, or proteins also found involved in other cellular devices, like transcription factors extracted from the transcriptional machinery , such as TFIIIC, or proteins involved in sister chromatid cohesion, like cohesin. But they share in common to generally act over long distances or between chromosomes, to structure a true intra- or inter-chromosomal genomic architecture, and to mainly act at the epigenetic level.

Their role in the control of gene expression is certified by the increasing number of pathologies to which their dysfunction contributes.

Table of Contents

04 ***Administration of Genetic Expression by Multi-Task Proteins and Long-Rangeaction***
Michèle Amouyal

06 ***Transcription Factor Binding At Enhancers: Shaping a Genomi Cregulatory Land-Scape in Flux***
Robert-Jan Palstra and Frank Grosveld

18 ***More Than Insulator: Multiple Roles of CTCF At the H19-Igf2 Imprinted Domain***
Purnima Singh, Dong-Hoon Lee and Piroska E. Szabó

27 ***The Role of CTCF Binding Sites in the 3 Immunog Lobulin Heavy Chain Regulatory Region***
Barbara k. Birshtein

35 ***Role of CTCF in the Regulation of Micro RNA Expression***
Yoshimasa Saito and Hidetsugu Saito

40 ***Regulation of Chromatin Structure by Poly (ADP-ribosyl) Ation***
Sascha Beneke

56 ***Diverse Developmental Disorders from the One Ring: Distinct Molecular Pathways Underlie the Cohesinopathies***
Julia A. Horsfield, Cristin G. Print and Maren Mönnich

71 ***Chromatin Loops, Gene Positioning, and Gene Expression***
Sjoerd Holwerda and Wouterde Laat

84 ***A Repetitive Elements Perspective in Polycomb Epigenetics***
Valentina Casa and Davide Gabellini

100 ***Me CP2 as a Genome-Wide Modulator: The Renewal of an Old Story***
Floriana Della Ragione, Stefania Filosa, Francesco Scalabrì and Maurizio D'Esposito

106 ***TFC6 (TFIIIC Subunit): A Bridge Between Prokaryotic and Eukaryotic Gene Regulation***
Michèle Amouyal

frontiers in
GENETICS

EDITORIAL
published: 10 April 2013
doi: 10.3389/fgene.2013.00052

Administration of genetic expression by multi-task proteins and long-range action

Michèle Amouyal*

Centre National de la Recherche Scientifique-Interactions à Distance, Paris, France
**Correspondence: michele.amouyal@club.fr*

Edited by:
Michael E. Symonds, The University of Nottingham, UK

Reviewed by:
Michael E. Symonds, The University of Nottingham, UK

In eukaryotes, a gene must be switched on or off at a given time of development, in a given tissue, in a given environment. To meet these specific obligations and accommodate their diversity, the gene mobilizes genetic elements all over the genome and factors which are not necessarily transcription-specific. The 10 articles of this book highlight the latest advances in the topic, focusing on the extraordinary adaptability of some of these proteins and their unexpected spectrum of competencies.

Historically, enhancers were the first elements found to act at-a-distance from the gene, as exhaustively reviewed by Palstra and Grosveld (2012), starting from the β-globin locus, a model for cellular differentiation. It is now well evidenced that these enhancers contact promoters with looping of in-between chromatin.

Other elements, such as insulators, assist this process. A prototype insulator is CTCF. At the β-globin locus (Palstra and Grosveld, 2012), the Igf2/H19 parent-of-origin imprinted locus (Singh et al., 2012), or at the immunoglobulin heavy chain antibodies locus (Birshtein, 2012), CTCF confers a basal folding to the genome, creating the proximity necessary for the productive contacts. CTCF also blocks RNA elongation at pause sites, sharing this feature with prokaryotic factors in addition to looping. Last, insulators are implicated in epigenetic regulation, counteract the spread of heterochromatin, decide of chromatin composition.

This 3-D genomic architecture allows to direct enhancer action from one gene to another, to coordinate expression of several genes or genetic loci simultaneously, to couple and control genes within one unit or several processes with the same factors, if necessary, sometimes delineating eukaryotic equivalents of prokaryotic operons. This structural scaffold provides strengthened, yet dynamic, interactions, stable enough to allow other contacts to take place, to resist to moving cellular tensions, even possibly to cell division.

However, the frontier between the different classes: (1) promoters/proximal elements, (2) enhancers, (3) insulators, are not clearly defined. The cell actually makes use of any element to ensure proper genetic expression. Hence, some promoters (for RNA polymerases II or III) and basal/proximal elements of the transcriptional machinery of initiation such as TFIIIC, indifferently act as insulators, as reported in (Amouyal, 2012; Holwerda and de Laat, 2012; Palstra and Grosveld, 2012). The LCR enhancer, of which the deletion leads to thalassemia, conversely illustrates this ambiguity. Once thought to only act by counteracting the spread of heterochromatin like some insulators, it also operates like classical enhancers by contacting promoter with specific factors (EKLF, GATA-1, FOG-1) and looping.

In addition, some factors perform other tasks than genetic expression, using the same structural device at the molecular level. Thus, according to cell cycle progression, but always by ensuring chromosomal cohesion, cohesin is either structurally involved in (1) sister chromatids cohesion/DNA damage repair, or in (2) gene transcription with chromosomal looping at several loci. Multi-functionality here explains the diversity of phenotypes in cohesinopathies due to defective cohesin, from Roberts to Cornelia de Lange syndromes, with all intermediates (Horsfield et al., 2012).

The extensive utilization by the cell of a peculiar skill for different applications is not restricted to DNA loopers. Thus PARP1 transcription factor (reviewed by Beneke, 2012) has the capacity to synthesize poly(ADP)ribose and to transfer it either covalently or non-covalently to other proteins. Addition of this polymeric sugar to CTCF presumably improves chromosomal looping by providing a dimerization interface and by stabilizing CTCF DNA-binding at several loci. Cancer marks a defective process. But PARP1 is also part of the basal RNA polymerase II machinery (as TFIIC), both a positive and negative cofactor of transcription, and mediates the response to DNA damage with the same tool. Thus, it loosens chromatin structure for the access of appropriate factors by the simple interaction of the poly(ADP-ribose) with histones.

In the same vein, CTCF regulates coding mRNAs as well as non-coding RNAs, in the same field of tumor suppression, control of cell cycle and proliferation, including embryonic stem cell differentiation for RNA regulators (Saito and Saito, 2012), which nicely corroborates the way CTCF acts in one case, anticipates it in the second one.

The cell also makes use of long-range action at different levels to assist gene expression. First thought to be confined to enhancer-promoter interaction, it has been extended with insulators to the structuring of a whole locus, and at an upper level, to genome-/cell-wide organization, by means of the same factors/auxiliaries of transcription, as if "he who can the least, can the most."

Holwerda and de Laat (2012) tackle the question of gene positioning within the nucleus in this context. The new technologies (Hi-C, *lac* operators tethered to lamina,...) indicate a susceptibility to gene silencing close to the nuclear periphery or at the heart of chromosome territories. Out of these

locations, up to 1 Mb domains of active chromatin are enriched at their border with insulators (CTCF, tRNAs, SINEs,. . .) instigating, at least contributing to this partitioning. These technically difficult and fully progressing *in vivo* researches are somewhat in line with earlier studies related to the nuclear matrix and genome attachment defined by chemical treatments.

Transition from DNA looping to high-order genomic organization is not surprising as the same elements (DNA repeats and protein apt to oligomerize in the simplest case) lead to intra-chromosomal looping, chromosomal clustering and condensation when reproduced, inter-chromosomal interactions, coating with arrays of tandem repeats.

In fact, genomic repetition is as common as enhancer occurrence and is extremely susceptible to genome rearrangements and pathogenic. In mammals, these genomic repeats would recruit the Polycomb/Trithorax proteins (reviewed by Casa and Gabellini, 2012) essential for cell identity and differentiation. Again, at an individual gene level, Polycomb proteins assist transcription factors for gene regulation. At the upper cellular scale, they (super)-structure the cell into compartments and convey information between them.

The Methyl-CpG-Binding-Protein-2 (Della Ragione et al., 2012) is another transcriptional auxiliary which is capable of oligomerization, DNA bridging and condensation, inducing drastic modifications of chromatin topology. Surprisingly, it is specifically over-expressed in neurons, in stoechiometric amounts with histone H1, and competes with H1 binding to nucleosomes. This neuronal chromatin plasticity is questioned in RETT syndrome, a neurodevelopment disorder with transient autistic features due to a defective MeCP2 protein, reversible in mice.

MeCP2 is also pluri-competent: it silences genes through preferential binding to methylated CpG dinucleotides *in vivo*, represses and activates genes independent of methylation, and is involved in RNA splicing.

TFIIIC is the last-born of genome-wide organizers. First known as a compound of basal RNA polymerase III machinery, it also binds separately to wide-spread sites on genome (ETC, COC, others). At a global level, it takes part in long-range action and high-order structures, is an enhancer blocker and counteracts the spreading of heterochromatin. At an individual gene level, this is a repressor of RNA polymerase II transcription, with several features of a prokaryotic factor, narrowing the frontier between prokaryotes and eukaryotes (Amouyal, 2012).

Clearly, genomic architecture and its influence on genetic expression still deserve further investigation. Also, the picture of a regulator is not complete if it is not traced throughout developmental or environmental changes, like CTCF in embryonic and fetal germ cells at the Igf2/H19 locus to specify its role in the setting-up of imprinting (Singh et al., 2012). Last, like emphasized by several articles of this volume, it is difficult to restrict some factors to a unique task. Thus, what some factors might do specifically with respect to genetic expression, other cellular factors might as well do it less specifically. In case of chromosomal looping for instance, any connection between two distant genomic sites might favor or disfavor specific enhancer-promoter interactions, generating a global network of connections at a given time, in a given cell line and a given environment, that the future will specify.

REFERENCES

Amouyal, M. (2012). TFC6 (TFIIIC Subunit): a bridge between Prokaryotic and Eukaryotic Gene regulation. *Front. Genet.* 3:64. doi: 10.3389/fgene.2012.00064

Beneke, S. (2012). Regulation of chromatin structure by poly(ADP-ribosyl)ation. *Front. Genet.* 3:169. doi: 10.3389/fgene.2012.00169

Birshtein, B. K. (2012). The role of CTCF binding sites in the 3' immunoglobulin heavy chain regulatory region. *Front. Genet.* 3:251. doi: 10.3389/fgene.2012.00251

Casa, V., and Gabellini, D. (2012). A repetitive elements perspective in Polycomb epigenetics. *Front. Genet.* 3:199. doi: 10.3389/fgene.2012.00199

Della Ragione, F., Filosa, S., Scalabrì, F., and D'Esposito, M. (2012). MeCP2 as a genome-wide modulator: the renewal of an old story. *Front. Genet.* 3:181. doi: 10.3389/fgene.2012.00181

Holwerda, S., and de Laat, W. (2012). Chromatin loops, gene positioning, and gene expression. *Front. Genet.* 3:217. doi: 10.3389/fgene.2012.00217

Horsfield, J. A., Print, C. G., and Mönnich, M. (2012). Diverse developmental disorders from the one ring: distinct molecular pathways underlie the cohesinopathies. *Front. Genet.* 3:171. doi: 10.3389/fgene.2012.00171

Palstra, R.-J., and Grosveld, F. (2012). Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Front. Genet.* 3:195. doi: 10.3389/fgene.2012.00195

Saito, Y., and Saito, H. (2012). Role of CTCF in the regulation of microRNA expression. *Front. Genet.* 3:186. doi: 10.3389/fgene.2012.00186

Singh, P., Lee, D.-H., and Szabó, P. E. (2012). More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. *Front. Genet.* 3:214. doi: 10.3389/fgene.2012.00214

Received: 24 February 2013; accepted: 20 March 2013; published online: 10 April 2013.

Citation: Amouyal M (2013) Administration of genetic expression by multi-task proteins and long-range action. Front. Genet. 4:52. doi: 10.3389/fgene.2013.00052

This article was submitted to Frontiers in Epigenomics and Epigenetics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

REVIEW ARTICLE
published: 28 September 2012
doi: 10.3389/fgene.2012.00195

Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux

Robert-Jan Palstra* and Frank Grosveld

Department of Cell Biology, Erasmus MC University Medical Center, Rotterdam, Netherlands

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France
Jim Davie, University of Manitoba, Canada

****Correspondence:***
Robert-Jan Palstra, Department of Cell Biology, Erasmus MC University Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, Netherlands.
e-mail: r.palstra@erasmusmc.nl

The mammalian genome is packed tightly in the nucleus of the cell. This packing is primarily facilitated by histone proteins and results in an ordered organization of the genome in chromosome territories that can be roughly divided in heterochromatic and euchromatic domains. On top of this organization several distinct gene regulatory elements on the same chromosome or other chromosomes are thought to dynamically communicate via chromatin looping. Advances in genome-wide technologies have revealed the existence of a plethora of these regulatory elements in various eukaryotic genomes. These regulatory elements are defined by particular *in vitro* assays as promoters, enhancers, insulators, and boundary elements. However, recent studies indicate that the *in vivo* distinction between these elements is often less strict. Regulatory elements are bound by a mixture of common and lineage-specific transcription factors which mediate the long-range interactions between these elements. Inappropriate modulation of the binding of these transcription factors can alter the interactions between regulatory elements, which in turn leads to aberrant gene expression with disease as an ultimate consequence. Here we discuss the bi-modal behavior of regulatory elements that act *in cis* (with a focus on enhancers), how their activity is modulated by transcription factor binding and the effect this has on gene regulation.

Keywords: enhancer, transcription factor, chromatin looping, transcription, *cis*-regulation

INTRODUCTION

Expression of genes is to a large extent directed by regulatory sequences within the promoters of genes. However, early transfection experiments led to the realization that promoters alone were not enough to direct the proper expression of genes. The first enhancers described were SV40 viral repeat sequences that are able to boost expression of a rabbit β-globin construct (Banerji et al., 1981). This enhancement of expression occurred independent of the orientation and location of the enhancer sequence within the reporter construct and this observation became the operational definition of enhancer elements. Soon after the description of viral enhancer sequences the first mammalian enhancer sequences were discovered within the human immunoglobulin heavy Chain locus (Banerji et al., 1983) and it turned out that this enhancer sequence acts in a tissue-specific fashion. Since the first enhancer discovery in humans, many more enhancers have been discovered in different organisms and it is estimated that over 1 million enhancers reside in the human genome (Heintzman et al., 2009). It is also becoming clear that enhancers are marked by the binding of specific chromatin modification factors and the presence of specific histone modifications (Maston et al., 2012). Recent work also suggests that not only protein-coding genes are under the influence of enhancers but that microRNA genes might also be under long-range developmental control (Sheng and Previti, 2011). Although we have come a long way in the 30 years since the first discovery of enhancers, their discovery still remains a challenging task and the mechanism of enhancer action is still largely unknown.

ENHANCER DISCOVERY BY MAPPING TRANSCRIPTION FACTOR BINDING SITES AND CHROMATIN MODIFICATIONS

Discovery of enhancers has always been a formidable task. DNAseI hypersensitivity mapping was the method of choice since it was observed that regulatory regions within the genome are hypersensitive to DNAseI digestion (Wu, 1980). However, this method was tedious, requiring careful titration of DNAseI concentration, restriction digestion, Southern blotting, and detection with labeled nucleotide probes which yielded only information on particular sequences or loci. The first attempts to identify enhancers on a genome-wide scale did not depend on DNaseI but involved enhancers traps (Hamada, 1986). In this method, a selectable reporter gene driven by an enhancer dependent promoter is randomly integrated in to the genome. Clones in which the reporter gene has integrated within the vicinity of an enhancer can be selected and the enhancer sequences isolated. Subsequent validation of enhancer activity can be done *in vitro* by transiently transfecting luciferase reporter constructs in cell lines or *in vivo* using reporter constructs in transgenic animals. However, this method remains a laborious procedure.

With the emergence of complete sequence information from many different model organisms attempts were made to identify regulatory sequences based on sequence conservation. These bioinformatics attempts were moderately successful (Meireles-Filho and Stark, 2009). However, it has become clear that not all conserved non-coding sequences have a detectable (enhancer) activity and not all enhancers are conserved at the sequence level (Blow et al., 2010; Royo et al., 2011). Recent advances

in genome-wide technologies like array technology and more recently high-throughput sequencing are proving to be a game changer for the genome-wide discovery of enhancers. More traditional techniques are currently combined with high-throughput sequencing technologies to identify enhancers on a genome-wide scale and novel approaches of enhancer discovery are introduced. One of the first techniques to be combined with array technology and later high-throughput sequencing as a read out was chromatin immunoprecipitation (ChIP; Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007) and even "old school" DNAseI hypersensitive site mapping has been combined with high-throughput sequencing in order to obtain genome-wide maps of "open" chromatin associated with regulatory regions (Sabo et al., 2006; Hesselberth et al., 2009; Bernstein et al., 2010).

Early genome-wide ChIP experiments found that enhancers are enriched in specific chromatin marks, especially high levels of H3K4me1 in combination with low levels of H3K4me3 appeared to mark enhancer sequences (Heintzman et al., 2007). Later it was found that acetylation of histone H3 at lysine 27 (H3K27Ac) specifically marks active enhancers (Creyghton et al., 2010) and recently it has been reported that in T-lymphocytes di- and trimethylation of histone H3 at lysine 4 are also correlated with active enhancers (Pekowska et al., 2011). As many more chromatin modifications have recently been identified (Tan et al., 2011), it is to be expected that several of these novel chromatin marks associate with enhancers (Kellner et al., 2012). Transcriptional co-activators like the acetyltransferase and transcriptional co-activator p300 (Visel et al., 2009a; May et al., 2011), the ATAC histone acetyl transferase complex (Krebs et al., 2011) and the ATP-dependent chromatin remodeler CHD7 (Schnetz et al., 2009, 2010) also appear to locate at enhancers. Clusters of tissue-specific transcription factors are hallmarks of enhancers and this fact has been exploited to identify enhancers. He et al. (2011) used a set of five cardiac-specific transcription factors to identify cardiac-specific enhancers that were distinct from p300 bound enhancers. Analysis of the binding of a set of three myogenic-specific transcription factors in combination with p300 binding and enhancer-associated chromatin marks before and after muscle differentiation allowed for the identification of muscle-specific enhancers (McCord et al., 2011). Furthermore, the mysterious highly occupied target (HOT) regions which are bound by many transcription factors but lack their consensus binding motif, function as spatial and temporal enhancers in transgenic assays (Kvon et al., 2012). Conversely, mapping of tissue-restricted enhancers via chromatin marks has lead to the discovery of specific transcription factor binding signatures that correspond to monocyte differentiation states (Pham et al., 2012).

Several laboratories have defined distinct chromatin signatures associated with specific regulatory elements based on the combinatorial analysis of multiple chromatin marks and transcription factor binding patterns (Wang et al., 2008; Ram et al., 2011; Bonn et al., 2012; Cotney et al., 2012; Hoffman et al., 2012), which allows to distinguish between specific enhancer states (Rada-Iglesias et al., 2011; Zentner et al., 2011; Bogdanovic et al., 2012; Cotney et al., 2012). Novel approaches to detect regulatory genomic regions are also emerging like formaldehyde-assisted isolation of regulatory elements (FAIRE) which identifies the more "open" chromatin state associated with enhancers based on differences in phenol extractability of these regions (Giresi et al., 2007). Analysis of different genome-wide data sets is also revealing novel properties of enhancers. Global nuclear run-on followed by high-throughput sequencing (GRO-seq) data revealed that enhancers display bidirectional expression of short transcripts (Melgar et al., 2011; Wang et al., 2011), while an in depth analysis of glucocorticoid receptor (GR)-regulated enhancers revealed that they are enriched in CpG dinucleotides and that their methylation status is cell type-specific and correlate with the accessibility of the enhancers (Wiench et al., 2011).

High-throughput genome-wide approaches have made enhancer discovery a more amendable task. To date, most of these studies have been performed on cell lines but the first attempts to follow enhancer dynamics during development have been successful (Bogdanovic et al., 2012; Cotney et al., 2012). Given the spatial and temporal specificity of enhancers the major challenge for the future will lie in obtaining the proper tissues at the right developmental stage or state of differentiation and performing reliable ChIP-seq on the often limiting amounts of these cells (Bonn et al., 2012).

TRANSCRIPTION FACTOR-MEDIATED LONG-RANGE ENHANCER–PROMOTER COMMUNICATION

One key feature of eukaryotic enhancers is that they can be located far away from the gene they regulate. How enhancers are able to communicate with their cognate promoters remained a mystery for about two decades. A number of models were proposed which included polymerase tracking, the spreading of chromatin structures, and direct contact between separated elements. The non-contact model (polymerase tracking and chromatin spreading) postulated a role for the intervening chromatin fiber which would propagate a "signal" from the enhancer to the promoter. The contact model, better known as the looping model, proposed that the active enhancer and promoter would reside in close proximity within the nucleus while the intervening chromatin loops out. Although early *in vitro* experiments in prokaryotic systems provided support for the contact model [reviewed in Amouyal (1991)], the first direct *in vivo* evidence in eukaryotes was provided by the phenomenon of transvection in *Drosophila* (Tartof and Henikoff, 1991). The contact model was subsequently experimentally tested by varying the position or distance of genes in a series of experiments using the human β-globin locus (Hanscombe et al., 1991; Dillon et al., 1997).

The subsequent development of new techniques like RNA TRAP (Carter et al., 2002) and chromosome conformation capture (3C; Dekker et al., 2002) and its application to mammalian loci (Tolhuis et al., 2002) allowed the mapping of chromatin folding of gene loci. These studies on the β-globin locus clearly demonstrated that the major regulatory element of the β-globin genes, the locus control region (LCR), resides in close proximity to the genes when active while the intervening chromatin and inactive genes loop out (Carter et al., 2002; Tolhuis et al., 2002). These interactions are developmental stage-specific (Palstra et al., 2003) and dependent on lineage-specific transcription factors (Drissen et al., 2004; Vakoc et al., 2005). Chromatin conformations similar to the ones

initially observed within the β-globin locus have been found in several other gene loci in different cell types generally confirming the looping model (de Wit and de Laat, 2012).

3C and its derivatives are currently the method of choice to demonstrate interactions between enhancers and their target genes (de Wit and de Laat, 2012). A major limitation of 3C is the fact that some knowledge of the location of the regulatory elements is needed to design primers. Combining 3C with high-throughput sequencing allows for the unbiased discovery of novel long-range interactions of a specific locus (Soler et al., 2010), especially when combined with ChIP-derived chromatin modifications or transcription factor binding profiles as was demonstrated in a study that identified adipocyte-specific enhancers (Mikkelsen et al., 2010) and a study which identified erythroid-specific enhancers for the *MYB* gene (Stadhouders et al., 2011). One of the remaining drawbacks of this approach is that it still relies on a single locus for a viewpoint and is therefore not truly unbiased. A Chia-PET approach that focuses on either enhancer marks (Chepelev et al., 2012) or promoter-associated RNA polymerase II (RNA pol II; Li et al., 2012) in part circumvents this limitation. A truly unbiased method like Hi-C could in principle detect all long-range enhancer–promoter interactions in a cell population although limitations in sequencing depth and limitations of the bioinformatic tools available currently restricts the resolution of this approach (Lieberman-Aiden et al., 2009). However, taking the fast developments in high-throughput sequencing and bioinformatics analysis into account it may be in the not too distant future that enhancer–promoter interactions are routinely identified using Hi-C. In fact, a first glimpse of tissue-specific promoter–enhancer interactions has been observed in a recent Hi-C study (Dixon et al., 2012).

Binding of lineage-specific transcription factors to enhancers and promoters plays a vital role in the establishment/maintenance of long-range promoter–enhancer interactions. There appears to be a distinct set of transcription factors that tend to bind to promoters and a distinct set that tend to bind at distal regulatory elements (Lan et al., 2012). Analysis of Hi-C and ENCODE data obtained in erythroid leukemia cells indicated that in general factors bound at promoters interact with factors bound at distal sites (Lan et al., 2012). For some transcription factors their role in chromatin looping has been studied in more detail. In a knock-out mouse model of the erythroid-specific transcription factor EKLF, no long-range interactions between the β-globin LCR and β-major gene are observed and the β-globin locus adopts a chromatin conformation reminiscent of the one observed in erythroid progenitor cells (Drissen et al., 2004). Re-introduction of EKLF restores LCR–β-globin interaction and this also occurs in the absence of protein synthesis demonstrating a direct involvement of EKLF in chromatin looping (Drissen et al., 2004). A similar study on the transcription factors GATA-1 and FOG1 has shown that these factors also play a vital role in LCR–β-globin gene interaction (Vakoc et al., 2005). The role of another erythroid transcription factor, the heterodimeric NF-E2 has been more controversial. One study demonstrated that chromatin looping was independent of NF-E2 in a knock-out mouse model of the NF-E2 p45 subunit (Kooren et al., 2007) while an other study demonstrated NF-E2-dependent chromatin looping in a cellular model system upon knock down of the MafK/NF-E2 p18 subunit (Du et al., 2008). Other lineage-specific factors that have been shown to play a role in chromatin looping are GATA3 and STAT6 in the T-cell lineage (Spilianakis and Flavell, 2004) and OCA-B in the B-cell lineage (Ren et al., 2011).

It is doubtful that lineage-specific DNA binding transcription factors are solely responsible for establishing enhancer–promoter interactions. Enhancer bound transcription factors recruit co-activators and general factors of which some have been shown to play a vital role in enhancer–promoter communication. One of the best studied factors is the widely expressed transcriptional cofactor Ldb1. The non-DNA-binding Ldb1 protein is able to interact with multiple transcription factors and mediates interactions between them (Matthews and Visvader, 2003). In erythroid cells, Ldb1 is part of a large complex that contains the core factors TAL1, LMO2, E2A, and GATA1 which is recruited to E boxes and GATA elements in, for example, the β-globin LCR and promoter (Wadman et al., 1997; Soler et al., 2010). Knock-down of Ldb1 in erythroid cells results in an impaired long-range interaction between the β-globin LCR and β-major promoter and a failure to activate β-major expression (Song et al., 2007). A recent report demonstrated that artificial tethering of the self association domain of Ldb1 to the β-globin promoter is able to induce a chromatin loop between the β-globin LCR and promoter and this was sufficient to induce expression of the β-globin gene (Deng et al., 2012b). Other general factors implicated in chromatin loop formation between enhancers and promoters are Brg1, the ATPase component of the SWI/SNF nucleosome remodeling complex (Kim et al., 2009) and the general transcription factor TFII-I (Ren et al., 2011). A different type but very interesting general nuclear factor involved in chromatin looping is cohesin. It is best known for its role in holding together sister chromatids during mitosis, but more recently it has been recognized that cohesin is intimately linked to transcription (Dorsett, 2011; Haering and Jessberger, 2012). The nuclear protein CCCTC-binding factor (CTCF) is thought to partition the genome in separate domains via chromatin loops preventing crosstalk between active and inactive regions (Weth and Renkawitz, 2011; Herold et al., 2012). Recently it was found that these CTCF-mediated chromatin loops are dependent on cohesin (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008; Wendt and Peters, 2009). Interestingly, in murine ES cells cohesin interacts with Mediator and the cohesin loading factor Nipbl and together they participate in chromatin loop formation between enhancers and promoters of ES cell-specific loci (Kagey et al., 2010). Similarly, upon differentiation of mouse erythroid leukemia (MEL) cells cohesin and Nipbl are recruited to the β-globin LCR and β-major promoter coinciding with an increase in transcription. Knock-down of one of these factors resulted in reduced chromatin looping between the β-globin LCR and promoter (Chien et al., 2012). Furthermore, the TBP core promoter associated factor TAF3 cooperates with CTCF and cohesin to mediate long-range chromatin loops between enhancers and promoters in the endoderm lineage (Liu et al., 2011).

The general picture that is emerging from these studies is that lineage-specific DNA binding transcription factors bound at promoters and enhancers recruit "looping" factors which setup

contacts between distal enhancers and promoters. Such factors appear to form loops within more "structural" loops mediated by general factors like CTCF (**Figure 1**).

The fact that active enhancers reside within close proximity of the active promoters they regulate is currently well recognized. How these interactions are established remains largely unknown. Whether the formation of a chromatin loop is an actively directed process or determined by random collisions has not been elucidated. Several studies suggest that polymerization of nuclear actin might be a driving force in bringing enhancers and promoters together. For example, actin polymerization is necessary for retonic acid induced recruitment of transcription factors to an enhancer element, for the induction of HoxB transcription (Ferrai et al., 2009), for the reactivation of OCT4 during reprogramming by oocytes (Miyamoto et al., 2012), and for the re-localization of gene loci in the interphase nucleus (Chuang et al., 2006; Dundr et al., 2007). Additionally, motor-proteins like nuclear Myosin I and dynein light chain-I have also been reported to be essential for nuclear receptor-induced co-localization of gene loci (Hu et al., 2008). The direct involvement of these factors in establishing enhancer–promoter chromatin loops has however not been shown. Some interpretations of the popular transcription factory hypothesis suggest an alternative actively directed process for bringing enhancers and promoters together (Papantonis and Cook, 2010; Deng et al., 2012a). In this view polymerases bound to enhancers would real-in the chromatin fiber until a promoter is encountered which is subsequently activated (West and Fraser, 2005). However, promoter–enhancer chromatin loops remain when RNA pol II transcription is pharmacologically inhibited, suggesting that such a scenario is unlikely (Mitchell and Fraser, 2008; Palstra et al., 2008).

ENHANCER MODE OF ACTION

How enhancers actually promote transcription of a gene when in close proximity remains poorly understood. What is clear is that enhancer bound transcription factors recruit co-activators either as part of an enhanceosome or flexible billboards (Alvarez et al., 2003; Arnosti and Kulkarni, 2005). In an enhanceosome a multiprotein complex is assembled at the enhancer and spacing of transcription factor binding sites is crucial for its function (Thanos and Maniatis, 1995). A similar model has been suggested for the multi-enhancer β-globin LCR where the hypersensitive

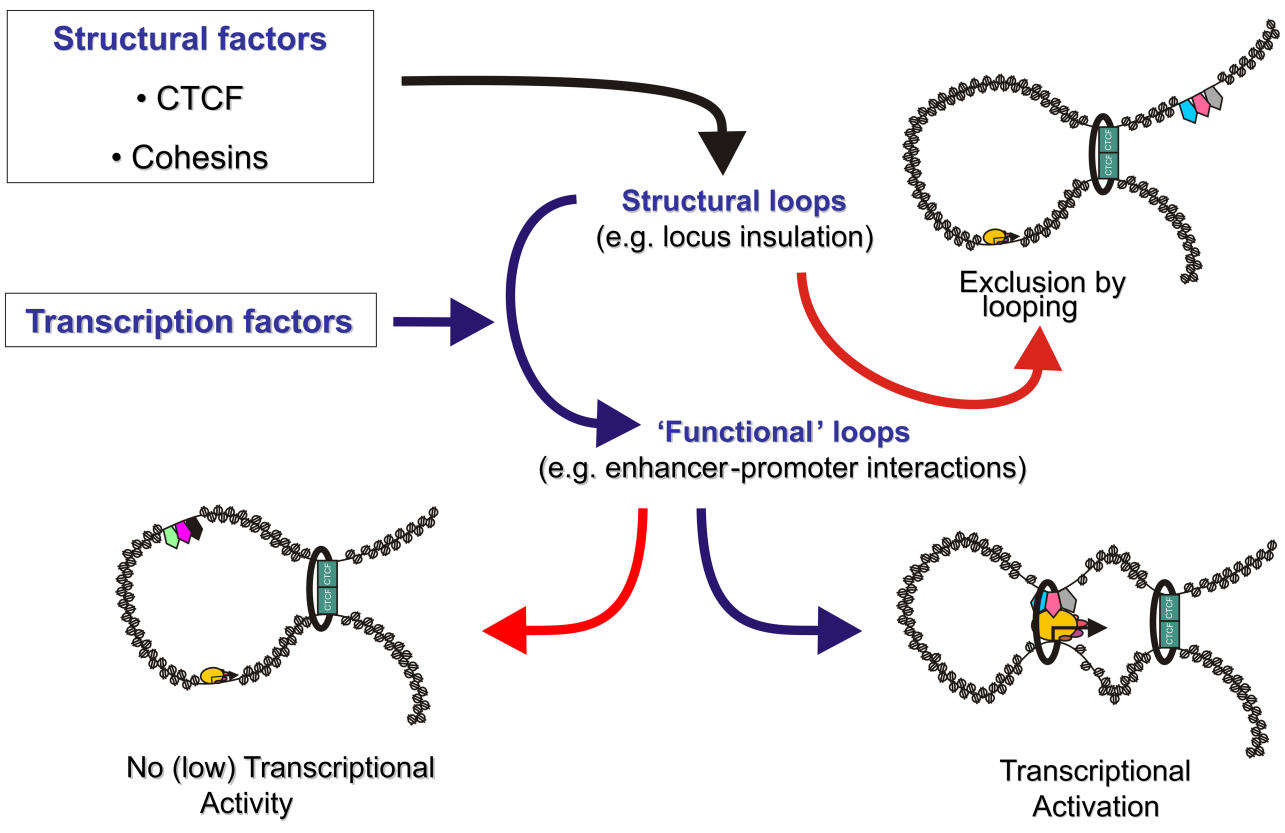


FIGURE 1 | The genome is organized in transcription factor mediated chromosome loops. Structural transcription factors like CTCF and cohesin (green rectangles and black ring) co-operate to partition the genome in looped domains. Depending on the location of the anchor points of these looped domains, enhancers are either excluded from a target gene's domain, effectively blocking activation (top right), or included in the target gene's topological domain (bottom). Tissue specific and ubiquitous transcription factors (pentagons) either induce enhancer–promoter communication via a chromatin loop (bottom right) or keep enhancers in a silent/poised state (bottom left). Rectangles depict CTCF, the black ring represents the cohesin complex, pentagons depict enhancer bound transcription factors and ovals depict components of the pre-initiation complex. Size of the icons depicts strength of binding or activity.

sites of the LCR are thought to form a holo complex (Ellis et al., 1996; Milot et al., 1996). Billboard enhancers are more flexible in their architecture since they consist of separate elements that individually are able to modulate transcription and the additive repressive or activating effects of these elements would determine the transcriptional outcome (Arnosti and Kulkarni, 2005).

Traditionally enhancers are thought to enhance recruitment of RNA pol II and the pre-initiation complex to promoters. It has been suggested that enhancers (or LCRs) function by simply increasing the local concentration of transcription factors, which in turn increases the efficiency of transcription (Palstra et al., 2003). Recent studies have shown that many genes contain stalled polymerases and that the transition from initiation to elongation appears to be a rate limiting step under stringent control (Nechaev and Adelman, 2011). It has therefore been suggested that enhancers play a role in facilitating this transition. Indeed, deletion of the β-globin LCR results in severely reduced phosphorylation of the RNA pol II C-terminal domain (CTD) and transcriptional elongation while pre-initiation complex (PIC) assembly and RNA pol II recruitment to the β-globin promoter was only reduced twofold (Sawado et al., 2003). The erythroid Myb gene enhancers are looped to a conserved CTCF binding site in the first intron of the Myb gene. The p-TEFb component Cdk9 is specifically recruited to the enhancer as part of the Ldb1 complex, and the conserved CTCF site in the intron marks a transition between pausing and elongating polymerases suggesting that enhancers are also essential in regulating transcriptional elongation (Stadhouders et al., 2011). Other results were obtained in a recent study were chromatin looping between the β-globin LCR and β-major gene was induced by tethering of a looping factor (Deng et al., 2012b). Recruitment of RNA pol II to the β-major promoter was restored upon induced chromatin looping while transcriptional elongation remained reduced. The lack of transcriptional elongation is in part explained by the failure to recruit and activate the P-TEFb elongation factor in this system which lacks the crucial erythroid-specific transcription factor GATA1 (Deng et al., 2012b). Together, these studies suggest that enhancers have a function in both PIC and RNA pol II recruitment or stabilization and facilitation of the transition between initiation and elongation.

Alternative mechanisms for enhancer function have also been proposed. Recent genome-wide studies have made clear that RNA pol II is recruited to enhancers (De Santa et al., 2010; Kim et al., 2010; Koch et al., 2011) and that these enhancers are transcribed (Melgar et al., 2011; Wang et al., 2011). A role for these transcripts in enhancer function has been suggested (Orom and Shiekhattar, 2011), however their exact role remains uncertain. Although some non-coding (nc) RNAs seem to behave like classical enhancers in reporter assays (Orom et al., 2010), other observations seem to refute a direct role for the generated ncRNA transcript. The activity of the human growth hormone enhancer is for example dependent on the level of enhancer transcription but not on the structure of its ncRNA (Yoo et al., 2012). Another example where non-coding transcripts are linked to enhancer function is the Kcnq1 imprinted domain (Korostowski et al., 2011). In this case, chromatin loop formation between regulatory elements prevents Kcnq1 promoter silencing by the non-coding Kcnq1ot1 transcript. An attractive but as yet untested possibility is that the ncRNAs are involved in promoting/stabilizing the interaction between the enhancer and its target promoter by RNA binding transcription factors at the enhancer and basic complex transcription factors at the promoter. The observation that RNA pol II complexes are recruited to enhancers has lead to a model in which enhancers are able to transfer RNA pol II to promoters either via direct transfer (Leach et al., 2001) or a tracking mechanism (Zhu et al., 2007). Transfer of polymerases from enhancer sequences to promoter sequences was indeed demonstrated in an *in vitro* assay (Vieira et al., 2004). Convincing *in vivo* data to support this model are however lacking and RNA pol II is still recruited to the β-major gene in the absence of an LCR (Sawado et al., 2003).

Enhancers also seem to play a role in polycomb eviction from developmental promoters containing CpG islands by recruiting the histone H3K27me3 demethylase JMJD3 to the promoter (Taberlay et al., 2011; Vernimmen et al., 2011). In fact, the activity of developmental enhancers itself appears to be kept under tight control by members of the polycomb complex and several other histone methyl transferases (Svotelis et al., 2011; Whyte et al., 2012; Zhu et al., 2012). In breast cancer cells, the poised enhancer of Bcl-2 is marked by H3K27me3. Activation of this enhancer requires the inactivation of the H3K27 methylase EZH2 a member of the polycomb complex and the simultaneous recruitment of the histone H3K27me3 demethylase JMJD3 which is under hormonal control (Svotelis et al., 2011). Several enhancers that have ubiquitous activities when tested in transgenic assays are repressed in non-permissive cells by the presence of flanking regions enriched in H3K9me3 at their endogenous location (Zhu et al., 2012). Cell type-specific recruitment of the H3K9 demethylase Jmjd2d alleviates this repression. Conversely, enhancers responsible for maintaining ES cell identity have to be silenced upon differentiation, which occurs through the recruitment of the H3K4/K9 histone demethylase LSD1 (Whyte et al., 2012).

On the other hand, enhancers that have to become active in a specific lineage are kept in a poised state upon stem cell differentiation via the sequential recruitment of lineage-restricted transcription factors. The transcription factor SOX2 is for example bound at neuron-specific regulatory elements in embryonic stem cells, and is replaced by SOX3 in neuronal progenitor cells and later by SOX11 in terminal differentiated neurons (Bergsland et al., 2011).

It is very well possible that enhancer action goes beyond just one activity and that enhancers perform different tasks sequentially during cellular differentiation. Initially, enhancers will keep gene loci in a transcriptionally competent state by sequential recruitment of progressively more lineage-restricted transcription factors. At a later stage, they will assemble and stabilize a pre-initiation complex at the gene promoter via chromatin looping and finally release paused polymerases through recruitment of elongation factors.

SPLIT PERSONALITIES OF REGULATORY ELEMENTS

As mentioned before, eukaryotic enhancers were operationally defined in transient transfection assays by the ability to activate a reporter gene irrespective of location and orientation relative to the promoter. This does not necessarily mean that these regulatory elements behave in a similar fashion at their native location

in the chromatin context of a cell which is subject to a variety of external signaling cues. The activity of enhancer like elements is regulated in a strict temporal and positional manner within a developing organism. A better approach to test the enhancer like abilities of a DNA sequence is to test it linked to a reporter gene via a transgenic approach. Besides the fact that enhancers can switch between multiple active, poised, and repressed states (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011), new studies indicate that a *cis*-regulatory element can have multiple properties simultaneously.

Depending on the assays used, multiple distinct classes of *cis*-regulatory elements can be recognized (Raab and Kamakaka, 2010). Promoters are bound by transcription factors, provide an assembly point for the RNA pol II holo complex and generally designate a more or less defined directional starting point of transcription. Enhancers recruit transcription factors, they can be transcribed and are able to boost expression from a distally located promoter often in a developmental stage and tissue-restricted manner. The action of enhancers can be counteracted by enhancer blockers when placed between the enhancer and promoter. On the other hand, silencers can suppress transcription from multiple positions relative to enhancers and promoters. Finally, insulators are genetic elements that counteract the spread of heterochromatin.

As discussed above genome-wide studies have demonstrated that many enhancers recruit RNA pol II and are transcribed (De Santa et al., 2010; Kim et al., 2010). Similar observations have been made almost two decades ago for hypersensitive site 2 of the β-globin LCR (Tuan et al., 1992). Most of these enhancer transcripts can be polyadenylated but remain short and are not elongated (Kim et al., 2010). Enhancers that are located intragenic however produce long spliced and polyadenylated transcripts and may therefore function as alternative promoters (Kowalczyk et al., 2012). Promoters of tRNA genes on the other hand have been shown to act as either insulators or enhancer blocking elements in yeast (Simms et al., 2008) and mammalian systems (Raab et al., 2011), which is mediated by binding of the general RNA Pol III transcription factor TFIIIC. In *Drosophila*, RNA pol II promoters containing stalled RNA pol II also act as enhancer blocking elements (Chopra et al., 2009). One model for enhancer blocking function, the decoy model, postulates that enhancer blockers interfere with enhancer–promoter interaction by producing inactive interactions between the enhancer blocking element and the promoter or the enhancer. *Drosophila* enhancer blocking elements indeed appear to form chromatin loops with promoters (Erokhin et al., 2011). Some enhancer blockers can also act as silencers in transient transfection assays suggesting that the distinction between these two elements depends on the assay involved (Petrykowska et al., 2008). Interestingly, it has been reported that the β-globin LCR, which is normally a very strong enhancer in erythroid cells, is able to act as a repressor when placed in the right genomic context (Feng et al., 2005). Specific repressors appear to act on enhancers by interfering with loop formation between enhancers and gene promoters (Chopra et al., 2012). Replacement of an activating loop by a repressive loop has also been observed. When the c-Kit gene is active in immature erythroid cells a GATA2-dependent chromatin loop is present between an upstream enhancer and the promoter (Jing et al., 2008). Upon erythroid maturation, GATA1 replaces GATA2 and the activating enhancer–promoter chromatin loop is replaced by a repressive chromatin loop between the promoter and a downstream silencer-like element. Interestingly, several genetic studies in drosophila have shown that enhancer blockers, when placed in the right context, can enhance enhancer–promoter communication or even act as enhancer elements (Rodin et al., 2007; Maksimenko et al., 2008; Soshnev et al., 2008; Fujioka et al., 2009). These observations indicate that enhancer blockers/silencers function, like enhancers, by means of long-range chromatin interactions. In mammalians, the major protein associated with enhancer blocking function is the 11 zinc-finger transcription factor CTCF (Bell et al., 1999), which is known to mediate long-range chromatin interactions (Splinter et al., 2006). Although CTCF is most famous for its role in enhancer blocking, the protein is also involved in gene activation (Weth and Renkawitz, 2011; Herold et al., 2012). Recent genome-wide analysis of enhancer–promoter interactions have indeed indicated that CTCF is associated with a proportion of enhancers and that CTCF mediates the interaction of these enhancers with their target promoters (Handoko et al., 2011; Li et al., 2012; Taslim et al., 2012).

In summary, it seems that the attempt to impose a strict definition on regulatory elements is much more complicated than expected: enhancers can behave like promoters, promoters can act as enhancer blockers, while enhancer blockers can function as enhancers, all dependent on the genomic context of the regulatory element and the specific set of transcription factors recruited.

ENHANCER TRANSCRIPTION FACTOR BINDING IN DEVELOPMENT, DISEASE, AND PHENOTYPE DIVERSITY

Tight control of transcription is crucial for the proper development of a multi-cellular organism. Enhancers play a crucial role in ensuring the proper spatio-temporal expression of genes by integrating the action of tissue-specific transcription factors and signaling cues (Buecker and Wysocka, 2012; Ong and Corces, 2012). Given the key role that enhancers play in the proper development of multi-cellular organisms it is of no surprise that disruption of enhancer function is a major contributor to pathological states. In fact, disease driven research has been crucial in the discovery and definition of mammalian enhancers. Investigation of γβ-thalassemia for example led to the discovery and characterization of the "super enhancer"-like β-globin LCR (Grosveld et al., 1987). In Dutch γβ-thalassemia, a large deletion removes 100 kb upstream of the β-globin gene but leaves the β-globin gene itself intact (Kioussis et al., 1983; Wright et al., 1984; Taramelli et al., 1986). The mutant locus is in a closed chromatin state and suffers from position effects. Further analysis of the region deleted in γβ-thalassemia revealed strong erythroid hypersensitive sites upstream of the ε-globin gene (Tuan et al., 1985). Cloning of these hypersensitive sites revealed that they impose position-independent, copy number-dependent high level expression on a β-globin transgene defining the operational properties of a LCR (Grosveld et al., 1987). Many other instances of disease causing enhancer disruptions are currently known (Kleinjan and Lettice, 2008). Translocations can either remove enhancer sequences from a locus (Kioussis et al., 1983) or place ectopic enhancers in the

vicinity of onco-genes as is observed in non-Hodgkin's lymphoma (Hayday et al., 1984). Smaller mutations in regulatory elements are also known to contribute to hereditary disease states. For example, several point mutations as well as insertions within the sonic hedgehog ZRS long-range enhancers cause several forms of preaxial polydactyly (Albuisson et al., 2011; Laurell et al., 2012). The effects of sequence variation in enhancer regions are not always catastrophic and can be quite subtle.

In the past decade, genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) which are statistically associated with phenotypic traits and disease states. The majority of the DNA variants identified in GWAS studies are located in non-coding regions without any known function while only a minority (∼30%) potentially disrupt the function of genes (Visel et al., 2009b; 1000 Genomes Project Consortium, 2010). Often linkage with unknown causal (non-synonymous coding) DNA variants within a haplotype block is assumed to explain association of non-coding DNA variants with a given trait. However meta-analysis demonstrated that 40% of the disease associated SNPs including their haplotype blocks exclusively involve non-coding sequence (Visel et al., 2009b) suggesting that these regions have a regulatory function. Moreover, a significant proportion of GWAS SNPs overlap with B, T, and ES cell enhancers (Teng et al., 2011), multiple sclerosis associated regions are located in chromatin regions that are active in B-cells (Disanto et al., 2012) and 80% of the colorectal cancer risk SNPs overlap with colon crypt enhancer marks (Akhtar-Zaidi et al., 2012).

One can easily imagine that the presence of a SNP might lead to differences in transcription factor binding at regulatory regions which could result in phenotypic changes and even disease (e.g., cancer) due to differences in transcriptional output of the associated genes (**Figure 2**). A study on 10 human lymphoblastic cell lines from different individuals indeed demonstrated that 7.5% of the binding sites for NF-κB and 25% of the RNA pol II binding sites differed between individuals (Kasowski et al., 2010). Differential binding occurred frequently at SNPs and structural variants and was often associated with changes in gene expression. Measurement of the genome-wide allelic imbalance of 24 transcription factors and the transcriptional co-factor p300 indicated that 5% of the binding sites for these factors vary depending on the sequence difference between alleles (Reddy et al., 2012). Chromatin accessibility to DNaseI also depends on genomic variation in lymphoblastoid cell lines and these differences in DNaseI hypersensitivity correlate with differences in transcription factor binding and changes in gene expression (Degner et al., 2012). These observations strongly suggest that many non-coding DNA variants are functional and mark for example enhancers for distally located genes which are involved in the trait under study. Identifying exactly which non-coding SNPs have a regulatory function has been cumbersome, mainly due to the presence of multiple linked non-coding SNPs within a haplotype block, the fact that enhancers are highly tissue- and developmental stage-specific and the lack of proper high-throughput assays to identify enhancer regions. Subsequent identification of the genes regulated by the causative SNPs

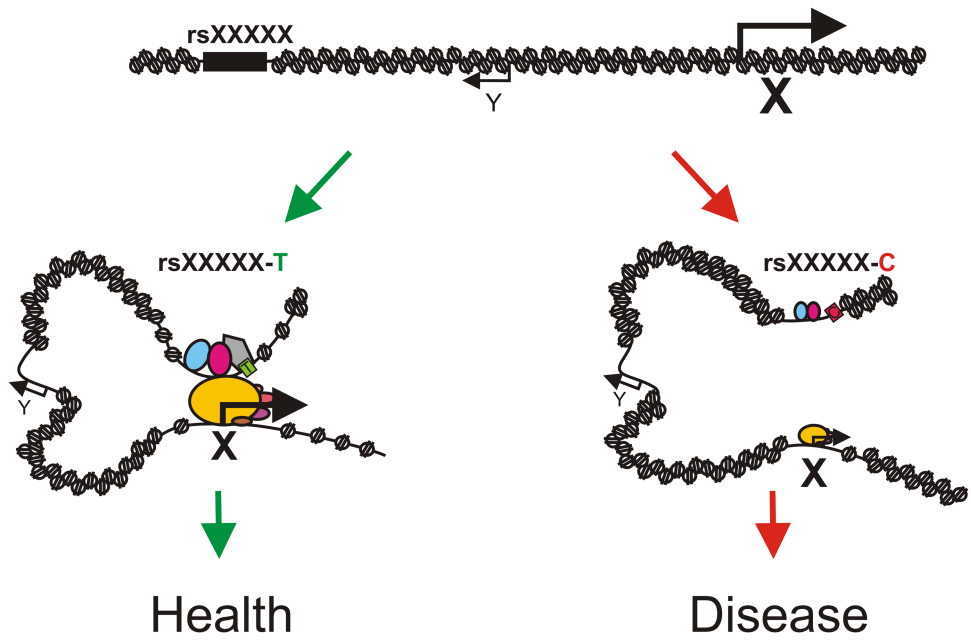


FIGURE 2 | Model depicting how sequence variation in distal regulatory elements might influence phenotypes or disease states. A phenotype associated SNP rsXXXXX is located in an enhancer for gene X. The T allele of rsXXXXX binds a transcription factor (the gray pentagon) with high affinity which allows for chromatin loop formation and proper activation of gene X resulting in normal development (left, green arrows). The C allele of rsXXXXX binds the transcription factor with a reduced affinity (light gray pentagon with dashed border) which leads to a less efficient enhancer, absence/reduction of looping and diminished expression of gene X resulting in aberrant or deviated development (right, red arrows). Note that gene Y is located in between the enhancer and gene X and is not regulated by the enhancer. Pentagons depict enhancer bound transcription factors and ovals depict components of the pre-initiation complex. Size of the icons depicts strength of binding or activity.

has proven to be even more difficult, since enhancers and their target genes are often separated by a significant extent of chromatin which can even contain non-target genes. The successful identification of regulatory SNPs and their linked target genes has therefore been limited to few isolated examples.

Several studies on specific risk loci support the notion that in several pathological states SNPs disrupt transcription factor binding sites within enhancers. For example, a risk allele for cleft lip disrupts an AP-2α binding site in an *IRF6* enhancer (Rahimov et al., 2008) and a variant linked to plasma low-density lipoprotein cholesterol and myocardial infarction creates a C/EBPα binding site which results in altered expression of the *SORT1* gene in hepatocytes (Musunuru et al., 2010). Studies on other disease associated loci have demonstrated chromatin loops between the regulatory variant and the genes they regulate. The variant rs6983267 is associated with an increased risk to develop various types of cancers and several studies have demonstrated that this SNP leads to altered TCF7L2 transcription factor binding, altered enhancer activity and that this region loops to the *MYC* proto-oncogene (Pomerantz et al., 2009; Ahmadiyeh et al., 2010; Wright et al., 2010). Similar observations have been made for e.g., variants associated with coronary artery disease (Harismendy et al., 2011), prostate cancer (Zhang et al., 2012), and COPD (Zhou et al., 2012). Not al disruptions of enhancers by SNPs lead to increased disease susceptibility, as they can also have non-pathological effects leading to phenotypic differences. Recently we could demonstrate that rs12913832, a SNP strongly associated with pigmentation in melanocytes, results in differential transcription factor binding at a melanocyte-specific enhancer. This difference in transcription factor binding leads to allele dependent attenuated looping between the enhancer and its target the *OCA2* pigment gene (Visser et al., 2012). Interestingly, allelic differences in enhancer activity are not always reflected in differential enhancer–promoter interactions (Wright et al., 2010), suggesting separate mechanisms for chromatin-loop formation and enhancer activity.

Combining genome-wide ChIP, FAIRE, and 3C high-throughput approaches with data derived from GWAS studies promises to boost the discovery of regulatory SNPs. These kinds of studies are crucial to obtain greater understanding of the impact of sequence variations on human health and disease (Chorley et al., 2008; Hawkins et al., 2010; Ernst et al., 2011) or (part of) the normal variation between individuals. Using these genome-wide approaches it will be possible to shift from just describing statistical associations between variants and traits to studies that actually discover the biology behind disease and phenotype associated non-coding variants.

CONCLUSIONS AND FUTURE PROSPECTS

Knowledge regarding enhancers and enhancer function has exploded in the past decades. Much of the early insight into enhancer function has been obtained from painstakingly dissecting single model loci. Due to the limited amount of loci investigated, the generality of the occurrence of enhancers and their mode of action remained unclear. With the recent advent of high throughput genome-wide techniques we are now able to address the generality of these early observations. Important insights regarding enhancer–promoter communication, the occurrence of enhancers and enhancer function have been obtained. Surprisingly, the regulatory landscape is far more complex and dynamic as anticipated and it appears that each cell type has thousands of enhancers of which many are cell type-specific. Chromatin looping between regulatory elements is widely observed and appears to be a general principle for long-range enhancer–promoter communication.

However, many challenges remain. Little is known about enhancer dynamics during cellular differentiation, how signaling cascades impact on enhancer function, the role of enhancers in evolution and disease susceptibility and how enhancers actually boost transcription. Further refinement of genome-wide techniques to study enhancer function will help to answer some of these questions. Tracking transcription factor binding and chromatin looping during differentiation will provide unprecedented insights into the dynamics of enhancer action. Although genome-wide approaches are currently in vogue to investigate enhancer function, answers to some of the remaining questions will still require the careful molecular dissection of selected model loci.

Even though progress in technologies has been impressive, several limitations remain. ChIP assays require knowledge regarding the factors involved in the regulation of gene loci and good quality antibodies against these factors are not always available. The genome-wide 3C spin offs currently lack resolution, which hampers the accurate determination of the exact contact points mediating enhancer–gene interactions. Furthermore, these methods all depend on protein–protein and protein–DNA cross-linking using formaldehyde requiring a certain amount of time, setting a limit on the temporal resolution of these methods. Information regarding cell-to-cell variability is still lacking, because the majority of the current methods to study enhancer function involve batch assays on many cells. Therefore, the field would greatly benefit from the development of single cell assays to study enhancer function. The integration of genome-wide data with focused, single locus data and single cell data will undoubtedly provide us with new exciting insights into the mechanisms that shape the genomic regulatory landscape in flux.

ACKNOWLEDGMENTS

We thank Ralph Stadhouders for carefully reading the manuscript. The literature regarding transcription factors and enhancers is overwhelming and the authors wish to apologize to those whose work we were not able to cite.

REFERENCES

Ahmadiyeh, N., Pomerantz, M. M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9742–9746.

Akhtar-Zaidi, B., Cowper-Sal-Lari, R., Corradin, O., Saiakhova, A., Bartels, C. F., Balasubramanian, D., et al. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. *Science* 336, 736–739.

Albuisson, J., Isidor, B., Giraud, M., Pichon, O., Marsaud, T., David, A., et al. (2011). Identification of two novel mutations in Shh long-range regulator associated with familial preaxial polydactyly. *Clin. Genet.* 79, 371–377.

Alvarez, M., Rhodes, S. J., and Bidwell, J. P. (2003). Context-dependent

transcription: all politics is local. *Gene* 313, 43–57.
Amouyal, M. (1991). The remote control of transcription, DNA looping and DNA compaction. *Biochimie* 73, 1261–1268.
Arnosti, D. N., and Kulkarni, M. M. (2005). Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* 94, 890–898.
Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729–740.
Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308.
Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837.
Bell, A. C., West, A. G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396.
Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* 25, 2453–2464.
Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* 28, 1045–1048.
Blow, M. J., Mcculley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* 42, 806–810.
Bogdanovic, O., Fernandez-Minan, A., Tena, J. J., De Lacalle-Mustienes, E., Hidalgo, C., Van Kruysbergen, I., et al. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res.* doi: 10.1101/gr.134833.111 [Epub ahead of print].
Bonn, S., Zinzen, R. P., Girardot, C., Gustafson, E. H., Perez-Gonzalez, A., Delhomme, N., et al. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* 44, 148–156.
Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* 28, 276–284.
Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). Long-range chromatin regulatory interactions *in vivo*. *Nat. Genet.* 32, 623–626.
Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Characterization of genome-wide enhancer–promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* 22, 490–503.
Chien, R., Zeng, W., Kawauchi, S., Bender, M. A., Santos, R., Gregson, H. C., et al. (2012). Cohesin mediates chromatin interactions that regulate mammalian {beta}-globin expression. *J. Biol. Chem.* 286, 17870–17878.
Chopra, V. S., Cande, J., Hong, J. W., and Levine, M. (2009). Stalled Hox promoters as chromosomal boundaries. *Genes Dev.* 23, 1505–1509.
Chopra, V. S., Kong, N., and Levine, M. (2012). Transcriptional repression via antilooping in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9460–9464.
Chorley, B. N., Wang, X., Campbell, M. R., Pittman, G. S., Noureddine, M. A., and Bell, D. A. (2008). Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat. Res.* 659, 147–157.
Chuang, C. H., Carpenter, A. E., Fuchsova, B., Johnson, T., De Lanerolle, P., and Belmont, A. S. (2006). Long-range directional movement of an interphase chromosome site. *Curr. Biol.* 16, 825–831.
Consortium, G. P. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.
Cotney, J., Leng, J., Oh, S., Demare, L. E., Reilly, S. K., Gerstein, M. B., et al. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res.* 22, 1069–1080.
Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21931–21936.
De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., et al. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 8, e1000384. doi: 10.1371/journal.pbio.1000384
de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24.
Degner, J. F., Pai, A. A., Pique-Regi, R., Veyrieras, J. B., Gaffney, D. J., Pickrell, J. K., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394.
Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.
Deng, B., Melnik, S., and Cook, P. R. (2012a). Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. *Semin. Cancer Biol.* doi: 10.1016/j.semcancer.2012.01.003 [Epub ahead of print].
Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., et al. (2012b). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244.
Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P., and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell* 1, 131–139.
Disanto, G., Sandve, G. K., Berlanga-Taylor, A. J., Morahan, J. M., Dobson, R., Giovannoni, G., et al. (2012). Genomic regions associated with multiple sclerosis are active in B cells. *PLoS ONE* 7, e32281. doi: 10.1371/journal.pone.0032281
Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
Dorsett, D. (2011). Cohesin: genomic insights into controlling gene transcription and development. *Curr. Opin. Genet. Dev.* 21, 199–206.
Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., et al. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 18, 2485–2490.
Du, M. J., Lv, X., Hao, D. L., Zhao, G. W., Wu, X. S., Wu, F., et al. (2008). MafK/NF-E2 p18 is required for beta-globin genes activation by mediating the proximity of LCR and active beta-globin genes in MEL cell line. *Int. J. Biochem. Cell Biol.* 40, 1481–1493.
Dundr, M., Ospina, J. K., Sung, M. H., John, S., Upender, M., Ried, T., et al. (2007). Actin-dependent intranuclear repositioning of an active gene locus *in vivo*. *J. Cell Biol.* 179, 1095–1103.
Ellis, J., Tan-Un, K. C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., et al. (1996). A dominant chromatin-opening activity in 5′ hypersensitive site 3 of the human beta-globin locus control region. *EMBO J.* 15, 562–568.
Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
Erokhin, M., Davydova, A., Kyrchanova, O., Parshikov, A., Georgiev, P., and Chetverina, D. (2011). Insulators form gene loops by interacting with promoters in *Drosophila*. *Development* 138, 4097–4106.
Feng, Y. Q., Warin, R., Li, T., Olivier, E., Besse, A., Lobell, A., et al. (2005). The human beta-globin locus control region can silence as well as activate gene expression. *Mol. Cell. Biol.* 25, 3864–3874.
Ferrai, C., Naum-Ongania, G., Longobardi, E., Palazzolo, M., Disanza, A., Diaz, V. M., et al. (2009). Induction of HoxB transcription by retinoic acid requires actin polymerization. *Mol. Biol. Cell* 20, 3543–3551.
Fujioka, M., Wu, X., and Jaynes, J. B. (2009). A chromatin insulator mediates transgene homing and very long-range enhancer–promoter communication. *Development* 136, 3077–3087.
1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.
Giresi, P. G., Kim, J., Mcdaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885.
Grosveld, F., Van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51, 975–985.
Haering, C. H., and Jessberger, R. (2012). Cohesin in determining chromosome architecture. *Exp. Cell Res.* 318, 1386–1393.
Hamada, H. (1986). Random isolation of gene activator elements from the human genome. *Mol. Cell. Biol.* 6, 4185–4194.
Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., et al.

(2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., et al. (1991). Importance of globin gene order for correct developmental expression. *Genes Dev.* 5, 1387–1394.

Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., et al. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* 470, 264–268.

Hawkins, R. D., Hon, G. C., and Ren, B. (2010). Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11, 476–486.

Hayday, A. C., Gillies, S. D., Saito, H., Wood, C., Wiman, K., Hayward, W. S., et al. (1984). Activation of a translocated human c-myc gene by an enhancer in the immunoglobulin heavy-chain locus. *Nature* 307, 334–340.

He, A., Kong, S. W., Ma, Q., and Pu, W. T. (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5632–5637.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.

Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. *Development* 139, 1045–1057.

Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., et al. (2009). Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat. Methods* 6, 283–289.

Hoffman, M. M., Buske, O. J., Wang, J., Weng, Z., Bilmes, J. A., and Noble, W. S. (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* 9, 473–476.

Hu, Q., Kwon, Y. S., Nunez, E., Cardamone, M. D., Hutt, K. R., Ohgi, K. A., et al. (2008). Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19199–19204.

Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X., et al. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell* 29, 232–242.

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., Van Berkum, N. L., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S. M., et al. (2010). Variation in transcription factor binding among humans. *Science* 328, 232–235.

Kellner, W. A., Ramos, E., Van Bortle, K., Takenaka, N., and Corces, V. G. (2012). Genome-wide phosphoacetylation of histone H3 at Drosophila enhancers and promoters. *Genome Res.* 22, 1081–1088.

Kim, S. I., Bultman, S. J., Kiefer, C. M., Dean, A., and Bresnick, E. H. (2009). BRG1 requirement for long-range interaction of a locus control region with a downstream promoter. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2259–2264.

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.

Kioussis, D., Vanin, E., Delange, T., Flavell, R. A., and Grosveld, F. G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* 306, 662–666.

Kleinjan, D. A., and Lettice, L. A. (2008). Long-range gene control and genetic disease. *Adv. Genet.* 61, 339–388.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18, 956–963.

Kooren, J., Palstra, R. J., Klous, P., Splinter, E., Von Lindern, M., Grosveld, F., et al. (2007). Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J. Biol. Chem.* 282, 16544–16552.

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.

Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. A., Sloane-Stanley, J. A., et al. (2012). Intragenic enhancers act as alternative promoters. *Mol. Cell* 45, 447–458.

Krebs, A. R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. *Mol. Cell* 44, 410–423.

Kvon, E. Z., Stampfel, G., Yanez-Cuna, J. O., Dickson, B. J., and Stark, A. (2012). HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev.* 26, 908–913.

Lan, X., Witt, H., Katsumura, K., Ye, Z., Wang, Q., Bresnick, E. H., et al. (2012). Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* 40, 7690–7704.

Laurell, T., Vandermeer, J. E., Wenger, A. M., Grigelioniene, G., Nordenskjold, A., Arner, M., et al. (2012). A novel 13 base pair insertion in the sonic hedgehog ZRS limb enhancer (ZRS/LMBR1) causes preaxial polydactyly with triphalangeal thumb. *Hum. Mutat.* 33, 1063–1066.

Leach, K. M., Nightingale, K., Igarashi, K., Levings, P. P., Engel, J. D., Becker, P. B., et al. (2001). Reconstitution of human beta-globin locus control region hypersensitive sites in the absence of chromatin assembly. *Mol. Cell. Biol.* 21, 2629–2640.

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98.

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

Liu, Z., Scannell, D. R., Eisen, M. B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 146, 720–731.

Maksimenko, O., Golovnin, A., and Georgiev, P. (2008). Enhancer–promoter communication is regulated by insulator pairing in a *Drosophila* model bigenic locus. *Mol. Cell. Biol.* 28, 5469–5477.

Maston, G. A., Landt, S. G., Snyder, M., and Green, M. R. (2012). Characterization of enhancer function from genome-wide analyses. *Annu. Rev. Genomics Hum. Genet.* 13, 29–57.

Matthews, J. M., and Visvader, J. E. (2003). LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep.* 4, 1132–1137.

May, D., Blow, M. J., Kaplan, T., Mcculley, D. J., Jensen, B. C., Akiyama, J. A., et al. (2011). Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* 44, 89–93.

McCord, R. P., Zhou, V. W., Yuh, T., and Bulyk, M. L. (2011). Distant cis-regulatory elements in human skeletal muscle differentiation. *Genomics* 98, 401–411.

Meireles-Filho, A. C., and Stark, A. (2009). Comparative genomics of gene regulation-conservation and divergence of cis-regulatory information. *Curr. Opin. Genet. Dev.* 19, 565–570.

Melgar, M. F., Collins, F. S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol.* 12, R113.

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

Mikkelsen, T. S., Xu, Z., Zhang, X., Wang, L., Gimble, J. M., Lander, E. S., et al. (2010). Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 143, 156–169.

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., De Boer, E., Langeveld, A., et al. (1996). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 87, 105–114.

Mitchell, J. A., and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* 22, 20–25.

Miyamoto, K., Pasque, V., Jullien, J., and Gurdon, J. B. (2012). Nuclear actin polymerization is required for transcriptional reprogramming of Oct4 by oocytes. *Genes Dev.* 25, 946–958.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466, 714–719.

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: regulating the transition from transcription initiation into productive elongation. *Biochim. Biophys. Acta* 1809, 34–45.

Ong, C. T., and Corces, V. G. (2012). Enhancers: emerging roles in cell fate specification. *EMBO Rep.* 13, 423–430.

Orom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143, 46–58.

Orom, U. A., and Shiekhattar, R. (2011). Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet.* 27, 433–439.

Palstra, R. J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and De Laat, W. (2008). Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS ONE* 3, e1661. doi: 10.1371/journal.pone.0001661

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and De Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35, 190–194.

Papantonis, A., and Cook, P. R. (2010). Genome architecture and the role of transcription. *Curr. Opin. Cell Biol.* 22, 271–276.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., et al. (2011). H3K4 trimethylation provides an epigenetic signature of active enhancers. *EMBO J.* 30, 4198–4210.

Petrykowska, H. M., Vockley, C. M., and Elnitski, L. (2008). Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. *Genome Res.* 18, 1238–1246.

Pham, T. H., Benner, C., Lichtinger, M., Schwarzfischer, L., Hu, Y., Andreesen, R., et al. (2012). Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood* 119, e161–e171.

Pomerantz, M. M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M. P., Doddapaneni, H., et al. (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.* 41, 882–884.

Raab, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., et al. (2011). Human tRNA genes function as chromatin insulators. *EMBO J.* 31, 330–350.

Raab, J. R., and Kamakaka, R. T. (2010). Insulators and promoters: closer than we think. *Nat. Rev. Genet.* 11, 439–446.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.

Rahimov, F., Marazita, M. L., Visel, A., Cooper, M. E., Hitchler, M. J., Rubini, M., et al. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat. Genet.* 40, 1341–1347.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., et al. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* 147, 1628–1639.

Reddy, T. E., Gertz, J., Pauli, F., Kucera, K. S., Varley, K. E., Newberry, K. M., et al. (2012). Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res.* 22, 860–869.

Ren, X., Siegel, R., Kim, U., and Roeder, R. G. (2011). Direct interactions of OCA-B and TFII-I regulate immunoglobulin heavy-chain gene transcription by facilitating enhancer–promoter communication. *Mol. Cell* 42, 342–355.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657.

Rodin, S., Kyrchanova, O., Pomerantseva, E., Parshikov, A., and Georgiev, P. (2007). New properties of *Drosophila* fab-7 insulator. *Genetics* 177, 113–121.

Royo, J. L., Hidalgo, C., Roncero, Y., Seda, M. A., Akalin, A., Lenhard, B., et al. (2011). Dissecting the transcriptional regulatory properties of human chromosome 16 highly conserved non-coding regions. *PLoS ONE* 6, e24824. doi: 10.1371/journal.pone.0024824

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., et al. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8309–8314.

Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., et al. (2006). Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays. *Nat. Methods* 3, 511–518.

Sawado, T., Halow, J., Bender, M. A., and Groudine, M. (2003). The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev.* 17, 1009–1018.

Schnetz, M. P., Bartels, C. F., Shastri, K., Balasubramanian, D., Zentner, G. E., Balaji, R., et al. (2009). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Res.* 19, 590–601.

Schnetz, M. P., Handoko, L., Akhtar-Zaidi, B., Bartels, C. F., Pereira, C. F., Fisher, A. G., et al. (2010). CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet.* 6, e1001023. doi: 10.1371/journal.pgen.1001023

Sheng, Y., and Previti, C. (2011). Genomic features and computational identification of human microRNAs under long-range developmental regulation. *BMC Genomics* 12, 270. doi: 10.1186/1471-2164-12-270

Simms, T. A., Dugas, S. L., Gremillion, J. C., Ibos, M. E., Dandurand, M. N., Toliver, T. T., et al. (2008). TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 7, 2078–2086.

Soler, E., Andrieu-Soler, C., De Boer, E., Bryne, J. C., Thongjuea, S., Stadhouders, R., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* 24, 277–289.

Song, S. H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol. Cell* 28, 810–822.

Soshnev, A. A., Li, X., Wehling, M. D., and Geyer, P. K. (2008). Context differences reveal insulator and activator functions of a Su(Hw) binding region. *PLoS Genet.* 4, e1000159. doi: 10.1371/journal.pgen.1000159

Spilianakis, C. G., and Flavell, R. A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* 5, 1017–1027.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., et al. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.

Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R. J., Bryne, J. C., Van Den Heuvel, A., et al. (2011). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO J.* 31, 986–999.

Svotelis, A., Bianco, S., Madore, J., Huppe, G., Nordell-Markovits, A., Mes-Masson, A. M., et al. (2011). H3K27 demethylation by JMJD3 at a poised enhancer of anti-apoptotic gene BCL2 determines ERalpha ligand dependency. *EMBO J.* 30, 3947–3961.

Taberlay, P. C., Kelly, T. K., Liu, C. C., You, J. S., De Carvalho, D. D., Miranda, T. B., et al. (2011). Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* 147, 1283–1294.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028.

Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J., et al. (1986). Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res.* 14, 7017–7029.

Tartof, K. D., and Henikoff, S. (1991). Trans-sensing effects from *Drosophila* to humans. *Cell* 65, 201–203.

Taslim, C., Chen, Z., Huang, K., Huang, T. H., Wang, Q., and Lin, S. (2012). Integrated analysis identifies a class of androgen-responsive genes regulated by short combinatorial long-range mechanism facilitated by CTCF. *Nucleic Acids Res.* 40, 4754–4764.

Teng, L., Firpi, H. A., and Tan, K. (2011). Enhancers in embryonic stem cells are enriched for transposable elements and genetic variations associated with cancers. *Nucleic Acids Res.* 39, 7371–7379.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091–1100.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. U.S.A.* 89, 11219–11223.

Tuan, D., Solomon, W., Li, Q., and London, I. M. (1985). The "beta-like-globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6384–6388.

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., et al. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* 17, 453–462.

Vernimmen, D., Lynch, M. D., De Gobbi, M., Garrick, D., Sharpe, J. A., Sloane-Stanley, J. A., et al. (2011). Polycomb eviction as a new distant enhancer function. *Genes Dev.* 25, 1583–1588.

Vieira, K. F., Levings, P. P., Hill, M. A., Crusselle, V. J., Kang, S. H., Engel, J. D., et al. (2004). Recruitment of transcription complexes to the beta-globin gene locus *in vivo* and *in vitro*. *J. Biol. Chem.* 279, 50350–50357.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009a). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858.

Visel, A., Rubin, E. M., and Pennacchio, L. A. (2009b). Genomic views of distant-acting enhancers. *Nature* 461, 199–205.

Visser, M., Kayser, M., and Palstra, R. J. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.* 22, 446–455.

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A., et al. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* 16, 3145–3157.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474, 390–394.

Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903.

Wendt, K. S., and Peters, J. M. (2009). How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Res.* 17, 201–214.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796–801.

West, A. G., and Fraser, P. (2005). Remote control of gene transcription. *Hum. Mol. Genet.* 14(Suppl. 1), R101–R111.

Weth, O., and Renkawitz, R. (2011). CTCF function is modulated by neighboring DNA binding factors. *Biochem. Cell Biol.* 89, 459–468.

Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., et al. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225.

Wiench, M., John, S., Baek, S., Johnson, T. A., Sung, M. H., Escobar, T., et al. (2011). DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* 30, 3028–3039.

Wright, J. B., Brown, S. J., and Cole, M. D. (2010). Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol. Cell. Biol.* 30, 1411–1420.

Wright, S., Rosenthal, A., Flavell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of beta-globin genes in murine erythroleukemia cells. *Cell* 38, 265–273.

Wu, C. (1980). The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* 286, 854–860.

Yoo, E. J., Cooke, N. E., and Liebhaber, S. A. (2012). An RNA-independent linkage of noncoding transcription to long-range enhancer function. *Mol. Cell. Biol.* 32, 2020–2029.

Zentner, G. E., Tesar, P. J., and Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 21, 1273–1283.

Zhang, X., Cowper-Sal Lari, R., Bailey, S. D., Moore, J. H., and Lupien, M. (2012). Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.* 22, 1437–1446.

Zhou, X., Baron, R. M., Hardin, M., Cho, M. H., Zielinski, J., Hawrylkiewicz, I., et al. (2012). Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum. Mol. Genet.* 21, 1325–1335.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res.* 35, 5532–5544.

Zhu, Y., Van Essen, D., and Saccani, S. (2012). Cell-type-specific control of enhancer activity by H3K9 trimethylation. *Mol. Cell* 46, 408–423.

Conflict of Interest Statement: 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.

Received: 09 July 2012; paper pending published: 23 July 2012; accepted: 12 September 2012; published online: 28 September 2012.

Citation: Palstra R-J and Grosveld F (2012) Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. Front. Gene. ***3****:195. doi: 10.3389/fgene.2012.00195*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

REVIEW ARTICLE
published: 15 October 2012
doi: 10.3389/fgene.2012.00214

More than insulator: multiple roles of CTCF at the *H19-Igf2* imprinted domain

Purnima Singh, Dong-Hoon Lee† and Piroska E. Szabó*

Department of Molecular and Cellular Biology, Beckman Research Institute, Duarte, CA, USA

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France
Patrick McGowan, Duke University, USA

***Correspondence:**
Piroska E. Szabó, Department of Molecular and Cellular Biology, Beckman Research Institute, City of Hope, 1500 E Duarte Rd, Duarte, CA 91010, USA.
e-mail: pszabo@coh.org

†Present address:
Dong-Hoon Lee, Department of Genetics and Biochemistry, Clemson University, Clemson, USA.

CTCF (CCCTC-binding factor)-mediated insulation at the *H19-Insulin-like growth factor 2 (Igf2)* imprinted domain is a classic example for imprinted gene regulation. DNA methylation difference in the imprinting control region (ICR) is inherited from the gametes and subsequently determines parental allele-specific enhancer blocking and imprinted expression in the soma. Recent genetic studies showed that proper monoallelic enhancer blocking at the *H19-Igf2* ICR is critical for development. Strict biallelic insulation at this locus causes perinatal lethality, whereas leaky biallelic insulation results in smaller size but no lethality. Apart from enhancer blocking, CTCF is also the master organizer of chromatin composition in the maternal allele along this imprinted domain, affecting not only histone tail covalent modifications but also those in the histone core. Additionally, CTCF binding in the soma protects the maternal allele from *de novo* DNA methylation. CTCF binding is not involved in the establishment of the gametic marks at the ICR, but it slightly delays *de novo* methylation in the maternally inherited ICR allele in prospermatogonia. This review focuses on the developmental and epigenetic consequences of CTCF binding at the *H19-Igf2* ICR.

Keywords: CTCF chromatin, imprinting, H19, Igf2, insulators, methylation, Zfp57, Trim28

CTCF (also known as CCCTC-binding factor) is a major organizer of the vertebrate genome and is essential for development (Moore et al., 2012). It is a versatile protein that regulates gene expression by binding to DNA via its multiple zinc fingers (Filippova, 2008; Ohlsson et al., 2010; Herold et al., 2012). CTCF plays roles in transcriptional activation and repression, insulation by enhancer blocking or chromosome barrier formation and organization of higher order chromatin by chromosomal looping and nuclear tethering (Phillips and Corces, 2009; Weth and Renkawitz, 2011; Barkess and West, 2012; Ghirlando et al., 2012). CTCF has been implicated in such diverse biological phenomena as genomic imprinting, X chromosome inactivation (Spencer et al., 2011), alternative splicing (Shukla et al., 2011), microsatellite instability (Libby et al., 2008), and V(D)J recombination (Guo et al., 2011). Several methodologies have been utilized for testing CTCF's function, including *in vitro* and cell culture assays, depletion or ablation of CTCF and its interactive partners, and deleting CTCF sites from episomal vectors, integrated transgenes or endogenous loci. The most direct functional test is to specifically inactivate the CTCF binding site(s) at an endogenous locus by point mutations. To date almost no such genetic studies exist in the latter category. One notable exception is the mouse *H19-Igf2* imprinted domain, which has been extensively studied in the past decade by several independent groups including ours. Precise point mutations have been made that inactivated the CTCF binding sites in the imprinting control region (ICR). In this review we will focus on some of the colorful roles that CTCF plays at the *H19-Igf2* imprinted locus. We will review that CTCF-mediated insulation controls reciprocal parental allele-specific expression of these two imprinted genes, emphasizing that correct monoallelic enhancer blocking at this locus is critical for normal fetal development. We will also summarize the roles CTCF plays in maintaining the epigenetic features of the maternal allele in the soma and, to some extent, in primordial germ cells (PGCs).

PARENTAL ALLELE-SPECIFIC ENHANCER INSULATION AT THE *H19-Igf2* IMPRINTED DOMAIN

CTCF-mediated insulation is a classic example for the regulation of genomic imprinting. Imprinted genes exhibit parental allele-specific expression (Ferguson-Smith, 2011; Abramowitz and Bartolomei, 2012). *Insulin-like growth factor* 2 (*Igf2*), and *H19* are neighboring genes, located on distal chromosome 7 in the mouse and expressed from the paternally or maternally inherited chromosome, respectively. Igf2 protein is important for promoting fetal and placental growth (DeChiara et al., 1990; Constancia et al., 2002) whereas the *H19* non-coding RNA moderates growth in the normal fetus (Gabory et al., 2009), puts the brake on the growth of the term placenta via its microRNA (Keniry et al., 2012) and also functions as a tumor suppressor (Yoshimizu et al., 2008). Both genes respond to the same endodermal enhancers that are distal to *H19* (Leighton et al., 1995) (**Figure 1A**). Between these two genes lies a 2.4 kb long differentially methylated region (DMR) that is required for the monoallelic expression of both the *H19* and *Igf2* genes, and therefore is called an ICR. Its deletion from the maternal allele results in biallelic *Igf2* expression and from the paternal allele in biallelic *H19* expression. Methylation of this DMR is exclusive to the paternally inherited chromosome and originates from the sperm (Tremblay et al., 1995,

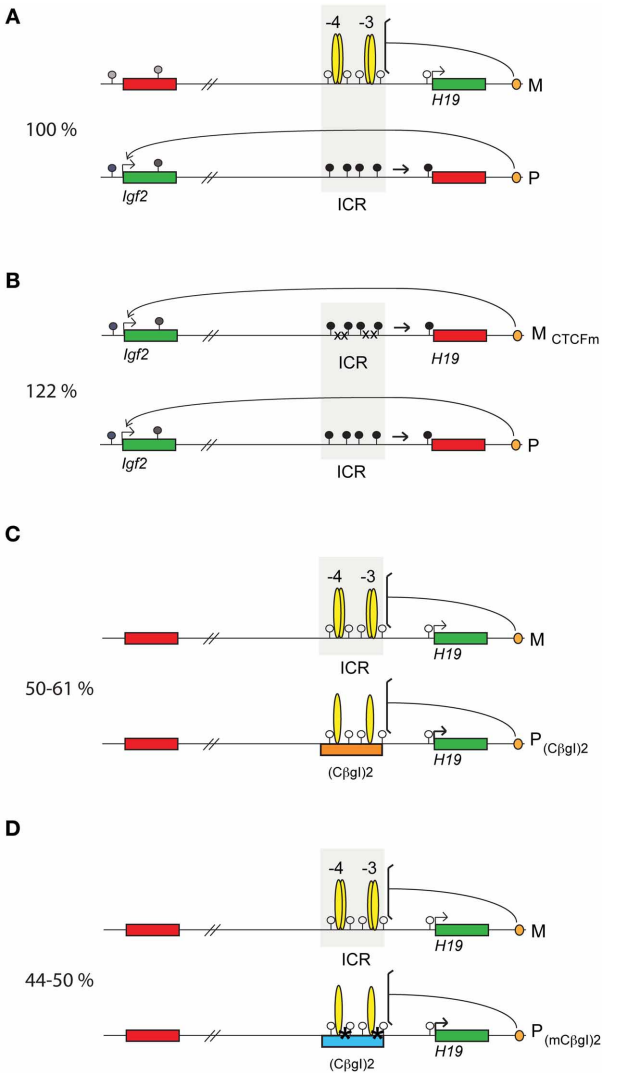


FIGURE 1 | Parental allele-specific enhancer insulation at the *H19-Igf2* imprinted domain. (A) Imprinted insulation at the *H19/Igf2* imprinted domain by CTCF binding in the ICR based on publications referenced in the text. Maternal chromosome (M): unmethylated (white lollipops) ICR (shaded area) is inherited from the oocyte. CTCF (yellow ovals) imparts insulator activity (bracket) between the *Igf2* promoters and the shared, downstream enhancers (orange oval). Initiation of *H19* expression depends on an unmethylated ICR during embryogenesis. Paternal chromosome (P): methylated (black lollipops) ICR is inherited from the sperm, CTCF cannot bind, hence ICR has no insulator activity thus, the *Igf2* promoters and enhancers can interact. During early postimplantation development, the *H19* promoter is inactivated by an ICR-dependent mechanism (horizontal arrow). **(B)** CTCF binding site mutations in the maternal ICR allele disrupt imprinted expression (Szabó et al., 2004). CTCF no longer binds in the mutant maternal chromosome (M_{CTCFm}), thus, the enhancers can access the *Igf2* promoter in both alleles. The mutant ICR is methylated and inactivates the *H19* promoter. **(C)** Non-imprinted insulation at the *H19/Igf2* locus by the chicken β-globin insulator duplex $(ChβGI)_2$ (orange rectangle) (Szabó et al., 2002). The $(ChβGI)_2$ is unmethylated and insulates the *Igf2* promoter from the shared enhancers when substituted for the ICR and transmitted maternally (not shown) or paternally (P), with 10% *Igf2* activity remaining. *H19* is overactivated 1.5-fold by the $(ChβGI)_2$ sequences in the paternal allele (bold arrow). **(D)** Biallelic insulation by the mutant chicken

(Continued)

FIGURE 1 | Continued
β-globin insulator duplex $(mChβGI)_2$ (turquoise rectangle) carrying mutations for boundary factor binding sites (stars) (Lee et al., 2010). Insulation is complete, with no detectable remaining *Igf2* expression. Relative fetus size for each genotype is shown to the left. Active genes and silent genes are depicted with green and red rectangles, respectively.

1997; Thorvaldsen et al., 1998). *Igf2* expression is also regulated by two additional paternally methylated DMRs. *Igf2* DMR1, upstream of the *Igf2* gene functions as a mesodermal silencer in the maternal allele (Constancia et al., 2000) while DMR2, in the sixth exon, functions as an enhancer in the paternal allele (Murrell et al., 2001).

To shed light on how the ICR regulates reciprocal expression of *Ig2* and *H19*, we used *in vivo* DNAseI, DMS footprinting and UV photofootprinting analysis of mouse embryo fibroblasts (MEFs) carrying maternal or paternal duplication of distal Chromosome 7 and discovered strong footprints at four consensus CTCF binding sites in the unmethylated maternal ICR allele but not in the methylated paternal allele. This provided evidence that the CTCF insulator protein blocks communication between the *Igf2* promoters and the shared downstream enhancers in the maternal chromosome (Szabó et al., 2000). At the same time, *in vitro* enhancer blocking, gelshift, episome assays, and *in vivo* ChIP assays confirmed that the *H19-Igf2* ICR acts as an enhancer blocker in the unmethylated maternal allele and CTCF binding is inhibited in the paternal ICR allele by DNA methylation, allowing *Igf2* promoter access to the enhancers (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000). To verify the enhancer blocker role of CTCF at this locus *in vivo*, CTCF-site mutations were introduced into the ICR allele in the mouse. Maternal transmission of these mutations resulted in biallelic *Igf2* expression and biallelic *H19* silencing (**Figure 1B**) (Pant et al., 2003; Schoenherr et al., 2003; Szabó et al., 2004; Han et al., 2008). CTCF has also been reported to be responsible at this locus for asynchronous replication of the two alleles: late replication of the maternal allele depends on CTCF binding (Bergstrom et al., 2007; Guibert et al., 2012). CTCF-dependent enhancer blocking requires cohesins (Rubio et al., 2008; Stedman et al., 2008; Nativio et al., 2009; Yao et al., 2010; Xiao et al., 2011) and involves regulating chromosome loop formation (Murrell, 2011).

Parental allele-specific CTCF binding has been detected recently at additional imprinted domains, at the *Rasgrf1* (Yoon et al., 2005), *Gtl2* (Lin et al., 2011), *Grb10* (Hikichi et al., 2003), *Kcnq1/Kcnq1ot1* (Fitzpatrick et al., 2007), and *Peg13* DMRs (Singh et al., 2011). It will be very interesting to test using genetic analyses whether these CTCF binding sites are required for regulating the allele-specific expression of imprinted transcripts by enhancer blocking.

MONOALLELIC INSULATION AT THE *H19-Igf2* ICR IS ESSENTIAL FOR NORMAL DEVELOPMENT

Genetic studies revealed that insulation strength of the *H19-Igf2* ICR has consequences to body size and viability. Insulation was absent at the *H19-Igf2* domain in mice carrying the ICR CTCF

site mutations in the maternal chromosome. This resulted in elevated *Igf2* expression and an overgrowth phenotype (**Figure 1B**). Prenatal fetuses were 122% heavier than their normal siblings (Szabó et al., 2004). We also noticed that adult males that carried the ICR CTCF site mutations became aggressive and fought frequently. Insulation was biallelic at this locus in mice where the ICR was replaced with two copies of the chicken beta globin insulator $(ChβGI)_2$ (**Figure 1C**) (Szabó et al., 2002). This introduced DNA fragment was of similar size to the ICR, had two CTCF binding sites, and also included sufficient number of CpG dinucleotides. The $(ChβGI)_2$ functioned as an enhancer blocker in the maternal allele. In the paternal allele, however, it behaved differently from the endogenous ICR. The $(ChβGI)_2$ did not attain *de novo* methylation in the male germ line and thus, it was not methylated in the paternally inherited allele in the somatic organs of $+/(ChβGI)_2$ fetuses. It consequently allowed biallelic CTCF binding and insulation of the *Igf2* promoters from the shared enhancers. *Igf2* expression was reduced to 10% of normal values and fetus size was reduced to 50–61% of normal littermates. *H19* expression was biallelic. Later a very similar mouse model was generated (Lee et al., 2010) that carried a mutant form of the $(mChβGI)_2$ sequences (**Figure 1D**). CTCF binding sites were retained in the $(mChβGI)_2$ but consensus sites for boundary proteins, USF1 (West et al., 2004; Yao et al., 2010) and VEZF1 (Clark et al., 1990; Dickson et al., 2010), were destroyed by point mutations. Although there was a slight, 32%, methylation at these sequences in the male germ line, paternal allele-specific methylation was not maintained in the soma. In $+/(mChβGI)_2$ offspring insulation was again biallelic, and even more strict than the insulation in $+/(ChβGI)_2$ fetuses. *Igf2* expression was undetectable and fetus size was reduced to 44–50% of normal littermates. Whereas the $+/(ChβGI)_2$ mice were viable, a fully penetrant perinatal lethality occurred in the $+/(mChβGI)_2$ genotype (**Figure 2A**). The absence of *Igf2* likely contributed to the lethality phenotype of $+/(mChβGI)_2$, but was not the sole cause, because *Igf2* homozygous mutant mice are small but viable (DeChiara et al., 1990). Similar conclusion was reached in the reciprocal experiment (**Figure 2B**), when perinatal lethality of mice carrying maternal duplication of distal chromosome 7 (MatDup.dist7) was rescued by introducing the CTCF site mutations into one allele of the *H19-Igf2* ICR (also called IC1) (Han et al., 2010). Correcting biallelic insulation of the *H19-Igf2* ICR was sufficient to rescue lethality, even though the duplicated chromosome region of MatDup.dist7 mice also carries the Kcnq1ot1 maternally methylated DMR (also called IC2), and additional misexpressed imprinted genes. These results have revealed that correct insulator dose and strength at the *H19-Igf2* ICR is required for perinatal viability: strict biallelic insulation at this imprinted locus is not tolerated in development.

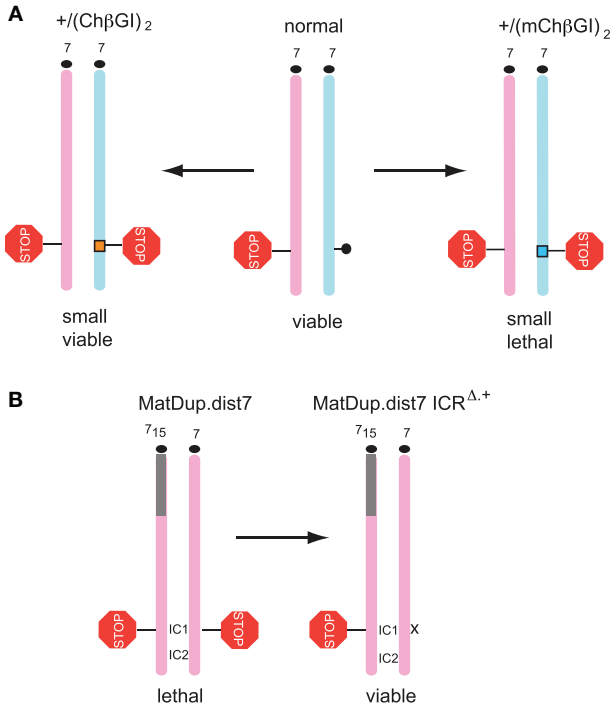


FIGURE 2 | Biallelic insulation at the ICR is not tolerated in development. (A) Introducing strict biallelic insulation to the ICR causes lethality. Substituting the paternal chromosome's (light blue) methylated (black lollipop) ICR of normal mice (middle) with the $(ChβGI)_2$ (Szabó et al., 2002) (orange box) or the $(mChβGI)_2$ (Lee et al., 2010) (turquoise box) has resulted in biallelic insulation (STOP signal). Lethality was observed in the $+/(mChβGI)_2$ but not in the $+/(ChβGI)_2$ genotype. The $+/(mChβGI)_2$ had strict insulation but the $+/(ChβGI)_2$ exhibited leaky insulation. **(B)** Maternal (pink) duplication of distal chromosome 7 (MatDup.dist7) fetuses that carry biallelic insulation at the ICR, also called imprinting control center 1 (IC1), have 40% body weight and die. The lethality phenotype is rescued by maternal transmission of one copy of the mutant IC1 (x) that lacks CTCF binding and insulator function (Han et al., 2010). The imprinting control center 2 (IC2) is bi-maternal. Correction of biallelic ICR insulation to monoallelic insulation is sufficient to rescue perinatal lethality of the MatDup.dist7 genotype.

CTCF IS THE MAJOR EPIGENETIC ORGANIZER OF THE MATERNAL ALLELE IN THE SOMA

CTCF is the master organizer of the maternal allele's chromatin (**Figure 3**). Utilizing single nucleotide polymorphisms (SNPs) between parental mouse lines and using quantitative allele-specific chromatin immunoprecipitation single nucleotide primer extension (SNuPE) assays, we measured the chromatin composition along the *H19/Igf2* imprinted domain in normal cells and cells with engineered mutations at the four ICR-CTCF binding sites. The chromatin composition showed great polarization along the *H19/Igf2* imprinted domain (Han et al., 2008; Singh et al., 2010a,b, 2011). Whereas the *H19* gene, promoter, and ICR were enriched in active chromatin marks, H3K4me2, H3K4me3, and H3K9ac in the maternal allele, the paternal allele of the same regions was enriched in repressive chromatin marks, such as H3K9me3 and H3K79me3. The ICR was slightly maternally biased for H3K4ac, H3K18ac, H3K36ac, H3K79ac, H4K5ac, H4K8ac, H4K12ac, and H4K91ac marks, but showed biallelic H3K27me3 enrichment. The *Igf2* promoter, DMR1 and DMR2 regions, were enriched in active marks, H3K4me2, H3K4me3, H3K9ac, H3K4ac, H3K18ac, H3K36ac, H3K79ac, H4K5ac, H4K8ac, H4K12ac, H4K91ac, H3K79me1, and H3K79me2 in the paternal allele but repressive marks, H3K27me3, H3K9me3 and repressive histone variant macroH2A1 in the maternal allele.

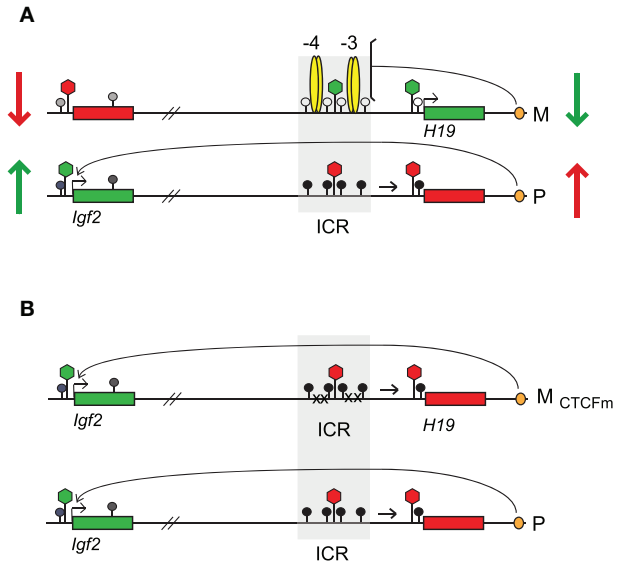


FIGURE 3 | CTCF is the major epigenetic organizer of the maternal allele in the soma. (A) Domain-wide allele-specific epigenetic features of the *H19*/*Igf2* imprinted domain. DNA methylation is paternal allele-specific in the ICR and at the *H19* promoter. Methylation is also paternally biased at the *Igf2* promoter and DMR2 (lollipops with shades of gray). Histone covalent modificatons are polarized along the domain. Active chromatin marks (green hexagon) exist at the active gene copies and in the maternal ICR but repressive marks (red hexagon) exist in the silent gene copies and the paternal ICR. **(B)** CTCF binding in the ICR is required for domain-wide epigenetic features. The maternal chromosome that carries CTCF binding site mutations (M_{CTCFm}) becomes very similar to the normal paternal chromosome in each epigenetic feature, DNA methylation and chromatin composition. Vertical arrows in **(A)** depict the changes in enrichment of active (green) and repressive (red) chromatin marks at the *Igf2* or *H19* regions that occur in response to CTCF site mutations.

Abolishing CTCF binding in the *H19-Igf2* ICR in the mutant cells resulted in a complete reorganization of the allele-specific chromatin composition (Han et al., 2008). In the maternal allele CTCF site mutant cells exhibited reduced H3K9ac, H3K4me2, and H3K4me3 at the *H19* ICR, promoter, gene body and reduced H3K27me3 at the *Igf2* P2 promoter and *Igf2* DMRs. These results revealed that ICR-CTCF binding is required for recruiting the maternal allele-specific active marks, H3K9ac, H3K4me2, and H3K4me3 at the *H19* locus and the maternal allele-specific repressing mark H3K27me3 and macroH2A1 at the *Igf2* locus. In agreement with these findings, it was shown that active histone tail modifications at the *H19* promoter depend on the activity state of the promoter (Verona et al., 2008) and that CTCF directly recruits the polycomb protein Suz12 to the *Igf2* locus to catalyze H3K27 trimethylation (Li et al., 2008a). In the paternal allele H3K27me3 and macroH2A1 levels increased and became biallelic in the CTCF site-mutant cells at the *H19* promoter while paternal H3K4me2 and H3K9ac increased and became biallelic at the *Igf2* DMRs. Indeed, histone acetylation at each lysine residue increased and became biallelic in the mutant cells at the *Igf2* DMR1, P2 promoter and DMR2, where it was paternal allele-specific in normal cells (Singh et al., 2010a). These results provided evidence that in the absence of CTCF binding, the mutant maternal chromosome accumulates histone marks that normally exist in the paternal chromosome. Therefore, CTCF binding in the ICR is required for excluding repressive chromatin from the *H19* region and excluding active chromatin, such as histone acetylation from the maternal allele at the *Igf2* locus at a distance.

When we examined how CTCF binding affects the histone globular domain modifications in the *H19-Igf2* imprinted domain (Singh et al., 2010b), we found that the ICR CTCF site point mutations caused a twofold increase in the heterochromatin mark H3K79me3 at the ICR sequences. Whereas it was strongly paternal allele-specific in normal cells, H3K79me3 became biallelic in the mutant cells at the ICR and at the *H19* promoter, providing evidence that at these sequences CTCF is required for excluding H3K79me3 from the maternal allele. The ICR CTCF site point mutations also caused a twofold increase of H3K79me1 and H3K79me2 levels in the mutant cells at the *Igf2* P2 promoter and *Igf2* DMRs where these paternal allele-specific activating chromatin marks became biallelic. H3K79me1 and H3K79me2 levels were low in abundance and biallelic at the *H19* locus and H3K79me3 levels were relatively high and biallelic at the *Igf2* regions, but these features did not change in response to the CTCF site mutations, indicating that CTCF-ICR binding is not responsible in the maternal allele for including H3K79me2 at the *H19* region and H3K79me3 at the *Igf2* locus. Taken together, with regard to globular domain modifications, the ICR CTCF site mutations have caused the paternalization of the maternal allele's chromatin composition along the *H19*/*Igf2* imprinted domain by exclusion: CTCF was responsible for the maternal allele's chromatin composition by excluding H4K91ac, H3K79me1, and H3K79me2 at the *Igf2* locus and by excluding H3K79me3 at the *H19* locus from the maternal allele.

In summary, with regard to histone tail modifications, in the maternal allele CTCF binding recruited active chromatin at the *H19* locus and repressive chromatin at the *Igf2* locus, and also excluded repressive chromatin at the *H19* locus and active chromatin from the *Igf2* locus (Han et al., 2008; Singh et al., 2010a). However, CTCF did not recruit globular domain modifications to the maternal allele, rather excluded them from the maternal allele at the *Igf2* locus (Singh et al., 2010b). It will be important to find out the mechanism of how CTCF interacts with different epigenetic modifiers in achieving the maternal allele's epiphenotype.

CONTROL OF DNA METHYLATION AT THE DMR

The key to all other parental allele-specific features at the *H19-Igf2* imprinted domain is the paternal-specific methylation of the ICR, because this determines monoallelic CTCF binding, and in turn CTCF binding determines monoallelic gene expression and maintenance of the polarized epigenetic features. It is important, therefore, to review here the imprint cycle of the ICR and discuss how this cycle is related to CTCF. The methylation mark in the *H19/Igf2* ICR is erased between generations in PGCs (Hajkova et al., 2002) and is subsequently reestablished specifically in male fetal germ cells (Davis et al., 1999, 2000; Ueda et al., 2000; Kato et al., 2007). After that ICR methylation is maintained throughout

spermatogenesis, fertilization, global epigenomic reprogramming in the zygote, preimplantation, and later during cell divisions in the soma (Li et al., 1993; Tucker et al., 1996; Hirasawa et al., 2008).

It is not known what initiates the paternal-specific methylation at the *H19-Igf2* DMR in the male germ line, but it depends on the *de novo* methyltransferase Dnmt3a and its cofactor, Dnmt3L (Bourc'his et al., 2001; Kato et al., 2007; Kaneda, 2011). Even though the CTCF binding sites maintain allele-specific methylation differences in the soma (see below), the same sites are not required for setting the gametic imprint in the germ line. The ICR that harbors CTCF site mutations is fully methylated in perinatal male fetal germ cells and is fully unmethylated in fetal female germ cells and ovulated oocytes (Schoenherr et al., 2003; Szabó et al., 2004). CTCF protein may affect the maintenance of unmethylated ICR in the oocyte indirectly, because CTCF-depleted oocytes exhibit increased methylation at that region (Fedoriw et al., 2004). The methylation imprinting process at the ICR in the male germ line appears to depend on two components, the ICR sequences and also the location of the ICR inside the *H19-Igf2* domain. The $(ChβGI)_2$ and the $(mChβGI)_2$ inserts (**Figures 1C** and **D**) attained only 11 and 32% methylation in place of the ICR in 18.5 days post-coitum (dpc) prospermatogonia, respectively, suggesting that ICR sequences are important for full methylation establishment in the male germ line (Szabó et al., 2002; Lee et al., 2010). When the ICR was introduced to other genomic locations, methylation imprint establishment did not occur in the male germ line, but paternal allele-specific methylation was acquired only later in the soma. However, when the ICR was placed downstream of the *H19* gene, it attained *de novo* methylation in the male germ line (Park et al., 2004; Tanimoto et al., 2005; Matsuzaki et al., 2009, 2010; Gebert et al., 2010). These studies suggested that the *H19-Igf2* domain's genomic location is also important for proper imprint establishment of the *H19-Igf2* ICR. It will be important to find the DNA sequences—inside and outside the ICR—that are necessary and sufficient for the mechanism of methylation imprint establishment of the ICR in prospermatogonia.

After imprint establishment the methylation of the *H19-Igf2* DMR is protected in the zygote's paternal pronucleus during the wave of zygotic reprogramming (Mayer et al., 2000; Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011) by the PGC7 protein (Nakamura et al., 2007). PGC7 is proposed to protect the *H19-Igf2* DMR from 5mC oxidation by Tet3 methylcytosine oxidase in a H3K9me2-dependent manner, similarly to how PGC7 protects the female pronucleus (Nakamura et al., 2012). H3K9me2 association at this locus is inherited from the sperm and may be sufficient to attract tight PGC7 binding, which in turn is expected to reduce Tet3 affinity to these regions (Nakamura et al., 2012). The repressor protein MBD3 is slightly biased toward the paternal allele of the ICR in ES cells and, according to MBD3 knockdown experiments, contributes to protecting CpG methylation of the paternal allele of the *H19-Igf2* DMR during preimplantation development (Reese et al., 2007). Genetic studies revealed that two additional proteins protect the ICR methylation during early development. Zfp57 transcription factor protects the ICR in ES cells (Zuo et al., 2012) and Trim28 (also known as KAP1) protects it in the embryo (Messerschmidt et al., 2012). Trim28 binds to the ICR in midgestation stage embryos (Messerschmidt et al., 2012). Both Zfp57 and Trim28 are associated with the methylated paternal allele of the ICR In ES cells (Quenneville et al., 2011). Zfp57-Trim28-Setdb1 triple occupied ChIP-sequencing peaks defined a consensus hexanucleotide sequence, TGC^{m}CGC where the CpG site is methylated (Quenneville et al., 2011). This consensus is present at each DMR, including the *H19-Igf2* ICR.

In somatic organs, the maternal allele's epigenetic profile at the *H19-Igf2* domain depends on CTCF binding in the ICR. CTCF binding is responsible for protecting the maternal allele from DNA methylation (**Figure 3**). Maternal inheritance of mutations in the CTCF binding sites resulted in highly elevated CpG methylation levels in somatic organs at the ICR (Pant et al., 2003; Schoenherr et al., 2003; Szabó et al., 2004), as well as the *H19* promoter, and *H19* gene body and even at the *Igf2* DMR1 and DMR2 sequences at ∼90-kb distance (Kurukuti et al., 2006; Han et al., 2008).

It is interesting to note that the Zfp57-Trim28-Setdb1 consensus sites overlap with three CTCF binding motifs in the ICR (**Figure 4**). At these sites the maternal allele has robust *in vivo* CTCF footprints in MEF. However, in MEFs no clear DNAseI footprints are discernable in the paternal allele (Szabó et al., 2000). Zfp57-Trim28 binding may only take place in the ICR at earlier time points, before the time of MEF derivation. Incidentally, the Zfp57-Trim28-Setdb1 consensus sites have been mutated in the *H19-Igf2* ICR (well before the consensus site was discovered) at the endogenous locus and in integrated transgenes (Engel et al., 2004; Matsuzaki et al., 2010). These mutations destroyed the Zfp57-Trim28-Setdb1 consensus sites such way that CTCF binding was not affected (**Figure 4**). As a result, methylation was reduced and insulator activity was gained in the mutant paternal ICR, likely because the reduced DNA methylation allowed CTCF binding. Zfp57-Trim28 may protect the ICR from demethylation by attracting repressing

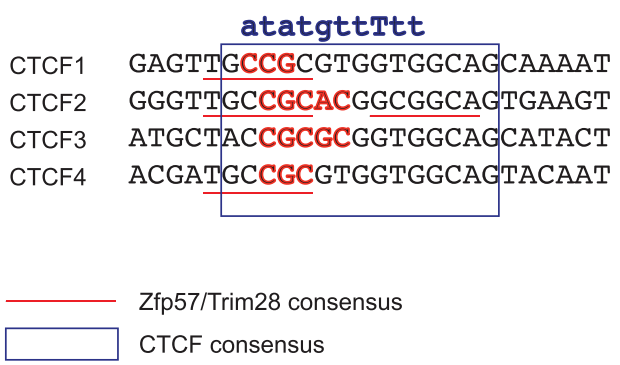


FIGURE 4 | Overlapping binding sites in the ICR for CTCF and Zfp57-Trim28-Setdb1 repressor complex explain their antagonistic roles at the ICR. The four CTCF binding sites (inside the blue rectangle) of the ICR are shown with the intertwined consensus sequences defined by Zfp57-Trim28-Setdb1 (underlined in red). The nucleotides that were mutated by point mutations (Engel et al., 2004) that specifically destroy Zfp57-Trim28-Setdb1 consensus sequence are in red. Point mutations that abolish CTCF binding and also destroy Zfp57-Trim28-Setdb1 consensus sites are written in the top line.

epigenetic modifiers and DNMTs to the target sequences and by facilitating heterochromatinization and DNA remethylation (Quenneville et al., 2011; Zuo et al., 2012), although this function may be redundant, because the Zfp57 null mutant midgestation embryos did not exhibit reduced ICR DNA methylation (Li et al., 2008b). It is interesting that Zfp57-Trim28-mediated protection of DNA methylation is required in the *H19-Igf2* ICR only when CTCF binding sites are present. When the CTCF consensus was destroyed together with the Zfp57-Trim28-Setdb1 consensus (**Figure 4**), DNA methylation maintenance was not affected (Szabó et al., 2004). Zfp57-Trim28's role at the ICR, therefore, is specific to preventing CTCF binding in the paternal allele by maintaining DNA methylation. One extension of this idea is that CTCF may protect the maternal allele from DNA methylation by preventing Zfp57-Trim28-Setdb1 binding. Therefore, the antagonistic roles (Engel et al., 2004) of the composite ICR CTCF sites are the following: to maintain the methylation-free status of the maternal chromosome through CTCF binding and to maintain DNA methylation in the paternal chromosome through Zfp57-Trim28-Setdb1 binding.

CTCF-DEPENDENT CHROMATIN BIAS DELAYS *de novo* METHYLATION OF THE MATERNAL ICR ALLELE IN MALE GERM CELLS

The process of methylation imprint erasure at the ICR is complete in PGCs by 13.5 dpc (**Figure 5**). Consequently, male fetal germ cells undergo *de novo* methylation at the ICR during fetal development, whereas female germ cells remain unmethylated till the end of oocyte maturation. It was noticed by several laboratories that the two ICR alleles are different in male germ cells with respect to the speed of *de novo* methylation. Methylation of the paternally inherited ICR allele precedes the maternally inherited allele (Davis et al., 1999, 2000; Ueda et al., 2000; Kato et al., 2007), implying that the two alleles are distinguished by an epigenetic mark, other than DNA methylation in 13.5 dpc prospermatogonia. We hypothesized that the chromatin composition may constitute this transient epigenetic memory and this in turn depends on maternal-allele-specific binding of CTCF in PGCs. In order to test our hypothesis we isolated fetal germ cells from mice that carry SNPs at the ICR to distinguish the parental chromosomes. Using allele-specific ChIP-SNuPE and real-time reverse-transcription PCR assays we found that CTCF was slightly biased toward the maternal allele, but it had a very low level of enrichment at 13.5 dpc at the ICR, suggesting that CTCF is almost completely removed from the ICR in germ cells before midgestation. The repressive histone mark, H3K9me3, was slightly biased toward the paternal allele at the ICR but its enrichment level was very low whereas the active mark, H3K4me2 was more abundant and it was slightly biased toward the maternal allele in prospermatogonia at 13.5 and 15.5 dpc. The level of H3K4me2 allelic bias was similar to the methylation bias between alleles (10–15%). When the maternal allele carried the CTCF site mutations in prospermatogonia, the chromatin bias was no longer observed at the ICR, suggesting that chromatin composition of the ICR depends on maternal-allele specific CTCF binding in PGCs, just like it does in somatic cells (Han et al., 2008; Singh et al., 2010a,b). The methylation bias was also absent between the parental alleles in the mutant prospermatogonia. These findings are consistent with the explanation that CTCF binding in PGCs is responsible for setting up a chromatin bias in PGCs, and that this chromatin is not fully erased in prospermatogonia before *de novo* methylation commences. Therefore, CTCF-dependent chromatin bias may influence the rate of DNA methylation in the parental alleles. We concluded

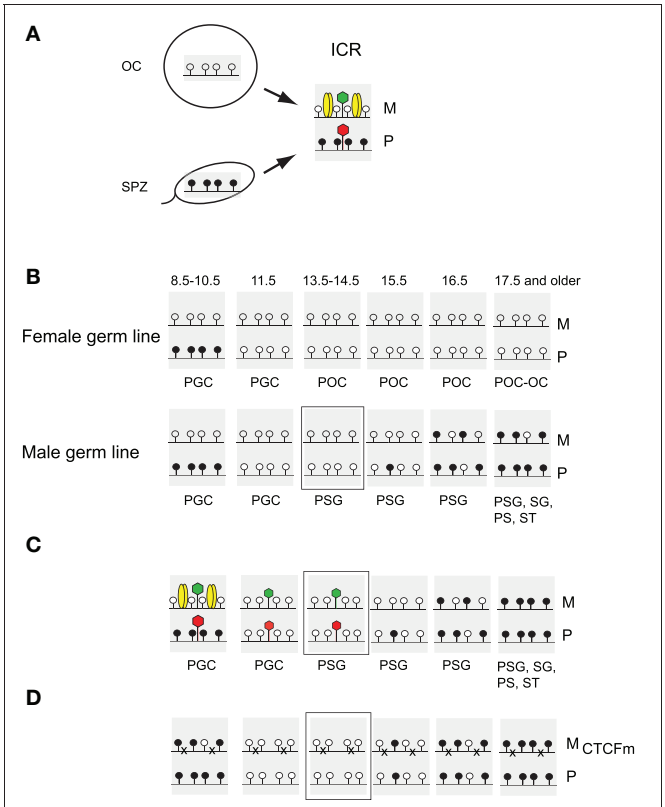


FIGURE 5 | CTCF binding delays *de novo* methylation of the maternal allele in male germ cells. (A) Differential methylation of the ICR is inherited from the gametes: methylation of the paternal allele (P) from spermatozoa (SPZ) and unmethylation of the maternal allele (M) from oocytes (OC). This primary methylation difference determines CTCF binding and chromatin composition in the soma and likely also in primordial germ cells (PGC), which exhibit imprinted *H19* and *Igf2* expression. Active or repressive chromatin (green or red hexagon) is present at respective alleles of the ICR. **(B)** Fate of the imprint in the female and male germ lines. Methylation status of the ICR is depicted in the primordial germ cells (PGC), primary oocytes (POC) and in prospermatogonia (PSG), spermatogonia (SG) pachytene spermatocytes (PS) and round spermatids (ST) with gestational stages in dpc. The developmental stage that appears epigenetically different without DNA methylation is marked with a rectangle. **(C)** Imprint establishment of the ICR in the normal male germ line. Expected CTCF binding and chromatin composition is depicted in primordial germ cells (PGC). Observed chromatin bias is depicted in prospermatogonia (PSG). Chromatin bias is observed in the normal ICR between the parental alleles in the absence of CpG methylation at 13.5–14.5 dpc. **(D)** Functional CTCF sites are required for chromatin bias and delayed methylation of the maternally inherited ICR allele. Maternal inheritance of the CTCF binding site mutations abolishes CTCF binding in the maternal allele in PGCs. No chromatin bias is observed between parental alleles at 13.5–14.5 dpc and the maternal allele's methylation is not delayed at 15.5–17.5 dpc.

that it is the H3K4me2 histone mark that most likely constitutes the epigenetic memory of the mother in prospermatogonia at 13.5–14.5 dpc and delays *de novo* CpG methylation in the maternal ICR allele. Indeed, removal of H3K4me2 by H3K4 demethylase KDM1B is required at least at certain maternal DMRs for the establishment of methylation imprints in oocytes (Ciccone et al., 2009). It is known that certain maternal DMRs exhibit delayed *de novo* methylation in the paternally inherited allele (Hiura et al., 2006). It will be interesting to find out using genetic analyses whether CTCF or other transcription factor provides transient epigenetic memory for those alleles.

In summary, CTCF plays complex roles at the *H19-Igf2* ICR. All of these roles may appear at first to depend on its major role at the domain, which is enhancer blocking. However, CTCF also protects the ICR from DNA methylation in the maternal allele and also sets up the maternal allele's chromatin composition in the soma and to some extent in PGCs. These functions at a single locus illuminate the versatility of CTCF in organizing gene expression and also in structuring the genome. It will be important to carry out similar genetic experiment by precisely inactivating the binding sites using point mutations to understand whether CTCF organizes local and domain-wide chromatin composition and/or maintains the unmethylated state at other loci in the genome, especially those that where insulator function has been shown (Herold et al., 2012). At least at one other locus, at the β-globin cluster 3′HS1, CTCF binding was shown to be required for recruiting active chromatin mark H3K9ac and repelling the repressing marks H3K9/27me3 (Splinter et al., 2006). We will be very curious to see whether CTCF binding sites in the Xist/Tsix RS14 region (Spencer et al., 2011) regulate the choice of X chromosome for inactivation by orchestrating local or domain-wide chromatin composition. Interestingly, mutations in the corresponding human sites either increase or decrease CTCF binding affinity and also reciprocally affect X inactivation skewing (Pugacheva et al., 2005). It will be especially critical to find out whether CTCF carries out its chromatin organizing activities parental allele-specifically at other imprinted domains and if proper CTCF binding at those DMRs is essential for development. We expect that this will be true at least at the *Dlk1-Gtl2* imprinted domain, because CTCF binding is allele-specific in a strategically important location at the *Gtl2* promoter (Lin et al., 2011) and because of the known lethality phenotypes associated with the misregulation of allele-specific expression at this imprinted domain (Lin et al., 2003; Wu et al., 2006; Takahashi et al., 2009, 2010).

ACKNOWLEDGMENTS

This work was supported by a Public Health Service grant (GM064378) from the National Institute of General Medicine to Piroska E. Szabó.

REFERENCES

Abramowitz, L. K., and Bartolomei, M. S. (2012). Genomic imprinting: recognition and marking of imprinted loci. *Curr. Opin. Genet. Dev.* 22, 72–78.

Barkess, G., and West, A. G. (2012). Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* 4, 67–80.

Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405, 482–485.

Bergstrom, R., Whitehead, J., Kurukuti, S., and Ohlsson, R. (2007). CTCF regulates asynchronous replication of the imprinted H19/Igf2 domain. *Cell Cycle* 6, 450–454.

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.

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., et al. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461, 415–418.

Clark, S. P., Lewis, C. D., and Felsenfeld, G. (1990). Properties of BGP1, a poly(dG)-binding protein from chicken erythrocytes. *Nucleic Acids Res.* 18, 5119–5126.

Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G., and Reik, W. (2000). Deletion of a silencer element in Igf2 results in loss of imprinting independent of H19. *Nat. Genet.* 26, 203–206.

Constancia, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., et al. (2002). Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature* 417, 945–948.

Davis, T. L., Trasler, J. M., Moss, S. B., Yang, G. J., and Bartolomei, M. S. (1999). Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* 58, 18–28.

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.

DeChiara, T. M., Efstratiadis, A., and Robertson, E. J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* 345, 78–80.

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., et al. (2010). VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* 6:e1000804. doi: 10.1371/journal.pgen.1000804

Engel, N., West, A. G., Felsenfeld, G., and Bartolomei, M. S. (2004). Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. *Nat. Genet.* 36, 883–888.

Fedoriw, A. M., Stein, P., Svoboda, P., Schultz, R. M., and Bartolomei, M. S. (2004). Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* 303, 238–240.

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

Filippova, G. N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. *Curr. Top. Dev. Biol.* 80, 337–360.

Fitzpatrick, G. V., Pugacheva, E. M., Shin, J. Y., Abdullaev, Z., Yang, Y., Khatod, K., et al. (2007). Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1. *Mol. Cell. Biol.* 27, 2636–2647.

Gabory, A., Ripoche, M. A., Le Digarcher, A., Watrin, F., Ziyyat, A., Forne, T., et al. (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* 136, 3413–3421.

Gebert, C., Kunkel, D., Grinberg, A., and Pfeifer, K. (2010). H19 imprinting control region methylation requires an imprinted environment only in the male germ line. *Mol. Cell. Biol.* 30, 1108–1115.

Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., Yao, H., et al. (2012). Chromatin domains, insulators, and the regulation of gene expression. *Biochim. Biophys. Acta* 1819, 644–651.

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

Guibert, S., Zhao, Z., Sjolinder, M., Gondor, A., Fernandez, A., Pant, V., et al. (2012). CTCF-binding sites within the H19 ICR differentially regulate local chromatin structures and cis-acting functions. *Epigenetics* 7, 361–369.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., et al. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15–23.

Han, L., Lee, D. H., and Szabó, P. E. (2008). CTCF is the master

organizer of domain-wide allele-specific chromatin at the H19/Igf2 imprinted region. *Mol. Cell. Biol.* 28, 1124–1135.

Han, L., Szabó, P. E., and Mann, J. R. (2010). Postnatal survival of mice with maternal duplication of distal chromosome 7 induced by a Igf2/H19 imprinting control region lacking insulator function. *PLoS Genet.* 6:e1000803. doi: 10.1371/journal.pgen.1000803

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.

Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. *Development* 139, 1045–1057.

Hikichi, T., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2003). Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res.* 31, 1398–1406.

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.

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.

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

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C. F., Wolffe, A., et al. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* 10, 853–856.

Kaneda, M. (2011). Genomic imprinting in mammals-epigenetic parental memories. *Differentiation* 82, 51–56.

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.

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., et al. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* 14, 659–665.

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., et al. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10684–10689.

Lee, D. H., Singh, P., Tsark, W. M., and Szabó, P. E. (2010). Complete biallelic insulation at the H19/Igf2 imprinting control region position results in fetal growth retardation and perinatal lethality. *PLoS ONE* 5:e12630. doi: 10.1371/journal.pone.0012630

Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L., and Tilghman, S. M. (1995). An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev.* 9, 2079–2089.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365.

Li, T., Hu, J. F., Qiu, X., Ling, J., Chen, H., Wang, S., et al. (2008a). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell. Biol.* 28, 6473–6482.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., et al. (2008b). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev. Cell* 15, 547–557.

Libby, R. T., Hagerman, K. A., Pineda, V. V., Lau, R., Cho, D. H., Baccam, S. L., et al. (2008). CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet.* 4:e1000257. doi: 10.1371/journal.pgen.1000257

Lin, S., Ferguson-Smith, A. C., Schultz, R. M., and Bartolomei, M. S. (2011). Nonallelic transcriptional roles of CTCF and cohesins at imprinted loci. *Mol. Cell. Biol.* 31, 3094–3104.

Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., et al. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat. Genet.* 35, 97–102.

Matsuzaki, H., Okamura, E., Fukamizu, A., and Tanimoto, K. (2010). CTCF binding is not the epigenetic mark that establishes post-fertilization methylation imprinting in the transgenic H19 ICR. *Hum. Mol. Genet.* 19, 1190–1198.

Matsuzaki, H., Okamura, E., Shimotsuma, M., Fukamizu, A., and Tanimoto, K. (2009). A randomly integrated transgenic H19 imprinting control region acquires methylation imprinting independently of its establishment in germ cells. *Mol. Cell. Biol.* 29, 4595–4603.

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

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.

Moore, J. M., Rabaia, N. A., Smith, L. E., Fagerlie, S., Gurley, K., Loukinov, D., et al. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. *PLoS ONE* 7:e34915. doi: 10.1371/journal.pone.0034915

Murrell, A. (2011). Setting up and maintaining differential insulators and boundaries for genomic imprinting. *Biochem. Cell Biol.* 89, 469–478.

Murrell, A., Heeson, S., Bowden, L., Constancia, M., Dean, W., Kelsey, G., et al. (2001). An intragenic methylated region in the imprinted Igf2 gene augments transcription. *EMBO Rep.* 2, 1101–1106.

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.

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

Nativio, R., Wendt, K. S., Ito, Y., Huddleston, J. E., Uribe-Lewis, S., Woodfine, K., et al. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.* 5:e1000739. doi: 10.1371/journal.pgen.1000739

Ohlsson, R., Bartkuhn, M., and Renkawitz, R. (2010). CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma* 119, 351–360.

Pant, V., Mariano, P., Kanduri, C., Mattsson, A., Lobanenkov, V., Heuchel, R., et al. (2003). The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.* 17, 586–590.

Park, K. Y., Sellars, E. A., Grinberg, A., Huang, S. P., and Pfeifer, K. (2004). The H19 differentially methylated region marks the parental origin of a heterologous locus without gametic DNA methylation. *Mol. Cell. Biol.* 24, 3588–3595.

Phillips, J. E., and Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.

Pugacheva, E. M., Tiwari, V. K., Abdullaev, Z., Vostrov, A. A., Flanagan, P. T., Quitschke, W. W., et al. (2005). Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Hum. Mol. Genet.* 14, 953–965.

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.

Reese, K. J., Lin, S., Verona, R. I., Schultz, R. M., and Bartolomei, M. S. (2007). Maintenance of paternal methylation and repression of the imprinted H19 gene requires MBD3. *PLoS Genet.* 3:e137. doi: 10.1371/journal.pgen.0030137

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., et al. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8309–8314.

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.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., et al. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.

Singh, P., Cho, J., Tsai, S. Y., Rivas, G. E., Larson, G. P., and Szabó,

P. E. (2010a). Coordinated allele-specific histone acetylation at the differentially methylated regions of imprinted genes. *Nucleic Acids Res.* 38, 7974–7990.

Singh, P., Han, L., Rivas, G. E., Lee, D. H., Nicholson, T. B., Larson, G. P., et al. (2010b). Allele-specific H3K79 Di- versus trimethylation distinguishes opposite parental alleles at imprinted regions. *Mol. Cell. Biol.* 30, 2693–2707.

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.

Spencer, R. J., del Rosario, B. C., Pinter, S. F., Lessing, D., Sadreyev, R. I., and Lee, J. T. (2011). A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. *Genetics* 189, 441–454.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., et al. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.

Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S., and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* 27, 654–666.

Szabó, P., Tang, S. H., Rentsendorj, A., Pfeifer, G. P., and Mann, J. R. (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr. Biol.* 10, 607–610.

Szabó, P. E., Tang, S. H., Reed, M. R., Silva, F. J., Tsark, W. M., and Mann, J. R. (2002). The chicken beta-globin insulator element conveys chromatin boundary activity but not imprinting at the mouse Igf2/H19 domain. *Development* 129, 897–904.

Szabó, P. E., Tang, S. H., Silva, F. J., Tsark, W. M., and Mann, J. R. (2004). Role of CTCF binding sites in the Igf2/H19 imprinting control region. *Mol. Cell. Biol.* 24, 4791–4800.

Takahashi, N., Kobayashi, R., and Kono, T. (2010). Restoration of Dlk1 and Rtl1 is necessary but insufficient to rescue lethality in intergenic differentially methylated region (IG-DMR)-deficient mice. *J. Biol. Chem.* 285, 26121–26125.

Takahashi, N., Okamoto, A., Kobayashi, R., Shirai, M., Obata, Y., Ogawa, H., et al. (2009). Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. *Hum. Mol. Genet.* 18, 1879–1888.

Tanimoto, K., Shimotsuma, M., Matsuzaki, H., Omori, A., Bungert, J., Engel, J. D., et al. (2005). Genomic imprinting recapitulated in the human beta-globin locus. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10250–10255.

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.

Tremblay, K. D., Duran, K. L., and Bartolomei, M. S. (1997). A 5′ 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol. Cell. Biol.* 17, 4322–4329.

Tremblay, K. D., Saam, J. R., Ingram, R. S., Tilghman, S. M., and Bartolomei, M. S. (1995). A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat. Genet.* 9, 407–413.

Tucker, K. L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P. W., Lei, H., et al. (1996). Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.* 10, 1008–1020.

Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., et al. (2000). The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* 5, 649–659.

Verona, R. I., Thorvaldsen, J. L., Reese, K. J., and Bartolomei, M. S. (2008). The transcriptional status but not the imprinting control region determines allele-specific histone modifications at the imprinted H19 locus. *Mol. Cell. Biol.* 28, 71–82.

West, A. G., Huang, S., Gaszner, M., Litt, M. D., and Felsenfeld, G. (2004). Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol. Cell* 16, 453–463.

Weth, O., and Renkawitz, R. (2011). CTCF function is modulated by neighboring DNA binding factors. *Biochem. Cell Biol.* 89, 459–468.

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

Wu, Q., Kumagai, T., Kawahara, M., Ogawa, H., Hiura, H., Obata, Y., et al. (2006). Regulated expression of two sets of paternally imprinted genes is necessary for mouse parthenogenetic development to term. *Reproduction* 131, 481–488.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.* 31, 2174–2183.

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R. D., and Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24, 2543–2555.

Yoon, B., Herman, H., Hu, B., Park, Y. J., Lindroth, A., Bell, A., et al. (2005). Rasgrf1 imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. *Mol. Cell. Biol.* 25, 11184–11190.

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.

Zuo, X., Sheng, J., Lau, H. T., McDonald, C. M., Andrade, M., Cullen, D. E., et al. (2012). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J. Biol. Chem.* 287, 2107–2118.

Conflict of Interest Statement: 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.

Received: 18 July 2012; paper pending published: 13 August 2012; accepted: 27 September 2012; published online: 15 October 2012.

Citation: Singh P, Lee D-H and Szabó PE (2012) More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. Front. Gene. ***3****:214. doi: 10.3389/fgene.2012.00214*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

REVIEW ARTICLE
published: 16 November 2012
doi: 10.3389/fgene.2012.00251

The role of CTCF binding sites in the 3′ immunoglobulin heavy chain regulatory region

Barbara K. Birshtein*

Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY, USA

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France
Félix Recillas-Targa, Universidad Nacional Autónoma de México, Mexico

***Correspondence:**
Barbara K. Birshtein, Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA.
e-mail: barbara.birshtein@einstein.yu.edu

The immunoglobulin heavy chain locus undergoes a series of DNA rearrangements and modifications to achieve the construction and expression of individual antibody heavy chain genes in B cells. These events affect variable regions, through VDJ joining and subsequent somatic hypermutation, and constant regions through class switch recombination (CSR). Levels of IgH expression are also regulated during B cell development, resulting in high levels of secreted antibodies from fully differentiated plasma cells. Regulation of these events has been attributed primarily to two *cis*-elements that work from long distances on their target sequences, i.e., an ∼1 kb intronic enhancer, Eμ, located between the V region segments and the most 5′ constant region gene, Cμ; and an ∼40 kb 3′ regulatory region (3′ RR) that is located downstream of the most 3′ C_H gene, Cα. The 3′ RR is a candidate for an "end" of B cell-specific regulation of the *Igh* locus. The 3′ RR contains several B cell-specific enhancers associated with DNase I hypersensitive sites (hs1–4), which are essential for CSR and for high levels of IgH expression in plasma cells. Downstream of this enhancer-containing region is a region of high-density CTCF binding sites, which extends through hs5, 6, and 7 and further downstream. CTCF, with its enhancer-blocking activities, has been associated with all mammalian insulators and implicated in multiple chromosomal interactions. Here we address the 3′ RR CTCF-binding region as a potential insulator of the *Igh* locus, an independent regulatory element and a predicted modulator of the activity of 3′ RR enhancers. Using chromosome conformation capture technology, chromatin immunoprecipitation, and genetic approaches, we have found that the 3′ RR with its CTCF-binding region interacts with target sequences in the V_H, Eμ, and C_H regions through DNA looping as regulated by protein binding. This region impacts on B cell-specific *Igh* processes at different stages of B cell development.

Keywords: immunoglobulin heavy chain gene locus, enhancers, insulators, CTCF, class switch recombination, Pax5, chromosome conformation capture (3C) assay

Igh GENES AND THEIR DNA REARRANGEMENTS AND MUTATION

The immunoglobulin heavy chain gene locus (*Igh*) undergoes an amazing array of DNA rearrangements and mutagenic events during B cell differentiation (reviewed in Max, 2008). A general question is how these DNA modifications are normally achieved during B cell development without mistakes that result in malignant transformation. Our studies have focused on a regulatory region that acts at long distances on target *Igh* sequences essential for these DNA rearrangement and mutagenic events (reviewed in Pinaud et al., 2011).

The *Igh* locus extends for ∼3 Mb and contains coding segments for constructing a diverse repertoire of variable region genes, through recombination of V_H (variable), D_H (diversity), and J_H (joining) segments, as well as for constant region (C_H) genes that, when translated, confer different functional capabilities on antibody molecules. During bone marrow B cell development, the locus undergoes sequential DNA rearrangement and mutational events that generate an enormous range of antibody heavy chain genes, each specifying individual antigen binding sites associated with specific constant regions. The initial event, i.e., recombinase-activator genes (RAG)-mediated V(D)J joining, involves first, a DJ join, and then V to DJ joining, both accompanied by deletions of intervening sequences; these lead to expression of a IgM heavy chain bearing a single variable region. Successful expression of one allele halts rearrangements on the other allele (allelic exclusion) and prompts VJ joining on the light chain allele. Upon leaving the bone marrow, the B cell with its H_2L_2 surface IgM is poised to receive signals through antigen and other receptors for T cell surface proteins and secreted cytokines that trigger further DNA targeted events, such as class switch recombination (CSR) and somatic hypermutation. CSR is initiated by germline transcription (GT) of the non-IgM C_H gene to which subsequent DNA rearrangement will occur. The DNA rearrangement event results in a shift of the VDJ gene segment from its position upstream of μ to upstream of γ, ε or α genes; as in VDJ joining, intervening DNA is deleted as a circle. V_H-hypermutation results, upon antigen selection, in B cells with higher affinity antigen-binding sites. Both CSR and somatic hypermutation depend on the activity of activation-dependent cytidine deaminase (AID). In fully differentiated plasma cells, heavy chain gene expression occurs at high levels. These multiple processes of VDJ joining, GT and CSR, and

increased *Igh* expression levels require tight regulation to contain these potentially mutagenic events within the confines of the *Igh* locus.

THE 3′ RR CONTAINS AN ENHANCER MODULE AND A HIGH-DENSITY CTCF-BINDING REGION

Two major long distance *Igh* control elements have been identified. Our focus here is on a large (∼50 kb) 3′ regulatory region (3′ RR), located downstream of the C_H genes (reviewed in Pinaud et al., 2011) and schematized in **Figure 1**. A second well-characterized control element is an ∼1 kb intronic enhancer, Eμ, positioned between the V, D, and J segments and the C_H genes, which is critical for VDJ joining (reviewed in Max, 2008).The murine 3′ RR contains a 5′ 28 kb segment, which has four enhancers that collectively support GT, CSR, and high levels of IgH expression in plasma cells. An ∼10 kb 3′ segment contains a region of high-density CTCF- and Pax5-binding sites with insulator activity. Pax5, a transcription factor essential for B cell identity (reviewed in Cobaleda et al., 2007), is associated with 3′ RR enhancers as well. Our studies have shown that the 3′ RR interacts at long distances with a number of *Igh* target sites, as part of its influence on CSR and regulation of *Igh* expression. This entire region is a candidate for a downstream "end" of B cell-specific regulation of the *Igh* locus. At the upstream V region end, the *Igh* locus begins in the general vicinity of telomeric sequences (mouse chr. 12, human chr. 14), suggestive of a natural boundary. At the 3′ C_H-end, beyond the terminus of the 3′ RR, *hole* (*Tmem121*), *Crip1/2*, and *mta1* are the nearest non-*Igh* downstream genes (all in the same inverted transcriptional orientation compared to the *Igh* locus) followed by the rest of the chromosome (Zhou et al., 2002a). There are multiple kinds of regulatory elements in this 3′ RR. Three of the four enhancers located in the 5′ segment of the murine 3′ RR form an ∼25 kb palindrome, in which the central hs1.2 enhancer is flanked by virtually identical terminal enhancers hs3A and hs3B (Saleque et al., 1997). A fourth enhancer, hs4, lies 3′ of hs3B in a separate 3 kb structural and functional unit (Michaelson et al., 1995; Saleque et al., 1997). Hs4 and the palindromic region vary in their acquisition of DNase I hypersensitivity during B cell maturation (Giannini et al., 1993); hs4 becomes hypersensitive early in B cell development and remains so throughout, while the palindromic enhancers become hypersensitive only later in B cell maturation. A similar 3′ RR (hs3, hs1.2, hs4) is located downstream of each of the two Cα genes in the human *Igh* locus (Chen and Birshtein, 1996, 1997; Mills et al., 1997; Sepulveda et al., 2004a,b; Frezza et al., 2009).

As a potential "end" of B cell-specific regulation of the *Igh* locus, how might the 3′ RR help to focus DNA rearrangement events on the *Igh* locus and prevent inherently mutagenic events like DNA rearrangements and mutations from encroaching into neighboring downstream genes? We predicted that the 3′ RR might house an insulator region with CTCF as a major functional contributor, similar to insulator regions found in other loci (Phillips and Corces, 2009; Amouyal, 2010; Yang and Corces, 2011). In fact, (and before the era of high-throughput genomic analyses), EMSA

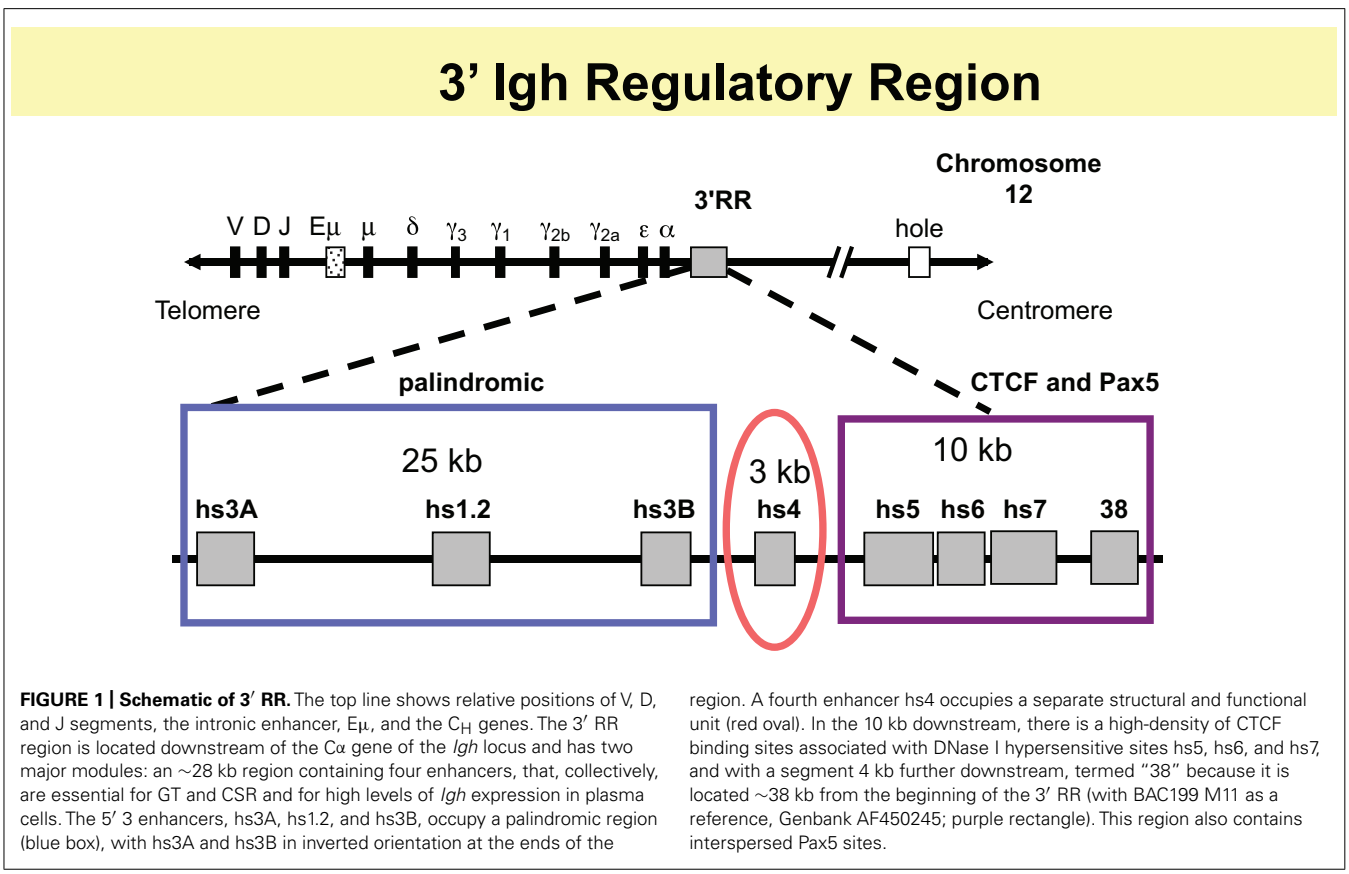


FIGURE 1 | Schematic of 3′ RR. The top line shows relative positions of V, D, and J segments, the intronic enhancer, Eμ, and the C_H genes. The 3′ RR region is located downstream of the Cα gene of the *Igh* locus and has two major modules: an ∼28 kb region containing four enhancers, that, collectively, are essential for GT and CSR and for high levels of *Igh* expression in plasma cells. The 5′ 3 enhancers, hs3A, hs1.2, and hs3B, occupy a palindromic region (blue box), with hs3A and hs3B in inverted orientation at the ends of the region. A fourth enhancer hs4 occupies a separate structural and functional unit (red oval). In the 10 kb downstream, there is a high-density of CTCF binding sites associated with DNase I hypersensitive sites hs5, hs6, and hs7, and with a segment 4 kb further downstream, termed "38" because it is located ∼38 kb from the beginning of the 3′ RR (with BAC199 M11 as a reference, Genbank AF450245; purple rectangle). This region also contains interspersed Pax5 sites.

with recombinant CTCF on 50 consecutive overlapping DNA fragments identified multiple CTCF sites (Garrett et al., 2005). These were associated with additional DNase I hs sites, hs5, 6, and 7, and with a segment 4 kb downstream of hs7, which because it is located 38 kb from the beginning of the 3′ RR has been termed "38." Hs5 and hs7 were shown to confer insulator activity in a cell line assay (Garrett et al., 2005). Analysis of EMSA with nuclear extracts from B cell lines using supershift studies with specific antibodies for CTCF and Pax5 showed that this entire hs5–7-"38" region contained interspersed CTCF- and Pax5-binding sites (Chatterjee et al., 2011). Because Pax5 is a regulator of 3′ RR enhancers, these data suggested that Pax5 might help coordinate the actions of the enhancer-containing region with the insulator region. Here we describe studies of the contribution of the 3′ RR to *Igh* regulation throughout B cell development; in levels of H chain expression in plasma cells, in GT and CSR in B cells, and in use of V_H genes in VDJ joining in pro- and pre-B cells. For the most part, these involve the formation of loops associating components of the 3′ RR with target *Igh* sequences, as described in the following sections and as schematized in **Figure 2**.

THE 3′ RR AFFECTS H CHAIN EXPRESSION IN PLASMA CELLS THROUGH CONTACTS WITH *Igh* TARGET SEQUENCES

A role for the 3′ RR in high levels of *Igh* expression in plasma cells was inferred when we established that the entire 3′ RR was deleted in a mouse plasma cell line that had lost 90% of its *Igh* expression levels (Gregor and Morrison, 1986; Michaelson et al., 1995). That the 3′ RR could loop to engage target *Igh* sequences was predicted from studies of another mouse plasma cell line, in which we detected an inversion of a segment extending from

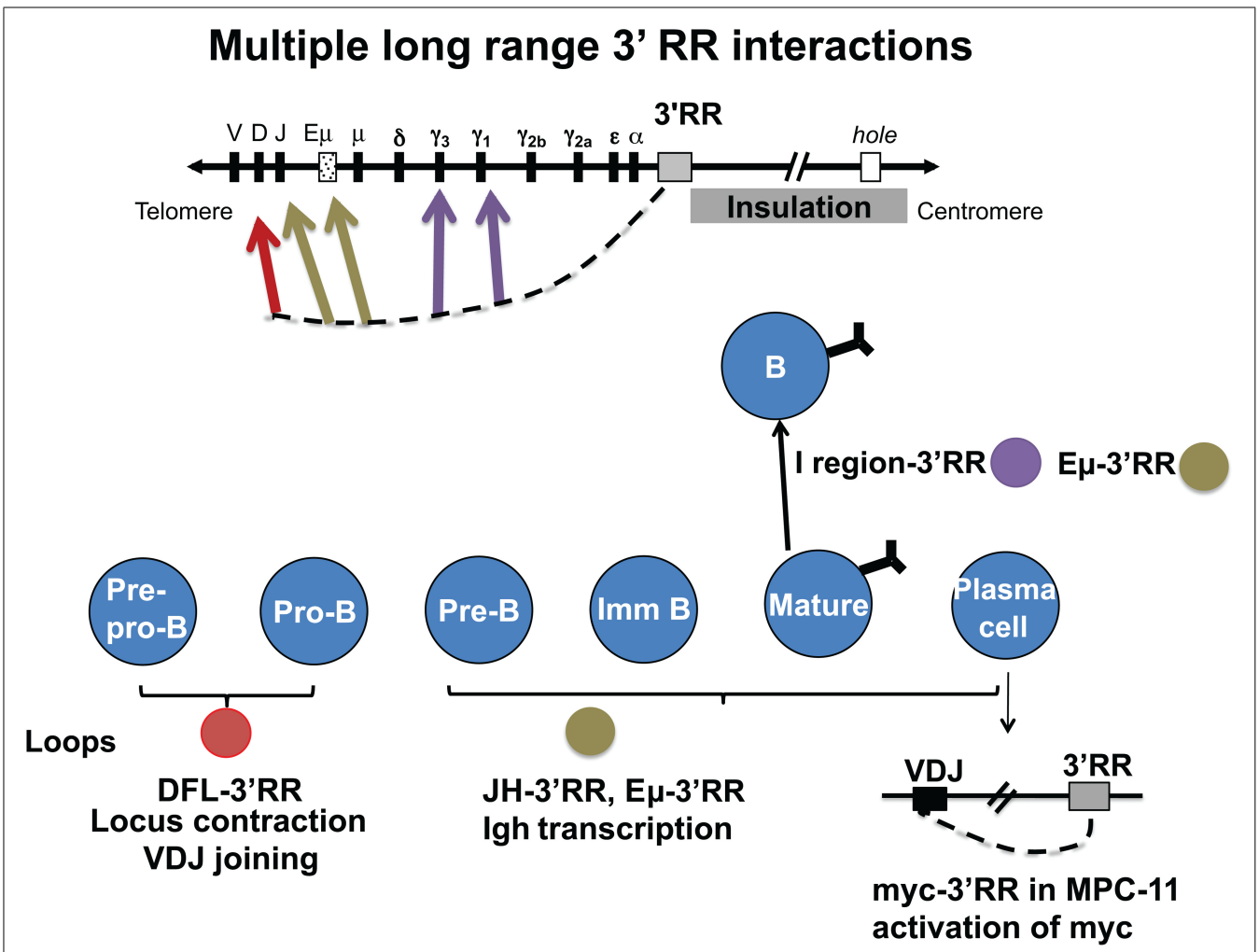


FIGURE 2 | The 3′ RR interacts via looping with many different target *Igh* sequences during B cell development. During VDJ joining in pro- and pre-B cells, the 3′ RR interacts with two CTCF sites upstream of D_H (red arrow). Also in pre-B cells, the 3′ RR is involved in allelic regulation (brown arrow). In B cells, interaction of the 3′ RR with I/switch regions is associated with GT and subsequent CSR (purple arrows). In plasma cells (lower right), the 3′ RR interacts with the expressed VDJ region and the intronic enhancer. Also in malignant plasma cells, the translocated *c-myc* oncogene interacts in *cis* by looping with the 3′ RR (Ju et al., 2007). Studies show some evidence that CTCF binding sites in the 3′ RR have some insulation activity (gray box) that is detected to impact as far as the "*hole*" gene. However, CTCF sites associated with downstream genes appear to provide an over-riding local influence (as discussed in the text).

expressed V_H gene sequences to the 3′ RR palindromic region (Calvo et al., 1991; Ju et al., 2007). Resolution of a loop formed by interactions between the V_H gene and the 3′ RR is the simplest intermediate to account for this inversion. Documentation of such a loop structure came upon implementation of the chromosome conformation capture (3C) method in a plasma cell line: here we showed physical interactions involving 3′ RR enhancers and its CTCF-binding region with the J_H sequence that was part of the expressed V_H gene. The adjacent Eμ sequences were not essential for this interaction (Ju et al., 2007). The contacts associated with chromatin loop formation were severely disrupted in a different MPC11 variant, whose expressed *Igh* gene had been rendered non-functional by substitution of the hs1.2 enhancer by the NeoR gene (Ju et al., 2007). These data implied that an intact 3′ RR was essential for H chain expression in plasma cells and that H chain expression depended on intact physical interaction in *cis* of the 3′ RR with the expressed V_H gene. An extension of these observations from cell lines to mice has derived from targeted deletion of the entire enhancer-containing region of the 3′ RR in mice, which has confirmed a critical role of the 3′ RR in promoting high levels of *Igh* expression in plasma cells (Vincent-Fabert et al., 2010).

Efforts to identify proteins that support loop formation and concomitant *Igh* expression used a loss-of-function strategy employing lentiviral-mediated shRNA directed against CTCF, Oct-2, and OBF-1/OCA-B (Ju et al., 2011) in the MPC11 plasma cell line. In no case did we see effects on *Igh* expression. We conclude that proteins other than those targeted were required to support H chain expression, or that residual levels of CTCF, Oct-2, and/or OBF-1/OCA-B remaining after the knock-down were sufficient, or that these factors act in a redundant fashion and that simultaneous knock down of multiple factors is required for a decrease of *Igh* chain expression.

TARGETED DELETIONS OF 3′ RR ENHANCERS REVEAL THEIR INVOLVEMENT IN GT AND CSR

The impact of targeted deletions of 3′ RR enhancers in mouse by a number of investigators has revealed their importance for two successive steps of the CSR process, i.e., transcription through C_H switch regions, followed by CSR. Deletion of the hs3B and hs4 region of the 3′ RR reduced switching to all isotypes except IgG1 (Cogne et al., 1994; Manis et al., 1998). The contribution to GT and CSR of the I/switch regions and of the 3′ RR enhancers has been fully demonstrated (reviewed in (Cogne and Birshtein, 2004). 3C studies on mature B cells undergoing CSR revealed interactions between the 3′ RR and switch regions through which transcription occurs prior to CSR (Wuerffel et al., 2007). These interactions were severely reduced in B cells from mice in which 3′ RR enhancers hs3B and hs4 were deleted. These data supported the importance of loop interactions between the 3′ RR and its target switch sequences for CSR. The distances involved range from ∼15 to ∼150 kb.

3C experiments also revealed cytokine-responsive chromosomal conformation involving the 3′ RR during GT and CSR (Wuerffel et al., 2007; Yan et al., 2011). Cytokine treatments that fostered switching to a particular isotype not only stimulate transcription of switch sequences of that isotype by activating the I region promoter upstream of switch sequences, but also result in specific increased 3C interactions between the 3′ RR and the isotype-specific switch region. Interestingly, a double deletion of hs3A and hs3B generated by the Eckhardt laboratory had no effect on either transcription or CSR (Yan et al., 2011). However, we found that in this doubly deleted mouse, isotype-specific interactions between switch regions and the 3′ RR ordinarily enhanced by cytokines were already at a high level in resting B cells, and there was a concomitant increase in interactions between the remaining 3′ RR enhancers, hs1.2 and hs4. These observations suggested that hs3A and hs3B modulate a functional hs1.2-hs4 3′ RR enhancer unit (Yan et al., 2011).

In fact, GT and CSR are generally unaffected after individual deletions of each of the four 3′ RR enhancers, including hs1.2 and hs4 (Manis et al., 1998; Vincent-Fabert et al., 2009; Bebin et al., 2010; Dunnick et al., 2011). Interestingly, a distinctive (but similarly functional) enhancer unit remains after each individual enhancer deletion, e.g., hs1.2, hs3B, hs4 (when hs3A is deleted); hs3A, hs3B, hs4 (when hs1.2 is deleted) and so on. This implies considerable flexibility in the structure and function of the 3′ RR enhancer unit, a point that is addressed further below. In all, the essential role of 3′ RR enhancers in GT and CSR can be met by their multiple alternative functional interactions with each other and with target switch sequences; these influence isotype-specific switching in response to cytokine signaling.

TARGETED DELETION OF 3′ RR CTCF BINDING SITES HS5–7

Our studies have shown that during GT and CSR, the multiple modules of the 3′ RR, i.e., enhancers and the CTCF-binding region hs5–7, interact with I/switch regions and with the Pax5 transcription factor. Pax5 (reviewed in (Cobaleda et al., 2007) is essential for B cell identity and, through reporter assays, was shown to play an important role in regulating murine 3′ RR enhancers (Singh and Birshtein, 1993, 1996). To determine the function of the CTCF-binding region, we generated hs5–7 KO mice (Volpi et al., 2012). B cells from hs5–7 KO mice showed essentially normal GT and CSR except for a modest increase in $IgG1^+$ cells upon switching in culture. One possibility to account for these observations is that interactions of *Igh* sequences with the CTCF/Pax5-binding site-rich hs5–7 region are secondary to the role of the 3′ RR enhancers and are not essential during CSR. Another possibility is that the deletion did not eliminate all candidate CTCF-binding sites. In fact, ChIP/Seq data (Degner et al., 2009) showed that the hs5–7 KO left behind a limited number of CTCF sites in the 3′ RR region, and other CTCF sites associated with each non-*Igh* downstream gene (R. Casellas, personal communication). Potentially, even a fraction of CTCF sites in this region or other CTCF-interacting sites are sufficient for appropriate biological activity. Similarly, we had anticipated that a reduction in insulator activity resulting from deletion of a large group of CTCF sites from the 3′ RR would enable the upstream unaffected 3′ RR enhancers to promote expression of downstream, non-*Igh* genes. However, our studies revealed only a modest increase in expression of the nearest downstream gene, *Tmem121*, while further downstream genes were unaffected (Volpi et al., 2012). It appears that local regulation of downstream genes by their own CTCF sites provides a back-up mechanism to restrain

inappropriately regulated activity of the *Igh* locus from inflicting damage on non-*Igh* genes.

INFLUENCE OF 3′ RR CTCF-BINDING REGION ON VDJ JOINING

CTCF has been described as a "master weaver of the genome" (Phillips and Corces, 2009). Thousands of genomic CTCF sites have been mapped, including those within the *Igh* locus (Garrett et al., 2005; Degner et al., 2009, 2011). Moving upstream (3′ to 5′) of the high-density CTCF-binding region in the hs5–7 region of the 3′ RR past the C_H and J_H regions that are devoid of CTCF sites, the CTCF sites that are closest to the 3′ RR are two sites located 5′ of the most 5′ D_H gene; by 3C, these have been shown to interact with the 3′ RR. Functional inactivation of the two D-associated CTCF sites abrogated normal VDJ joining (Guo et al., 2011); as a result, they have been named intergenic control region 1 (IGCR1). These studies imply a role of CTCF in VDJ joining. In fact, functional inactivation of CTCF in pro-B cells by shRNA (Degner et al., 2011) resulted in an increased distance between the interacting 3′ RR and D_H/CTCF sequences, i.e., a reduction in V_H-locus contraction, and an increase in anti-sense transcription in D_H and V_H regions. To determine, therefore, whether 3′ RR CTCF sites that bind to D_H/CTCF are critical for the role of D_H/CTCF in VDJ joining, we assessed a mouse with a targeted deletion of CTCF binding sites in the hs5–7 region of the 3′ RR (Volpi et al., 2012). Here, we were surprised to find essentially normal levels of VDJ joining in hs5–7 KO pro- and pre-B cells, except for a detectable increase in DQ52-J_H3 usage at multiple stages of B cell development. In addition, there was a modest, albeit statistically significant reduction in *Igh* locus contraction, and an increase by twofold over wild-type in the use of proximal V_H7183 genes while distal V_HJ558 usage was unaffected. Notably, allelic exclusion was correctly maintained. Although these data uncover an effect of the 3′ RR-CTCF-binding region on the *Igh* locus when VDJ joining is occurring presumptively through interactions of this region with D_H/CTCF, they also imply the presence of considerable backups for proper *Igh* regulation.

Pax5 AND CTCF AS REGULATORS OF THE 3′ RR DURING CSR

Pax5

As a step toward further understanding mechanisms that control the 3′ RR, we have identified transcription factors that regulate 3′ RR enhancer activity. Experiments showed that the four 3′ RR enhancers are regulated by a common set of transcription factors, namely Oct-binding proteins, NFκB, and Pax5 (Michaelson et al., 1996), which could synergize for concerted repression (Singh and Birshtein, 1996) or for concerted activation of 3′ RR enhancers (Michaelson et al., 1996). YY1 has also been implicated (Gordon et al., 2003). Importantly, Pax5 appears to regulate each of the 3′ RR enhancers as well as the CTCF-binding region. Using chromatin immunoprecipitation (ChIP), we found that as B cells are induced to switch by culture with LPS +/- IL4, Pax5 shifts in its association with modules of the 3′ RR (Chatterjee et al., 2011). In resting B cells, Pax5 binds predominantly to hs4. At 48 h when GT and switch region-3′ RR interactions are at a peak, Pax5 has shifted away from hs4 to bind to upstream enhancer (hs1.2) and downstream insulator (hs7) flanking sites. At 96 h, when CSR has been completed, Pax5 regains hs4 binding as seen in resting B cells. Regardless of whether switching to γ3 or γ2b occurred by stimulation with LPS, or to γ1 through stimulation by LPS + IL4, the Pax5 pattern of binding to the 3′ RR was similar.

When we compared B cells that successfully undergo sequential steps in switch recombination with those that are deficient in GT and/or CSR (Chatterjee et al., 2011), we found that the Pax5-binding pattern to the 3′ RR is mechanistically associated with CSR. For example, stimulation of NFκB $p50^{-/-}$ cells for 48 h with LPS + IL4 shows deficiency in normal GT; accordingly, the Pax5 profile is different from normal B cells. Pax5 continues to bind to hs4 although acquiring binding to hs1.2. In cells stimulated with anti-IgM + IL4, which undergo normal GT but fail to switch, the Pax5-binding pattern at 48 h is like that of cells stimulated by LPS + /-IL4, but at 96 h, the pattern is disrupted. Collectively, these data suggest that dynamic changes in Pax5 binding to the 3′ RR are supported by an isotype-independent scaffold on which GT and CSR occur.

CTCF

To determine whether changes in CTCF binding to the 3′ RR were similarly associated with CSR, we analyzed binding of CTCF and its cofactor cohesin, this latter consisting of multiple subunits, including Rad21 (Chatterjee et al., 2011). In contrast to changes in Pax5 binding, we found relatively stable interactions of CTCF with the high-density CTCF-binding region in hs5–7 and "38" throughout the steps in GT and CSR that occurred in cells cultured with LPS + /-IL4. Also as expected, together with CTCF, Rad21 bound preferentially to hs7 upon stimulation with either LPS + /-IL4 or with anti-IgM + IL4. However, in resting B cells and independent of CTCF, Rad21 additionally bound to hs1.2 at low levels, and then at substantially increased levels at 48 h of stimulation before binding at reduced levels again to hs1.2 at 96 h. A similar pattern of CTCF-independent Rad21 binding to hs1.2 was detected in cells stimulated with anti-IgM + IL4.

Collectively, these data showed that CTCF and cohesin binding to the 3′ RR, both to cognate CTCF sites and independent of known CTCF sites, appear to contribute to a framework for the 3′ RR, while Pax5 has dynamic interactions with its binding sites. We have proposed (Chatterjee et al., 2011) that the multiple Pax5-binding sites in 3′ RR enhancers could support a scaffold structure: various enhancer deletions or shifts in enhancer occupancy could take place, leaving behind varying constellations of functional Pax5 sites.

REGULATION OF 3′ RR BY DNA METHYLATION

We predicted that the 3′ RR is subject to epigenetic regulation as it acquires its functional capability. The 3′ RR essentially can be divided into two regions under separate epigenetic control, the 5′ palindromic enhancers and the more 3′ hs4-"38" region. Beginning in pro-B cells, the hs4-"38" region is associated with marks of active chromatin (Garrett et al., 2005) and with DNA demethylation (Giambra et al., 2008), which appear to be set in place by expression of Pax5 and linker histone H1. The upstream palindromic enhancers – hs3A-hs1.2-hs3B – acquire both epigenetic marks in B and plasma cells (Giambra et al., 2008).

BOX 1 | Regulation of methylation and chromatin modifications of 3′ RR during B cell development

1. In pro-B cells, the hs4 enhancer and the CTCF-binding region hs5-"38" are demethylated and show marks of active chromatin. These marks are retained during B cell development. In B and plasma cells, the palindromic enhancers hs3A-1.2-3B acquire both epigenetic marks.
2. A polymorphic region between hs4 and hs5 reveals demethylation specific for the expressed allele in pre-B cells.
3. The 3′ RR in resting B cells is mostly methylated. In B cells stimulated to undergo class switching, the 3′ RR becomes progressively demethylated with limited accompanying changes in chromatin marks.

The two *Igh* alleles in the mouse 70Z/3 pre-B cell line (C57Bl/6-derived and DBA/J-derived) can be distinguished by their stage during VDJ joining, their association with a polymorphic DNA segment that is subject to DNA demethylation (Giambra et al., 2008), and by the formation of loops involving the 3′ RR (Ju et al., 2011). The expressed VDJ-joined, C57Bl/6-derived, allele is associated in *cis* with a 3′ RR containing a deletion of hs3A-hs1.2 (with no apparent impact on *Igh* expression; Saleque et al., 1999). The polymorphic region located between hs4 and hs5 on this allele is demethylated. In contrast, the unexpressed DJ-joined allele (DBA/J-derived) fails to undergo looping in *cis* with its intact 3′ RR, and the hs4-hs5 sequence remains methylated. These data reinforce the role of the 3′ RR in *cis*-regulation of the *Igh* locus and imply that DNA demethylation in the 3′ RR, looping and *Igh* VDJ rearrangement and expression may be associated.

Interestingly, B cells stimulated for GT and CSR do not reveal any significant changes in chromatin marks of the 3′ RR (Garrett et al., 2005). Instead, we have identified progressive DNA demethylation of the 3′ RR (Giambra et al., 2008) and (Giambra, V., in preparation). These observations suggest that in resting B cells prior to stimulation for CSR, the 3′ RR is poised in its chromatin profile. We predict that DNA demethylation is associated with architectural changes by which the 3′ RR influences GT, CSR, and high levels of *Igh* expression in plasma cells. These epigenetic alterations of the 3′ RR during B cell development are summarized in **Box 1**.

DO *Igh* DNA REPLICATION PATTERNS SPECIFY ANOTHER TERMINUS OF THE *Igh* LOCUS?

Various landmarks might demarcate functional termini for the *Igh* locus; (1) the distinctive cluster of CTCF sites in hs5–7 that is located downstream of the C_H part of the locus and (2) ∼20 kb further downstream, the nearest non-*Igh* downstream gene, *Tmem121,* i.e., *hole.* In collaborative studies (Michaelson et al., 1997; Ermakova et al., 1999; Zhou et al., 2002a,b, 2005), we identified a replication origin downstream of *Tmem121* that is also a candidate for a functional B cell-specific terminus of *Igh* regulation. These studies showed that the *Igh* locus had different temporal patterns of DNA replication in non-B cells and at various stages of B cell development. In non-B cells, an origin of replication was identified ∼11 kb downstream of *Tmem121,* which is ∼30 kb downstream of the hs5–7 region and ∼76–79 kb downstream of the Cα gene. DNA sequences downstream of this landmark all replicated early in S. Beginning at this origin and moving upstream, i.e., 3′ to 5′, the 500 kb region within which C_H, J_H, D_H, and V_H7183 sequences were located replicated progressively later in S. This was consistent with the absence of activated origins of replication in this region. Sequences further upstream of the 500 kb transition region all replicated late in S. However, in pro-B and pre-B cells, the temporal transition region was eliminated as the entire *Igh* locus replicated early in S, indicative of the firing of multiple origins that were otherwise latent in non-B cells. Hence, this origin-containing region downstream of *Tmem121* appeared to demarcate upstream sequences that are under B cell-specific *Igh* regulation from downstream sequences under non-*Igh* control. Notably, in mature B cells and plasma cells, the temporal transition region was again evident and the replication pattern was similar to that seen in non-B cells. The change in replication was paralleled by a change in location of the *Igh* locus from a position at the nuclear periphery in non-B cells to away from the nuclear periphery in pro- and pre-B cells, with resumption of a nuclear periphery location in B and plasma cells. Analysis of replication dynamics in a cell line in which the 3′ RR enhancer region had been deleted, leaving behind the CTCF/Pax5-binding region and further downstream sequences, showed no difference compared to wild-type plasma cells (Michaelson et al., 1997). While these findings showed that the 3′ RR enhancer region is not essential for the timing of replication of the *Igh* locus in plasma cells, inferences about the role of the CTCF/Pax5-binding region in this process are not possible.

SUMMARY

Here we have discussed two major modules of the 3′ RR, which extends ∼40 kb beginning downstream of Cα. The 5′ 28 kb segment contains four enhancers, which, collectively, support GT, CSR, and high levels of IgH expression in plasma cells. The ∼10 kb 3′ segment contains a region of high-density CTCF- and Pax5-binding sites with insulator activity. During B cell development, the 3′ RR-its enhancers and CTCF-binding region – is involved, via loop formation, with various target *Igh* sequences. These include: (1) CTCF sites upstream of D_H that are essential for normal VDJ joining and allelic *Igh* expression in pre-B cells; (2) I/switch sequences required for GT and CSR in B cells, and c) J_H and Eμ, which support *Igh* expression in plasma cells. While 3′ RR enhancers are essential for GT and CSR, as demonstrated by targeted deletions, independent deletion of at least seven of an estimated nine CTCF sites in the 3′ RR resulted in only a mild phenotype (Volpi et al., 2012). We found essentially normal VDJ joining but with a slight decrease in V_H-locus contraction, a twofold increase in usage of proximal V_H7183 genes and an apparent increase in DQ52-J_H3 usage. Steps in GT and CSR appeared generally indistinguishable from wild-type, as was the chromosomal architecture of the 3′ RR assessed by 3C. In all, we conclude that the CTCF-binding region is a nidus for physical interactions with *Igh* targets of important biological consequence. However, there must be many back-ups that provide functional compensation to CTCF. These back-ups may include local regulators, such as CTCF sites associated with other neighboring genes, or proteins

other than CTCF and/or epigenetic regulators that terminate B cell-specific regulation of the *Igh* locus.

ONGOING KEY QUESTIONS

Which proteins/nucleic acids/other molecules are essential for loop formation? Are there different kinds of structural and functional loops? What do loops do? Do they engage the *Igh* locus in particular subnuclear domains for DNA rearrangements, mutation, etc. during different stages of B cell development? How does the 3′ RR function? What mediates architectural interactions among the 3′ RR enhancers themselves and between the enhancer and CTCF-binding modules? How does loop formation in the *Igh* locus relate to loops in other loci? What specifically does CTCF contribute to the structure and function of the *Igh* locus?

ACKNOWLEDGMENTS

This review focuses predominantly on work carried out in my laboratory, which was supported by NIH RO1AI13509 and reflects the efforts of multiple contributors from my laboratory. I thank Charles Calvo, Sandra Giannini, Jennifer Michaelson, Nancy Martinez, Nasrin Ashouian, Alexis Price-Whelan, Chaoqun Chen, Francine Garrett-Bakelman, Alejandro Sepulveda, Rabih Hassan, Vincenzo Giambra, Steven Gordon, Alexander Emelyanov, Zhongliang Ju, Sanjukta Chatterjee, and Sabrina Volpi. These individuals built on the efforts of students and research associates who preceded them. I also thank our collaborators at Einstein, Matthew Scharff, Sergio Roa, Carl Schildkraut, Randall Little, Olga Ermakova, Jie Zhou, Qiaoxin Yang, Winfried Edelmann, Harry Hou and Uwe Werling, Britta Will and Ulrich Steidl; laboratories elsewhere, headed by Laurel Eckhardt, Victor Lobanenkov, Roy Riblet, Domenico Frezza, and Ann Feeney; other colleagues who have studied the 3′ regulatory region, especially Michel Cogne, John Manis, Amy Kenter, Fred Alt, and Wesley Dunnick; and Richard Chahwan, Xiaohua Wang and Howard Steinman for critique of this paper. I regret that I can not acknowledge all those whose efforts have led to an increased understanding of the intriguing processes by which immunoglobulin heavy chain genes are regulated.

REFERENCES

Amouyal, M. (2010). Gene insulation. Part II: natural strategies in vertebrates. *Biochem. Cell. Biol.* 88, 885–898.

Bebin, A. G., Carrion, C., Marquet, M., Cogne, N., Lecardeur, S., Cogne, M., et al. (2010). *In vivo* redundant function of the 3′ IgH regulatory element HS3b in the mouse. *J. Immunol.* 184, 3710–3717.

Calvo, C.-F., Giannini, S. L., Martinez, N., and Birshtein, B. K. (1991). DNA sequences 3′ of the IgH chain cluster rearrange in mouse B cell lines. *J. Immunol.* 146, 1353–1360.

Chatterjee, S., Ju, Z., Hassan, R., Volpi, S. A., Emelyanov, A. V., and Birshtein, B. K. (2011). Dynamic changes in binding of immunoglobulin heavy chain 3′ regulatory region to protein factors during class switching. *J. Biol. Chem.* 286, 29303–29312.

Chen, C., and Birshtein, B. K. (1996). A region of 20 bp repeats lies 3′ of human Ig $C_{\alpha}1$ and $C_{\alpha}2$ genes. *Int. Immunol.* 8, 115–122.

Chen, C., and Birshtein, B. K. (1997). Virtually identical enhancers containing a segment of homology to murine 3′IgH-E(hs1,2) lie downstream of human Ig C alpha 1 and C alpha 2 genes. *J. Immunol.* 159, 1310–1318.

Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007). Pax5: the guardian of B cell identity and function. *Nat. Immunol.* 8, 463–470.

Cogne, M., and Birshtein, B. K. (2004). "Regulation of class switch recombination," in *Molecular Biology of B Cells*, eds. T. Honjo, F. W. Alt and M. Neuberger (San Diego, CA: Elsevier Academic Press), 289–305.

Cogne, M., Lansford, R., Bottaro, A., Zhang, J., Gorman, J., Young, F., et al. (1994). A class switch control region at the 3′ end of the immunoglobulin heavy chain locus. *Cell* 77, 737–747.

Degner, S. C., Verma-Gaur, J., Wong, T. P., Bossen, C., Iverson, G. M., Torkamani, A., et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9566–9571.

Degner, S. C., Wong, T. P., Jankevicius, G., and Feeney, A. J. (2009). Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J. Immunol.* 182, 44–48.

Dunnick, W. A., Shi, J., Zerbato, J. M., Fontaine, C. A., and Collins, J. T. (2011). Enhancement of antibody class-switch recombination by the cumulative activity of four separate elements. *J. Immunol.* 187, 4733–4743.

Ermakova, O. V., Nguyen, L. H., Little, R. D., Chevillard, C., Riblet, R., Ashouian, N., et al. (1999). Evidence that a single replication fork proceeds from early to late replicating domains in the IgH locus in a non-B cell line. *Mol. Cell* 3, 321–330.

Frezza, D., Giambra, V., Mattioli, C., Piccoli, K., Massoud, R., Siracusano, A., et al. (2009). Allelic frequencies of 3′ Ig heavy chain locus enhancer HS1,2-A associated with Ig levels in patients with schizophrenia. *Int. J. Immunopathol. Pharmacol.* 22, 115–123.

Garrett, F. E., Emelyanov, A. V., Sepulveda, M. A., Flanagan, P., Volpi, S., Li, F., et al. (2005). Chromatin architecture near a potential 3′ end of the igh locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. *Mol. Cell. Biol.* 25, 1511–1525.

Giambra, V., Volpi, S., Emelyanov, A. V., Pflugh, D., Bothwell, A. L., Norio, P., et al. (2008). Pax5 and linker histone H1 coordinate DNA methylation and histone modifications in the 3′ regulatory region of the immunoglobulin heavy chain locus. *Mol. Cell. Biol.* 28, 6123–6133.

Giannini, S. L., Singh, M., Calvo, C.-F., Ding, G., and Birshtein, B. K. (1993). DNA regions flanking the mouse Ig 3′a enhancer are differentially methylated and DNAse I hypersensitive during B cell differentiation. *J. Immunol.* 150, 1772–1780.

Gordon, S. J., Saleque, S., and Birshtein, B. K. (2003). Yin Yang 1 is a lipopolysaccharide-inducible activator of the murine 3′ Igh enhancer, hs3. *J. Immunol.* 170, 5549–5557.

Gregor, P. D., and Morrison, S. L. (1986). Myeloma mutant with a novel 3′ flanking region: loss of normal sequence and insertion of repetitive elements leads to decreased transcription but normal processing of the alpha heavy-chain gene products. *Mol. Cell. Biol.* 6, 1903–1916.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430.

Ju, Z., Chatterjee, S., and Birshtein, B. K. (2011). Interaction between the immunoglobulin heavy chain 3′ regulatory region and the IgH transcription unit during B cell differentiation. *Mol. Immunol.* 49, 297–303.

Ju, Z., Volpi, S. A., Hassan, R., Martinez, N., Giannini, S. L., Gold, T., et al. (2007). Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3′ regulatory region. *J. Biol. Chem.* 282, 35169–35178.

Manis, J. P., Van Der Stoep, N., Tian, M., Ferrini, R., Davidson, L., Bottaro, A., et al. (1998). Class switching in B cells lacking 3′ immunoglobulin heavy chain enhancers. *J. Exp. Med.* 188, 1421–1431.

Max, E. E. (2008). "Immunoglobulins: molecular genetics," in *Fundamental Immunology*, Chap. 6. (ed.) W. E. Paul (Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins), 192–236.

Michaelson, J. S., Ermakova, O., Birshtein, B. K., Ashouian, N., Chevillard, C., Riblet, R., et al. (1997). Regulation of the replication of the murine immunoglobulin heavy chain gene locus: evaluation of the role of the 3′ regulatory region. *Mol. Cell. Biol.* 17, 6167–6174.

Michaelson, J. S., Giannini, S. L., and Birshtein, B. K. (1995). Identification of 3′ alpha-hs4, a novel Ig heavy chain enhancer element regulated at multiple stages of B cell differentiation. *Nucleic Acids Res.* 23, 975–981.

Michaelson, J. S., Singh, M., Snapper, C. M., Sha, W. C., Baltimore, D., and Birshtein, B. K. (1996). Regulation of 3′ IgH enhancers by a common set of factors, including kappa B-binding proteins. *J. Immunol.* 156, 2828–2839.

Mills, F. C., Harindranath, N., Mitchell, M., and Max, E. E. (1997). Enhancer complexes located downstream of both human immunoglobulin C alpha genes. *J. Exp. Med.* 186, 845–858.

Phillips, J. E., and Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.

Pinaud, E., Marquet, M., Fiancette, R., Peron, S., Vincent-Fabert, C., Denizot, Y., et al. (2011). The IgH locus 3′ regulatory region: pulling the strings from behind. *Adv. Immunol.* 110, 27–70.

Saleque, S., Singh, M., and Birshtein, B. K. (1999). Ig heavy chain expression and class switching *in vitro* from an allele lacking the 3′ enhancers DNase I-hypersensitive hs3A and hs1,2. *J. Immunol.* 162, 2791–2803.

Saleque, S., Singh, M., Little, R. D., Giannini, S. L., Michaelson, J. S., and Birshtein, B. K. (1997). Dyad symmetry within the mouse 3′ IgH regulatory region includes two virtually identical enhancers (C alpha3′E and hs3). *J. Immunol.* 158, 4780–4787.

Sepulveda, M. A., Emelyanov, A. V., and Birshtein, B. K. (2004a). NF-kappaB and Oct-2 synergize to activate the human 3′ Igh hs4 enhancer in B cells. *J. Immunol.* 172, 1054–1064.

Sepulveda, M. A., Garrett, F. E., Price-Whelan, A., and Birshtein, B. K. (2004b). Comparative analysis of human and mouse 3′ Igh regulatory regions identifies distinctive structural features. *Mol. Immunol.* 42, 605–615.

Singh, M., and Birshtein, B. K. (1993). NF-HB (BSAP) is a repressor of the murine immunoglobulin heavy-chain 3′ alpha enhancer at early stages of B-cell differentiation. *Mol. Cell. Biol.* 13, 3611–3622.

Singh, M., and Birshtein, B. K. (1996). Concerted repression of an immunoglobulin heavy-chain enhancer, 3′ alpha E(hs1,2). *Proc. Natl. Acad. Sci. U.S.A.* 93, 4392–4397.

Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogne, N., Cogne, M., et al. (2010). Genomic deletion of the whole IgH 3′ regulatory region (hs3a, hs1,2, hs3b, hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. *Blood* 116, 1895–1898.

Vincent-Fabert, C., Truffinet, V., Fiancette, R., Cogne, N., Cogne, M., and Denizot, Y. (2009). Ig synthesis and class switching do not require the presence of the hs4 enhancer in the 3′ IgH regulatory region. *J. Immunol.* 182, 6926–6932.

Volpi, S. A., Verma-Gaur, J., Hassan, R., Ju, Z., Roa, S., Chatterjee, S., et al. (2012). Germline deletion of Igh 3′ regulatory region elements hs 5, 6, 7 (hs5–7) affects B cell-specific regulation, rearrangement, and insulation of the Igh locus. *J. Immunol.* 188, 2556–2566.

Wuerffel, R., Wang, L., Grigera, F., Manis, J., Selsing, E., Perlot, T., et al. (2007). S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 22, 711–722.

Yan, Y., Pieretti, J., Ju, Z., Wei, S., Christin, J. R., Bah, F., et al. (2011). Homologous elements hs3a and hs3b in the 3′ regulatory region of the murine immunoglobulin heavy chain (Igh) locus are both dispensable for class-switch recombination. *J. Biol. Chem.* 286, 27123–27131.

Yang, J., and Corces, V. G. (2011). Chromatin insulators: a role in nuclear organization and gene expression. *Adv. Cancer Res.* 110, 43–76.

Zhou, J., Ashouian, N., Delepine, M., Matsuda, F., Chevillard, C., Riblet, R., et al. (2002a). The origin of a developmentally regulated Igh replicon is located near the border of regulatory domains for Igh replication and expression. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13693–13698.

Zhou, J., Ermakova, O. V., Riblet, R., Birshtein, B. K., and Schildkraut, C. L. (2002b). Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells. *Mol. Cell. Biol.* 22, 4876–4889.

Zhou, J., Saleque, S., Ermakova, O., Sepulveda, M. A., Yang, Q., Eckhardt, L. A., et al. (2005). Changes in replication, nuclear location, and expression of the Igh locus after fusion of a pre-B cell line with a T cell line. *J. Immunol.* 175, 2317–2320.

Conflict of Interest Statement: 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.

Received: 02 July 2012; paper pending published: 27 August 2012; accepted: 26 October 2012; published online: 16 November 2012.

Citation: Birshtein BK (2012) The role of CTCF binding sites in the 3′ immunoglobulin heavy chain regulatory region. Front. Gene. ***3****:251. doi: 10.3389/fgene.2012.00251*

This article was submitted to Frontiers in Epigenomics and Epigenetics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

REVIEW ARTICLE
published: 25 September 2012
doi: 10.3389/fgene.2012.00186

Role of CTCF in the regulation of microRNA expression

Yoshimasa Saito* and Hidetsugu Saito

Division of Pharmacotherapeutics, Faculty of Pharmacy, Keio University, Tokyo, Japan

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Timothy Bredy, The University of Queensland, Australia
Félix Recillas-Targa, Universidad Nacional Autónoma de México, Mexico

****Correspondence:***
Yoshimasa Saito, Division of Pharmacotherapeutics, Faculty of Pharmacy, Keio University, 1-5-30, Shibakoen, Tokyo 105-8512, Japan.
e-mail: yoshimasa.saito@gmail.com

MicroRNAs (miRNAs) are small non-coding RNAs that regulate expression of various target genes. miRNAs are expressed in a tissue-specific manner and play important roles in cell proliferation, apoptosis, and differentiation. Epigenetic alterations such as DNA methylation and histone modification are essential for chromatin remodeling and regulation of gene expression including miRNAs. The CCCTC-binding factor, CTCF, is known to bind insulators and exhibits an enhancer-blocking and barrier function, and more recently, it also contributes to the three-dimensional organization of the genome. CTCF can also serve as a barrier against the spread of DNA methylation and histone repressive marks over promoter regions of tumor suppressor genes. Recent studies have shown that CTCF is also involved in the regulation of miRNAs such as *miR-125b1*, *miR-375,* and the *miR-290* cluster in cancer cells and stem cells. *miR-125b1* is a candidate of tumor suppressor and is silenced in breast cancer cells. On the other hand, *miR-375* may have oncogenic function and is overexpressed in breast cancer cells. CTCF is involved in the regulation of both *miR-125b1* and *miR-375*, indicating that there are various patterns of CTCF-associated epigenetic regulation of miRNAs. CTCF may also play a key role in the pluripotency of cells through the regulation of *miR-290* cluster. These observations suggest that CTCF-mediated regulation of miRNAs could be a novel approach for cancer therapy and regenerative medicine.

Keywords: microRNA, CTCF, cancer cell, embryonic stem cell, *miR-125b1*, *miR-375*, *miR-290* cluster

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs that regulate various target genes and play important roles in cell proliferation, apoptosis, and differentiation. One of the important mechanisms of miRNA expression is epigenetic alteration such as DNA methylation and histone modification. The CCCTC-binding factor, CTCF, is known to bind insulators and exhibits an enhancer-blocking and barrier function, and more recently, it also contributes to the three-dimensional organization of the genome. Although, there are a number of studies describing regulation of miRNA expression including epigenetic alterations, only a few studies have reported the association between miRNA expression and CTCF. In this report, we review recent studies regarding miRNAs and CTCF, and discuss about roles of CTCF in the regulation of miRNA expression.

miRNA

miRNAs are ∼22 nucleotide (nt) non-coding RNAs that can post-transcriptionally downregulate the expression of various target genes. Currently, ∼1500 human miRNAs have been identified in the human genome, and each miRNA potentially controls hundreds of target genes. In animals, miRNA genes are generally transcribed by RNA polymerase II (pol II) to form primary transcripts (pri-miRNAs). Pol II transcribed pri-miRNAs are capped with 7-methylguanosine and are polyadenylated. The nuclear RNase III enzyme Drosha and its co-factor DGCR8 process pri-miRNAs into ∼60 nt precursor miRNAs (pre-miRNAs), which form an imperfect stem-loop structure. Pre-miRNAs are transported into the cytoplasm by exportin 5 and are subsequently cleaved by Dicer into mature miRNAs which are then loaded into the RNA-induced silencing complex (RISC). The miRNA/RISC complex downregulates specific gene products by translational repression via binding to partially complementary sequences in the 3′ untranslated regions of the target mRNAs or by directing mRNA degradation via binding to perfectly complementary sequences. miRNAs are expressed in a tissue-specific manner and play important roles in metabolism, proliferation, apoptosis, and differentiation. Moreover, recent studies have shown a link between aberrant expression of miRNAs and the development of cancer (Calin and Croce, 2007; Cho, 2007; Saito et al., 2009).

Abbreviations: miRNA, microRNA; RISC, RNA-induced silencing complex; ERα, estrogen receptor α; ESC, embryonic stem cell; EEmiRC, early embryonic miRNA cluster; IE, intragenic enhancer.

EPIGENETIC REGULATION OF miRNA EXPRESSION

Since miRNAs can have large-scale effects through regulation of a variety of genes during mammalian development and carcinogenesis, an understanding of the regulatory mechanisms controlling miRNA expression is important. There are several reports of transcription factors binding to the promoter regions of specific miRNA genes and activating the transcription of pri-miRNAs, resulting in increased expression of mature miRNAs. *c-Myc* binds to the regulatory region of the *miR-17-92* cluster and increased expression of *c-Myc* leads to the activation of the miRNAs in the cluster (O'Donnell et al., 2005).

Epigenetic alterations such as DNA methylation and histone modification play critical roles in chromatin remodeling and regulation of gene expression in mammalian development and in human diseases. Many miRNAs are expressed in a tissue- and tumor-specific manner, implying that some miRNAs are subject to epigenetic control. We have shown that *miR-127*, which is embedded in a CpG island, is strongly induced by treatment with DNA methylation inhibitors and histone deacetylase inhibitors, indicating that some miRNA genes are controlled by epigenetic alterations in their promoter regions and can be activated by chromatin modifying drugs (Saito et al., 2006, 2009). Lujambio et al. (2007) compared miRNA expression profiling between the wild-type HCT116 colon cancer cell line and HCT116 after genetic disruption of both *DNA methyltransferase (DNMT) 1* and *DNMT3b* (DKO cells). They found that 18 out of 320 miRNAs are significantly upregulated in DKO cells. In particular, *miR-124a* is silenced by its own CpG island hypermethylation in human tumors, but can be activated by inhibition of DNA methylation. They also demonstrated that the oncogene *CDK6* is a target of *miR-124a* and that epigenetic silencing of *miR-124a* in cancer cells modulates *CDK6* activity. It has been reported that *miR-9-1* and *miR-9-3* are potential tumor suppressor miRNAs and are inactivated by epigenetic mechanisms in human cancers (Lehmann et al., 2008; Lujambio et al., 2008). *miR-34a* was identified as a target of p53 and induces a G(1) cell cycle arrest, senescence and apoptosis (He et al., 2007; Tazawa et al., 2007). *miR-34a* expression is silenced in several types of cancer including pancreatic cancer due to aberrant CpG methylation of its promoter. Re-expression of *miR-34a* in a pancreatic carcinoma cell line induced senescence and cell cycle arrest at least in part by targeting *CDK6*, indicating that *miR-34a* represents a tumor suppressor gene which is inactivated by CpG methylation in pancreatic cancer (Lodygin et al., 2008). *miR-34b* and *miR-34c* are also reported to be silenced by aberrant CpG island methylation in colorectal cancer (Toyota et al., 2008). Thus, a number of miRNAs are under epigenetic control and disruption of DNA methylation patterns and histone modification in the promoter regions of miRNAs might be associated with cancer development (Esteller, 2011).

The CCCTC-binding factor, CTCF, is known to bind insulators and exhibits an enhancer-blocking function. CTCF can also serve as a barrier against the spread of DNA methylation and histone repressive marks over promoter regions of tumor suppressor genes (Recillas-Targa et al., 2011). CTCF is a highly conserved multifunctional zinc finger protein involved in transcriptional repression and activation, insulation, epigenetic events such as imprinting of the *H19/IGF2* locus, and X-inactivation, and which binds preferentially to unmethylated DNA (Filippova, 2008; Phillips and Corces, 2009). Moreover, CTCF play important roles during carcinogenesis: epigenetic silencing of tumor suppressor genes such as *p16* and *Rb* (De La Rosa-Velazquez et al., 2007; Witcher and Emerson, 2009), apoptosis of breast cancer cells (Docquier et al., 2005), and regulation of important tumor suppressor genes such as *p53* (Recillas-Targa et al., 2011; Saldana-Meyer and Recillas-Targa, 2011). These findings suggest that CTCF may be involved in epigenetic regulation of non-coding RNAs including miRNAs as well as coding RNAs.

DISRUPTION OF CTCF BINDING AT THE *miR-125b1* CPG ISLAND IN HUMAN CANCERS

Recent studies have reported that expression of *miR-125b* is downregulated in various human cancers including glioblastoma, prostate cancer, ovarian cancer, and breast cancer (Scott et al., 2007; Zhang et al., 2011). In addition, *miR-125b* suppresses oncogenes such as EST1, ERBB2, ERBB3, and Bak1 as its targets, suggesting that *miR-125b* functions as a tumor suppressor. DNA hypermethylation at the CpG island of *miR-125b* was observed in cell lines and in tissue samples from patients with breast cancer (Zhang et al., 2011).

Soto-Reyes et al. (2012) investigated epigenetic alterations such as DNA methylation and histone modification, and association of CTCF at the locus of *miR-125b1* in breast cancer cells. They found aberrant DNA methylation of the *miR-125b1* CpG island and that disruption of CTCF binding correlated with incorporation of repressive histone modifications such as histone H3 lysine 9 (K9) trimethylation and histone H3 K27 trimethylation in cancer cells. In normal breast cells expressing *miR-125b1*, CTCF might prevent the recruitment of epigenetic silencing components, such as DNA methylation and repressive histone modifications, and also favors an open chromatin structure. In breast cancer cells, the loss of CTCF is associated with CpG island methylation and the gain of repressive histone modifications such as histone H3 K9 trimethylation and histone H3 K27 trimethylation. Disruption of CTCF binding at CpG island induces silencing of *miR-125b1* expression (**Figure 1**). These findings suggest that CTCF plays an important role in the regulation of the tumor suppressor *miR-125b1* in cooperation with DNA methylation and histone modification in breast cancer cells. A recent study has also demonstrated that *miR-125b1* can be silenced by DNA methylation, which may lead to activation of the ETS1 proto-oncogene and a worse prognosis in breast cancer

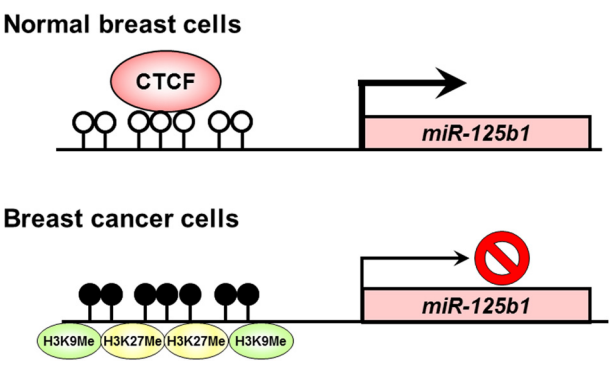


FIGURE 1 | Disruption of CTCF binding at the *miR-125b1* CpG island in human cancers. In normal breast cells, CTCF might prevent the recruitment of epigenetic silencing components, such as DNA methylation and repressive histone modifications, and also favors an open chromatin structure. Meanwhile, in breast cancer cells, the loss of CTCF is associated with CpG island methylation and the gain of repressive histone modifications such as histone H3 K9 trimethylation and histone H3 K27 trimethylation. Open circle, unmethylated DNA; filled circle, methylated DNA; H3K9Me, histone H3 K9 methylation; H3K27Me, histone H3 K27 methylation.

patients (Zhang et al., 2011). Reactivation of the tumor suppressor *miR-125b1* by epigenetic therapy using DNA methylation inhibitors may have clinical promise for the treatment of breast cancer patients.

ROLE OF CTCF IN THE REGULATION OF *miR-375* EXPRESSION IN BREAST CANCER CELLS

Breast cancer is the leading cause of cancer death in women worldwide. Estrogen receptor α (ER α) upregulation causes abnormal cell proliferation in approximately 70% of breast cancers (Shoker et al., 1999; Vargo-Gogola and Rosen, 2007). A recent study has reported that *miR-375* is overexpressed in ERα-positive breast cancer cell lines and plays an important role in cell proliferation (de Souza Rocha Simonini et al., 2010). There are CpG islands in the upstream region of the *miR-375* gene. DNA hypermethylation is observed in the CpG island of ERα-positive breast cancer cells showing high expression of *miR-375*, whereas DNA hypomethylation and histone H3 K9 dimethylation are observed in the CpG islands of ERα-negative breast cancer cells. CTCF binds to unmethylated DNA in the CpG islands of ERα-negative cells and induces silencing of *miR-375* expression. These findings suggest that overexpression of *miR-375* is caused by dissociation of CTCF from the CpG island of *miR-375* gene via loss of epigenetic marks including local DNA hypomethylation and histone H3 K9 dimethylation (de Souza Rocha Simonini et al., 2010) (**Figure 2**). It has been shown that *miR-375* suppresses *Ras dexamethasone-induced 1* (*RASD1*) as its potential target, and RASD1 can suppress the growth of breast cancer cells and down-regulate ERα expression (Vaidyanathan et al., 2004; de Souza Rocha Simonini et al., 2010). Thus the modulation of ERα expression by *miR-375* is achieved through the repression of RASD1. These observations provide a possibility that inhibition of *miR-375* could be a novel clinical approach for the treatment of ERα-positive breast cancer.

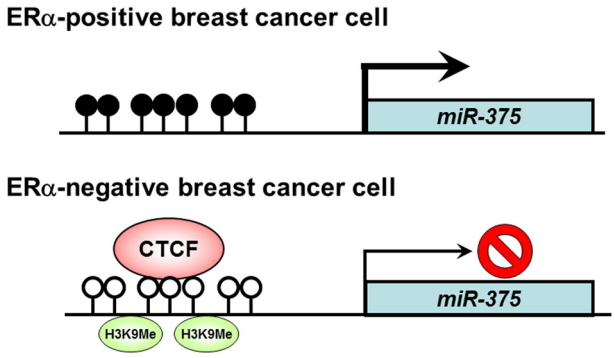


FIGURE 2 | Role of CTCF in the regulation of *miR-375* expression in breast cancer cells. In ERα-positive breast cancer cells, DNA hypermethylation is observed in the CpG island of the *miR-375* gene, and *miR-375* expression is activated. On the other hand, in ERα-negative breast cancer cells, DNA hypomethylation and histone H3 K9 dimethylation are observed, resulting in silencing of *miR-375* expression by binding of CTCF in the CpG island. Open circle, unmethylated DNA; filled circle, methylated DNA; H3K9Me, histone H3 K9 methylation.

CTCF MODULATES EXPRESSION OF THE EARLY EMBRYONIC miRNA CLUSTER

Human embryonic stem cells (ESCs) are derived from the inner cell mass of the human blastocyte and can be kept in an undifferentiated, self-renewing state indefinitely. ESCs have the advantage of being pluripotent, which endows them with the ability to differentiate into virtually every cell type in the human body. Thus, ESCs have gained popularity as a potentially ideal cell candidate for regenerative medicine. The early embryonic miRNA cluster (EEmiRC) has been identified in ESCs of mammals, and shows a remarkable cross-eutherian species conservation at the levels of both pre-miRNA hairpins and the core-promoter region (Houbaviy et al., 2003, 2005). EEmiRC encodes 7 miRNAs (*miR-290*, *-291a*, *-292*, *-291b*, *-293*, *-294* and *-295*), which have been labeled as ESC-specific/pluripotency-associated miRNAs controlling cell-cycle progression, proliferation, and DNA methylation in undifferentiated/pluripotent cells. Therefore, understanding the biology of ESCs requires detailed knowledge of the mechanisms regulating EEmRC expression.

Little is known about the molecular mechanisms underlying the regulation of the EEmiRC expression. Recent studies have showed that the sequences upstream to the EEmiRC promoter contains active binding sites for Nanog, Oct3/4, Sox2, Tcf3, c-Myc, and 4n-Myc. Histone H3 K4 trimethylation and histone H3 K27 trimethylation were observed in ESCs and in differentiated cells, respectively (Chen et al., 2008; Judson et al., 2009). However, attempts to activate EEmiRC expression by ectopic expression of these individual transcriptional factors in fibroblasts were unsuccessful, suggesting that EEmiRC expression is under epigenetic control (Judson et al., 2009). Tata et al. (2011) identified a 332-bp intragenic enhancer (IE) region within the EEmiRC, which is able to modulate the transcription of the mouse EEmiRC locus. These miRNAs involve pluripotency factors and epigenetic mechanisms in pluripotent and differentiated cells. The results of chromatin immunoprecipitation (ChIP) assays demonstrated that the level of occupancy of Oct3/4, Sox2, and CTCF in this region gradually and dramatically decreased during ESC differentiation, suggesting a functional role for these transcription factors in regulating EEmiRC expression. This IE also contains a CpG island showing a differential pattern of DNA and histone methylation marks during differentiation of ESCs. Since, *miR-290* cluster miRNAs have been shown to suppress Rbl2 as their target and Rbl2 modulates DNMTs (Benetti et al., 2008; Sinkkonen et al., 2008), EEmiRC may comprise a feedback loop with DNMTs. These findings indicate that this region plays a critical role in the regulation of EEmiRC expression, presumably through binding of transcription modulators such as Oct3/4, Sox2, and CTCF. Cohesin is a DNA-binding protein complex that is essential for sister chromatid cohesion and facilitates the repair of damaged DNA. Recent experiments have revealed that cohesin binds to the same sites in mammalian genomes as CTCF and cooperates with CTCF in regulating gene expression (Herold et al., 2012). Epigenetic effectors including CTCF and cohesin may modulate the pluripotency of cells through the regulation of *miR-290* cluster.

Table 1 | miRNAs associated with CTCF.

miRNA	Expression	Target genes	Association with CTCF	References
miR-125b1	Decreased in human cancers including glioblastoma, prostate cancer, ovarian cancer and breast cancer	EST1, ERBB2, ERBB3, Bak1	In breast cancer cells, disruption of CTCF binding at the *miR-125b1* CpG island correlated with DNA methylation and methylation of histone H3K9 and K27 induces silencing of *miR-125b1* expression	Scott et al., 2007; Zhang et al., 2011; Soto-Reyes et al., 2012
miR-375	Overexpressed in ERα-positive breast cancer cells	RASD1	In ERα-positive breast cancer cells, *miR-375* overexpression was caused by dissociation of CTCF from the *miR-375* promoter via loss of epigenetic marks including local DNA hypomethylation and histone H3 K9 methylation	de Souza Rocha Simonini et al., 2010
miR-290 cluster	*miR-290* cluster (*miR-290, -291a, -292, -291b, -293, -294 and -295*) have been identified as ESC-specific/pluripotency-associated miRNAs	Rbl2	CTCF binds to intragenic enhancer region within the early embryonic miRNA cluster (EEmiRC) and modulates the expression of the EEmiRC	Houbaviy et al., 2003, 2005; Benetti et al., 2008; Chen et al., 2008; Sinkkonen et al., 2008; Judson et al., 2009; Tata et al., 2011

PERSPECTIVES AND CONCLUSION

Table 1 shows a summary of the association between miRNAs and CTCF. These findings indicate that the insulator protein CTCF plays various roles in the regulation of miRNAs such as *miR-125b1*, *miR-375*, and the *miR-290* cluster during mammalian development and carcinogenesis. *miR-125b1* is a candidate of tumor suppressor and is silenced in breast cancer cells. On the other hand, *miR-375* may have oncogenic function and is overexpressed in breast cancer cells. CTCF is involved in the regulation of both *miR-125b1* and *miR-375*, indicating that there are various patterns of CTCF-associated epigenetic regulation of miRNAs. CTCF-mediated regulation of these miRNAs may provide a novel therapeutic approach for breast cancer. CTCF may also play a key role in the pluripotency of cells through the regulation of *miR-290* cluster. Since, the link between miRNAs and CTCF has only just begun to be understood, other miRNA genes regulated by CTCF will be identified. Further studies are necessary to investigate whether CTCF-mediated regulation of miRNAs could be a novel approach for cancer therapy and regenerative medicine.

ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Young Scientists A (23680090 to Yoshimasa Saito) and a Grant-in-Aid for Scientific Research C (24590993 to Hidetsugu Saito) from the Japan Society for the Promotion of Science (JSPS), Takeda Science Foundation (to Yoshimasa Saito), and Inaida Foundation (to Hidetsugu Saito).

REFERENCES

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P., Li, E., Serrano, M., Millar, S., Hannon, G., and Blasco, M. A. (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat. Struct. Mol. Biol.* 15, 268–279.

Calin, G. A., and Croce, C. M. (2007). Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *Clin. Invest. J.* 117, 2059–2066.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W. K., Clarke, N. D., Wei, C. L., and Ng, H. H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.

Cho, W. C. (2007). OncomiRs: the discovery and progress of microRNAs in cancers. *Mol. Cancer* 6, 60.

De La Rosa-Velazquez, I. A., Rincon-Arano, H., Benitez-Bribiesca, L., and Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTC. *Cancer Res.* 67, 2577–2585.

de Souza Rocha Simonini, P., Breiling, A., Gupta, N., Malekpour, M., Youns, M., Omranipour, R., Malekpour, F., Volinia, S., Croce, C. M., Najmabadi, H., Diederichs, S., Sahin, O., Mayer, D., Lyko, F., Hoheisel, J. D., and Riazalhosseini, Y. (2010). Epigenetically deregulated microRNA-375 is involved in a positive feedback loop with estrogen receptor alpha in breast cancer cells. *Cancer Res.* 70, 9175–9184.

Docquier, F., Farrar, D., D'Arcy, V., Chernukhin, I., Robinson, A. F., Loukinov, D., Vatolin, S., Pack, S., Mackay, A., Harris, R. A., Dorricott, H., O'Hare, M. J., Lobanenkov, V., and Klenova, E. (2005). Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Res.* 65, 5112–5122.

Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874.

Filippova, G. N. (2008). Genetics and epigenetics of the multifunctional protein CTC. *Curr. Top. Dev. Biol.* 80, 337–360.

He, L., He, X., Lim, L. P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., Jackson, A. L., Linsley, P. S., Chen, C., Lowe, S. W., Cleary, M. A., and Hannon, G. J. (2007). A microRNA component of the p53 tumour suppressor network. *Nature* 447, 1130–1134.

Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF:

insights into insulator function during development. *Development* 139, 1045–1057.

Houbaviy, H. B., Dennis, L., Jaenisch, R., and Sharp, P. A. (2005). Characterization of a highly variable eutherian microRNA gene. *RNA* 11, 1245–1257.

Houbaviy, H. B., Murray, M. F., and Sharp, P. A. (2003). Embryonic stem cell-specific MicroRNAs. *Dev. Cell* 5, 351–358.

Judson, R. L., Babiarz, J. E., Venere, M., and Blelloch, R. (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* 27, 459–461.

Lehmann, U., Hasemeier, B., Christgen, M., Muller, M., Romermann, D., Langer, F., and Kreipe, H. (2008). Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *Pathol. J.* 214, 17–24.

Lodygin, D., Tarasov, V., Epanchintsev, A., Berking, C., Knyazeva, T., Korner, H., Knyazev, P., Diebold, J., and Hermeking, H. (2008). Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle* 7, 2591–2600.

Lujambio, A., Calin, G. A., Villanueva, A., Ropero, S., Sanchez-Cespedes, M., Blanco, D., Montuenga, L. M., Rossi, S., Nicoloso, M. S., Faller, W. J., Gallagher, W. M., Eccles, S. A., Croce, C. M., and Esteller, M. (2008). A microRNA DNA methylation signature for human cancer metastasis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13556–13561.

Lujambio, A., Ropero, S., Ballestar, E., Fraga, M. F., Cerrato, C., Setien, F., Casado, S., Suarez-Gauthier, A., Sanchez-Cespedes, M., Git, A., Spiteri, I., Das, P. P., Caldas, C., Miska, E., and Esteller, M. (2007). Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res.* 67, 1424–1429.

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435, 839–843.

Phillips, J. E., and Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.

Recillas-Targa, F., de la Rosa-Velazquez, I. A., and Soto-Reyes, E. (2011). Insulation of tumor suppressor genes by the nuclear factor CTC. *Biochem. Cell Biol.* 89, 479–488.

Saito, Y., Liang, G., Egger, G., Friedman, J. M., Chuang, J. C., Coetzee, G. A., and Jones, P. A. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9, 435–443.

Saito, Y., Suzuki, H., and Hibi, T. (2009). The role of microRNAs in gastrointestinal cancers. *Gastroenterol. J.* 44(Suppl. 19), 18–22.

Saito, Y., Suzuki, H., Tsugawa, H., Nakagawa, I., Matsuzaki, J., Kanai, Y., and Hibi, T. (2009). Chromatin remodeling at Alu repeats by epigenetic treatment activates silenced microRNA-512-5p with downregulation of Mcl-1 in human gastric cancer cells. *Oncogene* 28, 2738–2744.

Saldana-Meyer, R., and Recillas-Targa, F. (2011). Transcriptional and epigenetic regulation of the p53 tumor suppressor gene. *Epigenetics* 6, 1068–1077.

Scott, G. K., Goga, A., Bhaumik, D., Berger, C. E., Sullivan, C. S., and Benz, C. C. (2007). Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA *miR-125a* or *miR-125b*. *Biol. Chem. J.* 282, 1479–1486.

Shoker, B. S., Jarvis, C., Clarke, R. B., Anderson, E., Hewlett, J., Davies, M. P., Sibson, D. R., and Sloane, J. P. (1999). Estrogen receptor-positive proliferating cells in the normal and precancerous breast. *Am. J. Pathol.* 155, 1811–1815.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 15, 259–267.

Soto-Reyes, E., Gonzalez-Barrios, R., Cisneros-Soberanis, F., Herrera-Goepfert, R., Perez, V., Cantu, D., Prada, D., Castro, C., Recillas-Targa, F., and Herrera, L. A. (2012). Disruption of CTCF at the miR-125b1 locus in gynecological cancers. *BMC Cancer* 12, 40.

Tata, P. R., Tata, N. R., Kuhl, M., and Sirbu, I. O. (2011). Identification of a novel epigenetic regulatory region within the pluripotency associated microcluster, RNA, EEmiRC. *Nucleic Acids Res.* 39, 3574–3581.

Tazawa, H., Tsuchiya, N., Izumiya, M., and Nakagama, H. (2007). Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15472–15477.

Toyota, M., Suzuki, H., Sasaki, Y., Maruyama, R., Imai, K., Shinomura, Y., and Tokino, T. (2008). Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. *Cancer Res.* 68, 4123–4132.

Vaidyanathan, G., Cismowski, M. J., Wang, G., Vincent, T. S., Brown, K. D., and Lanier, S. M. (2004). The Ras-related protein AGS1/RASD1 suppresses cell growth. *Oncogene* 23, 5858–5863.

Vargo-Gogola, T., and Rosen, J. M. (2007). Modelling breast cancer: one size does not fit all. *Nat. Rev. Cancer* 7, 659–672.

Witcher, M., and Emerson, B. M. (2009). Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol. Cell* 34, 271–284.

Zhang, Y., Yan, L. X., Wu, Q. N., Du, Z. M., Chen, J., Liao, D. Z., Huang, M. Y., Hou, J. H., Wu, Q. L., Zeng, M. S., Huang, W. L., Zeng, Y. X., and Shao, J. Y. (2011). miR-125b is methylated and functions as a tumor suppressor by regulating the ETS1 proto-oncogene in human invasive breast cancer. *Cancer Res.* 71, 3552–3562.

Conflict of Interest Statement: 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.

Received: 08 June 2012; paper pending published: 27 June 2012; accepted: 03 September 2012; published online: 25 September 2012.

*Citation: Saito Y and Saito H (2012) Role of CTCF in the regulation of microRNA expression. Front. Gene. **3**:186. doi: 10.3389/fgene.2012.00186*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in GENETICS

REVIEW ARTICLE
published: 03 September 2012
doi: 10.3389/fgene.2012.00169

Regulation of chromatin structure by poly(ADP-ribosyl)ation

Sascha Beneke*

Institute of Veterinary Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

Edited by:
Michèle Amouyal, CNRS, France

Reviewed by:
Paola Caiafa, Università Sapienza di Roma, Italy
Elena Klenova, University of Essex, UK

****Correspondence:***
Sascha Beneke, Institute of Veterinary Pharmacology and Toxicology, University of Zurich, Winterthurerstrasse 260, 8057 Zurich, Switzerland.
e-mail: sascha.beneke@vetpharm.uzh.ch

The interaction of DNA with proteins in the context of chromatin has to be tightly regulated to achieve so different tasks as packaging, transcription, replication and repair. The very rapid and transient post-translational modification of proteins by poly(ADP-ribose) has been shown to take part in all four. Originally identified as immediate cellular answer to a variety of genotoxic stresses, already early data indicated the ability of this highly charged nucleic acid-like polymer to modulate nucleosome structure, the basic unit of chromatin. At the same time the enzyme responsible for synthesizing poly(ADP-ribose), the zinc-finger protein poly(ADP-ribose) polymerase-1 (PARP1), was shown to control transcription initiation as basic factor TFIIC within the RNA-polymerase II machinery. Later research focused more on PARP-mediated regulation of DNA repair and cell death, but in the last few years, transcription as well as chromatin modulation has re-appeared on the scene. This review will discuss the impact of PARP1 on transcription and transcription factors, its implication in chromatin remodeling for DNA repair and probably also replication, and its role in controlling epigenetic events such as DNA methylation and the functionality of the insulator protein CCCTC-binding factor.

Keywords: poly(ADP-ribosyl)ation, PARP1, chromatin, recruitment, transcription, CTCF

POLY(ADP-RIBOSYL)ATION

Poly(ADP-ribosyl)ation as enzymatic reaction is known since the early sixties of the last century (Chambon et al., 1963). In the following 20 years it was related to several nuclear functions, i.e., histone modification (Aubin et al., 1982), differentiation (Farzaneh et al., 1982; Pekala and Moss, 1983), cell death (Sims et al., 1983), transcriptional regulation (Slattery et al., 1983) and DNA repair/genome stability (Davies et al., 1978; Durkacz et al., 1980). Also the major players were analyzed:

(1) Structure of the product poly(ADP-ribose) (PAR) (Chambon et al., 1966; Nishizuka et al., 1967; Reeder et al., 1967),
(2) Synthesizing enzyme poly(ADP-ribose) polymerase(−1) (PARP1) [(Tsopanakis et al., 1976), cDNA cloned in (Cherney et al., 1987; Suzuki et al., 1987)] and
(3) Degrading enzyme poly(ADP-ribose) glycohydrolase (PARG) [(Ueda et al., 1972), cDNA cloned in (Lin et al., 1997)].

In the enzymatic reaction NAD^+ is cleaved into nicotinamide and ADP-ribose, with the latter attached to glutamate or aspartate via an ester bond (Ogata et al., 1980b), and to lysine, forming a ketoamine by Schiff-Base and Amadori rearrangement (Altmeyer et al., 2009). Whereas esters are enzymatically easy to revert, ketoamines show substantial stability and may form a "modification-mark" on the respective protein. After attachment of the first ADP-ribose moiety, further units are rapidly added via α-gylcosidic bonds and branches can originate from the growing chain, depending on the synthesizing enzyme and interaction partner (Naegeli and Althaus, 1991).

PARPs are nowadays a family of 17 enzymes, but not all of them are active ADP-ribose transferases and only few show truly polymerizing activity (Hottiger et al., 2010). In case of PARP1, the product poly(ADP-ribose) displays a tree-like structure, forming a highly negative charged cloud at the covalently modified protein, which impacts on functionality probably through electrostatic repulsion of affected enzymes from DNA (Zahradka and Ebisuzaki, 1982). The main acceptor of PAR is PARP1 itself (Ogata et al., 1981), but also its interaction partners can be modified, as shown for several nuclear proteins *in vitro* and *in vivo*. Degradation of the polymer is performed by PARG in an endo- as well as exoglycosidic reaction, releasing PAR of different length as well as ADP-ribose monomers (Meyer-Ficca et al., 2004; Bonicalzi et al., 2005). Enzymatic activity of PARP1 is very low and PAR in unstimulated cells has an estimated half-life of up to several hours (Alvarez-Gonzalez and Althaus, 1989). After application of DNA strand-break inducing agents, PARP1 dimerizes at the break, leading to its activation (Mendoza-Alvarez and Alvarez-Gonzalez, 1993; Jorgensen et al., 2009; Langelier et al., 2012). PARP1 can also bind non-B-DNA structures (Soldatenkov et al., 2002; Lonskaya et al., 2005; Potaman et al., 2005). PAR synthesized in this process displays a much reduced half-life of less than a minute as high local concentrations of the polymer stimulate PARG activity (Alvarez-Gonzalez and Althaus, 1989).

Increased poly(ADP-ribosyl)ation (PARylation) metabolism is one of the first cellular responses following exposure to genotoxic stress (Haince et al., 2007, 2008). In addition to covalent

modification proteins can interact with PAR in a non-covalent fashion. So far, three different motifs have been described:

First, a sequence of basic and hydrophobic residues, the so called PAR-Binding-Motif (PBM) (Pleschke et al., 2000), which is present in many proteins involved in maintaining genomic stability, i.e., telomerase, p53, histones, base-excision-repair (BER) platform protein XRCC1, nucleotide-excision-repair (NER) protein XPA and many more.

Next, it was reported that the macro-domain binds in an end-capping mode to the tip of a PAR chain (Karras et al., 2005).

Finally, a PAR-Binding-Zinc finger (PBZ) was discovered in APLF, a histone chaperone (Ahel et al., 2008).

The wide-spread regulatory impact of PARylation has been described in a recent publication (Gagne et al., 2012). A large scale analysis of PAR-interacting proteins after application of genotoxic stress revealed that specific proteins are associated with PAR in a sequential way after challenge, with an early group representing repair complexes, followed by translation regulators and finally factors involved in RNA processing. Both principles, covalent and non-covalent interaction, can be present side-by-side within one protein. For example the tumor suppressor p53 displays three covalent as well as three non-covalent binding sites (Fahrer et al., 2007; Kanai et al., 2007). Interestingly, the interaction partner is one determinant that affects complexity of PAR, i.e., chain-length and branching (Naegeli and Althaus, 1991). Additionally, proteins differ in their ability to bind to different PAR structures (Fahrer et al., 2007).

In summary, PARP1 (respectively its product PAR) is able to change the surrounding environment by either excluding modified proteins from distinct sites, or by attracting factors containing PAR interaction-motifs.

PARP1 IN DNA-REPAIR AND REPLICATION

SINGLE-STRAND BREAK REPAIR AND HISTONE SHUTTLE

Activity of PARP1 has been correlated with DNA damage since it was discovered (Miller, 1975a,b). DNA strand-breaks are strong inducers of PARylation, stimulating the enzyme several hundred-fold. The exact cellular function of this energetic costly reaction was long unclear, but application of genotoxic agents with simultaneous suppression of PARylation led to increased persistence of breaks (Morgan and Cleaver, 1983), reduced repair (Yamamoto and Okamoto, 1982) and enhanced sister-chromatid-exchanges (Hori, 1981; Otsuka et al., 1983; Park et al., 1983; Meyer et al., 2000), indicating that PARP1 activity is intimately involved in maintaining genomic stability. As histones have been reported early as covalent acceptors of PAR (Aubin et al., 1982), disassembly of nucleosomes to facilitate repair was suggested. Soon after this theory, *in vitro* experiments showed that purified PAR added to polynucleosomes was able to relax their condensed structure (Poirier et al., 1982). This pointed to non-covalent interaction between at least the linker histone H1 and PAR. Indeed, affinity of H1 to polymer is strong enough to resist phenol partitioning (Panzeter et al., 1992). In addition, also core histones have been shown to be covalently (Ueda et al., 1975; Ogata et al., 1980a; Messner et al., 2010) and non-covalently (Adamietz and Rudolph, 1984; Kreimeyer et al., 1984) modified.

These data led to the assumption that one of the major tasks of PAR synthesis is to clear DNA from nucleosomes by direct modification as well as binding of histones to polymer, granting access of repair factors to the lesion (Mathis and Althaus, 1987; Realini and Althaus, 1992). The detection of PBMs in histones and many other proteins related to DNA repair and stress response, i.e., tumor suppressor p53, cyclin-dependent kinase inhibitor p21, base-excision- and single-strand break-repair protein XRCC1, nucleotide-excision repair protein XPA, DNA-Pol Σ, telomerase subunit TERT, Ku70 and mismatch-repair protein MSH6 (Pleschke et al., 2000), corroborated the hypothesis of PARP1 as a repair and cell cycle regulator. This was confirmed *in vivo* by the fact that the BER adaptor protein XRCC1 (X-ray repair cross-complementing protein 1) depends on PAR for its recruitment to lesions. Inhibition or knockout of PARP1 strongly impacts on XRCC1 enrichment at DNA strand breaks (El-Khamisy et al., 2003). XRCC1 interacts as shuttle with proteins necessary to perform the synthesis and resealing steps after incision as DNA Polβ, polynucleotide kinase and DNA ligase III. Direct interaction of PARP1 with DNA ligase III may help in formation and guiding of the productive complex (Leppard et al., 2003).

Thus, PARP1 and its activity are important regulators of DNA nick-repair. Shortage of the substrate NAD^+ or strong activation may limit efficiency of repair, as PARP1 binds tightly to DNA breaks if no auto-modification takes place (Satoh and Lindahl, 1992; Satoh et al., 1994), and hyperactivation may shift the spectrum of PARP1 protein-substrates. This is in line with studies showing increased genomic instability by application of PARP inhibitors, and at least *in vitro*, PARP1 is able to inhibit DNA polymerases α and β as well as DNA ligase II by covalent modification (Yoshihara et al., 1985). This could represent a regulatory mechanism to avoid futile repair attempts of cells suffering from a high burden of DNA damage. PARP1 also interacts and stimulates flap-endonuclease-1 (FEN1), responsible for cleaving exposed DNA single strands (flaps) derived from strand-displacement synthesis during BER or replication (Prasad et al., 2001). Finally, the chromatin remodeler Alc1 (Ahel et al., 2009; Gottschalk et al., 2009) and APLF1, a histone chaperone including AP-endonuclease activity (Eustermann et al., 2010; Mehrotra et al., 2011), are recruited and activated upon PAR binding, probably facilitating nucleosome disassembly and re-assembly before and after repair process (**Figure 1**).

DOUBLE-STRAND BREAK REPAIR AND REPLICATION

PARP1 also regulates signaling in double strand break repair (DSBR). Inhibition of PARylation hampers and delays activation of initiator PI3K-related kinase ATM (ataxia telangiectasia mutated) (Haince et al., 2007), and ATM forms a complex with PARP1 (Aguilar-Quesada et al., 2007). There is evidence that also DNA-PK directly interacts with and is stimulated by PARP1 (Ruscetti et al., 1998). The interaction of DNA-PK and PARP1 is strengthened by the observation that suppression of the activity of one of them negatively affects the functionality of the other *in vitro* (Veuger et al., 2004). In addition to these two important damage-signaling kinases, PARP1 has many overlapping interaction partners with WRN, a RecQ helicase with

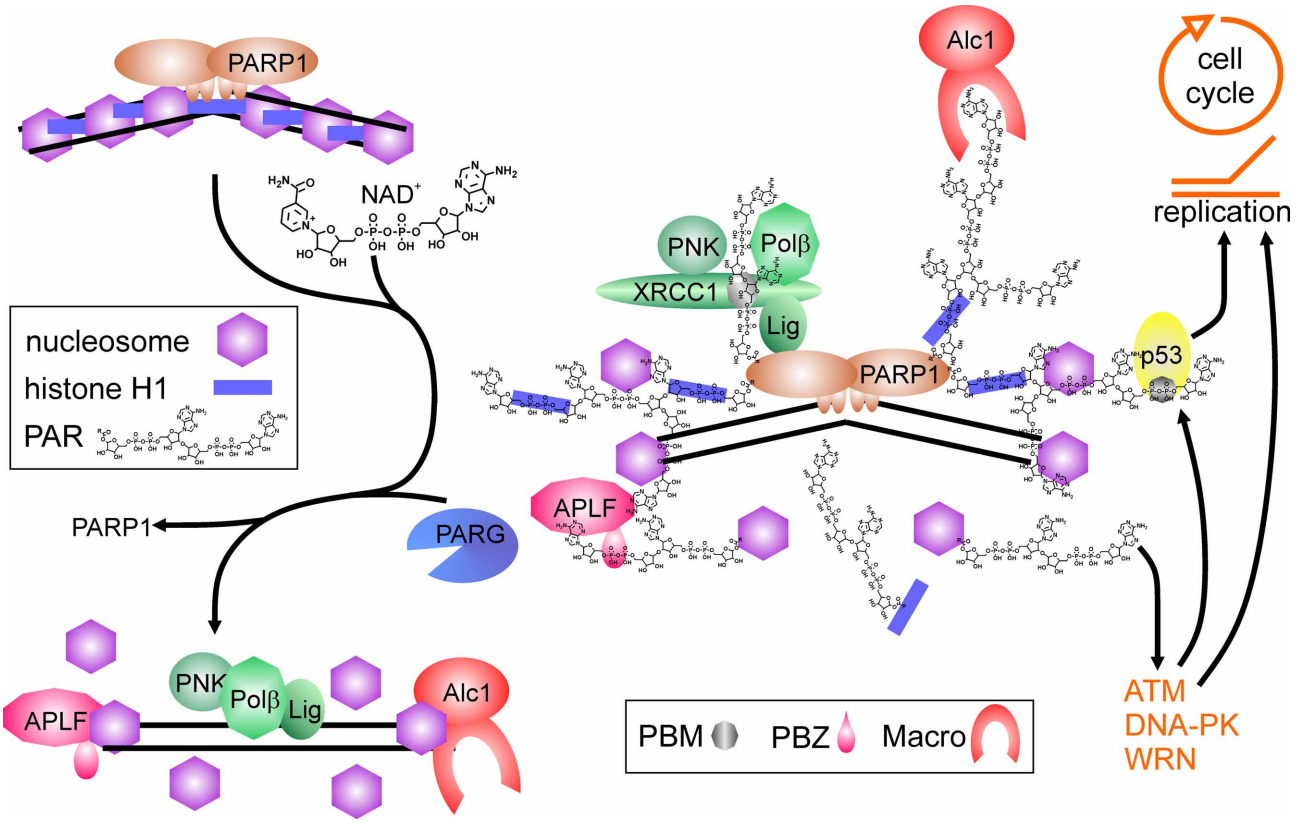


FIGURE 1 | Poly(ADP-ribose) polymerase1 in DNA repair. Binding to DNA breaks and dimerization activates PARP1, which synthesizes poly(ADP-ribose) (PAR) from NAD^+, covalently modifying itself and neighboring proteins, i.e., histones (blue bars = histone H1, purple hexagons = core histones). Proteins containing PAR interaction motifs (PBM = PAR binding motif, PBZ = PAR binding zinc finger; Macro = macro-domain) are recruited to the site of damage, whereas histones in the vicinity are displaced from DNA. Auto-modification of PARP1 abrogates PARP1 DNA binding. Recruited proteins like the XRCC1-complex containing PNK (polynucleotide kinase), Polβ (DNA polymerase β) and Lig (DNA ligase III) are released from PAR chains by degrading activity of PARG and can perform repair on the nucleosome-free DNA. Histone chaperones as APLF and Alc1 may help in disassembling and reassembling of histones on DNA. Binding of PAR to p53 (either covalent or non-covalent) as well as interaction of PARP1 and PAR with proteins like ATM, DNA-PK and WRN regulate cell cycle progression and replication.

exonuclease activity mutated in the Werner adult premature aging syndrome. WRN is responsible for resolving DNA structures such as Holliday junctions and repair intermediates. It participates in BER, DSBR, replication and maintenance of telomeres, the latter one by proper opening the protective *t*-loop. WRN and PARP1 directly interact and regulate each other (Adelfalk et al., 2003; von Kobbe et al., 2003, 2004), and are able to form a complex with the DNA-PK subunits K70/Ku80 (Li et al., 2004). In this regard, it is interesting to note that FEN1 also interacts with WRN in BER and at telomeres (Brosh et al., 2001; Sharma et al., 2003), where also PARP1 activity is needed to maintain proper length (Beneke et al., 2008). Another cellular site were all three proteins—FEN1, WRN, and PARP1—are located together is the replication complex (Sharma et al., 2004). It has been shown that PARP1 modifies at least 15 different proteins in the complex, most prominently DNA Polα, topoisomerase I (TopoI) and proliferating cell nuclear antigen (PCNA), but it is unclear if PARylation is needed for proper assembly of replication complex or for regulation of its functionality (Simbulan-Rosenthal et al., 1998). Poisoning of TopoI stalls replication forks, and reversal of this depends on PARP1 activity (Ray Chaudhuri et al., 2012), probably by reactivating TopoI and induction of repair (Malanga and Althaus, 2004).

PARP1 IN TRANSCRIPTION

PARP1 ACTIVITY AS NEGATIVE CONTROLLER OF TRANSCRIPTION

Transcription by RNA Pol II is regulated in multiple ways, i.e., by induced assembly of different specific transcription factor complexes at susceptible promoters. In addition, general transcription factors—named TFII followed by a letter—are needed for proper transcription of any gene [see Thomas and Chiang (2006) for review]. PARP1 has been isolated in 1983 as TFIIC, necessary for suppression of transcription initiation at nicked DNA (Slattery et al., 1983). Activated PARP1 abrogates formation of the pre-initiation complex (PIC) (Oei et al., 1998b) by PARylating the TATA-binding protein (TBP) (Oei et al., 1998a) and TFIIF (Rawling and Alvarez-Gonzalez, 1997) (**Figure 2A**). Similarly, specific transcription factors as YY1, p53, CREB, Sp1, and NFκB are prevented from binding to their respective recognition sequence if PARylated (Wesierska-Gadek et al., 1996; Oei et al.,

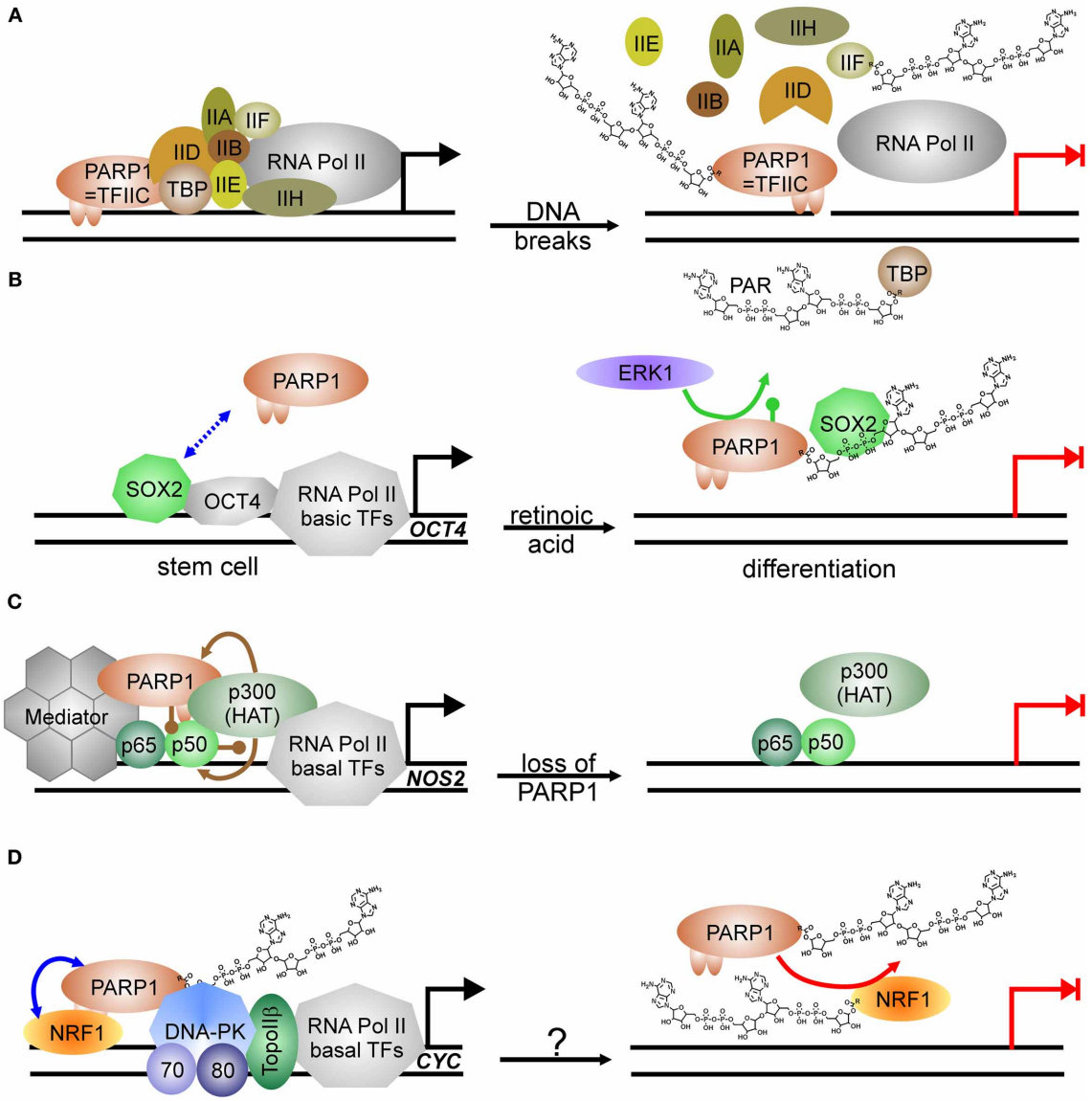


FIGURE 2 | PARP1-activity mediated suppression of transcription. **(A)** PARP1 as basal transcription factor TFIIC monitors DNA breaks in the vicinity of promoters. Transcription machinery is disassembled at pre-initiation complex formation due to modification of TBP (TATA-binding protein) and TFIIF with PAR after DNA damage induction. Transcription is blocked (switch from black arrow to blocked red arrow). **(B)** PARP1 in regulation of stem cell differentiation. SOX2 weakly interacts with PARP1 (dashed double-headed blue arrow). Phosphorylation (green lollypop) of PARP1 by kinase ERK1 leads to auto-modification of PARP1. SOX2 DNA-binding and dimerization with OCT4 is disrupted by interaction with PARylated PARP1. Transcription is abrogated (switch from black arrow to blocked red arrow). **(C)** Positive impact of PARP1 protein itself on transcription as co-activator of NFκB. At the *NOS2* promoter, PARP1 is acetylated (brown lollypops) by p300 HAT (histone acetyl-transferase), which also acetylates NFκB, and interacts thereafter with NFκB subunit p50. Binding of co-activator Mediator to the complex is stabilized by PARP1 and facilitates transcription. Loss of PARP1 and also putatively its activation disrupts transcription complex. Transcription is abrogated (switch from black arrow to blocked red arrow). **(D)** PARP1 as co-activator and PARP1 activity as repressor. PARP1 complexes with NRF1 irrespectively of its own modification status (blue double-headed arrow). Covalent modification of NRF1 with PAR (red arrow) disrupts the permissive transcription complex containing DNA-PKcs/Ku70/Ku80 and TopoIIβ, releasing NRF1 from DNA. Transcription is blocked (switch from black arrow to blocked red arrow). The respective stimulus needs to be determined (question mark).

1997; Chang and Alvarez-Gonzalez, 2001; Mendoza-Alvarez and Alvarez-Gonzalez, 2001). PARylation negatively controls also the function of transcription factors essential in sex-determination via SRY, and maintenance of "stem-ness" of cells via SOX2. SRY (sex-determining region of Y) is the master regulator in sex-determination and essential for testis development. SRY-mediated transcription is severely impaired upon PARP1 stimulation, as its covalent modification abrogates interaction with its cognate DNA-binding sequence (Li et al., 2006). SOX2 acts in concert with OCT4 in stem-cell maintenance. Both form a

complex on respective promoters/enhancers, i.e., *NANOG* and *SOX2* and *OCT4*, leading to positive feedback control [for review, see Kashyap et al. (2009)]. SOX2 interacts weakly with PARP1 on regulatory elements, but upon activation of PARP1, binding between both proteins is enhanced due to auto-modification of PARP1 (Lai et al., 2012) (**Figure 2B**). Although SOX2 is not a direct target of PARylation, SOX2 DNA-binding is inhibited, leading to disruption of SOX2/OCT4 transcriptional complexes and induction of differentiation. Hypothetically, this is achieved by SOX2-PAR interaction, but formal proof is missing yet. This sequence of events was described in embryonic stem cells treated with retinoic acid: exposure to RA led to activation of FGF/ERK1 pathway resulting in increased PARylation of PARP1, probably by phosphorylation of PARP1, which has been shown to activate the enzyme (Kauppinen et al., 2006; Cohen-Armon, 2007). Thereafter, binding between SOX2 and PARP1 is enhanced due to auto-modification, transactivator function of SOX2 is inhibited and subsequently, differentiation of ESC is induced.

PARP1 PROTEIN AS POSITIVE CO-FACTOR IN TRANSCRIPTION

On the other hand, PARP1 is also a general activator of transcription as it is identical with positive co-factor 1 (PC1) (Meisterernst et al., 1997). Supporting this, PARP1 has been shown to associate with RNA Pol II-dependent promoters in open chromatin, whereas H1 is mainly found in heterochromatic-like regions, making their presence on chromosomes mutually exclusive (Krishnakumar et al., 2008). Specifically, E2F1 interacts with PARP1 in order to induce expression of S-phase genes such as DNA Polα/DNA primase, RPA and E2F1 itself (Simbulan-Rosenthal et al., 1999). DNA-binding or PARP1 activity is not needed for this co-activator function (Simbulan-Rosenthal et al., 2003). Similar to E2F1, another important transcription factor depends on PARP1 protein for transactivator function: NFκB, the master-regulator of immune-responsive genes (Hassa and Hottiger, 1999) (**Figure 2C**). PARP1 and both subunits of NFκB, p50 and p65, form a ternary complex, and without PARP1, some genes targeted by NFκB are not expressed, for example *NOS2*, coding for inducible nitric oxide synthase (Hassa et al., 2001). PARP1 activity is dispensable for co-activator function and may even inhibit NFκB-dependent transcription due to interference with its DNA binding (Chang and Alvarez-Gonzalez, 2001). There is evidence that effective NFκB-mediated transactivation of genes has several layers of regulation. PARP1 acetylation by histone acetyl-transferase (HAT) p300 is a prerequisite for binding to NFκB subunit p50, and p300 also binds and activates NFκB directly (Hassa et al., 2005). Additionally, Mediator—another co-activator complex—interacts with both NFκB and PARP1, synergistically enhancing NFκB transactivator function.

A switch between co-activating and repressive function has been described in insulin producing β-cells. At the Reg protein promoter PARP1 presence is necessary for transcription, but activation by DNA strand breaks disrupts the complex and transcription is silenced (Akiyama et al., 2001). In line, the master transcriptional regulator of genes related to energy metabolism and mitochondrial function, NRF1 (nuclear respiratory factor), is also controlled by PARP1 activity (**Figure 2D**). NRF1 binds PARP1 irrespective of auto-modification status, and PARP1 recruits the DNA-PK/TopoIIβ complex to NRF1-regulated promoters for expression, i.e., of the cytochrome c gene (*CYC*). As soon as NRF1 becomes a target for PARP1 activity, NRF1 loses its ability to bind PARP1 and transcription of respective genes is shut down (Hossain et al., 2009).

Thus, it seems a general feature that PARP1 functions as a nuclear sensor of stress exposure, and upon stimulation of its enzymatic activity by DNA breaks or phosphorylation, it shuts down transcription. The PARP1 protein itself may act as positive regulator for expression. In this way, a broad range of genes can be repressed that are not necessary for proper response—or even contradictory—to the imposed stress.

PARP1 ACTIVITY AS POSITIVE CO-FACTOR IN TRANSCRIPTION

However, transcriptional regulation by PARP1 grew more complicated in 2002, when a groundbreaking work appeared in *Genes and Development* and a follow up 2003 in *Science*, using *D. melanogaster* as a model (Tulin et al., 2002; Tulin and Spradling, 2003). Here, PARP1 activity is described to facilitate transcription. *D. melanogaster* encodes in its genome only two PARPs, one is similar to PARP5 (tankyrase) and the other shares substantial degree of homology with PARP1 from other organisms. In *D. melanogaster*, PARylation is needed during larval development as well as in heat shock for activation of specific genes, i.e., heat-shock protein *Hsp70*. Employing polytene chromosomes it could be visualized that hormone application or heat shock induced PARP1 activity, and that the synthesized PAR opened chromatin structure, generating so called "puffs," which are areas of ongoing transcription. The mechanism was further elucidated by Petesch and Lis (Petesch and Lis, 2008, 2012). The heat shock factor (HSF) binds to the *Hsp70* promoter, where a stalled RNA Pol II resides, poised for transcription. HSF recruits the HAT Tip60, which acetylates histone H2A, leading to its exchange (**Figure 3A**). PARP1 resides dormant at the *Hsp70* promoter and its activity is rapidly induced by Tip60, either by the described histone switch or by direct acetylation. Subsequently, PARP1 modifies itself and is released from the promoter. Following this, histones are disassembled from the DNA and trapped in the growing polymer chain, paving the way for the RNA polymerase. Interestingly, mammalian cells contain the PARP1-suppressive histone macroH2A1.1 in *HSP70* genes responsive to heat shock, whereas constitutive *HSP70* promoters lack this variant (Ouararhni et al., 2006). In addition, heat shock induces expression of *HSP70* dependent on PAR synthesis, pointing to a very similar regulatory mechanism. Thus, PARP activity changes the surrounding chromatin by disengaging suppressive nucleosomal DNA binding. In the following years, this feature was extended to other factors than histones.

Similar to RA-mediated differentiation of ESC described above, PARP1 activity is involved in differentiation of neuronal stem cells, NSC, but this time as positive regulator of transcription (Ju et al., 2004) (**Figure 3B**). In NSC, transcription factor HES1 (Hairy/Enhancer of Split) is a negative regulator of gene expression. It interacts with the TLE (transducin-like Enhancer of split)/Groucho co-repressor complex. Groucho is able to recruit histone deacetylases, forming suppressive chromatin marks on

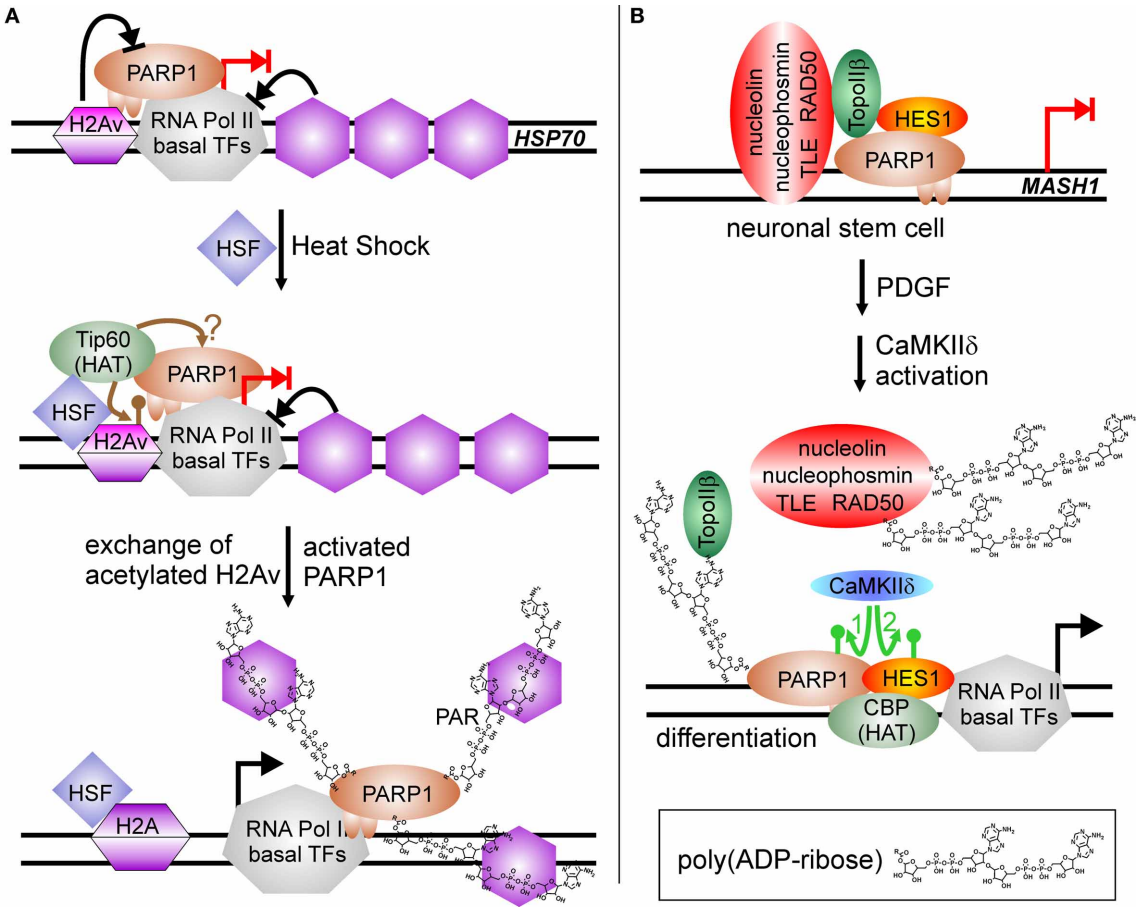


FIGURE 3 | PARP1-activity mediated stimulation of transcription. **(A)** Chromatin modulation by PARP(1) activity. *HSP70* promoter is silenced by incorporated H2A variant (H2Av), blocking PARP(1) activity (blocked black arrow from H2Av to PARP1). RNA Pol II complex is assembled at promoter and poised for transcription, but specific nucleosome positioning halts RNA Pol II activity (blocked black arrow from nucleosomes to RNA Pol II). Heat shock induces the translocation of HSF (heat shock factor) to the *HSP70* promoter and recruits the histone acetyl-transferase Tip60, which induces replacement of H2Av against standard H2A by acetylation (brown lollypop). Putatively, it could also target PARP1, similar to the situation at NFκB-regulated promoters. Activated PARP1 is released from the promoter and traps suppressive histones in the growing PAR chain, facilitating transcription (switch blocked red arrow to black arrow). **(B)** PARylation activates expression of differentiation-linked genes. Treatment of neuronal stem cells with PDGF (platelet derived growth factor) induces activity of the kinase CaMKIIδ. Phosphorylation of PARP1 (green arrow No. 1) stimulates PARylation, leading to disassembly of the large co-repressor complex including nucleolin, nucleophosmin, TLE, RAD50, TopoIIβ, PARP1 and HES1 at *MASH1* promoter. Auto-modified PARP1 and HES1 recruit histone acetyl-transferase CBP, and subsequent phosphorylation of repressor protein HES1 by CaMKIIδ (green arrow No. 2) initiates transcription (switch blocked red arrow to black arrow).

differentiation-linked promoters like *MASH1*. PARP1 is part of this repressor complex, together with TopoIIβ, nucleophosmin, nucleolin and Rad50. Initiation of signaling events inducing differentiation by platelet-derived growth factor (PDGF) leads to activation of calcium-dependent kinase CaMKIIδ, which in turn is recruited to the *MASH1* promoter and phosphorylates PARP1. Phosphorylation activates PARP1 resulting in PARylation of co-repressor proteins, i.e., TLE/Groucho, TopoIIβ, nucleophosmin, nucleolin, Rad50, and PARP1 itself. Polymer-modified proteins except PARP1 leave the complex and histone acetylase CBP is recruited. Subsequently, HES1 is also phosphorylated by CaMKIIδ, which turns this repressive transcription factor in an activator of *MASH1* expression. Addition of a PARP1 inhibitor or a PARP1 mutant lacking polymerization activity (Glu988 to Ala988) blocked differentiation.

Low levels of a similar repressor complex are found at the 17 β–estradiol (E_2)-sensitive *pS2* promoter, composed of PARP1, TopoIIβ, nucleophosmin, nucleolin and HSP70 (Ju et al., 2006). Treatment with E_2 leads to a rapid increase of TopoIIβ and PARP1 at the promoter, followed by recruitment of DNA-PK and co-activator CBP, whereas co-repressors are lost from *pS2* promoter (**Figure 4**). Formation of double-strand breaks (dsb) by TopoIIβ induces PARP1 activity and replacement of histone H1 with HMGB1/2, facilitating expression. Again, treatment with a PARP1 inhibitor or usage of the same catalytic mutant as above blocked *pS2* activation.

There are several more examples for PARP1 activity driven transcription. The repressor-activator switch has also been described in context of chromatin-modulator protein DEK (Gamble and Fisher, 2007). In a complex, DEK and PARP1

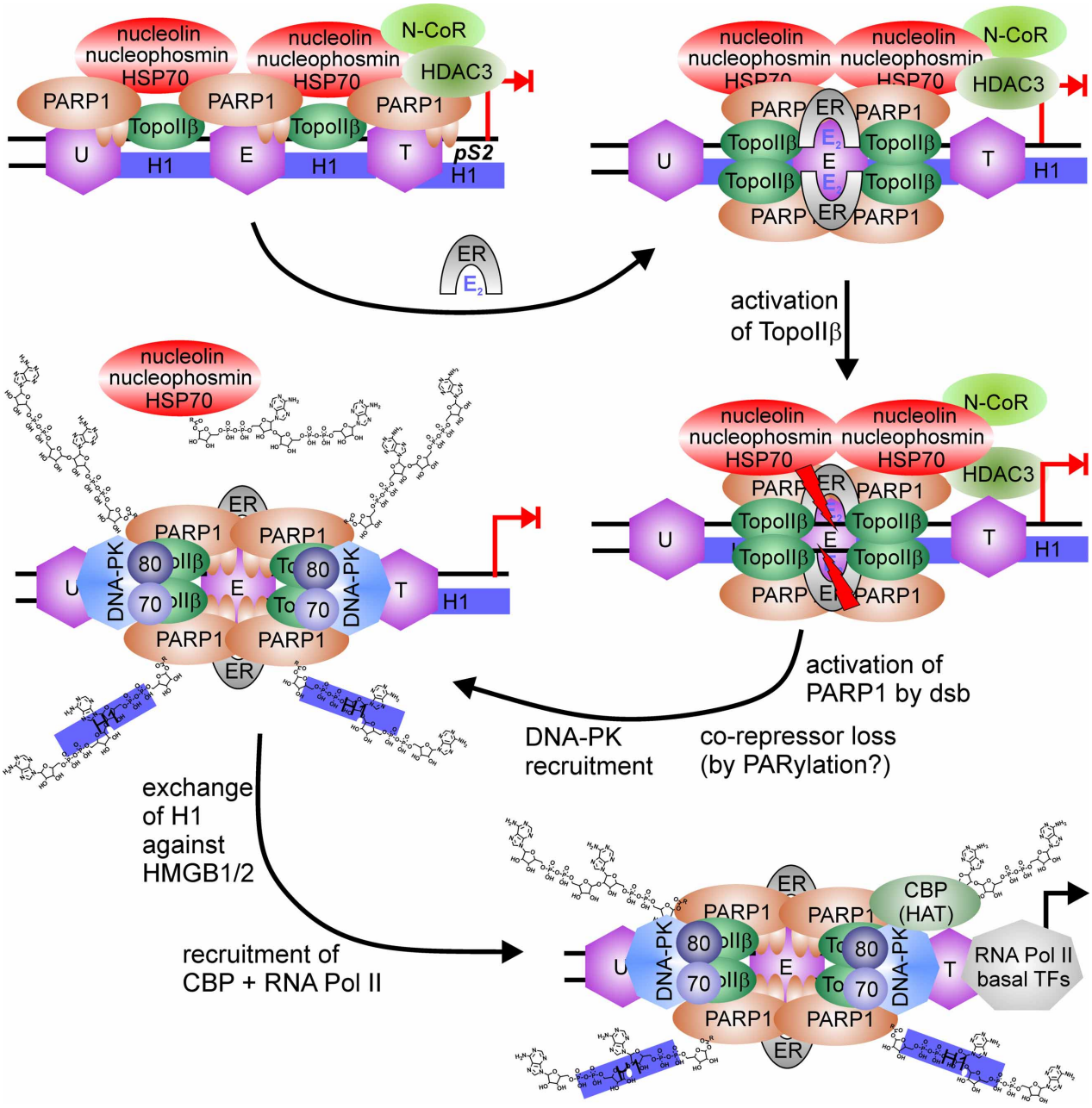


FIGURE 4 | PARP1-switch from repressor to activator of *pS2* expression. PARP1 is member of a large co-repressor complex with nucleolin, nucleophosmin, HSP70, TopoIIβ, PARP1, and N-CoR. 17β–estradiol (E_2) application recruits estrogen-receptor (ER) to the ER responsive element, where the nucleosome (E) is localaized. This induces enrichment of TopoIIβ and PARP1 at the same site (E) and loss at surrounding nucleosomes located upstream (U) or downstream at the TATA-box (T). TopoIIβ induces DNA double-strand break at ERE (red flashes), which activates PARP1 and leads to the disassembly of co-repressor complex, putatively by PARylation. Additionally, DNA-PK complex (DNA-PKcs/Ku70/Ku80) is bound to the interrupted DNA strand. PARylation induces exchange of suppressive linker histone H1 against permissive HMGB1/2 protein and recruits histone acetyl-transferase (HAT) CBP.

suppress transcription *in vitro* on chromatinized plasmid templates. Addition of NAD^+ relieves suppression as both DEK and PARP1 are lost from template due to modification with poly(ADP-ribose). This enables the recruitment of the Mediator co-activator complex and subsequent transcription. PARP1 is also localized at promoters of mitochondria-related nuclear genes for DNA repair and transcription (Lapucci et al., 2011). Treatment of cells with PARP inhibitors reduces mitochondrial DNA integrity and as a consequence, expression of respiratory genes and ATP production is compromised.

Of note, PARP1 regulates its own promoter, which resembles that of TATA-less housekeeping genes. Upstream of the

initiation site, there are racket-like inverted repeats, which are able to form alternative stem-loops. These structures can be bound and stabilized by PARP1, leading to abrogation of transcription. Activity of the enzyme is not necessary for repression, but would obviously release the suppression of the *PARP1* gene (Oei et al., 1994; Schweiger et al., 1995; Soldatenkov et al., 2002; Vidakovic et al., 2009). In this way, PARP1 protein keeps itself at a constant level.

POST-TRANSLATIONAL MODIFICATIONS OF PARP1 IN TRANSCRIPTION

In summary, PARP1 is able to regulate transcription at several levels. If PARP1 is in fact belonging to the group of general factors of RNA-PolII transcription (the missing TFIIC) may be questionable, but its interaction with several transactivator proteins is without doubt. It can act itself as a co-activator of gene expression, with the potential to abrogate transcription after activation. In this way, genes are transiently silenced that are either not needed for or may even interfere with an appropriate stress response in cells. Alternatively, PARP1 activity can rearrange nucleosomal organization and facilitate thereby accessibility of the promoter to transcription factors and RNA Pol II. In this setting, PARP1 can either be specifically recruited or may be switched from a co-repressor to a co-activator after stimulation by post-translational modification [for review, see also Kraus (2008)]. Indeed, PARP1 is targeted by many enzymatic activities. Most prominent is the auto-modification by PARylation, inhibiting DNA-binding as well as enzymatic reaction. Phosphorylation by ERK1/2 (Kauppinen et al., 2006; Cohen-Armon, 2007), AMPK (Walker et al., 2006) and CaMKIIδ (Ju et al., 2004) has been reported, stimulating PARP1. Acetylation of PARP1 also increases activity (Hassa et al., 2005), whereas SUMOylation seems to restrict protein-substrate targeting of PARP1 (Masson et al., 1997; Messner et al., 2009; Ryu et al., 2010). K48-Ubiquitination leads to degradation of PARP1 (Wang et al., 2008; Martin et al., 2009), which is probably induced by auto-modification of the enzyme (Kashima et al., 2012). Interestingly, there is crosstalk between these modifications, as SUMOylation inhibits PARP1 acetylation, thus diminishing its co-activator function in NFκB transcription (Messner et al., 2009), and for activation of the *HSP70.1* promoter in mammalian cells an ordered sequence of PARP1 modifications has been described (Martin et al., 2009): Heat shock induces activation and auto-modification of PARP1 residing at the *HSP70.1* promoter, which recruits SUMOylating enzymes Ubc9 and PIASy to this site, resulting in polySUMOylation of PARP1 and full transcriptional activation of the *HSP70.1* gene. SUMO-modification in turn attracts ubiquitin-ligase RNF4, which subsequently tags PARP1 for degradation. Promoters of inducible HSP70.1 and HSP70.2, but not of constitutive HSP70.8, are enriched of histone macroH2A1.1, which suppresses PARP1 activity. Heat shock relieves suppression (Ouararhni et al., 2006), putatively via Tip60-mediated acetylation of the histone as described in insect cells, thus facilitating PARylation reaction.

PARP1 AND CTCF IN EPIGENETIC CONTROLLING

First evidence that PARP1 plays a role in epigenetic mechanisms came from experiments utilizing PARP inhibitors. Treatment of fibroblasts with 3-aminobenzamide (3AB), a first generation PARP inhibitor with low potency, induced increased methylation of CpG islands in the *Htf9* promoter (Zardo and Caiafa, 1998), and cells displayed a rise in number and density of heterochromatic foci as well as genome-wide DNA-methylation (de Capoa et al., 1999). CCCTC-binding factor (CTCF) is known to bind regulatory regions that are hypomethylated, organizing chromatin domains as insulator and transcriptional regulator, a function which has been extensively described for the *IGF2-H19 ICR* (imprinting control region). Binding of CTCF to the non-methylated maternal *ICR*-allele facilitates *H19* transcription and silencing of *IGF2,* whereas the paternal *IGF2* gene is expressed. Loss of CTCF function increases methylation marks in respective sites and *vice versa* (CTCF is topic of several review in this special issue), i.e., in case of the *H19 ICR* not only the paternal allele, but also maternal allele is methylated. Using the *H19 ICR* as bait, CTCF was shown to be a prominent target of PARP1 activity, resulting in a molecular size shift from 130 kDa to 180 kDa (Yu et al., 2004). Covalent modification of CTCF did not interfere with its DNA-binding ability in contrast to many other proteins, but on the opposite, lack of PAR due to 3AB treatment abrogated its insulator function. Actually, CTCF bound to target sites was associated with a higher amount of PAR than free unbound CTCF.

Soon after, another link between CTCF, PARP1 and methylation has been discovered. It was shown that DNA-methyltransferase 1 (*DNMT1*) binds to PARP1, mainly if PARP1 is auto-modified. Binding to PAR—probably via two putative PBM—inhibits DNA methylation by *DNMT1*. Interestingly, *DNMT1* has a higher affinity to PAR than to DNA, as it is case for histones (Reale et al., 2005). CTCF binds to *DNMT1* itself, but is unable to block *DNMT1* activity, so it depends on recruited PARP1 to abrogate *DNMT1* function despite physical presence. CTCF stimulates PARP1 activity even without nicked DNA, leading to an increase in PARylated PARP1 and CTCF (Guastafierro et al., 2008). In addition, the 130 kDa form CTCF was shown to bind PAR in a non-covalent manner (**Figure 5**) (Zampieri et al., 2012). In contrast to the negative effect on *DNMT1* activity, there is evidence that PARP1 and PARylation are needed to maintain expression of *DNMT1* in mouse L929 fibroblasts. PARP1 and PAR were detected at the *DNMT1* promoter in conjunction with *DNMT1* but without CTCF, and loss of PAR by overexpression of the degrading enzyme PARG severely reduced *DNMT1* in cells by silencing through promoter-methylation (Zampieri et al., 2009). Thus, PARP1 activity maintains transcription at the *DNMT1* promoter by keeping it clear of DNA-methylation marks inserted by *DNMT1* itself. However, an earlier publication by the same group showed the opposite effect, even in the same cell system (Zardo et al., 2002). Treatment of L929 cells with 2 mM 3AB resulted in twofold increased expression of *DNMT1*. Thus, it seems that PARP1 inhibition and increased polymer degradation by PARG overexpression may not be the same. With 3AB, PAR formation is blocked, whereas increased PARG activity induces faster loss of *synthesized* PAR. It could also be the other way round, with low-dose 3AB not preventing basal PARylation and high PARG activity leading to degradation of basal polymers. Thus, results from these two approaches may not be directly comparable.

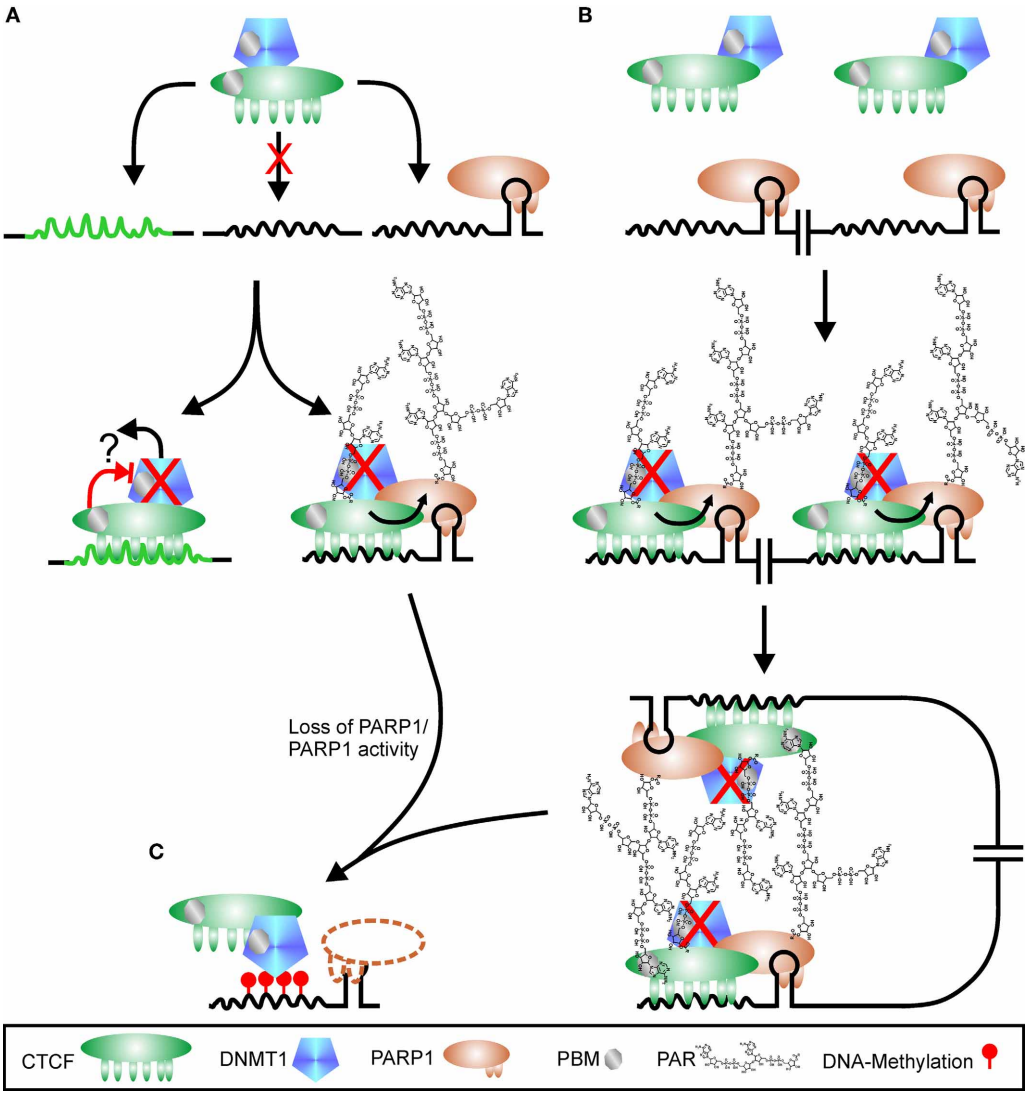


FIGURE 5 | Regulation of CTCF function by PARylation. (A) CTCF shows a high variability in putative binding sequences. There are probably high-affinity sites (green peaked line) and low-affinity sites (black peaked line), with the latter one hypothetically only used if additional signals are present, for example PARP1 bound to a stem loop. CTCF is in complex with *DNMT1* (DNA methyl-transferase 1). Binding to high-affinity sites may suppress *DNMT1* activity directly (red cross), either by altered interaction after DNA-binding (blocked red arrow) or by release of *DNMT1* from complex (black arrow and question mark). Interaction of CTCF with PARP1 on low-affinity sites stimulates PARP1 activity, which covalently modifies itself and CTCF. *DNMT1* is inhibited (red cross) by binding to PAR via a PBM (PAR binding motif). Loss of PARP1 (dashed outline of PARP1 protein) or polymer releases suppression of *DNMT1*, and the CTCF recognition site is *de novo* methylated (red lollipops), omitting further CTCF binding **(C)**. Restructuring chromatin domains may be achieved by simultaneous usage of two adjacent CTCF-PARP1 sites as shown in **(B)**. As both CTCF and *DNMT1* contain PBMs, PAR chains may serve as "glue" between the two complexes, stabilizing chromatin loops. Loss of PARP1 or its product PAR disrupts chromatin domain organization, facilitating *DNMT1* activity **(C)**.

The connection between the four players PARP1, PAR, CTCF, and *DNMT1* has been elucidated in more detail for the differentially methylated region 1 (DMR1) upstream of the *Igf2* promoter (Zampieri et al., 2012). The three proteins CTCF, PARP1, and *DNMT1* can dimerize with each other independently and form together a ternary complex, even without polymer. Most *DNMT1* is associated with CTCF, whereas only a fraction of cellular PARP1 is part of the complex. This complex binds to unmethylated CTCF target sites only. At the DMR1, all three proteins are detected, in conjunction with PAR. Overexpression of PARG leads to disruption of the complex, loss of PARP1 and CTCF and *de novo* methylation of DMR1 by the still bound *DNMT1*. The subcellular distribution of CTCF is also under control of polymer formation (Torrano et al., 2006). Differentiation of K562 myeloid cells induces translocation of CTCF from the nucleoplasm to the nucleolus, accompanied by reduction of rRNA synthesis and growth arrest. Fractionation experiments revealed that the 180 kDa (modified) form of CTCF was prevalent in nucleoli. Inhibition of PARylation by 3AB prevented relocalization of CTCF to nucleoli upon stimulus and

restored nucleolar transcription. Similar results regarding control of rDNA transcription and nulceolar organization by CTCF and PARylation have been described for Drosophila (Guerrero and Maggert, 2011).

There are several examples for the impact of PARylation on CTCF function. CTCF is necessary for proper expression of tumor suppressors p16 (*CDKN2A-INK4*) and E-cadherin (*CDH*) (Witcher and Emerson, 2009) and loss of CTCF or PARP1 represses transcription of these genes. Abrogating polymer synthesis induces hypermethylation, binding of CTCF to respective regulatory sequences is lost and p16 and E-cadherin genes are silenced. In contrast, c-Myc expression was not affected by abrogating PARP1 activity. Also another tumor suppressor, p19ARF, is under control of the CTCF-PARP1-PAR complex (Farrar et al., 2010). Mutation of the potential PARylation attachment sites in CTCF led to loss of insulator function in regulation of transcription and imprinting, similar to application of a PARP inhibitor. PARP1 binds wild-type and mutant CTCF with equal efficiency, but only the wild-type version was able to maintain p19 expression, as well as proper methylation pattern at the *H19 ICR*. The authors also showed that there are genomic hot spots of interaction between CTCF and PARP1. Despite earlier suggestions, it appeared that both isoforms of CTCF, i.e., 130 kDa as well as 180 kDa, are ADP-ribosylated, but to a different extent. Whereas the larger one contains long and putatively branched polymer, the small isoform contains oligo(ADP-ribose), detected only by an antibody with high affinity to short ADP-ribose chains. As not only cell cycle inhibitors p16 and p19 are controlled by CTCF, but also c-Myc (Lobanenkov et al., 1990; Gombert and Krumm, 2009), pRb (De La Rosa-Velazquez et al., 2007), p21 and p27 (Qi et al., 2003), loss of CTCF function may support cancer formation and indeed, 87.7% of tested breast tumors showed alterations in the ratio between PARylated 180 kDa and 130 kDa forms of CTCF. Whereas normal breast tissue contains only the large isoform, both can be detected in tumor tissue. Interestingly, there is transition from CTCF-180 to CTCF-130 in primary cultures from breast tissue upon stimulation of proliferation and *vice versa*, i.e., growth arrest induces CTCF-180 (Docquier et al., 2009). This is in line with the above described observation of (Torrano et al., 2006). Despite general interaction between CTCF and PARP1 independently from other factors, CTCF function is not on all sites impaired by abrogating PARylation.

DISCUSSION

PARP1 IN REPAIR

PARP1 regulating chromatin can be divided into two different major subsets: one is characterized by no or low levels of PARylation in unstimulated cells, the other by high levels of PAR as cellular stress response, but the border between these is somehow blurred. Stimulation by signaling pathways leading to phosphorylation of PARP1 at specific promoters may result in high local PARylation with no obvious change in overall polymer abundance. So, to which group does it belong? Nevertheless, massive PARylation after genotoxic stress results in changes in chromatin, which may be specific for the surrounding information or more general. Overall changes include the rearrangement of nucleosomal structure by modification of core and linker histones, which can be covalent (confined to the direct interaction with PARP1) and non-covalent, reaching beyond the proteins' localization by spreading of the PAR-"tree". Thus, PARP1 activity clears the way for repair enzymes and complexes (see **Figure 1**). Additionally, the polymer is capable of attracting factors if they contain one of the three PAR-interaction modules described so far, which many proteins in DNA-maintenance pathways do. Probably, binding to polymer traps and therefore enriches respective proteins at the site of DNA breaks, and subsequent release by PARG activity enables repair of the damage. By combination of these two functions in one enzyme, chromatin loosening and protein attraction, repair rates can be accelerated. Additionally, PAR-synthesis activates the initiator kinase ATM. It has been suggested that the shift from the catalytically inactive dimer to the active monomeric form of ATM may be induced by chromatin alterations due to DNA breaks (Khanna et al., 2001), and that interaction with the MRN complex (MRE11/RAD50/NBN)—which is also a downstream target of ATM—aids in this (Assenmacher and Hopfner, 2004). The discovery of a PBM in ATM, the modulation of kinase activation by PARP inhibition and the reported direct interaction between both proteins support the hypothesis that local PAR-formation initiates the respective signaling cascade, as polymer relaxes chromatin and is bound by ATM. Thus, blocking PARP1 activity obviously slows down repair.

PARP1 IN TRANSCRIPTION

A more specific way of mediating stress response by PARP1 activity is its participation in transcriptional regulation. Suppression of transcription in a generalized way helps to avoid additional damage induced by clash of complexes (RNA Pol II vs. DNA-repair) or possible sequence-loss caused by melting the double-strand during transcription in the vicinity of breaks. This may be facilitated by the proposed role of TFIIC/PARP1 as suppressor of nick-induced transcription via modification of basal TFs like TBP, blocking formation of PIC. But as most data supporting this came from *in vitro* experiments, this actually may be not the case in living cells. Alternatively, specific inhibition of certain promoters can be achieved in triggering PARP1 activity if the enzyme is present in the complex. Interaction with several transcription factors such as YY1, NFκB or others has been reported in several publications. Interestingly, there is mounting evidence that PARP1 acts as a switch in these complexes. For example, it is an essential co-factor of NFκB-mediated transcription, but PARylation disrupts the transcription machinery, at least *in vitro*. Similarly, polymer formation interferes with YY1 or p53 DNA binding. To complicate the whole situation, p53 displays not only three covalent attachment sites for PAR, but contains also three polymer-binding motifs. Covalent modification interferes with respective DNA binding, but strikingly abrogates nuclear export of p53 (Kanai et al., 2007); however, what is the purpose of p53 binding non-covalently to PAR? One suggestion may be the attraction and exchange of proteins at promoters. Aging and correlated oxidative stress in rat liver cells leads at the androgen receptor promoter to the exchange of positive co-factors including PARP1 against transcriptional suppressors including p53 (Shi et al., 2008). A hypothesis would be that stress-associated activation and auto-modification of PARP1 disrupts the permissive

complex, and p53 is attracted by binding to synthesized polymer, resulting in silencing of the androgen receptor gene. Alternatively, retention of p53 in the nucleus may be achieved by interaction with PAR without any direct modification.

In addition, PARP1 can be activated even in the absence of DNA breaks by post-translational modifications. Phosphorylation of PARP1 mediated by CaMKIIδ after PDGF stimulation of neuronal stem cells initiates PAR synthesis at HES1-suppressed promoters. As a result, co-repressor proteins Groucho/TLE, nucleolin, nucleophosmin and TopoIIβ are released and co-activators, for example CBP, are recruited, inducing differentiation. Interestingly, PARP1 can still be found at the promoter, suggesting localization of the protein independent of its DNA-binding ability (Ju et al., 2004). If TopoIIβ activity is needed in this sequence of events has not been determined. Exchanging specific factors mediated by PARP1 activity is also seen in response to other signaling events. TopoIIβ dependent transcriptional activation is intimately associated with PARylation upon strand-break formation and subtle changes in nucleosome-positioning (Ju et al., 2006). A PARP1/TopoIIβ/DNA-PK complex is recruited to the *pS2* promoter upon stimulation of cells by estradiol and induces a DNA break. This in turn activates the PARP1 protein residing at the promoter as part of the repressor complex and modification of histone H1, which is subsequently exchanged against HMGB1, facilitating transcription. Unfortunately, the authors did not show any data about if and when proteins are PARylated. Also, the authors did not dissect the order of observed events, i.e., which is first: dsb formation by TopoIIβ or PARylation? They proposed TopoIIβ as initiating enzyme, triggering PARP1 activity, but failed to provide evidence for that. It could also well be that binding of the ER-E_2 complex induces formation of an aberrant DNA structure by kinking the DNA, resulting in activation of PARP1. Poly(ADP-ribose) would in turn release co-repressors and H1 and recruit co-activators, i.e., DNA-PK. Subsequent dsb formation by TopoIIβ could be necessary to enable DNA binding of DNA-PK and integration of HMGB1/2 into the complex. Of note, the suppressive complex at the *pS2* promoter also contained nucleolin and nucleophosmin in addition to PARP1/TopoIIβ. Thus, these three proteins seem to be more general interacting partners of PARP1 in transcription, with nucleolin and nucleophosmin as suppressive factors, whereas PARP1 and TopoIIβ can act as switches. In addition, activity of TopoIIβ is dampened by PARP1 in mouse spermatogenesis. Inhibition of PARP1 increases double-strand break formation of TopoIIβ (Meyer-Ficca et al., 2011b), and necessary exchange of histones against protamine for compaction is disturbed, resulting in poor sperm quality and reduced fertility (Meyer-Ficca et al., 2011a). As it seems, TopoIIβ and PARP1 have a more intimate relationship in controlling chromatin and expression than thought before.

PARP1, CTCF, AND *DNMT1*

PARP activity is needed to prevent spreading of heterochromatic regions by inhibition of *DNMT1*. In addition, PARP1 interacts with chromatin-domain organizing insulator and transcription factor CTCF, which binds only to unmethylated DNA. This implies that epigenetic regulation is mediated by the interplay of PARP1, CTCF, and *DNMT1*. Lack of PAR/PARP1 or CTCF enhances the activity of *DNMT1*. Thus, the ternary complex is poised to change DNA-methylation patterns and subsequently expression profiles. Probably only basic polymer synthesis is needed for PARP1 mediated regulation of CTCF binding, as no publications are so far available that report increased CTCF localization to DNA after PARP1 activity stimulation. On the other hand, reducing PAR-levels has a dramatic impact on CTCFs DNA-binding, cellular localization and genomic methylation-pattern. If CTCF is a direct target of PARP1 or may only be recruited to PAR is still unsolved, as binding to PAR can be strong and resist general separation procedures. Alternatively, the two CTCF isoforms, i.e., 180 kDa and 130 kDa, may represent covalently modified and PAR-bound CTCF, respectively. The question is still unsolved why presence of CTCF on some genomic sites depends on poly(ADP-ribose) and on others not. Hypothetically, the high variability of CTCF binding sequences and the ability of PARP1 to bind to secondary structures may give an answer: binding of CTCF at weak interaction sites is only supported if next to the CTCF docking site a stem loop is present, bound by PARP1 (**Figure 5A**). Concomitant presence of the two proteins stabilizes the complex and triggers PARylation, directly stimulated by CTCF. *DNMT1* is in most cases found in association with CTCF and is therefore also recruited to the weak interaction site. Binding to the polymer abrogates *DNMT1* activity, but the enzyme is poised to methylate DNA as soon as the polymer-mark is lost (**Figure 5C**). At high-affinity sites, CTCF is able to bind on its own and may inhibit *DNMT1* directly or in conjunction with other proteins. Alternatively, binding of CTCF at this position may reduce affinity to *DNMT1* with subsequent loss of the methyl-transferase (**Figure 5A**). If two CTCF/PARP1 sites are located in close proximity due to chromatin domain organization, covalently modified CTCF can induce loop formation by interaction of its polymer-mark with the PBM of another CTCF molecule at the second position (**Figure 5B**), a hypothesis already raised in (Klenova and Ohlsson, 2005; Caiafa et al., 2009). It has been shown that loop-formation is one prominent feature of CTCF mediated chromatin restructuring (Yusufzai et al., 2004; Yusufzai and Felsenfeld, 2004). Auto-modified PARP1 in turn may assist in this. *DNMT1* could also be instrumental in domain formation as its own PAR-binding motif may aid in stabilizing the complex. If PARP1 or its product PAR is lost, *DNMT1* is no longer inhibited and can methylate the respective DNA sequence, abrogating CTCF binding. The hypothesis of CTCF docking sites with different affinities under putative control of PARP1 presence is supported by data presented in Witcher and Emerson (2009). Whereas the PARylation-independent CTCF-homology sequence in the *MYC* promoter displays only very weak PARP1 binding and no recruitment of TopoIIβ, PARP1 strongly interacts on its own with the PARylation-dependent *p16/INK4* promoter together with TopoIIβ. Alternative models have been suggested, in which CTCF is first bound to DNA and recruits in a second step PARP1 to specific sites (Caiafa and Zlatanova, 2009). CTCF-induced PARP1 activity in turn attracts *DNMT1* by binding to PAR chains. However, more recent data show that all three proteins, CTCF, PARP1, and *DNMT1*, independently

interact with each other, indicating putative complex formation even in the absence of DNA (Zampieri et al., 2012). In addition, the presence of PARP1 at the silenced *p16/INK4* promoter in the absence of CTCF (Witcher and Emerson, 2009) argues in favor of the hypothesis that PARP1 independently binds to sites in the vicinity of CTCF target sequences and regulates insulator function in cases where binding of CTCF is weak.

CONCLUDING REMARKS

One major disadvantage in many newer studies tackling PARylation in transcription and chromatin organization is the use of the first-generation low-potency PARP1 inhibitor 3-aminobenzamide, and this in high doses, at which unspecific effects cannot be excluded. There are several more suitable inhibitors available such as olaparib, which has been used also in clinical trials. On the other hand, high doses of PARP inhibitors may be needed to block also unstimulated physiological PARylation. So far, no inhibitor dose-response curves have been published, analyzing especially consequences for chromatin re-organization. Adding to this, even measuring PAR levels in unchallenged cells has not been possible so far.

A yet unsolved obstacle is the experimental discrimination between covalent and non-covalent modification of proteins by poly(ADP-ribose). Addition of chaotropic agents for separation of unbound PAR from proteins may not always be successful, as in some cases interaction is strong enough to resist phenol partitioning (Panzeter et al., 1992). Non-covalent interaction can be tested by using purified PAR and recombinant proteins employing affinity assays, but the question remains if the target is also covalently modified. *In vitro* approaches to solve this problem may yield false positives, as test-tube conditions are unlikely to mirror the situation in a cell. This brings up the next question: what defines a protein respectively a specific amino acid position as substrate for PARylation? No consensus sequence has been determined yet. This leaves room for speculation, for example if only appropriate amino acids exposed in a specific 3D environment are targeted by PARP1, independent of the actual primary sequence. Recently, a MS-based method turned out to be effective in detecting covalent modification of lysines in core histone tails (Messner et al., 2010). Surprisingly, glutamates have not been found as targets for PARylation, despite earlier work defining a specific glutamic acid residue in histone H1 and in H2B as covalently modified by poly(ADP-ribose) (Ogata et al., 1980a,b). This may result from differences in the experimental approaches. Mutational analysis of potential acceptor sites in p53 strongly suggests that at least some glutamates are targeted by PARP1 (Kanai et al., 2007). Nevertheless, using MS techniques seems to be the appropriate step toward unraveling the nature of polymer target sites. In this way, also changes in phosphorylation profiles of PARP1 and PARG have been defined (Gagne et al., 2009).

Another problem arises from the combination of DNA-damage dependent stimulation and activity-related chromatin-modulating properties within one enzyme. To monitor the interaction between proteins and DNA, the method of choice is chromatin immunoprecipitation (ChIP). The sample processing includes crosslinking of proteins to DNA by administering low concentrations (about 1%) of formaldehyde to cells for a short time, usually 10 min. We proved now in a recent publication, that this procedure induces DNA strand-breaks and damage signaling itself, as detected by massive increase in PARylation and phosphorylation of H2AX (Beneke et al., 2012). This impacted on the efficiency of immunoprecipitation as suppression of both γH2AX formation and PARylation, or even PARylation alone changed the obtained results. The observed reduction in ChIP yields was specifically dependent on the monitored combination of promoter and protein. Thus, data obtained so far may be only the tip of the iceberg, as more subtle changes could be blurred by ChIP-induced DNA breaks and resulting damage signaling.

ACKNOWLEDGMENTS

The author wishes to thank Prof. Alexander Bürkle and Prof. Felix Althaus for their support.

REFERENCES

Adamietz, P., and Rudolph, A. (1984). ADP-ribosylation of nuclear proteins *in vivo*. Identification of histone H2B as a major acceptor for mono- and poly(ADP-ribose) in dimethyl sulfate-treated hepatoma AH 7974 cells. *J. Biol. Chem.* 259, 6841–6846.

Adelfalk, C., Kontou, M., Hirsch-Kauffmann, M., and Schweiger, M. (2003). Physical and functional interaction of the Werner syndrome protein with poly-ADP ribosyl transferase. *FEBS Lett.* 554, 55–58.

Aguilar-Quesada, R., Munoz-Gamez, J. A., Martin-Oliva, D., Peralta, A., Valenzuela, M. T., Matinez-Romero, R., Quiles-Perez, R., Menissier-de Murcia, J., de Murcia, G., Ruiz de Almodovar, M., and Oliver, F. J. (2007). Interaction between ATM and PARP-1 in response to DNA damage and sensitization of ATM deficient cells through PARP inhibition. *BMC Mol. Biol.* 8, 29.

Ahel, I., Ahel, D., Matsusaka, T., Clark, A. J., Pines, J., Boulton, S. J., and West, S. C. (2008). Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature* 451, 81–85.

Ahel, D., Horejsi, Z., Wiechens, N., Polo, S. E., Garcia-Wilson, E., Ahel, I., Flynn, H., Skehel, M., West, S. C., Jackson, S. P., Owen-Hughes, T., and Boulton, S. J. (2009). Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* 325, 1240–1243.

Akiyama, T., Takasawa, S., Nata, K., Kobayashi, S., Abe, M., Shervani, N. J., Ikeda, T., Nakagawa, K., Unno, M., Matsuno, S., and Okamoto, H. (2001). Activation of Reg gene, a gene for insulin-producing beta - cell regeneration: poly(ADP-ribose) polymerase binds Reg promoter and regulates the transcription by autopoly(ADP-ribosyl)ation. *Proc. Natl. Acad. Sci. U.S.A.* 98, 48–53.

Altmeyer, M., Messner, S., Hassa, P. O., Fey, M., and Hottiger, M. O. (2009). Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. *Nucleic Acids Res.* 37, 3723–3738.

Alvarez-Gonzalez, R., and Althaus, F. R. (1989). Poly(ADP-ribose) catabolism in mammalian cells exposed to DNA-damaging agents. *Mutat. Res.* 218, 67–74.

Assenmacher, N., and Hopfner, K. P. (2004). MRE11/RAD50/NBS1, complex activities. *Chromosoma* 113, 157–166.

Aubin, R. J., Dam, V. T., Miclette, J., Brousseau, Y., Huletsky, A., and Poirier, G. G. (1982). Hyper(ADP-ribosyl)ation of histone H1. *Can. J. Biochem.* 60, 1085–1094.

Beneke, S., Cohausz, O., Malanga, M., Boukamp, P., Althaus, F., and Bürkle, A. (2008). Rapid regulation of telomere length is mediated by poly(ADP-ribose) polymerase-1. *Nucleic Acids Res.* 36, 6309–6317.

Beneke, S., Meyer, K., Holtz, A., Huttner, K., and Burkle, A. (2012). Chromatin composition is changed by poly(ADP-ribosyl)ation during chromatin immunoprecipitation. *PLoS ONE* 7:e32914. doi: 10.1371/journal.pone.0032914

Bonicalzi, M. E., Haince, J. F., Droit, A., and Poirier, G. G. (2005). Regulation of poly(ADP-ribose) metabolism by poly(ADP-ribose)

glycohydrolase: where and when? *Cell Mol. Life Sci.* 62, 739–750.

Brosh, R. M. Jr., von Kobbe, C., Sommers, J. A., Karmakar, P., Opresko, P. L., Piotrowski, J., Dianova, I., Dianov, G. L., and Bohr, V. A. (2001). Werner syndrome protein interacts with human flap endonuclease 1 and stimulates its cleavage activity. *EMBO J.* 20, 5791–5801.

Caiafa, P., Guastafierro, T., and Zampieri, M. (2009). Epigenetics: poly(ADP-ribosyl)ation of PARP-1 regulates genomic methylation patterns. *FASEB J.* 23, 672–678.

Caiafa, P., and Zlatanova, J. (2009). CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *J. Cell. Physiol.* 219, 265–270.

Chambon, P., Weill, J. D., Doly, J., Strosser, M. T., and Mandel, P. (1966). On the formation of a novel adenylic compound by enzymatic extracts of liver nuclei. *Biochem. Biophys. Res. Commun.* 25, 638–643.

Chambon, P., Weill, J. D., and Mandel, P. (1963). Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. *Biochem. Biophys. Res. Commun.* 11, 39–43.

Chang, W. J., and Alvarez-Gonzalez, R. (2001). The sequence-specific DNA binding of NF-kappa B is reversibly regulated by the automodification reaction of poly (ADP-ribose) polymerase 1. *J. Biol. Chem.* 276, 47664–47670.

Cherney, B. W., McBride, O. W., Chen, D. F., Alkhatib, H., Bhatia, K., Hensley, P., and Smulson, M. E. (1987). cDNA sequence, protein structure, and chromosomal location of the human gene for poly(ADP-ribose) polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8370–8374.

Cohen-Armon, M. (2007). PARP-1 activation in the ERK signaling pathway. *Trends Pharmacol. Sci.* 28, 556–560.

Davies, M. I., Halldorsson, H., Nduka, N., Shall, S., and Skidmore, C. J. (1978). The involvement of poly(adenosine diphosphate-ribose) in deoxyribonucleic acid repair. *Biochem. Soc. Trans.* 6, 1056–1057.

de Capoa, A., Febbo, F. R., Giovannelli, F., Niveleau, A., Zardo, G., Marenzi, S., and Caiafa, P. (1999). Reduced levels of poly(ADP-ribosyl)ation result in chromatin compaction and hypermethylation as shown by cell-by-cell computer-assisted quantitative analysis. *FASEB J.* 13, 89–93.

De La Rosa-Velazquez, I. A., Rincon-Arano, H., Benitez-Bribiesca, L., and Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res.* 67, 2577–2585.

Docquier, F., Kita, G. X., Farrar, D., Jat, P., O'Hare, M., Chernukhin, I., Gretton, S., Mandal, A., Alldridge, L., and Klenova, E. (2009). Decreased poly(ADP-ribosyl)ation of CTCF, a transcription factor, is associated with breast cancer phenotype and cell proliferation. *Clin. Cancer Res.* 15, 5762–5771.

Durkacz, B. W., Omidiji, O., Gray, D. A., and Shall, S. (1980). (ADP-ribose)n participates in DNA excision repair. *Nature* 283, 593–596.

El-Khamisy, S. F., Masutani, M., Suzuki, H., and Caldecott, K. W. (2003). A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Res.* 31, 5526–5533.

Eustermann, S., Brockmann, C., Mehrotra, P. V., Yang, J. C., Loakes, D., West, S. C., Ahel, I., and Neuhaus, D. (2010). Solution structures of the two PBZ domains from human APLF and their interaction with poly(ADP-ribose). *Nat. Struct. Mol. Biol.* 17, 241–243.

Fahrer, J., Kranaster, R., Altmeyer, M., Marx, A., and Bürkle, A. (2007). Quantitative analysis of the binding affinity of poly(ADP-ribose) to specific binding proteins as a function of chain length. *Nucleic Acids Res.* 35, e143.

Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A., and Klenova, E. (2010). Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosyl)ation. *Mol. Cell. Biol.* 30, 1199–1216.

Farzaneh, F., Zalin, R., Brill, D., and Shall, S. (1982). DNA strand breaks and ADP-ribosyl transferase activation during cell differentiation. *Nature* 300, 362–366.

Gagne, J. P., Moreel, X., Gagne, P., Labelle, Y., Droit, A., Chevalier-Pare, M., Bourassa, S., McDonald, D., Hendzel, M. J., Prigent, C., and Poirier, G. G. (2009). Proteomic investigation of phosphorylation sites in poly(ADP-ribose) polymerase-1 and poly(ADP-ribose) glycohydrolase. *J. Proteome Res.* 8, 1014–1029.

Gagne, J. P., Pic, E., Isabelle, M., Krietsch, J., Ethier, C., Paquet, E., Kelly, I., Boutin, M., Moon, K. M., Foster, L. J., and Poirier, G. G. (2012). Quantitative proteomics profiling of the poly(ADP-ribose)-related response to genotoxic stress. *Nucleic Acids Res.* doi: 10.1093/nar/gks486 [Epub ahead of print].

Gamble, M. J., and Fisher, R. P. (2007). SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery. *Nat. Struct. Mol. Biol.* 14, 548–555.

Gombert, W. M., and Krumm, A. (2009). Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. *PLoS ONE* 4:e6109. doi: 10.1371/journal.pone.0006109

Gottschalk, A. J., Timinszky, G., Kong, S. E., Jin, J., Cai, Y., Swanson, S. K., Washburn, M. P., Florens, L., Ladurner, A. G., Conaway, J. W., and Conaway, R. C. (2009). Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13770–13774.

Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L., and Caiafa, P. (2008). CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J. Biol. Chem.* 283, 21873–21880.

Guerrero, P. A., and Maggert, K. A. (2011). The CCCTC-binding factor (CTCF) of Drosophila contributes to the regulation of the ribosomal DNA and nucleolar stability. *PLoS ONE* 6:e16401. doi: 10.1371/journal. pone. 0016401

Haince, J. F., Kozlov, S., Dawson, V. L., Dawson, T. M., Hendzel, M. J., Lavin, M. F., and Poirier, G. G. (2007). Ataxia telangiectasia mutated (ATM) signaling network is modulated by a novel poly(ADP-ribose)-dependent pathway in the early response to DNA-damaging agents. *J. Biol. Chem.* 282, 16441–16453.

Haince, J. F., McDonald, D., Rodrigue, A., Dery, U., Masson, J. Y., Hendzel, M. J., and Poirier, G. G. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. *J. Biol. Chem.* 283, 1197–1208.

Hassa, P. O., Covic, M., Hasan, S., Imhof, R., and Hottiger, M. O. (2001). The enzymatic and DNA binding activity of PARP-1 are not required for NF-kappa B coactivator function. *J. Biol. Chem.* 276, 45588–45597.

Hassa, P. O., Haenni, S. S., Buerki, C., Meier, N. I., Lane, W. S., Owen, H., Gersbach, M., Imhof, R., and Hottiger, M. O. (2005). Acetylation of poly(ADP-ribose) polymerase-1 by p300/CREB-binding protein regulates coactivation of NF-kappaB-dependent transcription. *J. Biol. Chem.* 280, 40450–40464.

Hassa, P. O., and Hottiger, M. O. (1999). A role of poly (ADP-ribose) polymerase in NF-kappaB transcriptional activation. *Biol. Chem.* 380, 953–959.

Hori, T. (1981). High incidence of sister chromatid exchanges and chromatid interchanges in the conditions of lowered activity of poly(ADP-ribose)polymerase. *Biochem. Biophys. Res. Commun.* 102, 38–45.

Hossain, M. B., Ji, P., Anish, R., Jacobson, R. H., and Takada, S. (2009). Poly(ADP-ribose) polymerase 1 interacts with nuclear respiratory factor 1 (NRF-1) and plays a role in NRF-1 transcriptional regulation. *J. Biol. Chem.* 284, 8621–8632.

Hottiger, M. O., Hassa, P. O., Luscher, B., Schuler, H., and Koch-Nolte, F. (2010). Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* 35, 208–219.

Jorgensen, T. J., Chen, K., Chasovskikh, S., Roy, R., Dritschilo, A., and Uren, A. (2009). Binding kinetics and activity of human poly(ADP-ribose) polymerase-1 on oligo-deoxyribonucleotide substrates. *J. Mol. Recognit.* 22, 446–452.

Ju, B. G., Lunyak, V. V., Perissi, V., Garcia-Bassets, I., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* 312, 1798–1802.

Ju, B. G., Solum, D., Song, E. J., Lee, K. J., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (2004). Activating the PARP-1 sensor component of the groucho/ TLE1 corepressor complex mediates a CaMKinase IIdelta-dependent neurogenic gene activation pathway. *Cell* 119, 815–829.

Kanai, M., Hanashiro, K., Kim, S. H., Hanai, S., Boulares, A. H., Miwa, M., and Fukasawa, K. (2007). Inhibition of Crm1-p53 interaction and nuclear export of p53 by poly(ADP-ribosyl)ation. *Nat. Cell Biol.* 9, 1175–1183.

Karras, G. I., Kustatscher, G., Buhecha, H. R., Allen, M. D., Pugieux, C., Sait, F., Bycroft, M., and Ladurner,

A. G. (2005). The macro domain is an ADP-ribose binding module. *EMBO J.* 24, 1911–1920.

Kashima, L., Idogawa, M., Mita, H., Shitashige, M., Yamada, T., Ogi, K., Suzuki, H., Toyota, M., Ariga, H., Sasaki, Y., and Tokino, T. (2012). CHFR protein regulates mitotic checkpoint by targeting PARP-1 protein for ubiquitination and degradation. *J. Biol. Chem.* 287, 12975–12984.

Kashyap, V., Rezende, N. C., Scotland, K. B., Shaffer, S. M., Persson, J. L., Gudas, L. J., and Mongan, N. P. (2009). Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem Cells Dev.* 18, 1093–1108.

Kauppinen, T. M., Chan, W. Y., Suh, S. W., Wiggins, A. K., Huang, E. J., and Swanson, R. A. (2006). Direct phosphorylation and regulation of poly(ADP-ribose) polymerase-1 by extracellular signal-regulated kinases 1/2. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7136–7141.

Khanna, K. K., Lavin, M. F., Jackson, S. P., and Mulhern, T. D. (2001). ATM, a central controller of cellular responses to DNA damage. *Cell Death Differ.* 8, 1052–1065.

Klenova, E., and Ohlsson, R. (2005). Poly(ADP-ribosyl)ation and epigenetics. Is CTCF PARt of the plot? *Cell Cycle* 4, 96–101.

Kraus, W. L. (2008). Transcriptional control by PARP-1, chromatin modulation, enhancer-binding, coregulation, and insulation. *Curr. Opin. Cell Biol.* 20, 294–302.

Kreimeyer, A., Wielckens, K., Adamietz, P., and Hilz, H. (1984). DNA repair-associated ADP-ribosylation *in vivo*. Modification of histone H1 differs from that of the principal acceptor proteins. *J. Biol. Chem.* 259, 890–896.

Krishnakumar, R., Gamble, M. J., Frizzell, K. M., Berrocal, J. G., Kininis, M., and Kraus, W. L. (2008). Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science* 319, 819–821.

Lai, Y. S., Chang, C. W., Pawlik, K. M., Zhou, D., Renfrow, M. B., and Townes, T. M. (2012). SRY (sex determining region Y)-box2 (Sox2)/poly ADP-ribose polymerase 1 (Parp1) complexes regulate pluripotency. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3772–3777.

Langelier, M. F., Planck, J. L., Roy, S., and Pascal, J. M. (2012). Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science* 336, 728–732.

Lapucci, A., Pittelli, M., Rapizzi, E., Felici, R., Moroni, F., and Chiarugi, A. (2011). Poly(ADP-ribose) polymerase-1 Is a nuclear epigenetic regulator of mitochondrial DNA repair and transcription. *Mol. Pharmacol.* 79, 932–940.

Leppard, J. B., Dong, Z., Mackey, Z. B., and Tomkinson, A. E. (2003). Physical and functional interaction between DNA ligase IIIalpha and poly(ADP-Ribose) polymerase 1 in DNA single-strand break repair. *Mol. Cell. Biol.* 23, 5919–5927.

Li, B., Navarro, S., Kasahara, N., and Comai, L. (2004). Identification and biochemical characterization of a Werner's syndrome protein complex with Ku70/80 and poly(ADP-ribose) polymerase-1. *J. Biol. Chem.* 279, 13659–13667.

Lin, W., Ame, J. C., Aboul-Ela, N., Jacobson, E. L., and Jacobson, M. K. (1997). Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* 272, 11895–11901.

Li, Y., Oh, H. J., and Lau, Y. F. (2006). The poly(ADP-ribose) polymerase 1 interacts with Sry and modulates its biological functions. *Mol. Cell. Endocrinol.* 257–258, 35–46.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., and Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 1743–1753.

Lonskaya, I., Potaman, V. N., Shlyakhtenko, L. S., Oussatcheva, E. A., Lyubchenko, Y. L., and Soldatenkov, V. A. (2005). Regulation of poly(ADP-ribose) polymerase-1 by DNA structure-specific binding. *J. Biol. Chem.* 280, 17076–17083.

Malanga, M., and Althaus, F. R. (2004). Poly(ADP-ribose) reactivates stalled DNA topoisomerase I and Induces DNA strand break resealing. *J. Biol. Chem.* 279, 5244–5248.

Martin, N., Schwamborn, K., Schreiber, V., Werner, A., Guillier, C., Zhang, X. D., Bischof, O., Seeler, J. S., and Dejean, A. (2009). PARP-1 transcriptional activity is regulated by sumoylation upon heat shock. *EMBO J.* 28, 3534–3548.

Masson, M., Menissier-de Murcia, J., Mattei, M. G., de Murcia, G., and Niedergang, C. P. (1997). Poly(ADP-ribose) polymerase interacts with a novel human ubiquitin conjugating enzyme: hUbc9. *Gene* 190, 287–296.

Mathis, G., and Althaus, F. R. (1987). Release of core DNA from nucleosomal core particles following (ADP-ribose)n-modification *in vitro*. *Biochem. Biophys. Res. Commun.* 143, 1049–1054.

Mehrotra, P. V., Ahel, D., Ryan, D. P., Weston, R., Wiechens, N., Kraehenbuehl, R., Owen-Hughes, T., and Ahel, I. (2011). DNA repair factor APLF is a histone chaperone. *Mol. Cell* 41, 46–55.

Meisterernst, M., Stelzer, G., and Roeder, R. G. (1997). Poly(ADP-ribose) polymerase enhances activator-dependent transcription *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2261–2265.

Mendoza-Alvarez, H., and Alvarez-Gonzalez, R. (1993). Poly(ADP-ribose) polymerase is a catalytic dimer and the automodification reaction is intermolecular. *J. Biol. Chem.* 268, 22575–22580.

Mendoza-Alvarez, H., and Alvarez-Gonzalez, R. (2001). Regulation of p53 sequence-specific DNA-binding by covalent poly(ADP-ribosyl)ation. *J. Biol. Chem.* 276, 36425–36430.

Messner, S., Altmeyer, M., Zhao, H., Pozivil, A., Roschitzki, B., Gehrig, P., Rutishauser, D., Huang, D., Caflisch, A., and Hottiger, M. O. (2010). PARP1 ADP-ribosylates lysine residues of the core histone tails. *Nucleic Acids Res.* 38, 6350–6362.

Messner, S., Schuermann, D., Altmeyer, M., Kassner, I., Schmidt, D., Schar, P., Muller, S., and Hottiger, M. O. (2009). Sumoylation of poly(ADP-ribose) polymerase 1 inhibits its acetylation and restrains transcriptional coactivator function. *FASEB J.* 23, 3978–3989.

Meyer-Ficca, M. L., Ihara, M., Lonchar, J. D., Meistrich, M. L., Austin, C. A., Min, W., Wang, Z. Q., and Meyer, R. G. (2011a). Poly(ADP-ribose) metabolism is essential for proper nucleoprotein exchange during mouse spermiogenesis. *Biol. Reprod.* 84, 218–228.

Meyer-Ficca, M. L., Lonchar, J. D., Ihara, M., Meistrich, M. L., Austin, C. A., and Meyer, R. G. (2011b). Poly(ADP-ribose) polymerases PARP1 and PARP2 modulate topoisomerase II beta (TOP2B) function during chromatin condensation in mouse spermiogenesis. *Biol. Reprod.* 84, 900–909.

Meyer-Ficca, M. L., Meyer, R. G., Coyle, D. L., Jacobson, E. L., and Jacobson, M. K. (2004). Human poly(ADP-ribose) glycohydrolase is expressed in alternative splice variants yielding isoforms that localize to different cell compartments. *Exp. Cell Res.* 297, 521–532.

Meyer, R., Müller, M., Beneke, S., Küpper, J. H., and Bürkle, A. (2000). Negative regulation of alkylation-induced sister-chromatid exchange by poly(ADP-ribose) polymerase-1 activity. *Int. J. Cancer.* 88, 351–355.

Miller, E. G. (1975a). Stimulation of nuclear poly (adenosine diphosphate-ribose) polymerase activity from HeLa cells by endonucleases. *Biochim. Biophys. Acta* 395, 191–200.

Miller, E. G. (1975b). Effect of deoxyribonuclease I on the number and length of chains of poly(ADP-ribose) synthesized, *in vitro*. *Biochem. Biophys. Res. Commun.* 66, 280–286.

Morgan, W. F., and Cleaver, J. E. (1983). Effect of 3-aminobenzamide on the rate of ligation during repair of alkylated DNA in human fibroblasts. *Cancer Res.* 43, 3104–3107.

Naegeli, H., and Althaus, F. R. (1991). Regulation of poly(ADP-ribose) polymerase. Histone-specific adaptations of reaction products. *J. Biol. Chem.* 266, 10596–10601.

Nishizuka, Y., Ueda, K., Nakazawa, K., and Hayaishi, O. (1967). Studies on the polymer of adenosine diphosphate ribose. I. Enzymic formation from nicotinamide adenine dinuclotide in mammalian nuclei. *J. Biol. Chem.* 242, 3164–3171.

Oei, S. L., Herzog, H., Hirsch-Kauffmann, M., Schneider, R., Auer, B., and Schweiger, M. (1994). Transcriptional regulation and autoregulation of the human gene for ADP-ribosyltransferase. *Mol. Cell Biochem.* 138, 99–104.

Oei, S. L., Griesenbeck, J., Schweiger, M., Babich, V., Kropotov, A., and Tomilin, N. (1997). Interaction of the transcription factor YY1 with human poly(ADP-ribosyl) transferase. *Biochem. Biophys. Res. Commun.* 240, 108–111.

Oei, S. L., Griesenbeck, J., Schweiger, M., and Ziegler, M. (1998a). Regulation of RNA polymerase II-dependent transcription by poly(ADP-ribosyl)ation of transcription factors. *J. Biol. Chem.* 273, 31644–31647.

Oei, S. L., Griesenbeck, J., Ziegler, M., and Schweiger, M. (1998b). A novel function of poly(ADP-ribosyl)ation: silencing of RNA

polymerase II-dependent transcription. *Biochemistry* 37, 1465–1469.

Ogata, N., Ueda, K., and Hayaishi, O. (1980a). ADP-ribosylation of histone H2B. Identification of glutamic acid residue 2 as the modification site. *J. Biol. Chem.* 255, 7610–7615.

Ogata, N., Ueda, K., Kagamiyama, H., and Hayaishi, O. (1980b). ADP-ribosylation of histone H1. Identification of glutamic acid residues 2 14, and the COOH-terminal lysine residue as modification sites. *J. Biol. Chem.* 255, 7616–7620.

Ogata, N., Ueda, K., Kawaichi, M., and Hayaishi, O. (1981). Poly(ADP-ribose) synthetase, a main acceptor of poly(ADP-ribose) in isolated nuclei. *J. Biol. Chem.* 256, 4135–4137.

Otsuka, F., Ohno, H., Enomoto, T., Hanaoka, F., and Yamada, M. (1983). Independent induction of sister-chromatid exchanges by 3-aminobenzamide and ultraviolet radiation in HeLa cells. *Mutat. Res.* 107, 289–296.

Ouararhni, K., Hadj-Slimane, R., Ait-Si-Ali, S., Robin, P., Mietton, F., Harel-Bellan, A., Dimitrov, S., and Hamiche, A. (2006). The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. *Genes Dev.* 20, 3324–3336.

Panzeter, P. L., Realini, C. A., and Althaus, F. R. (1992). Noncovalent interactions of poly(adenosine diphosphate ribose) with histones. *Biochemistry* 31, 1379–1385.

Park, S. D., Kim, C. G., and Kim, M. G. (1983). Inhibitors of poly(ADP-ribose) polymerase enhance DNA strand breaks, excision repair, and sister chromatid exchanges induced by alkylating agents. *Environ. Mutagen.* 5, 515–525.

Pekala, P. H., and Moss, J. (1983). 3T3-L1 preadipocyte differentiation and poly(ADP-ribose) synthetase. *Mol. Cell. Biochem.* 53-54, 221–232.

Petesch, S. J., and Lis, J. T. (2008). Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* 134, 74–84.

Petesch, S. J., and Lis, J. T. (2012). Activator-induced spread of poly(ADP-ribose) polymerase promotes nucleosome loss at Hsp70. *Mol. Cell* 45, 64–74.

Pleschke, J. M., Kleczkowska, H. E., Strohm, M., and Althaus, F. R. (2000). Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins. *J. Biol. Chem.* 275, 40974–40980.

Poirier, G. G., de Murcia, G., Jongstra-Bilen, J., Niedergang, C., and Mandel, P. (1982). Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure. *Proc. Natl. Acad. Sci. U.S.A.* 79, 3423–3427.

Potaman, V. N., Shlyakhtenko, L. S., Oussatcheva, E. A., Lyubchenko, Y. L., and Soldatenkov, V. A. (2005). Specific binding of poly(ADP-ribose) polymerase-1 to cruciform hairpins. *J. Mol. Biol.* 348, 609–615.

Prasad, R., Lavrik, O. I., Kim, S. J., Kedar, P., Yang, X. P., Vande Berg, B. J., and Wilson, S. H. (2001). DNA polymerase beta -mediated long patch base excision repair. Poly(ADP-ribose)polymerase-1stimulates strand displacement DNA synthesis. *J. Biol. Chem.* 276, 32411–32414.

Qi, C. F., Martensson, A., Mattioli, M., Dalla-Favera, R., Lobanenkov, V. V., and Morse, H. C. 3rd. (2003). CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. *Proc. Natl. Acad. Sci. U.S.A.* 100, 633–638.

Rawling, J. M., and Alvarez-Gonzalez, R. (1997). TFIIF, a basal eukaryotic transcription factor, is a substrate for poly(ADP-ribosyl)ation. *Biochem. J.* 324(Pt 1), 249–253.

Ray Chaudhuri, A., Hashimoto, Y., Herrador, R., Neelsen, K. J., Fachinetti, D., Bermejo, R., Cocito, A., Costanzo, V., and Lopes, M. (2012). Topoisomerase I poisoning results in PARP-mediated replication fork reversal. *Nat. Struct. Mol. Biol.* 19, 417–423.

Reale, A., Matteis, G. D., Galleazzi, G., Zampieri, M., and Caiafa, P. (2005). Modulation of *DNMT1* activity by ADP-ribose polymers. *Oncogene* 24, 13–19.

Realini, C. A., and Althaus, F. R. (1992). Histone shuttling by poly(ADP-ribosylation). *J. Biol. Chem.* 267, 18858–18865.

Reeder, R. H., Ueda, K., Honjo, T., Nishizuka, Y., and Hayaishi, O. (1967). Studies on the polymer of adenosine diphosphate ribose. II. Characterization of the polymer. *J. Biol. Chem.* 242, 3172–3179.

Ruscetti, T., Lehnert, B. E., Halbrook, J., Le Trong, H., Hoekstra, M. F., Chen, D. J., and Peterson, S. R. (1998). Stimulation of the DNA-dependent protein kinase by poly(ADP-ribose) polymerase. *J. Biol. Chem.* 273, 14461–14467.

Ryu, H., Al-Ani, G., Deckert, K., Kirkpatrick, D., Gygi, S. P., Dasso, M., and Azuma, Y. (2010). PIASy mediates SUMO-2/3 conjugation of poly(ADP-ribose) polymerase 1 (PARP1) on mitotic chromosomes. *J. Biol. Chem.* 285, 14415–14423.

Satoh, M. S., and Lindahl, T. (1992). Role of poly(ADP-ribose) formation in DNA repair. *Nature* 356, 356–358.

Satoh, M. S., Poirier, G. G., and Lindahl, T. (1994). Dual function for poly(ADP-ribose) synthesis in response to DNA strand breakage. *Biochemistry* 33, 7099–7106.

Schweiger, M., Oei, S. L., Herzog, H., Menardi, C., Schneider, R., Auer, B., and Hirsch-Kauffmann, M. (1995). Regulation of the human poly(ADP-ribosyl) transferase promoter via alternative DNA racket structures. *Biochimie* 77, 480–485.

Sharma, S., Sommers, J. A., Driscoll, H. C., Uzdilla, L., Wilson, T. M., and Brosh, R. M. Jr. (2003). The exonucleolytic and endonucleolytic cleavage activities of human exonuclease 1 are stimulated by an interaction with the carboxyl-terminal region of the Werner syndrome protein. *J. Biol. Chem.* 278, 23487–23496.

Sharma, S., Otterlei, M., Sommers, J. A., Driscoll, H. C., Dianov, G. L., Kao, H. I., Bambara, R. A., and Brosh, R. M. Jr. (2004). WRN helicase and FEN-1 form a complex upon replication arrest and together process branchmigrating DNA structures associated with the replication fork. *Mol. Biol. Cell* 15, 734–750.

Shi, L., Ko, S., Kim, S., Echchgadda, I., Oh, T. S., Song, C. S., and Chatterjee, B. (2008). Loss of androgen receptor in aging and oxidative stress through Myb protooncoprotein-regulated reciprocal chromatin dynamics of p53 and poly(ADP-ribose) polymerase PARP-1. *J. Biol. Chem.* 283, 36474–36485.

Simbulan-Rosenthal, C. M., Rosenthal, D. S., Boulares, A. H., Hickey, R. J., Malkas, L. H., Coll, J. M., and Smulson, M. E. (1998). Regulation of the expression or recruitment of components of the DNA synthesome by poly(ADP-ribose) polymerase. *Biochemistry* 37, 9363–9370.

Simbulan-Rosenthal, C. M., Rosenthal, D. S., Luo, R., Samara, R., Espinoza, L. A., Hassa, P. O., Hottiger, M. O., and Smulson, M. E. (2003). PARP-1 binds E2F-1 independently of its DNA binding and catalytic domains, and acts as a novel coactivator of E2F-1-mediated transcription during re-entry of quiescent cells into S phase. *Oncogene* 22, 8460–8471.

Simbulan-Rosenthal, C. M., Rosenthal, D. S., Luo, R., and Smulson, M. E. (1999). Poly(ADP-ribose) polymerase upregulates E2F-1 promoter activity and DNA pol alpha expression during early S phase. *Oncogene* 18, 5015–5023.

Sims, J. L., Berger, S. J., and Berger, N. A. (1983). Poly(ADP-ribose) polymerase inhibitors preserve nicotinamide adenine dinucleotide and adenosine 5′-triphosphate pools in DNA-damaged cells: mechanism of stimulation of unscheduled DNA synthesis. *Biochemistry* 22, 5188–5194.

Slattery, E., Dignam, J. D., Matsui, T., and Roeder, R. G. (1983). Purification and analysis of a factor which suppresses nick-induced transcription by RNA polymerase II and its identity with poly(ADP-ribose) polymerase. *J. Biol. Chem.* 258, 5955–5959.

Soldatenkov, V. A., Chasovskikh, S., Potaman, V. N., Trofimova, I., Smulson, M. E., and Dritschilo, A. (2002). Transcriptional repression by binding of poly(ADP-ribose) polymerase to promoter sequences. *J. Biol. Chem.* 277, 665–670.

Suzuki, H., Uchida, K., Shima, H., Sato, T., Okamoto, T., Kimura, T., and Miwa, M. (1987). Molecular cloning of cDNA for human poly(ADP-ribose) polymerase and expression of its gene during HL-60 cell differentiation. *Biochem. Biophys. Res. Commun.* 146, 403–409.

Thomas, M. C., and Chiang, C. M. (2006). The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* 41, 105–178.

Torrano, V., Navascues, J., Docquier, F., Zhang, R., Burke, L. J., Chernukhin, I., Farrar, D., Leon, J., Berciano, M. T., Renkawitz, R., Klenova, E., Lafarga, M., and Delgado, M. D. (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *J. Cell Sci.* 119(Pt 9), 1746–1759.

Tsopanakis, C., McLaren, E. A., and Shall, S. (1976). Purification of poly(adenosine diphosphate ribose) polymerase from pig thymus. *Biochem. Soc. Trans.* 4, 774–777.

Tulin, A., and Spradling, A. (2003). Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. *Science* 299, 560–562.

Tulin, A., Stewart, D., and Spradling, A. C. (2002). The *Drosophila heterochromatic* gene encoding

poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. *Genes Dev.* 16, 2108–2119.

Ueda, K., Oka, J., Naruniya, S., Miyakawa, N., and Hayaishi, O. (1972). Poly ADP-ribose glycohydrolase from rat liver nuclei, a novel enzyme degrading the polymer. *Biochem. Biophys. Res. Commun.* 46, 516–523.

Ueda, K., Omachi, A., Kawaichi, M., and Hayaishi, O. (1975). Natural occurrence of poly(ADP-ribosyl) histones in rat liver. *Proc. Natl. Acad. Sci. U.S.A.* 72, 205–209.

Veuger, S. J., Curtin, N. J., Smith, G. C., and Durkacz, B. W. (2004). Effects of novel inhibitors of poly(ADP-ribose) polymerase-1 and the DNA-dependent protein kinase on enzyme activities and DNA repair. *Oncogene* 23, 7322–7329.

Vidakovic, M., Gluch, A., Qiao, J., Oumard, A., Frisch, M., Poznanovic, G., and Bode, J. (2009). PARP-1 expression in the mouse is controlled by an autoregulatory loop: PARP-1 binding to an upstream S/MAR element and to a novel recognition motif in its promoter suppresses transcription. *J. Mol. Biol.* 388, 730–750.

von Kobbe, C., Harrigan, J. A., May, A., Opresko, P. L., Dawut, L., Cheng, W. H., and Bohr, V. A. (2003). Central role for the Werner syndrome protein/poly(ADP-ribose) polymerase 1 complex in the poly(ADP-ribosyl)ation pathway after DNA damage. *Mol. Cell Biol.* 23, 8601–8613.

von Kobbe, C., Harrigan, J. A., Schreiber, V., Stiegler, P., Piotrowski, J., Dawut, L., and Bohr, V. A. (2004). Poly(ADP-ribose) polymerase 1 regulates both the exonuclease and helicase activities of the Werner syndrome protein. *Nucleic Acids Res.* 32, 4003–4014.

Walker, J. W., Jijon, H. B., and Madsen, K. L. (2006). AMP-activated protein kinase is a positive regulator of poly(ADP-ribose) polymerase. *Biochem. Biophys. Res. Commun.* 342, 336–341.

Wang, T., Simbulan-Rosenthal, C. M., Smulson, M. E., Chock, P. B., and Yang, D. C. (2008). Polyubiquitylation of PARP-1 through ubiquitin K48 is modulated by activated DNA, NAD+, and dipeptides. *J. Cell. Biochem.* 104, 318–328.

Wesierska-Gadek, J., Schmid, G., and Cerni, C. (1996). ADP-ribosylation of wild-type p53 *in vitro*: binding of p53 protein to specific p53 consensus sequence prevents its modification. *Biochem. Biophys. Res. Commun.* 224, 96–102.

Witcher, M., and Emerson, B. M. (2009). Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol. Cell* 34, 271–284.

Yamamoto, H., and Okamoto, H. (1982). Poly(ADP-ribose) synthetase inhibitors enhance streptozotocin-induced killing of insulinoma cells by inhibiting the repair of DNA strand breaks. *FEBS Lett.* 145, 298–302.

Yoshihara, K., Itaya, A., Tanaka, Y., Ohashi, Y., Ito, K., Teraoka, H., Tsukada, K., Matsukage, A., and Kamiya, T. (1985). Inhibition of DNA polymerase alpha, DNA polymerase beta, terminal deoxynucleotidyl transferase, and DNA ligase II by poly(ADP-ribosyl)ation reaction *in vitro*. *Biochem. Biophys. Res. Commun.* 128, 61–67.

Yusufzai, T. M., and Felsenfeld, G. (2004). The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8620–8624.

Yusufzai, T. M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* 13, 291–298.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, A. P., Lobanenkov, V., Klenova, E., and Ohlsson, R. (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat. Genet.* 36, 1105–1110.

Zahradka, P., and Ebisuzaki, K. (1982). A shuttle mechanism for DNA-protein interactions. The regulation of poly(ADP-ribose) polymerase. *Eur. J. Biochem.* 127, 579–585.

Zampieri, M., Guastafierro, T., Calabrese, R., Ciccarone, F., Bacalini, M. G., Reale, A., Perilli, M., Passananti, C., and Caiafa, P. (2012). ADP-ribose polymers localized on Ctcf-Parp1-*DNMT1* complex prevent methylation of Ctcf target sites. *Biochem. J.* 441, 645–652.

Zampieri, M., Passananti, C., Calabrese, R., Perilli, M., Corbi, N., De Cave, F., Guastafierro, T., Bacalini, M. G., Reale, A., Amicosante, G., Calabrese, L., Zlatanova, J., and Caiafa, P. (2009). Parp1 localizes within the *DNMT1* promoter and protects its unmethylated state by its enzymatic activity. *PLoS ONE* 4:e4717. doi: 10.1371/journal.pone.0004717

Zardo, G., and Caiafa, P. (1998). The unmethylated state of CpG islands in mouse fibroblasts depends on the poly(ADP-ribosyl)ation process. *J. Biol. Chem.* 273, 16517–16520.

Zardo, G., Reale, A., Passananti, C., Pradhan, S., Buontempo, S., De Matteis, G., Adams, R. L., and Caiafa, P. (2002). Inhibition of poly(ADP-ribosyl)ation induces DNA hypermethylation: a possible molecular mechanism. *FASEB J.* 16, 1319–1321.

Conflict of Interest Statement: 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.

Received: 04 July 2012; accepted: 17 August 2012; published online: 03 September 2012.

Citation: Beneke S (2012) Regulation of chromatin structure by poly(ADP-ribosyl)ation. Front. Gene. ***3****:169. doi: 10.3389/fgene.2012.00169*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

REVIEW ARTICLE
published: 12 September 2012
doi: 10.3389/fgene.2012.00171

Diverse developmental disorders from The One Ring: distinct molecular pathways underlie the cohesinopathies

***Julia A. Horsfield*[1]*, *Cristin G. Print*[2] *and* *Maren Mönnich*[1]**

[1] *Department of Pathology, Dunedin School of Medicine, The University of Otago, Dunedin, New Zealand*
[2] *Department of Molecular Medicine and Pathology, School of Medical Sciences, Bioinformatics Institute, The University of Auckland, Auckland, New Zealand*

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Jennifer Gerton, Stowers Institute for Medical Research, USA
Kyoko Yokomori, University of California Irvine, USA

***Correspondence:**
Julia A. Horsfield, Department of Pathology, Dunedin School of Medicine, University of Otago, PO Box 913, Dunedin, New Zealand.
e-mail: julia.horsfield@otago.ac.nz

The multi-subunit protein complex, cohesin, is responsible for sister chromatid cohesion during cell division. The interaction of cohesin with DNA is controlled by a number of additional regulatory proteins. Mutations in cohesin, or its regulators, cause a spectrum of human developmental syndromes known as the "cohesinopathies." Cohesinopathy disorders include Cornelia de Lange Syndrome and Roberts Syndrome. The discovery of novel roles for chromatid cohesion proteins in regulating gene expression led to the idea that cohesinopathies are caused by dysregulation of multiple genes downstream of mutations in cohesion proteins. Consistent with this idea, *Drosophila*, mouse, and zebrafish cohesinopathy models all show altered expression of developmental genes. However, there appears to be incomplete overlap among dysregulated genes downstream of mutations in different components of the cohesion apparatus. This is surprising because mutations in all cohesion proteins would be predicted to affect cohesin's roles in cell division and gene expression in similar ways. Here we review the differences and similarities between genetic pathways downstream of components of the cohesion apparatus, and discuss how such differences might arise, and contribute to the spectrum of cohesinopathy disorders. We propose that mutations in different elements of the cohesion apparatus have distinct developmental outcomes that can be explained by sometimes subtly different molecular effects.

Keywords: cohesin, gene expression regulation, animal models, CdLS, RBS

INTRODUCTION

The cohesin complex and proteins that regulate its interaction with chromatin have multiple roles in cell division, DNA damage repair, gene transcription, and chromosome architecture. Proteins that make up the cohesin complex have been characterized in several model systems (see **Table 1**). The mechanics of cell division has been well researched for decades, and the identity of the chromosome cohesion proteins that hold together sister chromatids after S phase and prior to mitosis has been known for 15 years (Guacci et al., 1997; Michaelis et al., 1997). Consequently, sister chromatid cohesion remains the best-characterized role for the cohesin complex and its regulators.

The first evidence that a transcriptional function existed for chromosome cohesion proteins emerged in 1999, when the *Nipped-B* gene was identified in a genetic screen for modifiers of long-range enhancer-promoter communication regulating *cut* gene expression in the *Drosophila* wing margin (Rollins et al., 1999). Further evidence that gene transcription isone of cohesin's crucial functions unfolded over the following years, and included a role for Scc1 in mating-type silencing in yeast (Lau et al., 2002), a transcriptional co-activation function for SA in mammalian cell lines (Lara-Pezzi et al., 2004), and complex long-range regulation of *cut* gene expression resulting from cohesin and *Nipped-B Drosophila* mutants (Rollins et al., 2004). Interest in the transcription function of cohesion proteins heightened when heterozygous mutations *NIPBL*, the human homolog of *Nipped-B*, were found to cause the human developmental disease, Cornelia de Lange Syndrome (CdLS; OMIM 122470; Krantz et al., 2004; Tonkin et al., 2004). Additional mutations causing CdLS were found in the cohesin subunits SMC1 (Musio et al., 2006; Deardorff et al., 2007) and SMC3 (Deardorff et al., 2007). Furthermore, homozygous mutations in *ESCO2*, which encodes a cohesion acetyltransferase (CoAT; Nasmyth, 2011; Higashi et al., 2012), were found to underlie a second human disorder, Robert's Syndrome (RBS; OMIM 268300; Vega et al., 2005). More recently, mutations in RAD21 have been found to cause a related developmental disorder that partially overlaps with CdLS (Deardorff et al., 2012b).

After the causative genes for CdLS and RBS were found, a flood of new results in vertebrates, from fish (Horsfield et al., 2007; Muto et al., 2011), mouse (Zhang et al., 2007, 2009; Kawauchi et al., 2009), and human cell lines (Liu et al., 2009), supported the notion that these syndromes could be caused by dysregulated expression of multiple developmental genes. This suggested that cohesin-related developmental disorders have related pathologies, and led to use of the term "cohesinopathies" to describe these disorders (Liu and Krantz, 2008; McNairn and Gerton, 2008).

Although the idea that cohesinopathies have a common causal basis in dysregulated gene expression is a popular one, it is clear that the output of gene regulation is different for each disorder. Human syndromes caused by *NIPBL*, *SMC1*, *SMC3*, *RAD21*, and

Table 1 | Nomenclature and function of cohesin subunits and cohesin regulators.

Chromosome cohesion regulator	*S. cerevisiae*	*S. pombe*	*D. melanogaster*	*X. laevis*	*D. rerio*	*H. sapiens*	Function
SMC subunits	Smc1	Psm1	SMC1	smc1a[1]	Smc1al, Smc1a[2]	SMC1A	Core cohesin subunit
				smc1b[1]	Smc1b	SMC1B	Cohesin subunit (meiosis)
	Smc3	Psm3	Cap/SMC3	smc3/cspg6[1]	Smc3	SMC3/CSPG6/ Bamacan	Core cohesin subunit
α-Kleisin subunit	Mcd1/Scc1	Rad21	Vtd/Rad21	rad21/mcd1/nxp1 /scc1[1]	Rad21a, Rad21b[2]	RAD21	Core cohesin subunit
	Rec8	Rec8	C(2)M	rec8	Rec8/zgc: 136888[1,3]	REC8	Cohesin subunit (meiosis)
	–	–	–		Rad21l1	RAD21L1/RAD21L	
Stromalin/SA subunit		Psc3	SA (stromalin)	stag1/sa1	Stag1[1,3]	STAG1/SA1/SCC3A	Cohesin subunit
			SA2 (stromalin-2)	stag2/sa2[1]	Stag2[1,3]	STAG2/SA2/SCC3B	
–		Rec11	–	stag3/sa3[1]	Stag3l3[1]	STAG3/SA3	Cohesin subunit (meiosis)
Interactors of α-kleisin and SA	Pds5	Pds5	Pds5	pds5a	Pds5a/zgc: 66331	PDS5A	Balancing cohesion establishment with cohesin dissociation
				pds5b/as3/aprin[1]	Pds5b[1]	PDS5B/APRIN/AS3	
	?	?	Dmt (Dalmatian)	cdca5/sororin[1]	Cdca5	CDCA5/SORORIN	
	Rad61/Wpl1	Wapl	Wapl	wapal	Wapl/KIAA 0261[1,3]	WAPAL/WAPL	
Kollerin	Scc2	Mis4	Nipped-B	nipbl/scc2/delangin	Nipbla/Scc2a, Nipblb/Scc2b	NIPBL/SCC2/ DELANGIN	Cohesin loading
	Scc4	Ssl3		mau2/scc4[1]	Mau2/zgc: 112338[1]	MAU2/SCC4	
Cohesin acetyl transferase (CoAT)	Eco1/Ctf7	Eso1	Eco/Deco San	esco1	Esco1[1]	ESCO1	Establishment of cohesion
				esco2/rbs/efo2[1]	Esco2	ESCO2	
Cohesin deacetylase (CoDAC)	Hos1	?	?	hdac8	Hdac8	HDAC8	Recycling of cohesin

[1] *Predicted/in silico annotated only.*

[2] *No functional data available.*

[3] *Duplicated (EnsemblZv9, release 68).*

?, protein not yet identified.

ESCO2 mutations share common features but appear to be clinically distinct. Here we revisit the theory that cohesinopathies result from dysregulated gene expression, and raise the question of whether subunits contributing to cohesin or its regulation can interact separately with distinct pathways leading to diverse phenotypic consequences.

OVERVIEW OF COHESIN STRUCTURE AND FUNCTION

The mitotic cohesin complex comprises two structural maintenance of chromosomes (SMC) subunits Smc1 and Smc3, which associate to form a tripartite ring incorporating an α-kleisin subunit, Mcd1/Scc1/Rad21. Smc1 and Smc3 are large rod-shaped proteins that dimerize at one end to form a "hinge" domain, and also interact at the other end via ATP-binding "heads," which in turn interact with the α-kleisin subunit (**Figure 1**). The α-kleisin interacts with additional subunits Scc3/Stromalin (SA), Pds5, and Wapl (Nasmyth, 2011; see **Table 1**). The formation of cohesin subunits into a large ring structure led to the theory that cohesin topologically entraps sister chromatids inside a single ring (Haering et al., 2008). Alternative models have been proposed for how cohesin physically holds two molecules of DNA together (Huang et al., 2005; Zhang et al., 2008b; Skibbens, 2010), although most are not compatible with the single ring theory (reviewed in Nasmyth, 2011).

The many functions of cohesin have been well described in recent reviews (Hirano, 2006; Nasmyth and Haering, 2009; Carretero et al., 2010; Nasmyth, 2011; Rhodes et al., 2011; Mehta et al., 2012). Cohesin turnover, recycling, loading onto chromosomes and residency there is controlled by several other proteins (**Figure 2**; **Table 1**). It was recently proposed that cohesin is loaded and unloaded from chromosomes by a "dual gate" mechanism (Nasmyth, 2011). The cohesin loading complex containing Scc2 (Nipped-B in *Drosophila* and NIPBL in human) and Scc4/MAU2, recently dubbed "kollerin" (Nasmyth, 2011), is

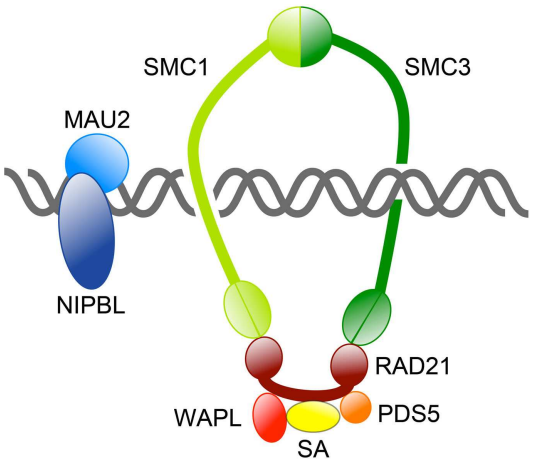


FIGURE 1 | Overview of the cohesin complex and its associated proteins. The cohesin complex consists of four core subunits: SMC1, SMC3, RAD21, and SA. Together these subunits form a large ring capable of topologically encircling DNA strands. Other proteins regulate cohesin's binding to DNA and its residency there. The NIPBL/MAU2 dimer loads cohesin onto DNA, whereas WAPL/PDS5 release cohesin from chromosomes by opening the SMC3-RAD21 interface.

responsible for loading cohesin onto chromosomes in G1 phase in yeast, and telophase in most other organisms. Kollerin directly loads cohesin onto the pre-replication complex (pre-RC) on chromatin *in vitro* in *Xenopus* extracts (Bermudez et al., 2012), indicating that it is likely to be necessary and sufficient for cohesin loading. Kollerin likely facilitates cohesin loading by enabling the transient opening of the Smc1-Smc3 hinge domains (**Figure 1**; Nasmyth, 2011). An opposing unloading activity is mediated by "releasin," a cohesion disestablishment complex containing Pds5 and Wapl that interacts with SA to unlock the cohesin ring (Gandhi et al., 2006; Kueng et al., 2006; Shintomi and Hirano, 2009). Releasin allows exit of DNA via the Smc1-Smc3 head domains by opening the Smc3-kleisin interface. In theory, cohesin snaps onto DNA via opening of the hinge domains, and exits DNA via opening the ring at the opposite end (Nasmyth, 2011).

Once loaded onto chromosomes, cohesin binds DNA with variable modes of stability (Gerlich et al., 2006; Gause et al., 2010) and is mobile, having the ability to translocate along chromosome arms (Lengronne et al., 2004; Hu et al., 2011), or readily detach via interaction with releasin. However during S phase, cohesin becomes stably bound for long enough to fulfill its function in sister chromatid cohesion. Stabilization of cohesin binding happens during the process of DNA replication (Skibbens et al., 1999; Kenna and Skibbens, 2003; Moldovan et al., 2006), and is mediated via acetylation of Smc3 by cohesin acetyl transferase (CoAT; Nasmyth, 2011). The known CoATs for Smc3 are Ctf7/Eco1 (yeast), or Esco1/2 (vertebrates; Skibbens et al., 1999; Ivanov et al., 2002; Hou and Zou, 2005).

CoAT-mediated acetylation of Smc3 generates the cohesive form of cohesin that holds together the sister chromatids from G2 until M phase (Ben-Shahar et al., 2008; Unal et al., 2008; Zhang et al., 2008a). In humans, both ESCO1 and ESCO2 CoATs are necessary for proper sister chromatid cohesion (Hou and Zou, 2005). However, it appears that ESCO2 CoAT is primarily required for cohesion in heterochromatic regions, and RBS patients who lack ESCO2 exhibit heterochromatin repulsion and precocious sister chromatid separation, particularly at centromeric regions (Vega et al., 2005). In human and *Drosophila* (but not yeast), the Sororin protein is additionally required to establish and maintain cohesion (Rankin et al., 2005; Schmitz et al., 2007; Nishiyama et al., 2010).

Once cohesion has been established in G2, cohesion-promoting and cohesin-releasing activities compete during chromosome condensation in prophase. The releasing activity that removes cohesin from chromosomes prevails along chromosome arms in a process known as the "prophase pathway," which involves phosphorylation of SA1/2 by Polo-like kinase (Plk) and Aurora B (Losada et al., 2002; Hauf et al., 2005) and complexing of SA and RAD21 by releasin (Gandhi et al., 2006; Kueng et al., 2006; Shintomi and Hirano, 2009). In the competing "establishment" activity, Sororin and CoAT function to antagonize releasin activity (Rowland et al., 2009; Sutani et al., 2009; Lafont et al., 2010; Nishiyama et al., 2010; Nasmyth, 2011) by a mechanism that also requires Pds5 (Vaur et al., 2012), and the phosphatase Ssu72 promotes cohesion by countering the phosphorylation of SA1/2 (Hauf et al., 2005; Kim et al., 2010b). By metaphase, most cohesin has been removed from chromosome arms, and the remaining, primarily centromeric cohesin, is protected from removal by Shugoshin (Wang and Dai, 2005).

At Anaphase, the remaining cohesin rings are opened, allowing chromosomes to separate (Craig and Choo, 2005). APC-mediated degradation of Securin (Salah and Nasmyth, 2000) releases the protease Separase, which cleaves the Rad21 subunit of cohesin (Waizenegger et al., 2000, 2002; Hornig et al., 2002). After telophase, Smc complexes can be recycled and reloaded onto chromatin. An important requirement for cohesin recycling is deacetylation of Smc3 by the class I histone deacetylase Hos1 (yeast) or HDAC8 (human; Beckouet et al., 2010; Borges et al., 2010; Xiong et al., 2010; Deardorff et al., 2012a). Thus, Smc3 deacetylation by Hos1 opposes Esco2's acetylation activity.

Cohesin has a further important role in DNA double strand break repair (reviewed in Dorsett and Strom, 2012; Wu and Yu, 2012). To effect double strand break repair, the cohesive form of cohesin must be established at the location of the break (Ball and Yokomori, 2008). Stabilization of cohesin at double strand breaks in budding yeast depends on acetylation of the Rad21/Mcd1p subunit by Eco1p, plus antagonism of the releasin complex containing Wpl (Heidinger-Pauli et al., 2009). Cohesin is recruited *de novo* at double strand breaks in G2 phase (Strom et al., 2007) in a Scc2/kollerin-dependent manner (Strom et al., 2004), and in vertebrates, this association also involves another SMC complex: the Smc5/6 complex (Strom and Sjogren, 2007; De Piccoli et al., 2009).

Other molecular events contribute to cohesin function in DSB repair. In budding yeast, it was shown that the phosphorylation of Mcd1p (Rad21) through ATR and Chk1 pathway is important for cohesion and DSB repair (Heidinger-Pauli et al., 2008). In

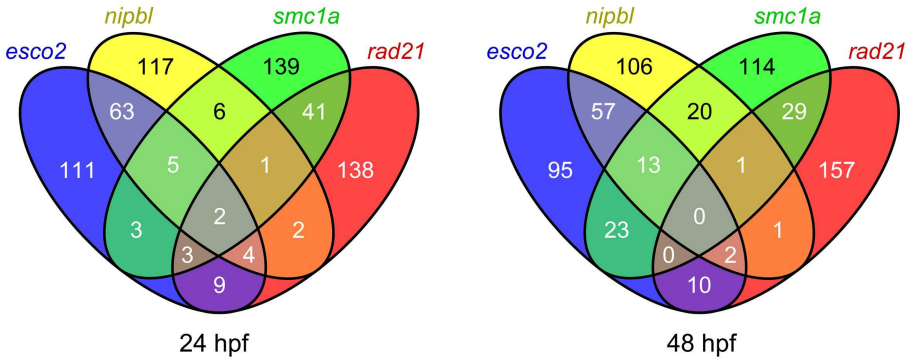


FIGURE 2 | Comparison of the top 200 affected probe sets in zebrafish embryos depleted of different cohesin subunits. Venn diagrams showing the overlap of the top 200 probe sets affected in zebrafish *esco2* and *nipbl* morphants, and *smc1a*^{hi1113a} and *rad21*^{nz171} mutants at 24 and 48 hpf (q value < 0.05).

human cells, cohesive cohesin at DSBs also depends on the pro-establishment activity of Sororin (Schmitz et al., 2007). Cohesin, but not chromosome cohesion, is required for activation of G1, intra-S, and G2–M DNA damage checkpoints (Jessberger, 2009; Watrin and Peters, 2009). In cancer cells, cohesin binding through the genome is reinforced following ionizing radiation (IR), in a process that requires ATM and SMC3 phosphorylation, and SMC3 acetylation by ESCO1. Both ESCO1 and SMC3 acetylation are required for intra-S phase checkpoint and cellular survival after IR (Kim et al., 2010a).

COHESIN AND MECHANISMS OF GENE TRANSCRIPTION

Despite good evidence that cohesin regulates gene expression directly and independently of cell division (Pauli et al., 2010; Dorsett, 2011), the mechanism(s) of transcriptional regulation by cohesin are not well understood. Cohesin binds to many sites throughout the genome, sometimes in combination with the CCCTC-binding factor (CTCF) insulator protein, which is known to mediate chromatin loop formation (Gondor and Ohlsson, 2008). Previous studies demonstrated that cohesin colocalizes with CTCF along chromosome arms, and is likely to cooperate with this protein in the regulation of gene expression or chromatin structure (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008). As well as CTCF, cohesin colocates genome-wide with other transcriptional regulators, such as estrogen receptor-α (Schmidt et al., 2010), and Mediator (Kagey et al., 2010) in a cell type-specific manner. Likely in combination with other factors, cohesin selectively binds genes with paused RNA polymerase. Although it is not involved in RNA polymerase pausing itself, cohesin can regulate transcription by determining the amount of elongating RNA polymerase on genes (Fay et al., 2011).

Regulation of many genes by cohesin appears to involve the three-dimensional (3D) organization of chromatin (Merkenschlager, 2010; Dorsett, 2011). A direct role for cohesin in chromatin looping has been demonstrated for several loci (Hadjur et al., 2009; Mishiro et al., 2009; Nativio et al., 2009; Hou et al., 2010; Chien et al., 2011; Kim et al., 2011) by studies showing that long-range interactions between regulatory sequences are reduced by cohesin knockdown. It is likely that cohesin regulates spatiotemporal gene expression in combination with diverse tissue-specific transcription factors, and by distinct modes of transcription regulation (Dorsett, 2011).

THE HUMAN COHESINOPATHIES

An overlapping spectrum of human syndromes can be attributed to mutations in cohesin subunits, or regulators of cohesin loading and unloading from chromosomes. The best known cohesinopathy is CdLS (OMIM 122470) also termed Brachmann de Lange syndrome (BdLS), a broad spectrum disorder with multiple developmental and cognitive abnormalities (de Lange, 1933; Opitz, 1985; Ireland et al., 1993; Jackson et al., 1993). CdLS patients are small in size and have a characteristic facial appearance, including arched eyebrows, hirsutism, synophrys, ptosis, long eyelashes, an upturned nose, a long philtrum, thin upper lip, and micrognathia. Developmental anomalies range from mild to severe, with more severe cases having upper limb truncations or limb differences. CdLS patients also frequently present with hearing loss, gastrointestinal defects, pyloric stenosis, genital abnormality, congenital diaphragmatic hernias, cardiac septal defects, and autistic behaviors (Jackson et al., 1993). All patients within the CdLS spectrum have neurodevelopmental delay and highly variable mental retardation (Deardorff et al., 2007).

More than half of CdLS cases (∼65%) are dominantly inherited, and caused by mutations in the *NIPBL* gene (OMIM 608667; Krantz et al., 2004; Tonkin et al., 2004), which encodes a crucial component of kollerin. Heterozygous truncating or non-sense *NIPBL* mutations are haploinsufficient, and strikingly, NIPBL protein levels need only be reduced by 15–30% to give rise to a CdLS phenotype (Krantz et al., 2004; Tonkin et al., 2004). This implies that the remaining intact *NIPBL* allele is upregulated in an attempt to compensate, and also that certain cell types and/or developmental processes are exquisitely sensitive to the levels of NIPBL. Missense mutations in *NIPBL* were also identified that may interfere with the interaction of NIPBL with its partner, MAU2, or other proteins (Braunholz et al., 2012).

Mutations in *SMC1A* (OMIM 300040) and *SMC3* (OMIM 606062) also give rise to syndromes that fall within the CdLS spectrum, and account for about 5% of CdLS cases (Musio et al., 2006;

Deardorff et al., 2007; Mannini et al., 2010). *SMC* mutations are heterozygous missense mutations and are thought to interfere with the structure of the SMC subunits such that functional interactions of the cohesin complex are disturbed, causing the disease pathology (Deardorff et al., 2007). In some cases missense mutations were shown to interfere with cohesin binding to DNA (Revenkova et al., 2009). Human developmental phenotypes resulting from *SMC* mutations are inclined to be milder than for *NIPBL* mutations; these individuals have fuller eyebrows and a prominence of the nasal bridge, with fewer structural abnormalities; however, all patients had some degree of mental retardation (Deardorff et al., 2007; Rohatgi et al., 2010). This suggests that brain development is particularly sensitive to disruption of SMC subunits.

RAD21 (OMIM 606462) mutations also cause a cohesinopathy syndrome (Deardorff et al., 2012b). Heterozygous deletions of *RAD21* and missense mutations, which included a dominant interfering mutation and one with essentially no function, gave rise to developmental anomalies with some overlap with CdLS. Patients with *RAD21* mutations have an even milder phenotype than those with *SMC* mutations. They have some divergence in the facial features and, most notably, they have extremely mild cognitive and physical abnormalities (Deardorff et al., 2012b). Consistent with RAD21 having a role in DNA damage response, lymphoblastoid cell lines from patients with RAD21 mutations exhibited radiation sensitivity. A gene transcription assay in zebrafish showed that *RAD21* missense mutations present in patients are not competent for proper regulation of gene expression (Deardorff et al., 2012b).

Homozygous recessive mutations in the *ESCO2* gene, which encodes a CoAT, cause another cohesinopathy, RBS (OMIM 268300; Schule et al., 2005; Vega et al., 2005, 2010; Gordillo et al., 2008). RBS is characterized by mild to severe growth deficiency, limb malformations (in particular, symmetric tetraphocomelia), multiple craniofacial abnormalities including cleft lip and/or cleft palate, microcephaly, and mental retardation. Mortality is high among severely affected pregnancies and newborns (Gordillo et al., 1993). A milder disorder with less marked limb reduction and survival to adulthood is known as SC phocomelia, but since both disorders arise from ESCO2 mutations with no apparent genotype/phenotype correlation (Schule et al., 2005; Vega et al., 2010), it has been proposed all *ESCO2* mutations be referred to as RBS (Vega et al., 2010). Unlike CdLS, cells from RBS patients exhibit precocious sister chromatid separation, particularly at heterochromatic regions of the chromosomes (Schule et al., 2005; Vega et al., 2005) leading to mitotic defects, lagging chromosomes, aneuploidy, and micronuclei formation. The acetyltransferase activity of ESCO2 appears to be crucial, since mutations in this domain are sufficient for the pathogenesis of RBS (Gordillo et al., 2008). While RBS features overlap with those of CdLS, there are appreciable differences. Whether gene regulation downstream of ESCO2 is responsible for RBS pathology is still under debate.

The wide spectrum of human developmental phenotypes owing to cohesin mutations characterized to date indicate that although these disorders have many features in common, there are also distinct differences. Gene expression and molecular studies in cells and in animal models have helped to uncover the common and divergent pathways that lie downstream of cohesinopathy mutations.

CHARACTERIZATION OF COHESINOPATHY MUTATIONS REVEALS THAT DISTINCT PATHWAYS ARE AFFECTED BY DIFFERENT COHESINOPATHY MUTATIONS

A comparison of the consequences of knocking down cohesin or its regulators in different animal model systems indicates there are a wide variety of outcomes for cell biology and gene expression. For mutations causing CdLS and similar cohesinopathies, it seems likely that specific developmental pathways are regulated downstream of the causative gene mutations. Several groups have conducted analyses of gene expression downstream of cohesinopathy mutations.

For some genes, it seems likely that small changes in the dose of cohesin or its regulators could have a large impact on transcription. In *Drosophila*, cohesin and Nipped-B bind to actively transcribed regions of the genome and are excluded from regions of polycomb group (PcG) silencing (Misulovin et al., 2008). For the rare genes where cohesin binding overlaps with PcG-mediated methylation of lysine 27 on histone 3 (H3K27me3), expression of those genes is hypersensitive to cohesin dose (Schaaf et al., 2009). In addition, cohesin ablation in post-mitotic neurons in the *Drosophila* mushroom body (Pauli et al., 2008; Schuldiner et al., 2008), or salivary glands (Pauli et al., 2010) affected the expression of specific loci including the gene encoding the ecdysone receptor. This suggests that some genes, perhaps in specific cell types, may dramatically change their transcriptional activity in response to a slight alteration of cohesin dose.

Intriguingly, it seems that the transcriptional response of some genes to cohesin or Nipped-B depletion is biphasic, and depends on the degree to which these proteins are depleted (Schaaf et al., 2009). The *Enhancer of split* gene complex (E(spl)-C) in *Drosophila* is exquisitely responsive to Rad21 and Nipped-B levels. Furthermore, when mRNA encoding these proteins is depleted in BG3 cells, the direction in which some E(spl)-C are regulated depends on the length of time of RNAi treatment, and the degree of Rad21 or Nipped-B knockdown. For example, E(spl)-C transcripts decrease after 3 days of Nipped-B RNAi, but increase by day 6 (Schaaf et al., 2009). These findings have implications for genome-wide gene expression studies in cohesinopathy models. Which genes are altered in expression is likely to depend on tissue type, developmental stage and degree to which cohesinopathy gene function has been knocked down.

On the other hand, loss of Nipbl also appears to result in low ($\leq$2) fold changes in the expression of a great many genes. Liu et al. (2009) analyzed gene expression and genome-wide binding of cohesin in lymphoblastoid cell lines from CdLS probands with mutations in *NIPBL* or in the cohesin subunit *SMC1A*, and found that $\sim$1500 genes (FDR $\leq$ 0.05) were dysregulated compared with controls. Dysregulated gene expression in the mutant cell lines was conserved, and correlated with disease severity and cohesin binding at misexpressed genes (Liu et al., 2009). Significantly, a panel of 23 genes could differentiate *NIPBL* mutations from *SMC1A* and *ESCO2* mutations indicating that *NIPBL* mutations have a distinguishable effect on gene expression.

Heterozygous mice carrying a gene-trap insertion into the *Nipbl* locus show many features overlapping with CdLS, and microarray analyses indicated that reducing Nipbl dose resulted in small changes in expression of a great many genes. These mice also had

severe developmental phenotypes, including craniofacial dysmorphology and heart defects, resembling CdLS. Of note was the altered expression of genes involved in fat metabolism, which could account for the lean habitus observed in mice and in CdLS patients (Kawauchi et al., 2009).

Mice with mutations in *Pds5a* and *5b* have also been generated. Mice homozygous null for *Pds5b* died shortly after birth, with multiple congenital anomalies, including heart defects, cleft palate, skeletal defects, gut defects, abnormal migration and axonal projections of sympathetic neurons, and germ cell depletion (Zhang et al., 2007). Mice null for *Pds5a* exhibit many of the same multiple abnormalities that were previously observed in *Pds5b*-deficient mice, plus additional abnormalities including renal agenesis (Zhang et al., 2009). Elimination of both Pds5a and 5b gave an additional lens phenotype not observed in single null mice, and resulted in embryonic lethality (Zhang et al., 2009). Gene expression studies in the Pds5 mice have not been published.

Most recently, significant knowledge about cohesin function was gained by generating mice deficient for cohesin subunit SA1 (Cuadrado et al., 2012; Remeseiro et al., 2012a,b). Loss of SA1 results in embryonic lethality, and heterozygous animals have shorter lifespan and increased aneuploidy as a result of chromosome segregation defects. Segregation defects arose from compromised telomere replication, which requires cohesion mediated specifically by cohesin-SA1. The resulting aneuploidy in *SA1* heterozygotes is thought to lead to early onset of tumorigenesis in these animals (Remeseiro et al., 2012a).

Interestingly, gene expression and genome-wide distribution of cohesin binding are dramatically altered in *SA1* null mice, with important implications for CdLS. Location of cohesin to gene promoters and CTCF binding sites appears to depend on SA1. Furthermore, SA1 ablation led to altered cohesin binding at particular gene clusters accompanied by dysregulation of their transcription (Remeseiro et al., 2012b). These studies highlight the function of SA1 in multiple processes, and identify a key transcriptional role that is distinct from the function of SA2 in centromeric chromosome cohesion.

Zebrafish models have also shed light on the role of cohesin and Nipbl in gene expression. In fact, the first published evidence that cohesin regulates gene expression in a vertebrate model system came from a forward genetic screen in zebrafish. This screen identified the *Rad21* subunit as a tissue-specific regulator of *runx1*, which encodes a hematopoietic transcription factor (Horsfield et al., 2007). In *rad21* mutants at 12 h post-fertilization (hpf), *runx1* expression was retained in Rohon–Beard neurons, but was absent from a discrete population of cells in the hematopoietic mesoderm. Importantly, the hematopoietic mesoderm precursor cell population was still present in mutants, and expressed the dimerization partner for Runx1, *cbfb*, although not *runx1* itself. Cohesin probably targets other *runx* genes in a cell type-specific manner, since *rad21* mutants also lacked expression of *runx3* in Rohon–Beard neurons and the lateral line primordia (Horsfield et al., 2007). Unfortunately, the onset of *runx2* expression (∼48 hpf) in zebrafish embryos is too late to determine its involvement, since *rad21* mutants arrest in development at 35 hpf. Like in *Drosophila*, cohesin is likely to regulate expression of genes in zebrafish brain; cohesin subunits are expressed in non-proliferating neurons of zebrafish brain implying a non-cell cycle role for cohesin in this tissue (Monnich et al., 2009).

A zebrafish model of *NIPBL*-mediated CdLS revealed much about the multifactorial origins of this developmental syndrome. Zebrafish have two copies of the *nipbl* gene, and depletion of both versions by morpholino oligonucleotides to create "morphants" also led to small-scale dysregulation of a large number of genes in early embryogenesis (up to 6 hpf; Muto et al., 2011). Because gene expression changes were measured at early gastrula stages it is likely that many are directly caused by reduced Nipbl function rather than by secondary effects. Interestingly, genes involved in endoderm development and left-right axial patterning including *sox17* and *foxa2*, were specifically downregulated in endoderm. Dysregulation of the endoderm-specifying hierarchy of Sox32, Sox17, and Foxa2 by Nipbl depletion is likely to contribute to the heart looping defects and gut tube defects observed at later stages in Nipbl-depleted zebrafish embryos (Muto et al., 2011). The zebrafish pathologies recapitulate heart and gastrointestinal tract abnormalities observed in CdLS, thereby allowing insight into the etiology of CdLS developmental defects.

Our own group conducted Affymetrix microarray analyses at a later stages (24 and 48 hpf) of zebrafish development in *rad21* mutants (Rhodes et al., 2010), *esco2* morphants (Monnich et al., 2011), *smc1a* mutants (available as part of an insertion mutant collection; Amsterdam et al., 2004) and *nipbl* morphants (Maren Mönnich, Cristin G. Print, Julia A. Horsfield, unpublished data). Interestingly, we found that the *eomes* gene, a master regulator of endoderm formation, is consistently downregulated in *rad21* and *smc1a* mutants, and *nipbl* morphants (FDR < 0.02), supporting a role for cohesin and Nipbl in endoderm formation. *Eomes* expression is regulated by pluripotency factors Nanog, Oct4, and Sox2 (Teo et al., 2011), all of which are transcriptional targets of cohesin and Nipbl in embryonic stem cells (Kagey et al., 2010). It is enticing to speculate that cohesin and Nipbl could participate in the initial specification of germ layers from stem cell precursors through modulating the expression of pluripotency factors.

We expected our microarray analyses of zebrafish cohesinopathy mutants and morphants to result in similar lists of up- or downregulated genes, since embryos were analyzed at similar stages and cohesinopathy genes would be predicted to have similar roles in gene expression. Therefore we were surprised to find only modest overlap between regulated gene sets (example in **Figure 2**).

Strikingly, a comparison of *rad21* mutant microarray data with *esco2* morphant microarray data revealed that there is scant overlap between genes regulated downstream of these mutations (Monnich et al., 2011). For example, the *myca* gene, which is downregulated in *rad21* zebrafish mutants and other species as well, is actually slightly upregulated in *esco2* morphants. Most of the genes regulated downstream of *esco2* are involved in cell proliferation or apoptosis, whereas many genes affected by the *rad21* mutation are developmental regulators (Monnich et al., 2011). What could be the reason for these differences? We concluded that while Esco2 and Rad21 have related roles in sister chromatid cohesion, they do not have the same input into the regulation of gene expression. We found that although *esco2* depletion has mild effects on neural crest cell migration, it does not induce patterning defects. Instead, even modest *esco2* depletion results in robust activation of

caspases, *p53/mdm2* upregulation, and massive cell death (Monnich et al., 2011). Loss of jaw elements and fin stunting in *esco2* morphants, which resemble RBS features, are therefore likely to be due to insufficient cells to contribute to the affected structures. In agreement with results from a conditional mouse knockout of *Esco2* (Whelan et al., 2012b), it appears that developmental defects observed in *esco2* morphant zebrafish arise from problems with cell survival rather than dysregulation of developmental genes.

Our microarray data of *nipbl* morphants was conducted under very mild knock down conditions of both *nipbl* genes at 24 and 48 hpf (**Table 2**; Maren Mönnich, Cristin G. Print, Julia A. Horsfield, unpublished data). We observed regulation of different sets of genes than those found by Muto et al. (2011) at the earlier timepoint of 6 hpf, which is not unexpected due the different developmental stage at which embryos were analyzed. We did not find any Gene Ontology categories of significance other than elevated expression of a network of genes related to p53. It is possible that degree of *nipbl* gene knockdown could also contribute to differences observed in regulated genes as discussed above, since at least some gene expression is likely to be sensitive to the dose of Nipbl protein (Schaaf et al., 2009).

Many genes that have altered regulation in response to depletion of cohesinopathy genes are different, raising the possibility that cohesin subunits and regulators have different functions in various pathways. However, genome-wide analyses of gene expression identified some commonly regulated pathways/genes such as those involved in endoderm development (*eomes*, *sox17*, *foxa3*), the *myc* transcription factor (except in *esco2* morphants), and downstream effectors of Notch signaling such as *hey1*, *her4.2*, and *ascl1*.

COMMON PATHWAYS REGULATED BY COHESINOPATHY GENES

Despite varying outcomes for gene expression and development identified using animal models of the cohesinopathies, some pathways seem more likely to be affected than others downstream of cohesinopathy genes. Common themes of pathways regulated by cohesinopathy genes are outlined below.

GROWTH, METABOLISM, AND PLURIPOTENCY

Perhaps not surprisingly, several studies have found links between cohesin and its regulators, and the control of pathways that underpin cell growth and proliferation. Somewhat more surprisingly, the level at which cohesin regulates growth and metabolism includes transcriptional control of specific gene targets. For example, the *Myc* oncogene is positively regulated by Nipbl and all cohesin subunits investigated to date (Misulovin et al., 2008; Kawauchi et al., 2009; Liu et al., 2009; Rhodes et al., 2010; Remeseiro et al., 2012b). Myc is a pluripotency factor, and it is probably significant that genes encoding other pluripotency factors Oct4, Nanog, and Sox2, are also bound and regulated by cohesin (Kagey et al., 2010; Nitzsche et al., 2011). Interestingly, pluripotency factors, e.g., Oct4 (Kim et al., 2011) and Nanog (Nitzsche et al., 2011) in turn appear to combine with cohesin to both positively and negatively regulate other target genes. These findings raise the interesting possibility that cohesin-mediated transcription is pivotal to cell fate decisions that determine the balance between pluripotency and differentiation (Dorsett, 2010).

Cohesinopathy genes regulate other growth pathways as well. In yeast, cohesinopathy mutations, including an Eco1RBS mutation (W216G), block transcription of ribosomal RNA genes thereby directly influencing ribosome biogenesis, protein translation and the cell's ability to grow (Bose et al., 2012). This finding links cohesin function to metabolism and growth through a role in rDNA transcription and translation regulation. Since Myc, a transcriptional target of cohesin, also regulates ribosome biogenesis (Eilers and Eisenman, 2008), cohesin appears to be a central regulator of growth by transcriptional control of multiple pathways. In $Nipbl^{+/-}$ mice, genes controlling fat metabolism are dysregulated (Kawauchi et al., 2009), indicating a direct involvement in regulation of another metabolic pathway. Consistent with dysregulated growth and metabolism, CdLS patients are small and lean (Liu and Krantz, 2009). It is possible that many of the large number of dysregulated genes in CdLS are targets of MYC, which regulates 10–20% of genes in the genome.

Transcriptional regulation of cell growth and proliferation pathways by cohesin could be elegantly intertwined with its role in the cell cycle, where it mediates sister chromatid cohesion. Transcriptional pathways promoting growth are tightly linked to cell division, and it is entirely possible that cohesin and its regulators have central roles in making these links.

NEURONAL DEVELOPMENT AND THE TRANSCRIPTION OF NEURONAL GENES

Neurodevelopmental disorders are among the most conserved features of the cohesinopathies (Deardorff et al., 2007). It is possible that these neurodevelopmental pathologies have a common molecular basis. Several lines of evidence suggest that cohesinopathy proteins influence the Notch signaling pathway, although the exact mechanisms are unknown. A recent study suggested that Esco2 physically interacts with Notch to antagonize Notch signaling, suggesting that one possible mechanism includes direct interaction with Notch receptor(s; Leem et al., 2011).

Our microarray analyses of zebrafish "cohesinopathy" embryos depleted for Rad21, Smc1a, Nipbl, or Esco2 identified conserved regulation of selected gene targets of the Notch signaling pathway. Notably, we found that the *ascl1* gene is downregulated in both *rad21* mutants (Horsfield et al., 2007; Rhodes et al., 2010) and *esco2* morphants (Monnich et al., 2011), as well as Nipbl-depleted embryos and *smc1a* mutants (**Table 2**; Maren Mönnich, Cristin G. Print, Julia A. Horsfield, unpublished data). In 48 hpf *rad21* heterozygous embryos (which are phenotypically normal), *ascl1* is significantly downregulated (Rhodes et al., 2010), indicating that *ascl1* expression is highly sensitive to even a slight reduction of Rad21 (heterozygotes have 60–70% of wild type *rad21* mRNA levels). Such sensitivity could have high functional significance. *Ascl1* is a potent neuronal lineage-specifying gene, being one of three genes sufficient to convert fibroblasts into iPN cells (Vierbuchen et al., 2010). Furthermore, Pds5b depletion altered *Ascl1* expression and blocked neuronal differentiation in a stem cell model (Denes et al., 2010).

We also found that certain Notch signaling targets of the *hairy/enhancer of split* family (such as *her4*, *hey1*) were consistently

Table 2 | Top 20 probe sets affected in zebrafish cohesinopathy microarrays at 24 h post-fertilization ($p < 0.05$).

rad21 Mutant			*smc1a* Mutant			*nipbl* Morphant			*esco2* Morphant		
Affymetrix probe ID	**Gene**	**Log_2 change**	**Affymetrix probe ID**	**Gene**	**Log_2 change**	**Affymetrix probe ID**	**Gene**	**Log_2 change**	**Affymetrix probe ID**	**Gene**	**Log_2 change**
EXPRESSION LEVELS "UP"											
Dr.24216.1.S1_at	cki	7.862	Dr.7787.1.S1_at	RPS27	1.056	Dr.21935.1.A1_at	wu:fc84a08	2.927	Dr.17659.1.S1_at		3.483
Dr.5211.1.A1_at		2.374	Dr.4314.1.A1_a_at	wu:fb95d03	0.729	Dr.7787.1.S1_at	RPS27	2.687	Dr.21935.1.A1_at	wu:fc84a08	3.267
Dr.19471.1.A1_at	scamp5	2.116	Dr.15033.1.S1_at		0.647	Dr.13570.1.A1_at	zc3h14	2.591	Dr.11242.1.A1_at	phlda3	3.159
Dr.14046.1.S1_at	UBE2D2	1.918	Dr.2727.1.A1_at		0.646	Dr.11242.1.A1_at	phlda3	2.538	Dr.10334.1.S1_at	casp8	2.926
Dr.4716.2.A1_at	nrarpa	1.813	Dr.4806.1.A1_at	TSTA3	0.642	Dr.11479.1.A1_at	lnx1	2.476	Dr.11479.1.A1_at	lnx1	2.786
Dr.4314.1.A1_x_at	wu:fb95d03	1.635	Dr.14044.1.A1_at	gpr137bb	0.614	Dr.17659.1.S1_at		2.452	Dr.12986.1.A1_a_at	fos	2.636
Dr.25322.1.S1_at	lin7c	1.533	Dr.1190.1.S1_at	anxa1b	0.578	Dr.10334.1.S1_at	casp8	2.406	Dr.23587.1.A1_at	gadd45al	2.632
Dr.17340.1.S1_at	hnmt	1.476	Dr.12502.3.S1_x_at	zgc:171781	0.57	Dr.198.1.S1_at	fst	2.382	Dr.7787.1.S1_at	RPS27	2.628
Dr.7532.1.A1_at	hm:zeh0402	1.397	Dr.5231.1.S1_at	hist2h2l	0.56	Dr.17275.1.A1_at	WHSC2	2.298	Dr.12986.1.A1_at	fos	2.581
Dr.26538.1.A1_at	gpr177	1.388	Dr.5129.1.S1_at	sesn3	0.557	Dr.26372.1.A1_at		2.225	Dr.5925.1.A1_at	wu:fi04f09	2.558
Dr.9457.1.A1_at	egln3	1.287	Dr.1999.1.S1_at	DF	0.55	Dr.542.1.S1_at	mdm2	2.212	Dr.542.1.S1_at	mdm2	2.499
Dr.9478.1.S1_at	cyp1a	1.239	Dr.5820.1.S1_at	ctsll	0.544	Dr.23406.1.S1_at	rpz5	2.168	Dr.11481.1.A1_at	rspo1	2.359
Dr.3432.1.S1_at	capg	1.231	Dr.8453.1.A1_at		0.543	Dr.7768.1.A1_at	shfm1	2.143	Dr.23406.1.S1_at	rpz5	2.219
Dr.24923.2.A1_at		1.204	Dr.15162.3.A1_a_at	pofut1	0.528	Dr.5925.1.A1_at	wu:fi04f09	2.141	Dr.6820.1.A1_at	gtpbp1l	2.173
Dr.18513.2.S1_a_at	sccpdhb	1.164	Dr.5211.1.A1_at		0.519	Dr.10083.1.S1_at		2.126	Dr.2052.1.S1_at	Tp53	2.158
Dr.7738.1.A1_at	aldh18a1	1.157	Dr.3499.3.A1_at	cldni	0.502	Dr.15033.1.S1_at		2.044	Dr.8209.1.S2_at	foxo5	2.098
Dr.5987.1.A1_at	msl1l1	1.139	Dr.15162.1.A1_a_at	pofut1	0.499	Dr.307.1.S1_at	dvr1	2.044	Dr.15033.1.S1_at		2.041
Dr.10735.1.S1_at	caspb	1.107	Dr.2536.1.S1_at	itm2bb	0.499	Dr.14044.1.A1_at	gpr137bb	2.012	Dr.19794.1.A1_at		2.025
Dr.15033.1.S1_at		1.1	Dr.2059.1.A1_at	slc2a2	0.492	Dr.24938.1.S1_a_at	zgc:158463	1.958	Dr.21979.1.A1_at	wu:fc92e10	2.023
Dr.4314.1.A1_a_at	wu:fb95d03	1.088	Dr.18631.2.A1_at	zgc:123295	0.49	Dr.20198.1.S1_a_at	hsp70	1.92	Dr.12986.2.S1_at	fos	1.997
EXPRESSION LEVELS "DOWN"											
Dr.5662.1.S1_at	rad21	−3.585	Dr.25729.1.S1_at	CRYGB	−1.511	Dr.5479.1.S1_at	rbp4	−2.219	Dr.20010.14.S1_at		−1.925
Dr.582.1.S1_a_at	cx43	−3.558	Dr.20020.1.S1_at	smc1a	−1.503	Dr.15054.1.S1_at	rbp2a	−1.949	Dr.18151.1.S1_at		−1.512
Dr.10102.2.S1_at	fam212aa	−3.2	Dr.13843.1.S1_at	bhlhe22	−1.485	Dr.5112.1.S3_at	sox11b	−1.813	Dr.16312.1.S1_at	sb:cb25	−1.476
Dr.10102.1.A1_at	fam212aa	−2.432	Dr.4407.1.A1_at	smc1a	−1.322	Dr.22100.1.A1_at	wu:fd13e05	−1.774	Dr.5372.9.S1_s_at	her4.2	−1.401
Dr.1.1.S1_at	myca	−2.271	Dr.15372.1.S1_x_at	CRYGB	−1.22	Dr.8342.1.A1_at	slc6a11	−1.731	Dr.3211.1.A1_at	GRIK2	−1.351
Dr.23067.1.S1_at	myhz1	−2.158	Dr.24771.1.A1_at	smc1a	−0.984	DrAffx.2.105.S1_at		−1.705	Dr.5372.7.A1_x_at	CH73-21G5.3	−1.326
Dr.4812.1.S1_s_at	myhz2	−2.147	Dr.12486.1.S1_at	CLDN11	−0.973	Dr.2778.1.S1_at	STX4A	−1.683	DrAffx.2.105.S1_at		−1.297
Dr.13843.1.S1_at	bhlhe22	−2.118	Dr.10102.2.S1_at	fam212aa	−0.952	Dr.4797.1.S1_at	zgc:123103	−1.682	Dr.5372.1.S1_x_at	her4.2	−1.289
Dr.10343.1.S1_at	atp1a1a.2	−2.074	Dr.22360.1.A1_at	AQP1	−0.879	Dr.21660.1.A1_at		−1.597	Dr.5372.7.A1_at	CH73−21G5.3	−1.246
Dr.22360.1.A1_at	AQP1	−1.857	Dr.19658.1.A1_at		−0.8	Dr.25195.1.S1_at	zgc:158494	−1.563	Dr.20850.1.S1_at	fabp7a	−1.238

(Continued)

Table 2 | Continued

rad21 Mutant			*smc1a* Mutant			*nipbl* Morphant			*esco2* Morphant		
Affymetrix probe ID	**Gene**	**Log_2 change**	**Affymetrix probe ID**	**Gene**	**Log_2 change**	**Affymetrix probe ID**	**Gene**	**Log_2 change**	**Affymetrix probe ID**	**Gene**	**Log_2 change**
Dr.2155.1.S1_at	plk3	−1.797	Dr.4299.1.S1_at	wu:fb83d05	−0.789	Dr.2426.1.S1_at	ambpl	−1.538	Dr.5434.1.S3_at	plp1a	−1.215
Dr.314.1.S1_at	ascl1a	−1.787	Dr.582.1.S1_a_at	cx43	−0.743	Dr.3891.1.A1_at	septin 6	−1.505	Dr.5434.1.S1_at	plp1a	−1.171
Dr.21790.1.A1_at	PCDH18	−1.724	Dr.10102.1.A1_at	fam212aa	−0.724	Dr.5434.1.S1_at	plp1a	−1.502	Dr.2970.1.S1_at	apoea	−1.121
Dr.1280.1.A1_at	cebpd	−1.66	Dr.25173.1.S1_at	smc1a	−0.719	Dr.3004.1.A1_at	AL954182.2	−1.495	Dr.20083.1.A1_at	ccng2	−1.111
Dr.737.1.A1_at	junbl	−1.653	Dr.20850.1.S1_at	fabp7a	−0.666	Dr.18151.1.S1_at		−1.492	Dr.5434.1.S4_at	plp1a	−1.107
Dr.13879.1.A1_at	islr2	−1.638	Dr.6081.1.S1_at	fam212ab	−0.664	Dr.22360.1.A1_at	AQP1	−1.441	AFFX-r2-Bs-dap-3_at		−1.029
Dr.20185.1.S1_at	lpl	−1.637	Dr.21790.1.A1_at	PCDH18	−0.661	Dr.5434.1.S2_at	plp1a	−1.395	Dr.25729.1.S1_at	CRYGB	−1.016
Dr.16053.1.S1_at	hbegfa	−1.577	Dr.3282.1.S1_at	s1pr1	−0.656	Dr.2132.1.A1_at	hao1	−1.388	AFFX-r2-Bs-dap-M_at		−1.012
Dr.318.1.A1_at	ascl1b	−1.558	Dr.314.1.S1_at	ascl1a	−0.655	Dr.13750.1.S1_at		−1.338	Dr.13843.1.S1_at	bhlhe22	−0.997
Dr.20850.1.S1_at	fabp7a	−1.541	Dr.8587.1.A1_at	igfbp1a	−0.652	Dr.4867.1.A1_at	hp	−1.338	Dr.8118.1.A1_at	otpb	−0.995

affected in our zebrafish cohesinopathy models (Rhodes et al., 2010), in agreement with cohesin/Nipbl regulation of the (E(spl)-C) in *Drosophila* (Schaaf et al., 2009). In combination with previous gene expression studies from *Drosophila* (Dorsett, 2009), strong evidence supports a link between cohesin-mediated transcription and cell fate in neuronal linages.

In addition to neuronal cell fate, it appears that cohesin together with CTCF could contribute to maintaining neuronal identity. Several studies show that cohesin and CTCF regulate expression of protocadherin genes (Kawauchi et al., 2009; Monahan et al., 2012; Remeseiro et al., 2012b). Cohesin-SA1 binds to the promoter of protocadherin genes and positively regulates their expression (Remeseiro et al., 2012b). Interestingly, CTCF and cohesin were recently found to modulate isoform expression of *Pcdhα* in a mouse neuroblastoma cell line (Monahan et al., 2012), by a mechanism assumed to involve enhancer-promoter communication. Cohesinopathy mutations could therefore have significant consequences for neuronal recognition of "self," and the capacity to make functional synaptic connections (Dekker, 2012), since protocadherins are key players in these processes (Frank and Kemler, 2002; Esumi et al., 2005).

Evidence suggests that the widespread disruption of neuronal gene expression found in cohesinopathy mutants results in abnormal behavior and function of neurons. As discussed previously, localized disruption of cohesin subunits causes failure of axon pruning in the *Drosophila* mushroom body (Pauli et al., 2008; Schuldiner et al., 2008). Other model systems have highlighted a role for cohesinopathy proteins in axon pathfinding and/or migration. For example, Mau2, the Scc4 homolog that binds to Nipbl, is necessary for proper axon guidance and migration in *C. elegans* (Seitan et al., 2006). Consistent with a requirement of cohesin for migration, enteric neurons derived from neural crest cells failed to migrate in mice mutant for cohesin subunit *Pds5b* (Zhang et al., 2007). Furthermore, in *esco2* morphant zebrafish, we observed defects that were consistent with abnormal neural crest cell migration (Monnich et al., 2011). In zebrafish mutant for *rad21*, we observed that while the trigeminal ganglia of the brain are specified, the axons clump together rather than extending forward (**Figure 3**). It is very likely that more subtle defects that are not

FIGURE 3 | Flat-mount staining (anti-HNK-1) of trigeminal ganglia in wild type (left) and *rad21* mutant (right) zebrafish embryos. In *rad21* mutants, central neuronal clumping occurred (arrow), and axons failed to migrate and populate anterior regions (red oval).

so easily observed (for example, problems with neuronal connectivity) take place in the central nervous system of cohesinopathy patients and animal models.

Growth, metabolism, and development of the central nervous system appear to be processes that are universally sensitive to cohesinopathy mutations. Much of this pathology is likely to be caused by cohesin's role in the regulation of gene expression. However, cohesin has another important role in the repair of DNA damage, and its loss is likely to trigger DNA damage checkpoints (Jessberger, 2009; Watrin and Peters, 2009). Activation of cell cycle checkpoints by cohesin depletion may represent additional biological processes contributing to cohesinopathies.

COHESINOPATHY GENES, DNA DAMAGE, AND CELL CYCLE CHECKPOINTS

When damaged DNA is detected, cells respond by coordinating cell cycle arrest, DNA repair, and programmed cell death (Ciccia and Elledge, 2010). The crucial roles of cohesin and its regulators in DNA damage repair have been recently and comprehensively reviewed elsewhere (Dorsett and Strom, 2012; Wu and Yu, 2012). Inability to repair DNA and proceed through the cell cycle is accompanied by activation of cell cycle checkpoints, followed by cell death in the absence of repair.

Interestingly, mutations in other genes responsible for the DNA damage response underlie human syndromes with phenotypes that overlap the cohesinopathies (Ciccia and Elledge, 2010). Overlapping phenotypes include microcephaly, growth defects, neurological disorders, and facial/skeletal dysmorphology. These features are among the most conserved between the cohesinopathies, and raise the possibility that defects in the DNA damage response pathway might contribute to the etiology of cohesinopathy syndromes. In support of this, a patient with a mutation in a gene encoding the DNA helicase DDX11/ChlR1 had microcephaly, premature sister chromatid separation, and genome instability. This patient had features of both Fanconi Anemia (associated with other DNA helicases involved in DNA damage repair, XPD, and FANCJ) and RBS, in which *ESCO2* is mutated. The syndrome, known as Warsaw Breakage Syndrome, is considered to reside at an interface between DNA damage repair and sister chromatid cohesion (van der Lelij et al., 2010).

It is possible that the CoAT ESCO2 has a particularly crucial role in DNA damage repair, since mutations in ESCO2 appear to resemble mutations in DNA damage repair pathways more than the other cohesinopathies do. Indeed, ESCO2-depleted cells are hypersensitive to DNA damaging agents such as Mitomycin C (van der Lelij et al., 2009; Whelan et al., 2012a). Acetylation of SMC3 is necessary for S phase checkpoint activation and cell survival (Kim et al., 2010a), which might explain the absolute requirement for ESCO2 at this stage of the cell cycle.

Other cohesinopathy mutations also have potential to compromise DNA damage repair. Mice heterozygous for a *Rad21* null mutation are hypersensitive to IR, and exhibit problems with integrity and maintenance of the gastrointestinal tract and hematopoietic system post-irradiation (Xu et al., 2010). In humans, patients with *RAD21* mutations also have impaired DNA damage repair (Deardorff et al., 2012b), and knock down of *RAD21* sensitizes breast cancer cells to chemical agents that damage DNA (Atienza et al., 2005; Xu et al., 2011). Therefore, full dosage and function of the *Rad21* gene is crucial for DNA damage repair. In addition, depletion of SMC1 sensitizes HeLa cells to DNA damage (Bauerschmidt et al., 2010). Interestingly, the cohesin regulator PDS5B (APRIN) and the cohesin subunits RAD21 and SMC3 were recently found to associate with the BRCA2 protein. PDS5B appears to have an essential function in both the DNA damage response and homologous recombination (Brough et al., 2012).

It is not clear to what extent DNA damage repair defects contribute the pathology of cohesinopathies (Dorsett and Strom, 2012), but evidence suggests that most cohesinopathy mutations are likely to impact on the cell cycle in intra-S and G2 phases, when DNA damage repair takes place. Insufficiency of DNA damage repair should lead to checkpoint activation and cell death, potentially resulting in a paucity of cells for adequate development. However, many cohesinopathy mutations give rise to altered transcription of developmental regulators rather than cell cycle phenotypes, raising the question of how distinct outcomes arise from mutations in proteins with a related function in the cell cycle.

A MODEL TO EXPLAIN DIVERSE COHESINOPATHY PHENOTYPES

We propose a model to explain the diverse phenotypes observed downstream of cohesinopathy genes, in which different phenotypes emerge according to the "phase" of the cohesin cycle that is *most* affected by a particular cohesinopathy mutation in a given population of cells (**Figure 4**). In this model, mutations affecting cohesin loading and its residency times on chromatin in interphase have a higher potential to influence the regulation of gene expression, since this function can be exquisitely sensitive to cohesin dose. Alternatively, mutations affecting the "cohesive" form of cohesin have more potential to impact on cell division, DNA damage repair, and cell cycle checkpoints. The consequences are that the latter mutations will affect sister chromatid cohesion, and initiate cell death pathways. Shared phenotypes such as microcephaly, craniofacial defects, and cognitive impairment are likely to lie at the interface between these two pathways.

It is important to note that cohesin subunits and the cohesin loading kollerin complex participate in *all* cohesin-related processes, including transcription regulation during interphase, chromatid cohesion during S phase, and DNA damage repair. Therefore mutations in genes encoding these proteins have potential to disrupt all the processes shown in the model (**Figure 4**). However, diverse outcomes from different cohesinopathy mutations could result if certain processes have differential sensitivity to loss of cohesin components and regulators, in distinct cell populations.

For example, zebrafish embryos zygotic null for *rad21* contain heavy maternal loading of Rad21 protein and are able to develop for about 20 h before cell cycle deficiencies halt growth.

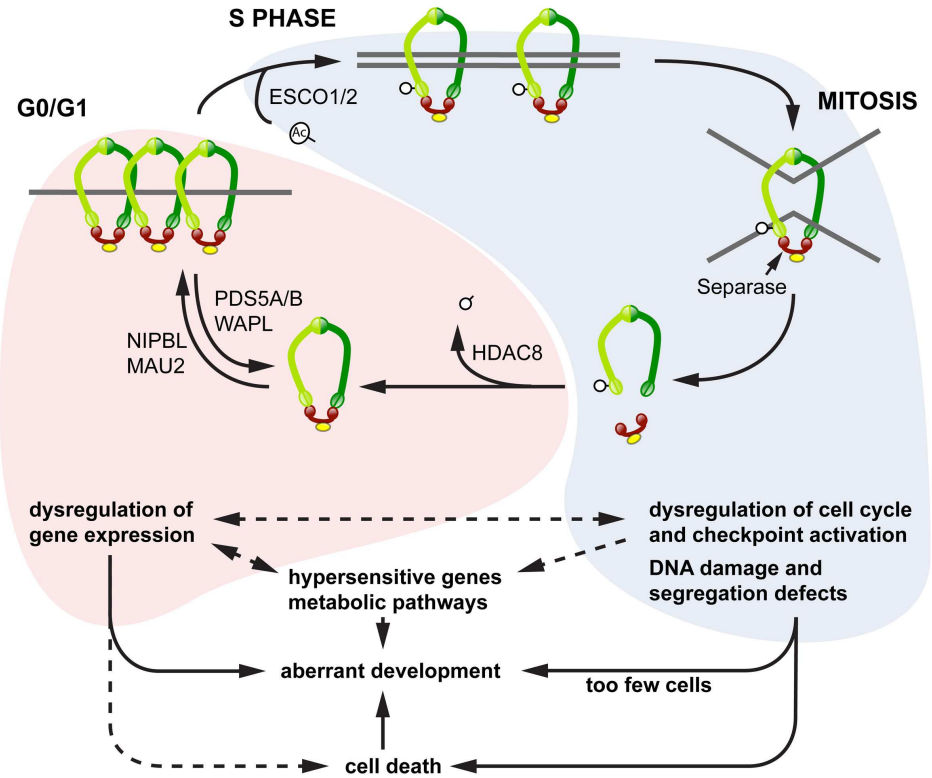


FIGURE 4 | Model for diverse cohesinopathy phenotypes. In interphase (red shading), cohesin binding to chromatin is dynamic, with varying residency times. Interphase cohesin binding is likely to be cell type-specific and to contribute toward regulating developmental genes. Mutations in cohesin subunits and their key interphase regulators (e.g., Nipbl, Hdac8) primarily impact on the regulation of gene expression, including transcriptional regulation of growth pathways. This results in syndromic developmental defects that derive from dysregulated transcription, with the possibility of cell death as a contributing factor. From S phase to G2/M (blue shading), the overriding function of cohesin involves sister chromatid cohesion and DNA damage repair. Key regulators in this process include the CoAT ESCO2 and other DNA damage repair proteins. Mutations in these regulators result in chromosome segregation defects, genomic instability, and cell death. Increased cell death and reduced cell proliferation results in too few cells to make up body structures, leading to a different class of developmental defects and dysregulation of metabolic pathways. Transcription of a small subset of hypersensitive genes, including some in the Notch signaling pathway, appears to be sensitive to both interphase and S/G2 modes of cohesin binding.

However, well before cell cycle defects have any impact, *rad21* null embryos fail to activate *runx1* expression in the hematopoietic mesoderm (Horsfield et al., 2007). Thus, there is a threshold level of cohesin essential for *runx1* expression that is below the level necessary to sustain cell division. The primary impact of suboptimal levels of Rad21 is that of altered gene expression in a subpopulation of cells, and the secondary impact of cell cycle arrest is not observed until Rad21 levels are further depleted. Radiation sensitivity observed in *Rad21* heterozygous mice (Xu et al., 2010) and in cells of patients with compromised RAD21 function (Deardorff et al., 2012b) indicates other functions of Rad21 are also dose-sensitive.

In summary, the phenotypic outcome of cohesinopathy mutations may differ between cell populations and in any given cell population, depend upon the degree of sensitivity of gene expression to cohesin levels, the requirement for cell proliferation, and the presence of environmental stressors such as DNA damaging agents. There is likely to be significant overlap in these contributing factors to cohesin-related developmental disorders.

CONCLUSION

Although considerable progress has been made over the last 10 years in the understanding of cohesin function in the cell cycle, transcription, and human developmental disease, important questions remain. How is the transcriptional role of cohesin coordinated with its role in genome organization, cell division, and DNA repair? Why do some cohesinopathy mutations lead to developmental gene dysregulation, while others lead to chromosome segregation defects and cell death? Human syndromes and animal models have potential to lend important insight into the integration of cohesin functions in cell division and development. Continued research will be vital for understanding the pathology of cohesinopathy syndromes, and the development of future potential for clinical management or therapy.

ACKNOWLEDGMENTS

The authors wish to thank Matt Deardorff for helpful discussions. Research on developmental roles for cohesin in the authors' laboratory is supported by the Royal Society of NZ Marsden Fund, Lottery Health Research NZ, the KD Kirkby Trust, and Cure Kids.

REFERENCES

Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S., and Hopkins, N. (2004). Identification of 315 genes essential for early zebrafish development. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12792–12797.

Atienza, J. M., Roth, R. B., Rosette, C., Smylie, K. J., Kammerer, S., Rehbock, J., Ekblom, J., and Denissenko, M. F. (2005). Suppression of RAD21 gene expression decreases cell growth and enhances cytotoxicity of etoposide and bleomycin in human breast cancer cells. *Mol. Cancer Ther.* 4, 361–368.

Ball, A. R. Jr., and Yokomori, K. (2008). Damage-induced reactivation of cohesin in postreplicative DNA repair. *Bioessays* 30, 5–9.

Bauerschmidt, C., Arrichiello, C., Burdak-Rothkamm, S., Woodcock, M., Hill, M. A., Stevens, D. L., and Rothkamm, K. (2010). Cohesin promotes the repair of ionizing radiation-induced DNA double-strand breaks in replicated chromatin. *Nucleic Acids Res.* 38, 477–487.

Beckouet, F., Hu, B., Roig, M. B., Sutani, T., Komata, M., Uluocak, P., Katis, V. L., Shirahige, K., and Nasmyth, K. (2010). An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Mol. Cell* 39, 689–699.

Ben-Shahar, T. R., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., and Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science* 321, 563–566.

Bermudez, V. P., Farina, A., Higashi, T.L., Du, F., Tappin, I., Takahashi, T. S., and Hurwitz, J. (2012). In vitro loading of human cohesin on DNA by the human Scc2-Scc4 loader complex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9366–9371.

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Rolef Ben-Shahar, T., and Uhlmann, F. (2010). Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. *Mol. Cell* 39, 677–688.

Bose, T., Lee, K. K., Lu, S., Xu, B., Harris, B., Slaughter, B., Unruh, J., Garrett, A., Mcdowell, W., Box, A., Li, H., Peak, A., Ramachandran, S., Seidel, C., and Gerton, J. L. (2012). Cohesin proteins promote ribosomal RNA production and protein translation in yeast and human cells. *PLoS Genet.* 8, e1002749. doi:10.1371/journal.pgen.1002749

Braunholz, D., Hullings, M., Gil-Rodriguez, M. C., Fincher, C. T., Mallozzi, M. B., Loy, E., Albrecht, M., Kaur, M., Limon, J., Rampuria, A., Clark, D., Kline, A., Dalski, A., Eckhold, J., Tzschach, A., Hennekam, R., Gillessen-Kaesbach, G., Wierzba, J., Krantz, I. D., Deardorff, M. A., and Kaiser, F. J. (2012). Isolated NIBPL missense mutations that cause Cornelia de Lange syndrome alter MAU2 interaction. *Eur. J. Hum. Genet.* 20, 271–276.

Brough, R., Bajrami, I., Vatcheva, R., Natrajan, R., Reis-Filho, J. S., Lord, C. J., and Ashworth, A. (2012). APRIN is a cell cycle specific BRCA2-interacting protein required for genome integrity and a predictor of outcome after chemotherapy in breast cancer. *EMBO J.* 31, 1160–1176.

Carretero, M., Remeseiro, S., and Losada, A. (2010). Cohesin ties up the genome. *Curr. Opin. Cell Biol.* 22, 781–787.

Chien, R., Zeng, W., Kawauchi, S., Bender, M. A., Santos, R., Gregson, H. C., Schmiesing, J. A., Newkirk, D., Kong, X., Ball, A. R. Jr., Calof, A. L., Lander, A. D., Groudine, M. T., and Yokomori, K. (2011). Cohesin mediates chromatin interactions that regulate mammalian {beta}-globin expression. *J. Biol. Chem.* 286, 17870–17878.

Ciccia, A., and Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* 40, 179–204.

Craig, J. M., and Choo, K. H. (2005). Kiss and break up – a safe passage to anaphase in mitosis and meiosis. *Chromosoma* 114, 252–262.

Cuadrado, A., Remeseiro, S., Gomez-Lopez, G., Pisano, D. G., and Losada, A. (2012). The specific contributions of cohesin-SA1 to cohesion and gene expression: implications for cancer and development. *Cell Cycle* 11, 2233–2238.

de Lange, C. (1933). Sur un type nouveau de dégénération (typus Amstelodamensis). *Archives de médecine des enfants* 36, 713–719.

De Piccoli, G., Torres-Rosell, J., and Aragon, L. (2009). The unnamed complex: what do we know about Smc5-Smc6? *Chromosome Res.* 17, 251–263.

Deardorff, M., Bando, M., Nakato, R., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K., De Baere, E., Decroos, C., Ernst, S., Francey, L., Gyftodimou, Y., Hirashima, K., Hullings, M., Ishikawa, Y., Kaur, M., Kiyono, T., Lombardi, P., Mortier, G., Nozaki, N., Petersen, M., Seimiya, H., Siu, V., Suzuki, Y., Takagaki, K., Tyshchenko, N., Wilde, J., Willems, P., Gillessen-Kaesbach, G., Christianson, D., Kaiser, F., Jackson, L., Hirota, T., Krantz, I., and Shirahige, K. (2012a). HDAC8 mutations in Cornelia de Lange syndrome provide insight into the cohesin acetylation cycle. *Nature.* doi:10.1038/nature11316

Deardorff, M. A., Wilde, J. J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Monnich, M., Yan, Y., Xu, W., Gil-Rodriguez, M. C., Clark, D., Hakonarson, H., Halbach, S., Michelis, L. D., Rampuria, A., Rossier, E., Spranger, S., Van Maldergem, L., Lynch, S. A., Gillessen-Kaesbach, G., Ludecke, H. J., Ramsay, R. G., Mckay, M. J., Krantz, I. D., Xu, H., Horsfield, J. A., and Kaiser, F. J. (2012b). RAD21 mutations cause a human cohesinopathy. *Am. J. Hum. Genet.* 90, 1014–1027.

Deardorff, M. A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A. D., Wilson, M., Lillquist, K., Siu, V., Ramos, F. J., Musio, A., Jackson, L. S., Dorsett, D., and Krantz, I. D. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *Am. J. Hum. Genet.* 80, 485–494.

Dekker, J. (2012). CTCF and cohesin help neurons raise their self-awareness. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8799–9599.

Denes, V., Pilichowska, M., Makarovskiy, A., Carpinito, G., and Geck, P. (2010). Loss of a cohesin-linked suppressor APRIN (Pds5b) disrupts stem cell programs in embryonal carcinoma: an emerging cohesin role in tumor suppression. *Oncogene* 29, 3446–3498.

Dorsett, D. (2009). Cohesin, gene expression and development: lessons from Drosophila. *Chromosome Res.* 17, 185–200.

Dorsett, D. (2010). Gene regulation: the cohesin ring connects developmental highways. *Curr. Biol.* 20, R886–R888.

Dorsett, D. (2011). Cohesin: genomic insights into controlling gene transcription and development. *Curr. Opin. Genet. Dev.* 21, 199–206.

Dorsett, D., and Strom, L. (2012). The ancient and evolving roles of cohesin in gene expression and DNA repair. *Curr. Biol.* 22, R240–R250.

Eilers, M., and Eisenman, R. N. (2008). Myc's broad reach. *Genes Dev.* 22, 2755–2766.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., Kitsukawa, T., Hamada, S., and Yagi, T. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* 37, 171–176.

Fay, A., Misulovin, Z., Li, J., Schaaf, C. A., Gause, M., Gilmour, D. S., and Dorsett, D. (2011). Cohesin selectively binds and regulates genes with paused RNA polymerase. *Curr. Biol.* 21, 1624–1634.

Frank, M., and Kemler, R. (2002). Protocadherins. *Curr. Opin. Cell Biol.* 14, 557–562.

Gandhi, R., Gillespie, P. J., and Hirano, T. (2006). Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr. Biol.* 16, 2406–2417.

Gause, M., Misulovin, Z., Bilyeu, A., and Dorsett, D. (2010). Dosage-sensitive regulation of cohesin chromosome binding and dynamics by Nipped-B, Pds5, and Wapl. *Mol. Cell. Biol.* 30, 4940–4951.

Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M., and Ellenberg, J. (2006). Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Curr. Biol.* 16, 1571–1578.

Gondor, A., and Ohlsson, R. (2008). Chromatin insulators and cohesins. *EMBO Rep.* 9, 327–329.

Gordillo, M., Vega, H., and Jabs, E. W. (1993). "Roberts syndrome," in *GeneReviews*, eds R. A. Pagon, T. D. Bird, C. R. Dolan, K. Stephens and M. P. Adam (Seattle: University of Washington). Available at: http://www.ncbi.nlm.nih.gov/books/NBK1153/

Gordillo, M., Vega, H., Trainer, A. H., Hou, F., Sakai, N., Luque, R., Kayserili, H., Basaran, S., Skovby, F., Hennekam, R. C., Uzielli, M. L., Schnur, R. E., Manouvrier, S., Chang, S., Blair, E., Hurst, J. A., Forzano, F., Meins, M., Simola, K. O., Raas-Rothschild, A., Schultz, R. A., McDaniel, L. D., Ozono, K., Inui, K., Zou, H., and Jabs, E. W. (2008). The molecular mechanism underlying Roberts syndrome involves loss of ESCO2 acetyltransferase activity. *Hum. Mol. Genet.* 17, 2172–2180.

Guacci, V., Koshland, D., and Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* 91, 47–57.

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410–413.

Haering, C. H., Farcas, A. M., Arumugam, P., Metson, J., and Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. *Nature* 454, 297–301.

Hauf, S., Roitinger, E., Koch, B., Dittrich, C. M., Mechtler, K., and Peters, J. M. (2005). Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol.* 3, e69. doi:10.1371/journal.pbio.0030069

Heidinger-Pauli, J. M., Unal, E., Guacci, V., and Koshland, D. (2008). The kleisin subunit of cohesin dictates damage-induced cohesion. *Mol. Cell* 31, 47–56.

Heidinger-Pauli, J. M., Unal, E., and Koshland, D. (2009). Distinct targets of the Eco1 acetyltransferase modulate cohesion in S phase and in response to DNA damage. *Mol. Cell* 34, 311–321.

Higashi, T. L., Ikeda, M., Tanaka, H., Nakagawa, T., Bando, M., Shirahige, K., Kubota, Y., Takisawa, H., Masukata, H., and Takahashi, T. S. (2012). The prereplication complex recruits XEco2 to chromatin to promote cohesin acetylation in Xenopus egg extracts. *Curr. Biol.* 22, 977–988.

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell Biol.* 7, 311–322.

Hornig, N. C., Knowles, P. P., Mcdonald, N. Q., and Uhlmann, F. (2002). The dual mechanism of separase regulation by securin. *Curr. Biol.* 12, 973–982.

Horsfield, J., Anagnostou, S., Hu, J. K. H., Cho, K. H.-Y., Geisler, R., Lieschke, G., Crosier, K., and Crosier, P. (2007). Cohesin-dependent regulation of runx genes. *Development* 134, 2639–2649.

Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3651–3656.

Hou, F., and Zou, H. (2005). Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. *Mol. Biol. Cell* 16, 3908–3918.

Hu, B., Itoh, T., Mishra, A., Katoh, Y., Chan, K. L., Upcher, W., Godlee, C., Roig, M. B., Shirahige, K., and Nasmyth, K. (2011). ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Curr. Biol.* 21, 12–24.

Huang, C. E., Milutinovich, M., and Koshland, D. (2005). Rings, bracelet or snaps: fashionable alternatives for Smc complexes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 537–542.

Ireland, M., Donnai, D., and Burn, J. (1993). Brachmann-de Lange syndrome. Delineation of the clinical phenotype. *Am. J. Med. Genet.* 47, 959–964.

Ivanov, D., Schleiffer, A., Eisenhaber, F., Mechtler, K., Haering, C. H., and Nasmyth, K. (2002). Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion. *Curr. Biol.* 12, 323–328.

Jackson, L., Kline, A. D., Barr, M. A., and Koch, S. (1993). de Lange syndrome: a clinical review of 310 individuals. *Am. J. Med. Genet.* 47, 940–946.

Jessberger, R. (2009). Cohesin's dual role in the DNA damage response: repair and checkpoint activation. *EMBO J.* 28, 2491–2494.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., Van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., and Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.

Kawauchi, S., Calof, A. L., Santos, R., Lopez-Burks, M. E., Young, C. M., Hoang, M. P., Chua, A., Lao, T., Lechner, M. S., Daniel, J. A., Nussenzweig, A., Kitzes, L., Yokomori, K., Hallgrimsson, B., and Lander, A. D. (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange syndrome. *PLoS Genet.* 5, e1000650. doi:10.1371/journal.pgen.1000650

Kenna, M. A., and Skibbens, R. V. (2003). Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. *Mol. Cell. Biol.* 23, 2999–3007.

Kim, B. J., Li, Y., Zhang, J., Xi, Y., Yang, T., Jung, S. Y., Pan, X., Chen, R., Li, W., Wang, Y., and Qin, J. (2010a). Genome-wide reinforcement of cohesin binding at pre-existing cohesin sites in response to ionizing radiation in human cells. *J. Biol. Chem.* 285, 22784–22792.

Kim, H. S., Baek, K. H., Ha, G. H., Lee, J. C., Kim, Y. N., Lee, J., Park, H. Y., Lee, N. R., Lee, H., Cho, Y., and Lee, C. W. (2010b). The hsSsu72 phosphatase is a cohesin-binding protein that regulates the resolution of sister chromatid arm cohesion. *EMBO J.* 29, 3544–3557.

Kim, Y. J., Cecchini, K. R., and Kim, T. H. (2011). Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7391–7396.

Krantz, I. D., Mccallum, J., Descipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., Nowaczyk, M. J., Toriello, H., Bamshad, M. J., Carey, J. C., Rappaport, E., Kawauchi, S., Lander, A. D., Calof, A. L., Li, H. H., Devoto, M., and Jackson, L. G. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat. Genet.* 36, 631–635.

Kueng, S., Hegemann, B., Peters, B. H., Lipp, J. J., Schleiffer, A., Mechtler, K., and Peters, J. M. (2006). Wapl controls the dynamic association of cohesin with chromatin. *Cell* 127, 955–967.

Lafont, A. L., Song, J., and Rankin, S. (2010). Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20364–20369.

Lara-Pezzi, E., Pezzi, N., Prieto, I., Barthelemy, I., Carreiro, C., Martinez, A., Maldonado-Rodriguez, A., Lopez-Cabrera, M., and Barbero, J. L. (2004). Evidence of a transcriptional co-activator function of cohesin STAG/SA/Scc3. *J. Biol. Chem.* 279, 6553–6559.

Lau, A., Blitzblau, H., and Bell, S. P. (2002). Cell-cycle control of the establishment of mating-type silencing in S. cerevisiae. *Genes Dev.* 16, 2935–2945.

Leem, Y. E., Choi, H. K., Jung, S. Y., Kim, B. J., Lee, K. Y., Yoon, K., Qin, J., Kang, J. S., and Kim, S. T. (2011). Esco2 promotes neuronal differentiation by repressing Notch signaling. *Cell. Signal.* 23, 1876–1884.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* 430, 573–578.

Liu, J., and Krantz, I. D. (2008). Cohesin and human disease. *Annu. Rev. Genomics Hum. Genet.* 9, 303–320.

Liu, J., and Krantz, I. D. (2009). Cornelia de Lange syndrome, cohesin, and beyond. *Clin. Genet.* 76, 303–314.

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., Spinner, N. B., Vega, H., Jackson, L. G., Shirahige, K., and Krantz, I. D. (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol.* 7, e1000119. doi:10.1371/journal.pbio.1000119

Losada, A., Hirano, M., and Hirano, T. (2002). Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. *Genes Dev.* 16, 3004–3016.

Mannini, L., Liu, J., Krantz, I. D., and Musio, A. (2010). Spectrum and consequences of SMC1A mutations: the unexpected involvement of a core component of cohesin in human disease. *Hum. Mutat.* 31, 5–10.

McNairn, A. J., and Gerton, J. L. (2008). Cohesinopathies: one ring, many obligations. *Mutat. Res.* 647, 103–111.

Mehta, G. D., Rizvi, S. M., and Ghosh, S. K. (2012). Cohesin: a guardian of genome integrity. *Biochim. Biophys. Acta.* 1823, 1324–1342.

Merkenschlager, M. (2010). Cohesin: a global player in chromosome biology with local ties to gene regulation. *Curr. Opin. Genet. Dev.* 20, 555–561.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* 91, 35–45.

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., and Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* 28, 1234–1245.

Misulovin, Z., Schwartz, Y. B., Li, X. Y., Kahn, T. G., Gause, M., Macarthur, S., Fay, J. C., Eisen, M. B., Pirrotta, V., Biggin, M. D., and Dorsett, D. (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. *Chromosoma* 117, 89–102.

Moldovan, G. L., Pfander, B., and Jentsch, S. (2006). PCNA controls establishment of sister chromatid cohesion during S phase. *Mol. Cell* 23, 723–732.

Monahan, K., Rudnick, N., Kehayova, P., Pauli, F., Newberry, K., Myers, R., and Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin

in the generation of single-cell diversity of Protocadherin-Œ± gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9125–9155.

Monnich, M., Banks, S., Eccles, M., Dickinson, E., and Horsfield, J. (2009). Expression of cohesin and condensin genes during zebrafish development supports a non-proliferative role for cohesin. *Gene Expr. Patterns* 9, 586–594.

Monnich, M., Kuriger, Z., Print, C. G., and Horsfield, J. A. (2011). A zebrafish model of Roberts syndrome reveals that esco2 depletion interferes with development by disrupting the cell cycle. *PLoS ONE* 6, e20051. doi:10.1371/journal.pone.0020051

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat. Genet.* 38, 528–530.

Muto, A., Calof, A. L., Lander, A. D., and Schilling, T. F. (2011). Multifactorial origins of heart and gut defects in nipbl-deficient zebrafish, a model of Cornelia de Lange syndrome. *PLoS Biol.* 9, e1001181. doi:10.1371/journal.pbio.1001181

Nasmyth, K. (2011). Cohesin: a catenase with separate entry and exit gates? *Nat. Cell Biol.* 13, 1170–1177.

Nasmyth, K., and Haering, C. H. (2009). Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* 43, 525–558.

Nativio, R., Wendt, K. S., Ito, Y., Huddleston, J. E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J. M., and Murrell, A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.* 5, e1000739. doi:10.1371/journal.pgen.1000739

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A. A., Mechtler, K., and Peters, J. M. (2010). Sororin mediates sister chromatid cohesion by antagonizing wapl. *Cell* 143, 737–749.

Nitzsche, A., Paszkowski-Rogacz, M., Matarese, F., Janssen-Megens, E. M., Hubner, N. C., Schulz, H., De Vries, I., Ding, L., Huebner, N., Mann, M., Stunnenberg, H. G., and Buchholz, F. (2011). RAD21 Cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS ONE* 6, e19470. doi:10.1371/journal.pone.0019470

Opitz, J. M. (1985). The Brachmann-de Lange syndrome. *Am. J. Med. Genet.* 22, 89–102.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B. S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A. G., and Merkenschlager, M. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433.

Pauli, A., Althoff, F., Oliveira, R. A., Heidmann, S., Schuldiner, O., Lehner, C. F., Dickson, B. J., and Nasmyth, K. (2008). Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Dev. Cell* 14, 239–251.

Pauli, A., Van Bemmel, J. G., Oliveira, R. A., Itoh, T., Shirahige, K., Van Steensel, B., and Nasmyth, K. (2010). A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands. *Curr. Biol.* 20, 1787–1798.

Rankin, S., Ayad, N. G., and Kirschner, M. W. (2005). Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. *Mol. Cell* 18, 185–200.

Remeseiro, S., Cuadrado, A., Carretero, M., Martinez, P., Drosopoulos, W. C., Canamero, M., Schildkraut, C. L., Blasco, M. A., and Losada, A. (2012a). Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. *EMBO J.* 31, 2076–2089.

Remeseiro, S., Cuadrado, A., Gomez-Lopez, G., Pisano, D. G., and Losada, A. (2012b). A unique role of cohesin-SA1 in gene regulation and development. *EMBO J.* 31, 2090–2102.

Revenkova, E., Focarelli, M. L., Susani, L., Paulis, M., Bassi, M. T., Mannini, L., Frattini, A., Delia, D., Krantz, I., Vezzoni, P., Jessberger, R., and Musio, A. (2009). Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. *Hum. Mol. Genet.* 18, 418–427.

Rhodes, J. M., Bentley, F. K., Print, C. G., Dorsett, D., Misulovin, Z., Dickinson, E. J., Crosier, K. E., Crosier, P. S., and Horsfield, J. A. (2010). Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. *Dev. Biol.* 344, 637–649.

Rhodes, J. M., Mcewan, M., and Horsfield, J. A. (2011). Gene regulation by cohesin in cancer: is the ring an unexpected party to proliferation? *Mol. Cancer Res.* 9, 1587–1607.

Rohatgi, S., Clark, D., Kline, A. D., Jackson, L. G., Pie, J., Siu, V., Ramos, F. J., Krantz, I. D., and Deardorff, M. A. (2010). Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey. *Am. J. Med. Genet. A* 152A, 1641–1653.

Rollins, R. A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. *Mol. Cell. Biol.* 24, 3100–3111.

Rollins, R. A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* 152, 577–593.

Rowland, B. D., Roig, M. B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., Beckouet, F., Underwood, P., Metson, J., Imre, R., Mechtler, K., Katis, V. L., and Nasmyth, K. (2009). Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. *Mol. Cell* 33, 763–774.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8309–8314.

Salah, S. M., and Nasmyth, K. (2000). Destruction of the securin Pds1p occurs at the onset of anaphase during both meiotic divisions in yeast. *Chromosoma* 109, 27–34.

Schaaf, C. A., Misulovin, Z., Sahota, G., Siddiqui, A. M., Schwartz, Y. B., Kahn, T. G., Pirrotta, V., Gause, M., and Dorsett, D. (2009). Regulation of the Drosophila enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. *PLoS ONE* 4, e6202. doi:10.1371/journal.pone.0006202

Schmidt, D., Schwalie, P., Ross-Innes, C. S., Hurtado, A., Brown, G., Carroll, J., Flicek, P., and Odom, D. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* 20, 578–588.

Schmitz, J., Watrin, E., Lenart, P., Mechtler, K., and Peters, J. M. (2007). Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr. Biol.* 17, 630–636.

Schuldiner, O., Berdnik, D., Levy, J. M., Wu, J. S., Luginbuhl, D., Gontang, A. C., and Luo, L. (2008). piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev. Cell* 14, 227–238.

Schule, B., Oviedo, A., Johnston, K., Pai, S., and Francke, U. (2005). Inactivating mutations in ESCO2 cause SC phocomelia and Roberts syndrome: no phenotype-genotype correlation. *Am. J. Hum. Genet.* 77, 1117–1128.

Seitan, V. C., Banks, P., Laval, S., Majid, N. A., Dorsett, D., Rana, A., Smith, J., Bateman, A., Krpic, S., Hostert, A., Rollins, R. A., Erdjument-Bromage, H., Tempst, P., Benard, C. Y., Hekimi, S., Newbury, S. F., and Strachan, T. (2006). Metazoan Scc4 homologs link sister chromatid cohesion to cell and axon migration guidance. *PLoS Biol.* 4, e242. doi:10.1371/journal.pbio.0040242

Shintomi, K., and Hirano, T. (2009). Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. *Genes Dev.* 23, 2224–2236.

Skibbens, R. V. (2010). Buck the establishment: reinventing sister chromatid cohesion. *Trends Cell Biol.* 20, 507–513.

Skibbens, R. V., Corson, L. B., Koshland, D., and Hieter, P. (1999). Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev.* 13, 307–319.

Strom, L., Karlsson, C., Lindroos, H. B., Wedahl, S., Katou, Y., Shirahige, K., and Sjogren, C. (2007). Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* 317, 242–245.

Strom, L., Lindroos, H. B., Shirahige, K., and Sjogren, C. (2004). Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell* 16, 1003–1015.

Strom, L., and Sjogren, C. (2007). Chromosome segregation and double-strand break repair – a complex connection. *Curr. Opin. Cell Biol.* 19, 344–349.

Sutani, T., Kawaguchi, T., Kanno, R., Itoh, T., and Shirahige, K. (2009). Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. *Curr. Biol.* 19, 492–497.

Teo, A. K., Arnold, S. J., Trotter, M. W., Brown, S., Ang, L. T., Chng, Z., Robertson, E. J., Dunn, N. R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* 25, 238–250.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly

Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* 36, 636–641.

Unal, E., Heidinger-Pauli, J. M., Kim, W., Guacci, V., Onn, I., Gygi, S. P., and Koshland, D. E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. *Science* 321, 566–569.

van der Lelij, P., Chrzanowska, K. H., Godthelp, B. C., Rooimans, M. A., Oostra, A. B., Stumm, M., Zdzienicka, M. Z., Joenje, H., and De Winter, J. P. (2010). Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1. *Am. J. Hum. Genet.* 86, 262–266.

van der Lelij, P., Godthelp, B. C., van Zon, W., van Gosliga, D., Oostra, A. B., Steltenpool, J., de Groot, J., Scheper, R. J., Wolthuis, R. M., Waisfisz, Q., Darroudi, F., Joenje, H., and de Winter, J. P. (2009). The cellular phenotype of Roberts syndrome fibroblasts as revealed by ectopic expression of ESCO2. *PLoS ONE* 4, e6936. doi:10.1371/journal.pone.0006936

Vaur, S., Feytout, A., Vazquez, S., and Javerzat, J. P. (2012). Pds5 promotes cohesin acetylation and stable cohesin-chromosome interaction. *EMBO Rep.* 13, 645–652.

Vega, H., Trainer, A. H., Gordillo, M., Crosier, M., Kayserili, H., Skovby, F., Uzielli, M. L., Schnur, R. E., Manouvrier, S., Blair, E., Hurst, J. A., Forzano, F., Meins, M., Simola, K. O., Raas-Rothschild, A., Hennekam, R. C., and Jabs, E. W. (2010). Phenotypic variability in 49 cases of ESCO2 mutations, including novel missense and codon deletion in the acetyltransferase domain, correlates with ESCO2 expression and establishes the clinical criteria for Roberts syndrome. *J. Med. Genet.* 47, 30–37.

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., Van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., Jabs, E. W., Inui, K., and Joenje, H. (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nat. Genet.* 37, 468–470.

Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Sudhof, T. C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035–1041.

Waizenegger, I., Gimenez-Abian, J. F., Wernic, D., and Peters, J. M. (2002). Regulation of human separase by securin binding and autocleavage. *Curr. Biol.* 12, 1368–1378.

Waizenegger, I. C., Hauf, S., Meinke, A., and Peters, J. M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* 103, 399–410.

Wang, X., and Dai, W. (2005). Shugoshin, a guardian for sister chromatid segregation. *Exp. Cell Res.* 310, 1–9.

Watrin, E., and Peters, J. M. (2009). The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *EMBO J.* 28, 2625–2635.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., and Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796–801.

Whelan, G., Kreidl, E., Peters, J. M., and Eichele, G. (2012a). The non-redundant function of cohesin acetyltransferase Esco2: some answers and new questions. *Nucleus* 3. PMID: 22614755. [Epub ahead of print].

Whelan, G., Kreidl, E., Wutz, G., Egner, A., Peters, J. M., and Eichele, G. (2012b). Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. *EMBO J.* 31, 71–82.

Wu, N., and Yu, H. (2012). The Smc complexes in DNA damage response. *Cell Biosci.* 2, 5.

Xiong, B., Lu, S., and Gerton, J. L. (2010). Hos1 is a lysine deacetylase for the Smc3 subunit of cohesin. *Curr. Biol.* 18, 1660–1665.

Xu, H., Balakrishnan, K., Malaterre, J., Beasley, M., Yan, Y., Essers, J., Appeldoorn, E., Thomaszewski, J. M., Vazquez, M., Verschoor, S., Lavin, M. F., Bertonchello, I., Ramsay, R. G., and Mckay, M. J. (2010). Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice. *PLoS ONE* 5, e12112. doi:10.1371/journal.pone.0012112

Xu, H., Yan, M., Patra, J., Natrajan, R., Yan, Y., Swagemakers, S., Tomaszewski, J. M., Verschoor, S., Millar, E. K., Van der Spek, P., Reis-Filho, J. S., Ramsay, R. G., O'Toole, S. A., Mcneil, C. M., Sutherland, R. L., Mckay, M. J., and Fox, S. B. (2011). Enhanced RAD21 cohesin expression confers poor prognosis and resistance to chemotherapy in high grade luminal, basal and HER2 breast cancers. *Breast Cancer Res.* 13, R9.

Zhang, B., Chang, J., Fu, M., Huang, J., Kashyap, R., Salavaggione, E., Jain, S., Shashikant, K., Deardorff, M. A., Uzielli, M. L., Dorsett, D., Beebe, D. C., Jay, P. Y., Heuckeroth, R. O., Krantz, I., and Milbrandt, J. (2009). Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies. *PLoS ONE* 4, e5232. doi:10.1371/journal.pone.0005232

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R. O., Erlich, J. M., Jay, P. Y., and Milbrandt, J. (2007). Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. *Development* 134, 3191–3201.

Zhang, J., Shi, X., Li, Y., Kim, B. J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S. Y., Wang, Y., Zhang, P., Kim, S. T., Pan, X., and Qin, J. (2008a). Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. *Mol. Cell* 31, 143–151.

Zhang, N., Kuznetsov, S. G., Sharan, S. K., Li, K., Rao, P. H., and Pati, D. (2008b). A handcuff model for the cohesin complex. *J. Cell Biol.* 183, 1019–1031.

Conflict of Interest Statement: 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.

Received: 08 July 2012; paper pending published: 23 July 2012; accepted: 17 August 2012; published online: 12 September 2012.

Citation: Horsfield JA, Print CG and Mönnich M (2012) Diverse developmental disorders from The One Ring: distinct molecular pathways underlie the cohesinopathies. Front. Gene. ***3****:171. doi: 10.3389/fgene.2012.00171*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

REVIEW ARTICLE
published: 17 October 2012
doi: 10.3389/fgene.2012.00217

Chromatin loops, gene positioning, and gene expression

Sjoerd Holwerda and Wouter de Laat*

Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, University Medical Center Utrecht, Utrecht, Netherlands

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Piroska E. Szabo, Beckman Research Institute of City of Hope, USA
Michel Cogne, Limoges University/Centre National de la Recherche Scientifique, France

***Correspondence:**
Wouter de Laat, Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, Netherlands.
e-mail: w.delaat@hubrecht.eu

Technological developments and intense research over the last years have led to a better understanding of the 3D structure of the genome and its influence on genome function inside the cell nucleus. We will summarize topological studies performed on four model gene loci: the α- and β-globin gene loci, the antigen receptor loci, the imprinted H19–Igf2 locus and the Hox gene clusters. Collectively, these studies show that regulatory DNA sequences physically contact genes to control their transcription. Proteins set up the 3D configuration of the genome and we will discuss the roles of the key structural organizers CTCF and cohesin, the nuclear lamina and the transcription machinery. Finally, genes adopt non-random positions in the nuclear interior. We will review studies on gene positioning and propose that cell-specific genome conformations can juxtapose a regulatory sequence on one chromosome to a responsive gene on another chromosome to cause altered gene expression in subpopulations of cells.

Keywords: chromatin domains, gene expression, nuclear organization, genome structure, nuclear periphery

INTRODUCTION

Only a few percent of the 3.2 billion base pairs of our genome is coding sequence. The remainder is intronic and intergenic sequences, long considered to be junk DNA, but now realized to contain hundreds of thousands of sequence modules with the potential to regulate gene expression (Shen et al., 2012). This greatly outnumbers the ∼25,000 genes that we carry in our genome. For the great majority of regulatory sites we do not know though whether they really exert a function *in vivo* and, if so, to which target gene they direct their activity. Studies into the shape of our genome provided evidence that regulatory DNA sequences can control transcription over distance by physically contacting target genes via chromatin looping. Initially such work was primarily done on individual gene loci. We will highlight findings on some of the most studied model gene systems, including the α- and β-globin gene loci, the immunoglobulin and other antigen receptor gene loci, the imprinted H19–Igf2 locus and the Hox gene clusters. Collectively, these studies showed how local DNA topology can change dynamically in time and place to accommodate developmental gene expression. It also uncovered some of the trans-acting factors that fold the chromatin. We will discuss the role of the nuclear lamina, CTCF, cohesin, and RNA polymerase II (RNAPII), being currently the most intensively studied general organizers of chromosome topology. Collectively, all studies emphasize the relationship between genome structure and genome function. Consensus seems to have reached now for shape being crucial for function within the ∼1 Mb scale. Here, regulatory sequences need to physically get in contact with genes to control their transcription. Beyond this level of organization, it is not as obvious how relevant the nuclear position and/or genomic environment of genes will be. Studies manipulating the nuclear location of genes start to provide insight in this and will be discussed. Finally, we propose that the probabilistic nature of nuclear positioning implies that we need to move from cell population-based to single cell studies to understand how remote genomic sequences can influence each other's function.

FUNCTIONALLY RELEVANT DNA INTERACTIONS BETWEEN GENES AND REGULATORY SEQUENCES

The realization that sequence information required for proper gene expression may sometimes reside at a large chromosomal distance away from the gene body came from observations in patients, showing that the deletion of sequences away from the β-globin genes proper caused thalassemia (Kleinjan and van Heyningen, 2005). For a long time, the mechanisms behind long-range gene activation remained enigmatic. Although still not entirely understood it is now clear that it involves physical contacts between such remote regulatory sequences and the genes that they control. This discovery relied mostly on the development of chromosome conformation capture (3C) technology, a method invented 10 years ago (Dekker et al., 2002) that allows quantitative measurements of DNA contact frequencies between pairs of selected genomic sites. Here, we will highlight observations made by 3C technology on four gene clusters (the globin gene loci, the antigen receptor loci, the imprinted H19–Igf2 locus and the Hox gene loci) that serve as model systems for varying types of gene regulation.

THE α- AND β-GLOBIN LOCI

Early evidence for chromatin looping being involved in mammalian gene regulation comes from studies on the β-globin locus. This is perhaps unsurprising as the globin loci have always been the subject of intense gene expression studies: their misregulation underlies thalassemia and the α- and β-globin genes serve as model systems to study developmental gene regulation. As pointed out, the observation that the deletion of sequences away from, but not affecting, the genes proper caused thalassemia (Van der Ploegh et al., 1980) first suggested that gene transcription was controlled by remote regulatory sequences. A series of remote regulatory sites were then demonstrated to exist in

these loci, the most important ones in the β-globin locus collectively referred to as a locus control region (LCR). The LCR controls expression of multiple β-globin genes which are arranged on the chromosome in order of their timed expression during development: embryonic β-globin genes are closest to and adult genes are furthest away from the LCR (**Figure 1A**). Proximity on the linear DNA template therefore clearly matters, but the exact mode of LCR action over distance long remained elusive. 3D proximity was implicated in transcription regulation when it was found that linear proximity is no longer important when two genes are positioned together at a large distance from the LCR (Hanscombe et al., 1991; Dillon et al., 1997). In 2002, first direct evidence for chromatin looping and spatial contacts between the LCR and an active β-globin gene was obtained, in studies using RNA TRAP (Carter et al., 2002) and 3C technology (Tolhuis et al., 2002). 3C technology in particular appeared extremely useful for further investigations on the topology of the β-globin locus.

The 3D configuration of the β-globin locus was found to dynamically follow the changes in gene expression that occur during development and during red blood cell differentiation. LCR–gene contacts are not detectable in tissue where the globins are inactive. During development, the LCR switches its contacts from embryonic to adult β-globin genes to ensure their activation at the appropriate developmental stage (Palstra et al., 2003). Proteins were shown to set up the chromatin loops in the locus. Transcription factors such as EKLF, GATA1, and Ldb1, that are important for proper globin gene expression and that bind to both the LCR and gene promoter regions, all appear necessary for stable LCR–gene interactions (Drissen et al., 2004; Vakoc et al., 2005; Song et al., 2007). Another transcription factor, CTCF, forms chromatin loops between binding sites surrounding the locus (**Figure 1A**). These CTCF-mediated loops precede LCR–gene contacts during red blood cell maturation (Palstra et al., 2003). The spatial entity formed in red blood cells as a consequence of LCR–gene and CTCF-mediated DNA interactions was referred to as an active chromatin hub (Tolhuis et al., 2002).

An outstanding question is whether gene activity follows locus conformation or vice versa. The inhibition of transcription was found to not change the chromatin loops, suggesting that function follows structure in the β-globin locus (Mitchell and Fraser, 2008; Palstra et al., 2008). More direct evidence that transcriptional enhancement is a consequence of looping has recently been provided. Ldb1 requires GATA1 for recruitment to the β-globin promoter, but binds to the LCR in a GATA1 independent manner. In an elegant assay employing artificial zinc fingers (ZFs) in GATA1-null cells, the tethering of ZF-Ldb1 to the β-globin promoter was shown to induce LCR–gene contacts and chromatin looping, and to activate β-globin gene expression. Without the LCR, loops were absent and gene expression was not activated (Deng et al., 2012). This data supports the idea that looping toward target genes is crucial for distal enhancers to activate transcription. Interestingly, a truncated version of Ldb1 composed of only its self-association domain was already sufficient to induce chromatin looping and activate transcription initiation, suggesting that Ldb1 multimerization may stabilize contacts between remote globin DNA sequences.

Similar to the β-globin locus, the mammalian α-globin genes are controlled by distal enhancer elements (Sharpe et al., 1993; Gourdon et al., 1994; Higgs et al., 1998). Active histone marks and erythroid-specific transcription factors are present at the locus before the occupancy by RNAPII is measurable (Anguita et al., 2004), suggesting that there is a role for these factors in recruitment of RNA polymerases to the α-globin gene promoters. Looping of the key enhancer elements to the α-globin promoters, with intervening DNA sequences looping out, has been demonstrated (Vernimmen et al., 2007, 2009). Timing of looping coincides with the binding of the pre-initiation complex and elongation factors (Vernimmen et al., 2007). Protein factors like GATA1, Ldb1, and Sp/XKLF also bind to the α-globin genes and regulatory sequences, and can be expected to perform similar roles in chromatin looping and transcription regulation as seen for β-globin.

ANTIGEN RECEPTOR GENE LOCI

The immunoglobulin loci, which are active in B cells, and the T cell receptor (TCR) loci that are active in T cells, generally stretch over large chromosomal regions of up to 3 Mb and are subdivided into different regions (V, D, J, and C) that each contain multiple gene segments. Particularly the V region is often extremely large. DNA rearrangement via V(D)J recombination is required to combine the different gene segments and assemble a functional antigen receptor that is unique in every B or T cell (Jung and Alt, 2004). The RAG proteins carry out V(D)J recombination and need to physically hold together two target sequences to cut and paste them together (Schatz and Ji, 2011). The 3D topology of the antigen receptor loci therefore must play a role in their regulation. 3D FISH studies were originally performed to search for topological features of the recombining loci. Indeed it was shown that the two ends of the receptor loci spatially come together prior to rearrangement (Kosak et al., 2002; Fuxa et al., 2004). The simultaneous visualization of intervening sequences then allowed demonstrating that locus contraction was not just a consequence of compaction but the result of chromatin looping, with intervening sequences looping out (Roldan et al., 2005; Sayegh et al., 2005; Jhunjhunwala et al., 2008). Multiple proteins including Pax5, YY1, CTCF, cohesin, and ikaros have been implicated in the spatial organization of these gene loci. Initial evidence for this was based on the observation that their depletion reduced contraction of the locus and lead to altered usage of the V genes during recombination (Roldan et al., 2005; Sayegh et al., 2005; Liu et al., 2007; Reynaud et al., 2008; Degner et al., 2009). More recently, 3C-based evidence was provided for looping between CTCF and cohesin bound chromatin sites across the antigen receptor loci (**Figure 1B**). Long-range chromatin interactions with three regulatory sequences in particular, the 3′ regulatory region (3′RR), the Eμ-intronic enhancer and the recently discovered intergenic control region 1 (IGCR1), seem important for proper rearrangement of the IgH locus. These loops may facilitate the inclusion of distal Vgenes, thereby enhancing the diversity of choice in usage of coding V elements during V(D)J recombination (Degner et al., 2011; Guo et al., 2011a,b; Ribeiro de Almeida et al., 2011; Seitan et al., 2011). Additionally, CTCF and cohesin may regulate chromatin accessibility and transcription in sub-regions of the loci, thereby directing the recombination machinery. As was

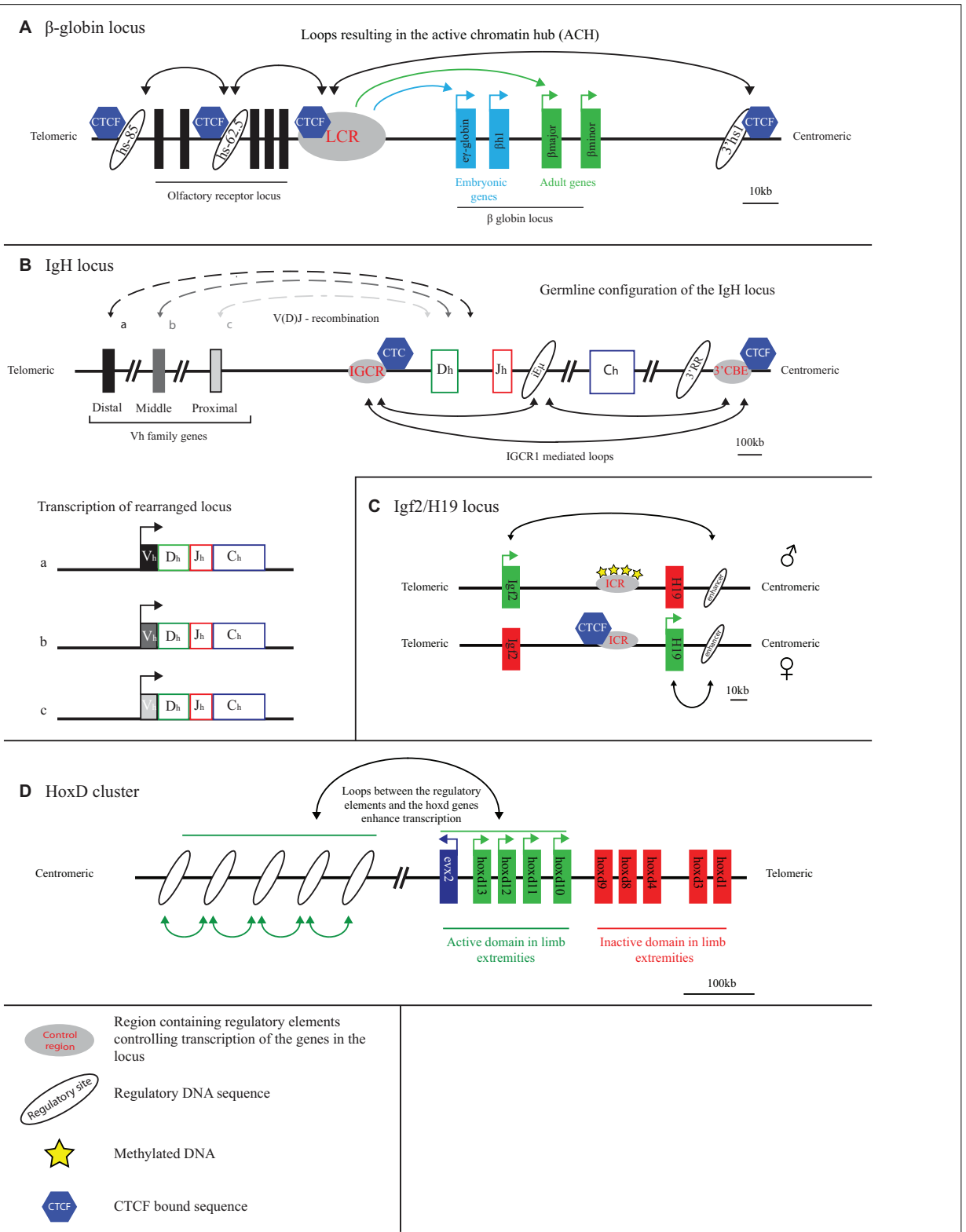


FIGURE 1 | Long-range transcriptional regulation at model gene loci. **(A)** At the active β-globin locus, LCR–gene contacts and interactions between flanking CTCF sites set up an active chromatin hub (ACH). **(B)** The IGCR1 contacts the 3′ regulatory region and the intronic enhancer of the IgH locus in pro-B cells. Inclusion of the distal V genes is influenced by the presence of the IGCR1. **(C)** CTCF blocks the interaction of the Igf2/H19 enhancer with the Igf2 gene on the maternal allele. Methylation of the ICR prevents CTCF binding and enables Igf2 expression from the paternal allele. **(D)** A "regulatory archipelago" controls the expression of the hoxd13–hoxd10 genes over distance in limb extremities.

pointed out, while multiple proteins that shape the conformation of the antigen receptor loci are known now, there is as yet no evidence that they act directly to promote synapsis between distal gene segments (Seitan and Merkenschlager, 2012). Whether such activity exists, or whether the overall spatial structure of the antigen receptor loci is already sufficient to direct such interactions and warrant usage of the full repertoire of gene segments, remains to be investigated.

H19/Igf2 LOCUS

The H19/Igf2 locus is an imprinted locus, with the H19 gene being expressed from the maternal and the Igf2 gene from the paternal allele. Both genes are under the control of a shared enhancer located on one side of the locus, 3′ of the H19 gene. The targeting of this enhancer to either one of the genes is determined by an imprinting control region (ICR) located in between Igf2 and H19 (Bartolomei et al., 1993; Ferguson-Smith et al., 1993; Leighton et al., 1995; Thorvaldsen et al., 1998). This ICR, which contains multiple CTCF binding sites, is methylated when paternally inherited and unmethylated when derived from the mother (Bartolomei et al., 1993; Ferguson-Smith et al., 1993). CTCF can only bind to the unmethylated, hence the maternally inherited, ICR (**Figure 1C**) (Bell and Felsenfeld, 2000; Hark et al., 2000).

Using an elegant approach that involved the site-specific integration of ectopic Gal-binding sites near the ICR it was shown that the ICR separates the H19 and the Igf2 gene in different chromatin compartments (Murrell et al., 2004). Because of the distinct capacity to bind CTCF, ICR contacts differ between the alleles such that enhancers are enabled to contact the Igf2 gene on the paternal allele but not on the maternal allele (Murrell et al., 2004). Subsequent studies based on 3C technology came to similar but not identical conclusions (Kurukuti et al., 2006; Yoon et al., 2007). Whereas one study reported bi-allelic interactions between the ICR and the enhancers (Kurukuti et al., 2006), another reported this interaction to be specific for the maternal allele. This study also showed that the CTCF-bound ICR promiscuously contacted enhancers and promoters, suggesting that such contacts are important for insulators to block effective enhancer–promoter communication (Yoon et al., 2007). In addition to its insulator function, the ICR appears required to initiate H19 gene expression: upon deletion of the four CTCF binding sites in the ICR, H19 transcripts were hardly detectable in the early embryo (Engel et al., 2006). In summary, studies on the H19/Igf2 locus confirm that gene competition for a shared enhancer involves competition for physical promoter–enhancer interactions. Moreover, they show that insulators bound by CTCF can hamper this interaction, possibly by physically competing for these contacts.

3D ORGANIZATION OF THE Hox GENES

When it comes to developmental gene regulation, the Hox gene clusters are among the most fascinating gene clusters. In mammals, four of these clusters are present (HoxA–D), each containing roughly a dozen genes that are expressed during development in a temporal and spatial manner that is co-linear with their genomic context (Kmita and Duboule, 2003). The HoxD gene cluster, but also other Hox clusters, is flanked on both sides by large gene-poor chromosomal regions. The Hox genes encode for transcription factors and are important for body axis formation as well as proper formation of the extremities. Correct spatiotemporal expression along the body axis appears controlled within the gene cluster proper, independent of surrounding gene sequences. As was shown by 4C technology, here the genes show little specific interactions with surrounding sequences, but fold into a distinct active and inactive compartment. When moving posteriorly along the axis, the number of genes contained within the active compartment increases, in agreement with their progressive activation and corresponding change of histone modifications (Noordermeer et al., 2011a). It was suggested that this topological separation can mediate the temporal expression pattern of the HoxD genes. In the extremities, in this case the developing limb bud, a different mechanism of transcriptional control is in place, with a correspondingly different 3D conformation of the gene cluster. The HoxD genes depend on distinct long-range regulatory sequences for their expression in the proximal and distal parts of the limb bud (**Figure 1D**). These sequences are present in the gene-poor regions located on the telomeric and centromeric side of the gene cluster, respectively (Spitz et al., 2003; Gonzalez et al., 2007). The active, much more than the inactive, HoxD genes loop toward these sides to contact the regulatory DNA sequences. Based on the DNA contact profiles of the active HoxD13 gene, as generated by 4C technology, new enhancers were identified in the gene desert that showed correct spatiotemporal reporter gene expression in transgenic mice (Montavon et al., 2011). The emerging picture from these studies is that Hox gene expression in the limb bud is under the control of a complex regulatory landscape with many enhancers spread over hundreds of kilobases of flanking DNA working in concert (Montavon et al., 2011). This picture seems confirmed by a recent high-resolution FISH study, which also revealed that further fine-tuning of the contacts between HoxD genes and flanking regulatory sequences takes place along the anterior–posterior axis of the limb bud (Williamson et al., 2012). A 5C analysis of the HoxA gene cluster in human primary fibroblasts taken from different anatomical sites revealed yet another dimension of Hox gene regulation. Contacts were identified with a site 5′ of the cluster that expresses a long intergenic non-coding RNA (lincRNA), named HOTTIP (Wang et al., 2011). HOTTIP RNA was reported to recruit proteins (WDR5) necessary to modify the histones and activate transcription of the genes contacted by the lincRNA locus (Wang et al., 2011). Thus, proper spatiotemporal Hox gene expression appears to be controlled by a very complex network of proximal and distal regulatory sequences that loop in a developmentally controlled manner toward specific Hox genes to physically confront them with activating protein and RNA molecules.

THE OVERALL SHAPE OF THE 3D GENOME

The initial 3C studies discussed above focused on individual genes and gene clusters, highlighting the functional importance of local chromatin loops and uncovering proteins that determine the topology of these gene loci (Splinter and de Laat, 2011). However, the genome is structurally organized also beyond the level of individual gene clusters. Original evidence that overall chromatin in the nucleus is not organized in a random fashion and that nuclear organization is related to transcriptional activity

comes from microscopy observations. It showed the separation of densely packed inactive chromatin and loosely packed active chromatin and demonstrated that chromosomes occupy individual chromosome territories (CTs; Branco and Pombo, 2006; Joffe et al., 2010). It also demonstrated that larger chromosomes tend to occupy more peripheral positions in the nucleus, while smaller ones often reside more in the nuclear interior. A recurrent theme in nuclear organization is that folding and positioning follow probabilistic rules. Thus, a given chromosome will have a preferred nuclear position, but this does not imply that it occupies this exact position in every cell (Bolzer et al., 2005). In other words: all genomes in a population of cells can be expected to fold according to the same probabilistic rules, yet every single cell likely has a different genome structure. Thanks to the development of more genome-wide versions of 3C technology (de Wit and de Laat, 2012; Dostie and Bickmore, 2012), the underlying, probabilistic, rules for genome folding are now rapidly being uncovered.

The most dominant force shaping the 3D genome seems the spatial separation between active and inactive chromatin. First observed under the microscope as a general feature of nuclear organization, it was then confirmed to also be relevant for the folding of individual chromosome segments (Shopland et al., 2006) and, at much higher resolution, for the genomic environments of individual genes (Simonis et al., 2006). The latter observation made by 4C technology for a few selected chromosomal sites was confirmed to apply to regions across the genome by recent Hi-C studies. In Hi-C, all versus all interactions of the genome are mapped, with the resolution of contact maps depending on the depth of sequencing, the size of the genome, and the complexity of the sample analyzed (Lieberman-Aiden et al., 2009; Yaffe and Tanay, 2011; Dixon et al., 2012; Kalhor et al., 2012). Hi-C studies showed that chromosomes are subdivided into topological domains that cover 0.2–1 Mb. The domains mark chromosomal regions within which DNA contacts are confined. They generally demarcate regions with a defined gene density and activity, and with corresponding chromatin accessibility, histone modifications, and replication timing. Preferred contacts among two types of topological domains are seen, the active and inactive topological domains, with the separation of active and inactive chromatin in the nucleus as a consequence (Lieberman-Aiden et al., 2009; Yaffe and Tanay, 2011; Dixon et al., 2012; Kalhor et al., 2012; Nora et al., 2012). In Drosophila in particular, an additional domain type hallmarked by the association of polycomb group (PcG) proteins is observed, which also shows preferred contacts with other PcG-bound topological domains (Tolhuis et al., 2011; Sexton et al., 2012). Marks for active chromatin (DNase I sensitivity, H3K4me1 and -me3, RNAPII) were enriched for regions showing also interchromosomal DNA contacts (Yaffe and Tanay, 2011; Kalhor et al., 2012), suggesting that open and active chromatin most easily reaches out of the CT. Boundaries of the domains were found enriched for CTCF, H3K4me1, transcriptional start sites (TSSs) and housekeeping genes, tRNA genes and SINE elements (Yaffe and Tanay, 2011; Dixon et al., 2012; Sexton et al., 2012). Interestingly, during cellular differentiation the topological domains appear to largely remain intact and structural changes mostly occur within the domains, suggesting that the domain boundaries are largely conserved between cell types (Dixon et al., 2012; **Figure 2**). The active and inactive compartments each seem to organize themselves independently. This was shown in studies on the active and inactive X chromosome in mammalian female

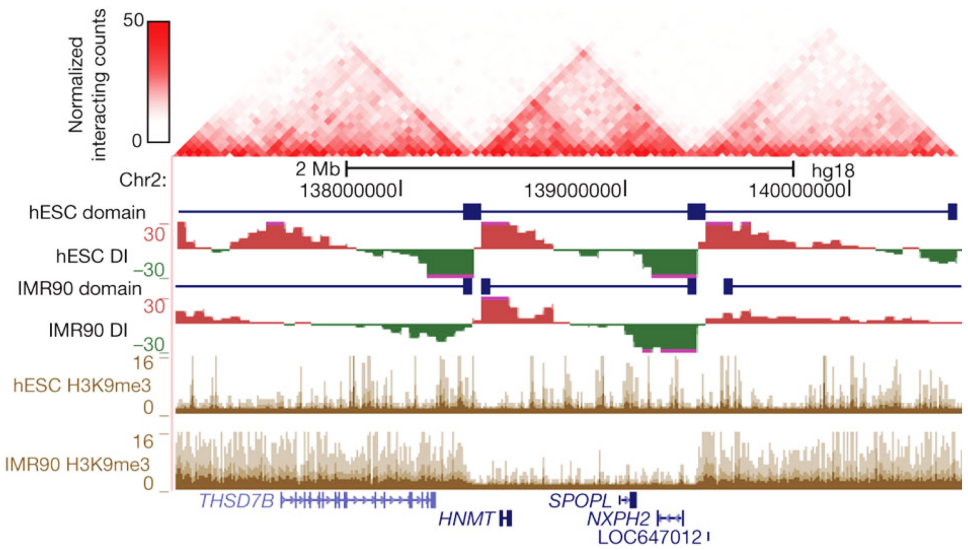


FIGURE 2 | Topological boundaries can act as barriers for spreading of heterochromatin. The 2D heat map shows the Hi-C interaction frequency in human ES cells. Underneath is indicated the directionality index (DI) in hESCs and IMR90 cells. The DI is a Hi-C measure showing a site's preference to engage in unidirectional contacts with downstream (red) or upstream (green) sequences. Borders of the topological domains are defined by a change in the directionality of interactions (transition from green to red). The UCSC Genome Browser shots show the distribution of H3K9me3, a measure for heterochromatin formation. Note that in IMR90 cells heterochromatin stops at the topological boundaries. Reprinted by permission from Macmillan Publishers Ltd (Dixon et al., 2012), copyright (2012).

cells, where the inactive X chromosome showed normal contacts between active chromatin regions but was found to specifically lack long-range contacts between inactive chromatin domains. Interestingly, these latter contacts were restored when the non-coding RNA Xist, which coats the inactive X chromosome, was deleted, implicating a role also for non-coding RNA in chromosome topology (Splinter et al., 2011).

Whether RNA plays a general role in the topological organization of chromosomes remains to be demonstrated. Proteins, however, are known to shape the configuration of the genome inside the cell. Nuclear lamina proteins, CTCF, cohesin, and RNAPII are best recognized as general organizers of the 3D genome and will be discussed below.

PROTEINS SHAPING THE GENOME

LAMINS AND THE NUCLEAR PERIPHERY

The nuclear periphery of mammalian cells is known to be enriched for inactive chromatin and to correlate with relatively low gene expression levels (Brown et al., 1997, 1999; Skok et al., 2001; Zink et al., 2004). The inner part of the nuclear membrane is coated with a protein network called the nuclear lamina. Lamina-associated domains (LADs), spanning 0.1–10 Mb, were identified across the genome based on an elegant approach called DamID, which takes advantage of DNA adenine methylase (DAM) fused in this case to lamin B1, a component of the nuclear lamina (Guelen et al., 2008). Characterization of the genomic content enriched in LADs showed that they are generally gene poor, transcriptionally inactive, depleted for active transcription marks such as RNAPII and active histone marks. At LAD borders, promoters transcribing away from LADs are found enriched, as well as CTCF binding sites (Guelen et al., 2008). Dynamic interaction of the genome with the nuclear lamina was seen during neural differentiation of embryonic stem cells (ESCs). Some, but certainly not all, regions in the genome that were transcriptionally activated or repressed during this process changed their association to the nuclear lamina accordingly (Peric-Hupkes et al., 2010). Furthermore, mis-expressed genes were correlated with a change in nuclear localization of these genes in cells carrying disease related lamin A mutations (Mewborn et al., 2010). Recently, mapping of the lamin A-interacting genes showed that lamin A is similarly, involved in anchoring silent genes to the nuclear lamina. Intriguingly though, depletion of lamin A changed the nuclear positioning of the lamin A bound genes but was not enough to change the expression of these genes (Kubben et al., 2012). Oppositely, as discussed below, the artificial tethering of genes to the nuclear lamina sometimes, but not always, leads to their silencing. Clearly, the nuclear lamina is involved in the spatial organization of the genome in a manner that at least reflects transcriptional activity. To what extent a peripheral positioning also determines gene activity still remains to be investigated.

CTCF

CTCF is probably the best characterized structural organizer of the genome to date. From the first description of the protein (Lobanenkov et al., 1990), it has been shown to be a versatile protein having direct transcriptional effects (Filippova et al., 1996; Vostrov and Quitschke, 1997; Yang et al., 1999) as well as effects on transcription over distance (Bell et al., 1999). The approximately 40,000 CTCF binding sites in the human and murine genome preferentially locate to intergenic regions and show high conservation between different cell types (Barski et al., 2007; Kim et al., 2007; Chen et al., 2008; Hou et al., 2010). CTCF is ubiquitously expressed and an essential protein (Heath et al., 2008). It has a well established role in chromatin folding at the β-globin locus, and in chromatin folding and gene expression at the H19/Igf2 locus and the antigen receptor loci, as described above. Also at other loci, including the human major histocompatibility complex (MHC) class II locus and the Kcnq5 gene, CTCF-mediated chromatin loops were found involved in gene regulation (Majumder et al., 2008; Majumder and Boss, 2010; Ren et al., 2012). At a more genome-wide scale, CTCF binding sites were found enriched at borders between the topological domains identified by Hi-C (Yaffe and Tanay, 2011; Dixon et al., 2012) as well as at LAD borders (Guelen et al., 2008), further hinting at an important role for this protein in organizing the 3D structure of chromosomes. Interest in the protein was raised even further when cohesin was found to co-occupy genomic sites with, and be positioned by, CTCF (see below; Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008).

ChIA-PET is a technology that combines chromatin immunoprecipitation (ChIP) with a 3C approach, to direct DNA topology studies specifically to the genomic sites that are bound by a protein of interest (Fullwood et al., 2009). ChIA-PET was applied to CTCF to study its DNA interactome (Handoko et al., 2011). Mostly intrachromosomal and a few interchromosomal interactions between CTCF-bound sequences were identified, with the intrachromosomal loop sizes ranging from 10–200 kb. The loops appeared to serve different purposes (**Figure 3**). They can isolate an active chromatin region from surrounding inactive chromatin or bring together enhancers and promoters in a single loop. Yet other loops formed by CTCF seem to isolate undefined chromatin from a flanking active and inactive chromosomal region (Handoko et al., 2011). Only a few percent of the total number of CTCF sites was found engaged in loop formation. This suggests that ChIA-PET only uncovers the tip of the topological iceberg. Alternatively, the majority of CTCF-bound sites is not involved in long-range chromatin interactions. If the latter is true, it would be interesting to understand what determines whether a CTCF binding site is engaged or not in a chromatin loop.

COHESIN

Cohesin is a multiprotein complex that forms a ring-like structure which captures and holds together the two DNA double-strand helices of sister chromatids after DNA replication. The discovery that cohesin binds to CTCF binding sites also in G1 phase of the cell cycle suggested that it has an additional role besides keeping sister chromatids together. Without CTCF, cohesin still binds to chromatin but is no longer found at specific locations along the chromosome arms, suggesting that CTCF positions cohesin on the chromatin (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008). Given its shape and function, cohesin was obviously considered an attractive protein for chromatin loop formation (Nasmyth and Haering, 2009). Indeed, cohesin was found to mediate chromatin looping at CTCF binding sites in several loci including the immunoglobulin locus (Degner et al., 2009), the interferon

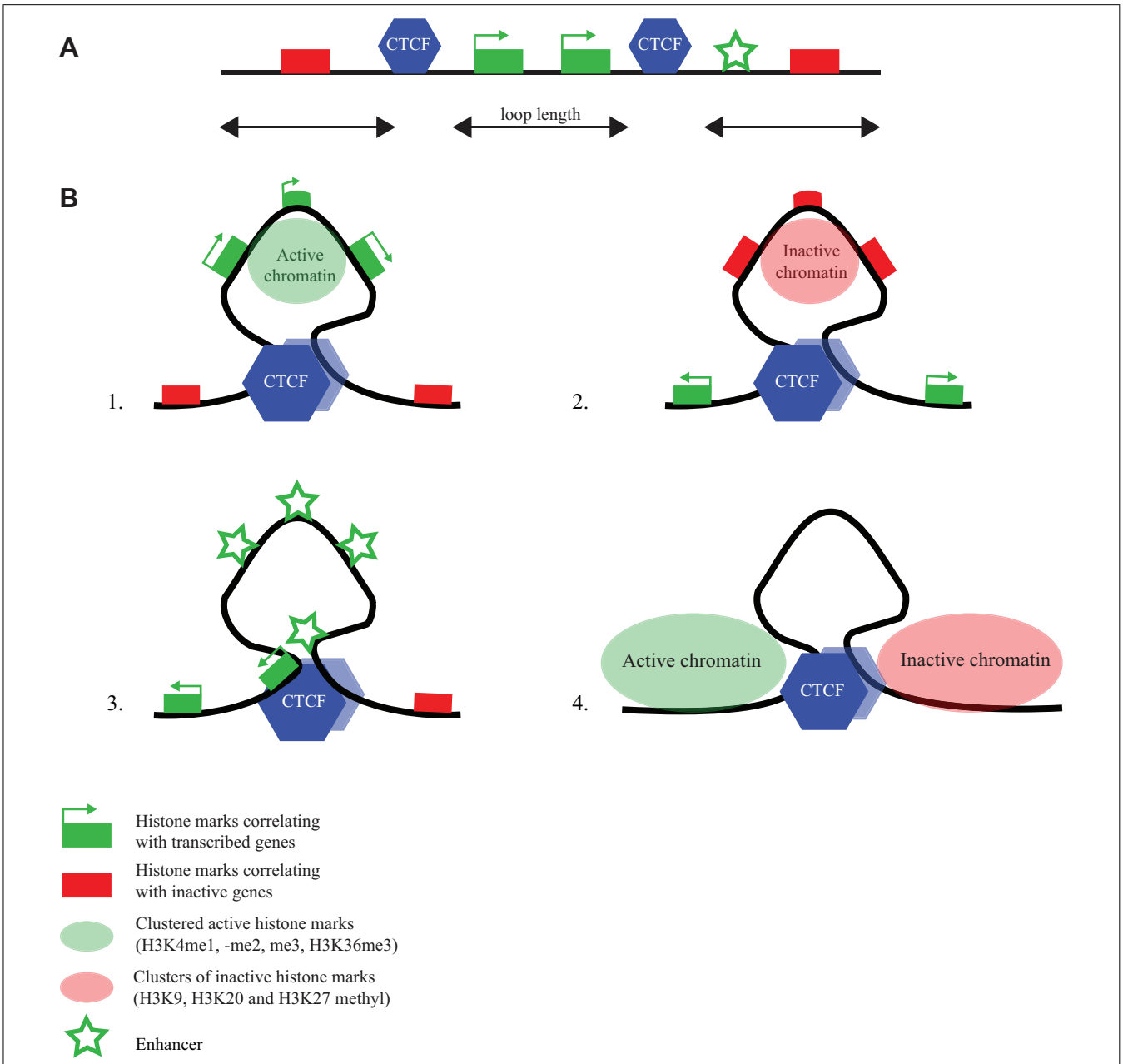


FIGURE 3 | CTCF flanks chromatin marked by specific histone modifications. (A) Linear representation of a chromosomal region with active and inactive genes, CTCF binding sites and an enhancer (for explanation of symbols, see bottom figure). **(B)** ChIA-PET reveals different chromatin loops formed by CTCF (Handoko et al., 2011): CTCF loops demarcate regions **(1)** with active chromatin marks, **(2)** with inactive chromatin marks, **(3)** with enhancers and promoters, and **(4)** with undefined chromatin surrounded by regions with opposing chromatin signatures.

gamma locus (Hadjur et al., 2009), the HoxA locus (Kim et al., 2011), the MHC class II locus (Majumder and Boss, 2011), the β-globin locus (Hou et al., 2010; Chien et al., 2011), and the H19/Igf2 locus (Nativio et al., 2009). Interestingly, at several sites bound by CTCF across different cell types, cohesin association was found to differ in a cell-dependent manner, with topological changes and altered gene expression changing accordingly (Chien et al., 2011; Kim et al., 2011). This suggests that possibly the co-recruitment of additional factors like cohesin determines whether a given CTCF binding site is engaged in a chromatin loop in a given cell type. A CTCF-independent role for cohesin in transcription regulation was also demonstrated, in a study that revealed cohesin and estrogen receptor co-binding near upregulated genes upon estrogen treatment of MCF-7 cells (Schmidt et al., 2010). Cohesin binding was enriched at sites demonstrated by ChIA-PET to form ER-mediated loops (Fullwood et al., 2009), suggesting that cohesin may help ER to mediate transcriptional responses via long-range DNA interactions (Schmidt et al., 2010). A further

CTCF-independent role of cohesin was observed in ESCs, where cohesin association was detected at sites bound by mediator and RNAPII, but not CTCF (Kagey et al., 2010). Enhancer promoter interactions of tissue-specific genes were shown by 3C technology to be mediated by the interaction with mediator and the cohesin loading factor, Nipbl. Cohesin and mediator together share distinct genomic sites in different tissues, unlike the shared binding sites between CTCF and cohesin which seem largely conserved between cell types (Kagey et al., 2010). Thus, cohesin may have CTCF-dependent and -independent roles in chromosome topology and gene regulation during development (Kagey et al., 2010; Schmidt et al., 2010).

RNA pol II

Transcription, and in particular the nuclear localization of RNA polymerase, has always been considered an attractive candidate to shape the 3D genome (Fraser and Bickmore, 2007). It may explain why active chromatin comes together in the nuclear space. Clusters of RNAPII, termed transcription factories, have been identified in the nucleus by electron microscopy and immunofluorescence (Jackson et al., 1993; Iborra et al., 1996; Grande et al., 1997; Jackson et al., 1998). It is difficult to assess the number of factories per cell as this appears to differ between cell types and is also dependent on the microscopy method used (Osborne et al., 2004). The concept assumes that genes need to migrate to pre-existing protein factories where multiple genes are transcribed simultaneously. In a more extreme model there may even be dedicated transcription factories that contain specific combinations of transcription factors and therefore need to be visited by defined categories of co-regulated genes (Xu and Cook, 2008; Schoenfelder et al., 2010). Does form indeed follow function, as suggested by these models? Not all observations necessarily support this idea. Live cell imaging with fluorescently tagged RNAPII so far has not provided convincing evidence for the existence of transcription factories (Kimura et al., 2002; Zobeck et al., 2010), nor for movement of genes upon transcriptional activation (Zobeck et al., 2010). Inhibition of transcription caused most RNA polymerase to dissociate from active genes, yet had no appreciable impact on their contacts with other active genes, as assessed by 4C technology, nor interfered with enhancer–gene contacts (Palstra et al., 2008). The recent demonstration that loop formation in the β-globin locus precedes transcriptional activation also suggests that function follows form (Deng et al., 2012). Possibly, shape and function both influence each other. It was proposed that initiating RNA polymerases that are close together in the nuclear space may aggregate to form the observed transcription factories. This is easiest envisioned to happen between genes that are proximal on the linear chromosome, as these per definition are close together in the nuclear space, rather than involving genes searching for distant co-regulated genes (Razin et al., 2011). Indeed, a ChIA-PET study focusing on chromatin loops formed between RNAPII-bound chromatin sites recently demonstrated the clustering of active gene promoters that neighbor each other on the chromosomes (Li et al., 2012).

ChIA-PET enables an unbiased genome-wide assessment of contacts formed by the genomic sites bound by a protein of interest. Remarkably, for all proteins studied so far, ChIA-PET primarily identifies local contacts between sites close together on the linear chromosome. On the one hand this probably emphasizes the importance of local chromatin loops for the expression of genes involved in these loops. On the other hand it raises the question: how important is the position of a gene relative to other chromosomal regions elsewhere in the genome? So far, mostly microscopy studies have tried to address this.

GENE POSITIONING IN THE CELL NUCLEUS

One of the earliest studies that followed the positioning of individual genes focused on the Ikaros proteins, required for the development of cells of the lymphoid lineage (Brown et al., 1997, 1999). Highly expressed lymphoid genes like CD45 and CD19 were not found associated with Ikaros in B cells, but stage-specific genes showed differential association with Ikaros during differentiation (Brown et al., 1997). When bound by Ikaros, these genes were found to be silenced and repositioned to pericentromeric heterochromatin (PCH). It was proposed that PCH-association facilitated heritable gene silencing during B cell differentiation (Brown et al., 1997, 1999). Subsequently, also other genes were found to occupy particular nuclear locations in relation to their status of transcription, and again this has been studied most notably for the forementioned model gene loci. The IgH locus, for example, was found to adopt a peripheral position in cells not transcribing the gene. When active in B cells, it adopts a more internal nuclear position (Kosak et al., 2002). In mature B cells, the non-productive IgH allele was reported to be frequently associated with PCH, perhaps to ensure its silencing (Skok et al., 2001; Roldan et al., 2005). Repositioning of loci to PCH is also important during lineage choice in T cells (Merkenschlager et al., 2004; Collins et al., 2011), where repositioning of the CD8 locus to PCH is seen in $CD4^+$ T cells and vice versa. Here localization was stated to be predictive for the developmental state of the T cell (Merkenschlager et al., 2004). Localization of inactive genes to the nuclear periphery was also found for the human CFTR locus (Zink et al., 2004; Ballester et al., 2008) and the casein cluster in mammary glands (Kress et al., 2011).

Similar observations were done on the β-globin locus. During erythroid maturation, which is accompanied by LCR-mediated transcriptional activation, the locus was observed to move from the periphery to the interior. Expression at the periphery was found, but it occurred more frequently in the nuclear interior, and the inward movement was dependent on the LCR (Ragoczy et al., 2006). Whereas one study reported preferred clustering of the active β-globin genes with other active erythroid genes (Schoenfelder et al., 2010), two other studies did not find this (Simonis et al., 2006; Brown et al., 2008). A different type of movement was observed for the Hox gene clusters. Induction of Hox gene expression influenced the position of the Hoxb1 and Hoxb9 genes relative to their CTs (Chambeyron et al., 2005). Expression was associated with a position more outside of the CT. This nuclear organization was dynamic as hoxb1 and -b9 could be repositioned in different stages of differentiation, in agreement with their transcriptional state (Chambeyron and Bickmore, 2004; Chambeyron et al., 2005). Similarly, Hoxd genes were looped outside their CT in the tailbud of e9.5 mice (Morey et al., 2007). In the forelimb bud, where Hoxd9 is also expressed (Tarchini and Duboule, 2006), no

looping out of the CT for this gene is found (Morey et al., 2007). Moreover, neighboring genes that are dragged along outside the CT not necessarily show bystander upregulation of gene expression (Noordermeer et al., 2008; Morey et al., 2009). Thus, these studies show that genes can, but do not need to move away from their CT and that looping out of the CT is not sufficient for gene activation.

To better understand the consequences of nuclear repositioning, tethering experiments can be done. These are based on the genomic integration of repeats of DNA binding sites (often bacterial LacO or TetO sequences) and the simultaneous expression in eukaryotic cells of cognate bacterial proteins (LacR or TetR) fused to a protein of interest. Fusion to fluorescent GFP enables following the genomic integration sites in live cell imaging studies (Robinett et al., 1996; Tumbar et al., 1999) and revealed that individual gene loci show limited movement during the interphase of mammalian cells (Chubb et al., 2002). Recruitment of transcriptional activators caused locus decondensation concomitant with increased transcription and histone acetylation, but neither was required to maintain the decondensed chromatin state (Tumbar et al., 1999; Ye et al., 2001; Nye et al., 2002; Chen et al., 2004). The targeting of heterochromatin protein 1 (HP1) to a non-heterochromatic locus reduced gene expression, induced locus condensation, and resulted in local H3K9me3 modifications, indicative of heterochromatin formation (Verschure et al., 2005; Hathaway et al., 2012).

Several studies used fusions of lamina components to address the consequences of recruitment to the nuclear periphery. In one study, which also enabled simultaneous visualization of nascent transcripts, the association of lamin B1 to a reporter locus caused repositioning, but only after cell division. Here, the kinetics of gene activation were similar to that at internal locations, indicating that loci maintain their transcriptional competence at the nuclear periphery (Kumaran and Spector, 2008). In another study, however, repositioning through the recruitment of emerin (EMD) was found to be accompanied by reporter gene silencing (Reddy et al., 2008). A third study measured chromosome-wide gene expression differences after tethering of the chromosome to the inner nuclear membrane. A few genes, some nearby and some at great distance from the integrated LacO cassettes, showed repressed transcription, but expression was not incompatible with peripheral location (Finlan et al., 2008). Interestingly, in a recent study it was demonstrated that the ectopic integration of LAD sequences can also reposition surrounding chromosomal regions to the periphery, and negatively influences the expression of surrounding genes (Zullo et al., 2012). GAGA motifs were found enriched in LADs and demonstrated to be responsible for peripheral recruitment. They are targets for the transcriptional repressor cKrox and the associated HDAC3 and Lap2β proteins, which were found to be necessary for peripheral recruitment (Zullo et al., 2012). Collectively, these studies suggest that nuclear compartmentalization and gene expression are coupled, but also emphasize the probabilistic nature of nuclear organization: genes positioned at the periphery of the cell nucleus do not necessarily lose their capacity to be transcribed, but appear more susceptible to transcriptional repression than at more internal nuclear positions.

CONCLUDING REMARKS

Over the last years research has made major progress in understanding the relationship between structure and function of the genome. Studies on model gene systems such as those discussed here have shown that local DNA interactions between regulatory sites and genes are important for transcriptional control. In mammals, such regulatory interactions can take place over chromosomal distances as large as a megabase. Transcription factors bound to these chromatin sites seem responsible for setting up the chromatin loops in chromosomal segments. Others, such as CTCF, appear capable to modify chromatin topology such that it hampers these interactions. Beyond this local scale of structural organization, genome folding seems to follow more probabilistic rules. Active and inactive chromatin separate, some chromosomal regions have an increased chance of being at the periphery than others, and, when assayed across large cell populations, all individual gene loci appear to have many different contact partners. Together this suggests that the exact genome conformation will differ from cell to cell. As a consequence, a given contact between two dispersed genomic regions will only occur in a subset of cells. If this contact influences the expression of the associated genes, this may not have an impact on the entire cell population, but can be important for the individual cells involved, as was shown recently (Noordermeer et al., 2011b). To study the functional consequences of cell to cell differences in genome conformation we therefore probably need to analyse form and function at the single cell level, with the exciting possibility to discover that the overall shape of our genome can determine cell fate decisions of individual cells.

ACKNOWLEDGMENTS

This work was financially supported by grant no. 935170621 from the Dutch Scientific Organization (NWO) and a European Research Council Starting Grant (209700, "4C") to Wouter de Laat.

REFERENCES

Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G., and Higgs, D. R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J.* 23, 2841–2852.

Ballester, M., Kress, C., Hue-Beauvais, C., Kieu, K., Lehmann, G., Adenot, P., et al. (2008). The nuclear localization of WAP and CSN genes is modified by lactogenic hormones in HC11 cells. *J. Cell. Biochem.* 105, 262–270.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837.

Bartolomei, M. S., Webber, A. L., Brunkow, M. E., and Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev.* 7, 1663–1673.

Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405, 482–485.

Bell, A. C., West, A. G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., et al. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 3, e157. doi: 10.1371/journal.pbio.0030157

Branco, M. R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS*

Biol. 4, e138. doi: 10.1371/journal.pbio.0040138

Brown, J. M., Green, J., das Neves, R. P., Wallace, H. A., Smith, A. J., Hughes, J., et al. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* 182, 1083–1097.

Brown, K. E., Baxter, J., Graf, D., Merkenschlager, M., and Fisher, A. G. (1999). Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell* 3, 207–217.

Brown, K. E., Guest, S. S., Smale, S. T., Hahm, K., Merkenschlager, M., and Fisher, A. G. (1997). Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 91, 845–854.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* 32, 623–626.

Chambeyron, S., and Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* 18, 1119–1130.

Chambeyron, S., Da Silva, N. R., Lawson, K. A., and Bickmore, W. A. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* 132, 2215–2223.

Chen, D., Belmont, A. S., and Huang, S. (2004). Upstream binding factor association induces large-scale chromatin decondensation. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15106–15111.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.

Chien, R., Zeng, W., Kawauchi, S., Bender, M. A., Santos, R., Gregson, H. C., et al. (2011). Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J. Biol. Chem.* 286, 17870–17878.

Chubb, J. R., Boyle, S., Perry, P., and Bickmore, W. A. (2002). Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* 12, 439–445.

Collins, A., Hewitt, S. L., Chaumeil, J., Sellars, M., Micsinai, M., Allinne, J., et al. (2011). RUNX transcription factor-mediated association of Cd4 and Cd8 enables coordinate gene regulation. *Immunity* 34, 303–314.

Degner, S. C., Wong, T. P., Jankevicius, G., and Feeney, A. J. (2009). Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J. Immunol.* 182, 44–48.

Degner, S. C., Verma-Gaur, J., Wong, T. P., Bossen, C., Iverson, G. M., Torkamani, A., et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9566–9571.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., et al. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24.

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P., and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell* 1, 131–139.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.

Dostie, J., and Bickmore, W. A. (2012). Chromosome organization in the nucleus – charting new territory across the Hi-Cs. *Curr. Opin. Genet. Dev.* 22, 125–131.

Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., et al. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 18, 2485–2490.

Engel, N., Thorvaldsen, J. L., and Bartolomei, M. S. (2006). CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. *Hum. Mol. Genet.* 15, 2945–2954.

Ferguson-Smith, A. C., Sasaki, H., Cattanach, B. M., and Surani, M. A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* 362, 751–755.

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., et al. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* 16, 2802–2813.

Finlan, L. E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., et al. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* 4, e1000039. doi: 10.1371/journal.pgen.1000039

Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. *Nature* 447, 413–417.

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

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* 18, 411–422.

Gonzalez, F., Duboule, D., and Spitz, F. (2007). Transgenic analysis of Hoxd gene regulation during digit development. *Dev. Biol.* 306, 847–859.

Gourdon, G., Sharpe, J. A., Wells, D., Wood, W. G., and Higgs, D. R. (1994). Analysis of a 70 kb segment of DNA containing the human zeta and alpha-globin genes linked to their regulatory element (HS-40) in transgenic mice. *Nucleic Acids Res.* 22, 4139–4147.

Grande, M. A., van der Kraan, I., de Jong, L., and van Driel, R. (1997). Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J. Cell Sci.* 110(Pt 15), 1781–1791.

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.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., and Selimyan, R. (2011a). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* 147, 332–343.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., et al. (2011b). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430.

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., et al. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410–413.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., et al. (1991). Importance of globin gene order for correct developmental expression. *Genes Dev.* 5, 1387–1394.

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.

Hathaway, N. A., Bell, O., Hodges, C., Miller, E. L., Neel, D. S., and Crabtree, G. R. (2012). Dynamics and memory of heterochromatin in living cells. *Cell* 149, 1447–1460.

Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., et al. (2008). CTCF regulates cell cycle progression of alpha-beta T cells in the thymus. *EMBO J.* 27, 2839–2850.

Higgs, D. R., Sharpe, J. A., and Wood, W. G. (1998). Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin. Hematol.* 35, 93–104.

Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3651–3656.

Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* 109(Pt 6), 1427–1436.

Jackson, D. A., Hassan, A. B., Errington, R. J., and Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12, 1059–1065.

Jackson, D. A., Iborra, F. J., Manders, E. M., and Cook, P. R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* 9, 1523–1536.

Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J., et al. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* 133, 265–279.

Joffe, B., Leonhardt, H., and Solovei, I. (2010). Differentiation and large scale spatial organization of the genome. *Curr. Opin. Genet. Dev.* 20, 562–569.

Jung, D., and Alt, F. W. (2004). Unraveling V(D)J recombination; insights into gene regulation. *Cell* 116, 299–311.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* 30, 90–98.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., et al. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245.

Kim, Y. J., Cecchini, K. R., and Kim, T. H. (2011). Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7391–7396.

Kimura, H., Sugaya, K., and Cook, P. R. (2002). The transcription cycle of RNA polymerase II in living cells. *J. Cell Biol.* 159, 777–782.

Kleinjan, D. A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* 76, 8–32.

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301, 331–333.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., et al. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158–162.

Kress, C., Kieu, K., Droineau, S., Galio, L., and Devinoy, E. (2011). Specific positioning of the casein gene cluster in active nuclear domains in luminal mammary epithelial cells. *Chromosome Res.* 19, 979–997.

Kubben, N., Adriaens, M., Meuleman, W., Voncken, J. W., van Steensel, B., and Misteli, T. (2012). Mapping of lamin A- and progerin-interacting genome regions. *Chromosoma* 121, 447–464.

Kumaran, R. I., and Spector, D. L. (2008). A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.* 180, 51–65.

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., et al. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10684–10689.

Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L., and Tilghman, S. M. (1995). An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev.* 9, 2079–2089.

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98.

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

Liu, H., Schmidt-Supprian, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., et al. (2007). Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev.* 21, 1179–1189.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., et al. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 1743–1753.

Majumder, P., and Boss, J. M. (2010). CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. *Mol. Cell. Biol.* 30, 4211–4223.

Majumder, P., and Boss, J. M. (2011). Cohesin regulates MHC class II genes through interactions with MHC class II insulators. *J. Immunol.* 187, 4236–4244.

Majumder, P., Gomez, J. A., Chadwick, B. P., and Boss, J. M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* 205, 785–798.

Merkenschlager, M., Amoils, S., Roldan, E., Rahemtulla, A., O'Connor, E., Fisher, A. G., et al. (2004). Centromeric repositioning of coreceptor loci predicts their stable silencing and the CD4/CD8 lineage choice. *J. Exp. Med.* 200, 1437–1444.

Mewborn, S. K., Puckelwartz, M. J., Abuisneineh, F., Fahrenbach, J. P., Zhang, Y., MacLeod, H., et al. (2010). Altered chromosomal positioning, compaction, and gene expression with a lamin A/C gene mutation. *PLoS ONE* 5, e14342. doi: 10.1371/journal.pone.0014342

Mitchell, J. A., and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* 22, 20–25.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., et al. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147, 1132–1145.

Morey, C., Kress, C., and Bickmore, W. A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome Res.* 19, 1184–1194.

Morey, C., Da Silva, N. R., Perry, P., and Bickmore, W. A. (2007). Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* 134, 909–919.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat. Genet.* 36, 889–893.

Nasmyth, K., and Haering, C. H. (2009). Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* 43, 525–558.

Nativio, R., Wendt, K. S., Ito, Y., Huddleston, J. E., Uribe-Lewis, S., Woodfine, K., et al. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.* 5, e1000739. doi: 10.1371/journal.pgen.1000739

Noordermeer, D., Branco, M. R., Splinter, E., Klous, P., van Ijcken, W., Swagemakers, S., et al. (2008). Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region. *PLoS Genet.* 4, e1000016. doi: 10.1371/journal.pgen.1000016

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011a). The dynamic architecture of Hox gene clusters. *Science* 334, 222–225.

Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., et al. (2011b). Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat. Cell Biol.* 13, 944–951.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.

Nye, A. C., Rajendran, R. R., Stenoien, D. L., Mancini, M. A., Katzenellenbogen, B. S., and Belmont, A. S. (2002). Alteration of large-scale chromatin structure by estrogen receptor. *Mol. Cell. Biol.* 22, 3437–3449.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36, 1065–1071.

Palstra, R. J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and de Laat, W. (2008). Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS ONE* 3, e1661. doi: 10.1371/journal.pone.0001661

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35, 190–194.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W., Solovei, I., Brugman, W., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* 38, 603–613.

Ragoczy, T., Bender, M. A., Telling, A., Byron, R., and Groudine, M. (2006). The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.* 20, 1447–1457.

Razin, S. V., Gavrilov, A. A., Pichugin, A., Lipinski, M., Iarovaia, O. V., and Vassetzky, Y. S. (2011). Transcription factories in the context of the nuclear and genome organization. *Nucleic Acids Res.* 39, 9085–9092.

Reddy, K. L., Zullo, J. M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452, 243–247.

Ren, L., Wang, Y., Shi, M., Wang, X., Yang, Z., and Zhao, Z. (2012). CTCF mediates the cell-type specific spatial organization of the Kcnq5 locus and the local gene regulation. *PLoS ONE* 7, e31416. doi: 10.1371/journal.pone.0031416

Reynaud, D., Demarco, I. A., Reddy, K. L., Schjerven, H., Bertolino, E., Chen, Z., et al. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat. Immunol.* 9, 927–936.

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M. J. W., Bergen, I. M., Thongjuea, S., Lenhard, B., et al. (2011). The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. *Immunity* 35, 501–513.

Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., et al. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* 135, 1685–1700.

Roldan, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., et al. (2005). Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat. Immunol.* 6, 31–41.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., et al. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8309–8314.

Sayegh, C. E., Jhunjhunwala, S., Riblet, R., and Murre, C. (2005). Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev.* 19, 322–327.

Schatz, D. G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* 11, 251–263.

Schmidt, D., Schwalie, P. C., Ross-Innes, C. S., Hurtado, A., Brown, G. D., Carroll, J. S., et al. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* 20, 578–588.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61.

Seitan, V. C., and Merkenschlager, M. (2012). Cohesin and chromatin organisation. *Curr. Opin. Genet. Dev.* 22, 93–100.

Seitan, V. C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K. E., et al. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476, 467–471.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472.

Sharpe, J. A., Summerhill, R. J., Vyas, P., Gourdon, G., Higgs, D. R., and Wood, W. G. (1993). Role of upstream DNase I hypersensitive sites in the regulation of human alpha globin gene expression. *Blood* 82, 1666–1671.

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A map of the *cis*-regulatory sequences in the mouse genome. *Nature* 488, 116–120.

Shopland, L. S., Lynch, C. R., Peterson, K. A., Thornton, K., Kepper, N., Hase, J., et al. (2006). Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *J. Cell Biol.* 174, 27–38.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38, 1348–1354.

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.

Song, S. H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol. Cell* 28, 810–822.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 113, 405–417.

Splinter, E., and de Laat, W. (2011). The complex transcription regulatory landscape of our genome: control in three dimensions. *EMBO J.* 30, 4345–4355.

Splinter, E., de Wit, E., Nora, E. P., Klous, P., van de Werken, H. J., Zhu, Y., et al. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* 25, 1371–1383.

Tarchini, B., and Duboule, D. (2006). Control of Hoxd genes' collinearity during early limb development. *Dev. Cell* 10, 93–103.

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.

Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., et al. (2011). Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* 7, e1001343. doi: 10.1371/journal.pgen.1001343

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465.

Tumbar, T., Sudlow, G., and Belmont, A. S. (1999). Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J. Cell Biol.* 145, 1341–1354.

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., et al. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* 17, 453–462.

Van der Ploegh, L. H., Konings, A., Oort, M., Roos, D., Bernini, L., and Flavell, R. A. (1980). gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* 283, 637–642.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G., and Higgs, D. R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* 26, 2041–2051.

Vernimmen, D., Marques-Kranc, F., Sharpe, J. A., Sloane-Stanley, J. A., Wood, W. G., Wallace, H. A., et al. (2009). Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). *Blood* 114, 4253–4260.

Verschure, P. J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A. E., Belmont, A. S., et al. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Mol. Cell. Biol.* 25, 4552–4564.

Vostrov, A. A., and Quitschke, W. W. (1997). The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J. Biol. Chem.* 272, 33353–33359.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796–801.

Williamson, I., Eskeland, R., Lettice, L. A., Hill, A. E., Boyle, S., Grimes, G. R., et al. (2012). Anterior–posterior differences in HoxD chromatin topology in limb development. *Development* 139, 3157–3167.

Xu, M., and Cook, P. R. (2008). Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* 181, 615–623.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* 43, 1059–1065.

Yang, Y., Quitschke, W. W., Vostrov, A. A., and Brewer, G. J. (1999). CTCF is essential for up-regulating expression from the amyloid precursor protein promoter during differentiation of primary hippocampal neurons. *J. Neurochem.* 73, 2286–2298.

Ye, Q., Hu, Y. F., Zhong, H., Nye, A. C., Belmont, A. S., and Li, R. (2001). BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. *J. Cell Biol.* 155, 911–921.

Yoon, Y. S., Jeong, S., Rong, Q., Park, K. Y., Chung, J. H., and Pfeifer, K. (2007). Analysis of the H19ICR insulator. *Mol. Cell. Biol.* 27, 3499–3510.

Zink, D., Amaral, M. D., Englmann, A., Lang, S., Clarke, L. A., Rudolph, C., et al. (2004). Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J. Cell Biol.* 166, 815–825.

Zobeck, K. L., Buckley, M. S., Zipfel, W. R., and Lis, J. T. (2010). Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol. Cell* 40, 965–975.

Zullo, J. M., Demarco, I. A., Piqué-Regi, R., Gaffney, D. J., Epstein, C. B., Spooner, C. J., et al. (2012). DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* 149, 1474–1487.

Conflict of Interest Statement: 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.

Received: 17 July 2012; paper pending published: 17 August 2012; accepted: 01 October 2012; published online: 17 October 2012.
Citation: Holwerda S and de Laat W (2012) Chromatin loops, gene positioning, and gene expression. Front. Gene. ***3****:217. doi: 10.3389/fgene.2012.00217*
This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in GENETICS

REVIEW ARTICLE
published: 08 October 2012
doi: 10.3389/fgene.2012.00199

A repetitive elements perspective in Polycomb epigenetics

Valentina Casa[1,2] and Davide Gabellini[1]*

[1] Division of Regenerative Medicine, Stem Cells, and Gene Therapy, Dulbecco Telethon Institute and San Raffaele Scientific Institute, Milano, Italy
[2] Università Vita-Salute San Raffaele, Milano, Italy

Edited by:
Michèle Amouyal, CNRS, France

Reviewed by:
Igor Kovalchuk, University of Lethbridge, Canada
Michèle Amouyal, CNRS, France

***Correspondence:**
Davide Gabellini, Division of Regenerative Medicine, Stem Cells, and Gene Therapy, Dulbecco Telethon Institute and San Raffaele Scientific Institute, DIBIT 2, 5A3-44, Via Olgettina 58, 20132 Milano, Italy.
e-mail: gabellini.davide@hsr.it

Repetitive elements comprise over two-thirds of the human genome. For a long time, these elements have received little attention since they were considered non-functional. On the contrary, recent evidence indicates that they play central roles in genome integrity, gene expression, and disease. Indeed, repeats display meiotic instability associated with disease and are located within common fragile sites, which are hotspots of chromosome re-arrangements in tumors. Moreover, a variety of diseases have been associated with aberrant transcription of repetitive elements. Overall this indicates that appropriate regulation of repetitive elements' activity is fundamental. Polycomb group (PcG) proteins are epigenetic regulators that are essential for the normal development of multicellular organisms. Mammalian PcG proteins are involved in fundamental processes, such as cellular memory, cell proliferation, genomic imprinting, X-inactivation, and cancer development. PcG proteins can convey their activity through long-distance interactions also on different chromosomes. This indicates that the 3D organization of PcG proteins contributes significantly to their function. However, it is still unclear how these complex mechanisms are orchestrated and which role PcG proteins play in the multi-level organization of gene regulation. Intriguingly, the greatest proportion of Polycomb-mediated chromatin modifications is located in genomic repeats and it has been suggested that they could provide a binding platform for Polycomb proteins. Here, these lines of evidence are woven together to discuss how repetitive elements could contribute to chromatin organization in the 3D nuclear space.

Keywords: repeats, Polycomb, non-protein-coding RNA, nuclear structure, FSHD muscular dystrophy

INTRODUCTION

In the last decade, when the genomic sequences of *Homo sapiens* and several model organisms became available, there was the realization that the number of protein-coding genes does not correlate with organism complexity. In fact, worms or flies have approximately the same number of protein-coding genes as mice or humans (Taft et al., 2007). On the other hand, the non-protein coding component of the genomic DNA, and in particular repetitive elements, represent a progressively larger proportion of the genome in organisms with increasing complexity (Neguembor and Gabellini, 2010). Recent estimations indicate that repetitive sequences could account for up to 66–69% of the human genome (De Koning et al., 2011). While this strongly suggests that it might significantly contribute to higher eukaryotes sophistication, the repetitive fraction of the genome is largely ignored.

The advent of next generation sequencing (NGS) has permitted a genome-wide view to gene expression and chromatin structure. However, NGS-based studies often take into account only reads for which a unique genomic alignment can be obtained, thus discarding data deriving from repetitive DNA (Myers et al., 2011). Despite this, there is increasing evidence of the peculiar functions of the repeated (epi)genome. For example, the role of DNA repeats in chromosome structural organization, gene regulation, genome integrity, and evolution has been described (Kidwell and Lisch, 2000; Lander et al., 2001; Waterston et al., 2002; Feschotte, 2008; Ting et al., 2011; Zhu et al., 2011).

DNA repeats can be also transcribed, frequently in a cell and tissue-specific fashion. Analyses based on Cap Analysis of Gene Expression (CAGE) technology from the Functional Annotation of Mouse (FANTOM) project, revealed an unexpectedly large proportion of capped-transcripts initiating from repetitive units. It has been suggested that these can provide regulatory elements to protein-coding genes, such as alternative promoters, exons, or polyadenylation sites, and ncRNAs, thus significantly expanding the regulatory capability of higher eukaryote genomes (Wang et al., 2007; Bourque et al., 2008; Faulkner et al., 2009; Tyekucheva et al., 2011). Moreover, binding sites for important regulatory factors such as CTCF or TP53 are often associated with genomic repeats (Wang et al., 2007; Bourque et al., 2008; Chadwick, 2008; Simeonova et al., 2012).

Repetitive elements can either mobilize or rearrange in somatic tissues, thus providing an unexpected dynamic dimension to the normal physiology of the soma, but also contributing to the etiopathogenesis of diseases (Kazazian et al., 1988; Ting et al., 2011; Zhu et al., 2011). For the role they can play in genome plasticity, repeats need to be finely tuned. To accomplish this, epigenetic mechanisms including RNA interference (RNAi), DNA methylation, and histone modifications are used to deal with the potentially dangerous effects of repeat transpositions and rearrangements (Slotkin and Martienssen, 2007; Maksakova et al., 2008).

Polycomb group proteins (PcG) are epigenetic repressors with the important function of maintaining the memory of transcriptional programs during development and differentiation (Morey and Helin, 2010; Schuettengruber et al., 2011). However, PcG role appears to go far beyond gene regulation, as they have been associated with many other important nuclear processes, including the regulation of higher order genome architecture and structure (Bantignies and Cavalli, 2011). Importantly, the vast majority of mammalian PcG proteins bind to non-coding DNA, and in particular repetitive elements, which for their intrinsic feature of being present in several copy number, may constitute binding platforms for Polycomb binding in mammals (Cabianca et al., 2012).

In this review, the biological role of DNA repeats and their epigenetic regulation is summarized with the hope of fostering new investigations of this largely unexplored region of the human genome.

GENETICS AND EPIGENETICS OF REPETITIVE ELEMENTS

Using classical annotation processes, about 50% of a typical mammalian genome is annotated as DNA repeats, 5–10% as genes and functional elements and the remaining 40–45% as DNA of unknown function. One caveat with traditional repeat annotation is that DNA repeat identification approaches, e.g., the RepeatMasker program (Smit et al., 1996–2004), use well-curated libraries of known repeat family consensus sequences. By doing so, ancient or divergent DNA repeat classes fail to be identified as repeats. Recently, using a highly sensitive alternative strategy, it was predicted that there may be more than 840 Mbp of additional repetitive sequences in the human genome, thus suggesting that up to 70% of the total genome is composed of repeats (De Koning et al., 2011).

DNA repeats can be present in different arrangements and sizes: they can be widely interspersed repeats (**Table 1**) or they can be located one next to another to form tandem repeats (**Table 2**). Repeats can range in size from 1 to 2 bases to millions of bases and might comprise just two copies or millions of copies (Batzer and Deininger, 2002; Jurka et al., 2007; Kim et al., 2008; Britten, 2010; Hua-Van et al., 2011).

INTERSPERSED REPEATS

Interspersed repeats are the results of ancient or present activity of mobile genetic elements. These elements can mediate their own mobilization either by a cut-and-paste mechanism, as DNA transposons, or by a copy-and-paste process, like retro-transposons (Solyom and Kazazian, 2012). While DNA transposons are now considered immobile, some retro-transposon elements are able to mobilize themselves and other elements. Retrotransposons are composed of long terminal repeat (LTR) and non-LTR containing elements. The LTR retrotransposons are endogenous retroviruses (ERVs) that have lost the ability to go outside the host cell due to a non-functional envelope gene. Non-LTR retrotransposons can be subdivided into long interspersed elements (LINEs), short interspersed elements (SINEs) and, in hominid genomes, medium sized SVAs (SINE-R/VNTR (variable number of tandem repeat)/Alu). In humans, the most important LINE is the RNA polymerase II transcribed LINE-1 (L1), while SINEs are essentially represented by the RNA polymerase III transcribed Alus. L1 is the only element able to encode the proteins required for mobilization. Hence, these are the only known autonomously active human retrotransposons. L1s are also responsible for the mobilization of the non-autonomous Alus, SVAs and processed pseudogenes (cellular mRNAs that become substrates of the reverse transcriptases and are inserted into the genome).

Mobile elements have a significant role in evolution and in generating genetic diversity. For example, the genome fraction occupied by mobile elements varies in different species and each eukaryote displays a specific mobile element complement, suggesting that mobile elements are important players during speciation and evolution (Faulkner, 2011). Being significant contributors to the copy number variation present in humans, mobile elements are also an important source of genetic variation

Table 1 | Major features of the most represented interspersed repetitive elements in the human genome.

Repeat type				Estimated number of copies	Average length	Mobility	Estimated % genome coverage	
Interspersed	Retrotransposons	LTR	LTR (Long terminal repeat) or ERV (Endogenous retroviruses) (MaLR, ERV, ERV1, ERV-K, ERV-L, etc.)	200,000	6–11 kb	Autonomous retrotransposition (retroviral-like)	8%	42%
		Non-LTR	LINE (Long interspersed element) (L1, L2, CR1, etc.)	500,000	6 kb	Autonomous retrotransposition	20%	
			SINE (Short interspersed element) (Alu, MIR, etc.)	1,000,000	0.3 kb	L-1 dependent Retrotransposition	13%	
			SVA SINE-R/VNTR/Alu	2700	2–5 kb	L-1 dependent Retrotransposition		
	DNA transposons	DNA transposons (MER1, MER2, Mariner, Merlin, etc.)		300,000	1–3 kb	inert	2–3%	

Number of copies and genome coverage are estimated values based on current genome coverage.

Table 2 | Major features of the most represented tandem repeats in the human genome.

Repeat type			Unit length	Array length	Estimated % genome coverage
Tandem	Satellite	Alpha-Satellite	171 bp	3–5 Mb	22–25%
		Satellite II (HsatII)	23–26 bp or multiple	10–70 kb	
		Satellite III (GAATGn- simple sequence)	5 bp or multiples up to 70 bp	7.5–100 kb	
		Beta-Satellite	68 bp	2–14.5 kb	
		Gamma-Satellite	220 bp	10–200 kb	
	VNTR (Variable number of tandem repeats)	Microsatellite (Short tandem repeat)	1–13 bp	Hundreds bp	
		Minisatellite (including telomeric repeats)	6–100 bp	1–15 kb or more	
		Macrosatellite	2–12 kb or more	Tens up to hundreds bp	

Genome coverage is estimated on the basis of current genome coverage.

(Brouha et al., 2003; Bennett et al., 2004; Mills et al., 2007; Iskow et al., 2010; Ekram et al., 2012). Moreover, mobile elements can display differential activity in different tissues of the soma, suggesting that every individual is a genetic mosaic variegated by the differential insertion of mobile elements (Muotri et al., 2005, 2010). Finally, retrotransposons have recently been identified as a major source of epigenetic variations in the mammalian genome (Ekram et al., 2012). Retrotransposition, with only few exceptions such as V(D)J recombination (Brack et al., 1978), is an almost unique source of somatic genetic mosaicism, leading not only to heritable genetic variation but also to intra-individual variability. This represents a revolutionary concept that is changing the view of this class of repetitive elements (Faulkner, 2011).

Due to their nature, mobile elements have the potential to affect common diseases, through structural variation, deregulated transcriptional activity or epigenetic effects. Moreover, their transposition can directly cause insertional mutagenesis, as proved by the existence of nearly 100 examples of mobile element insertions causing disease (Lee et al., 2012; Solyom and Kazazian, 2012).

For their genotoxic potential, mobile elements are usually kept repressed by epigenetic mechanisms. DNA methylation represents one of the major players in the repression of repetitive elements (Liang et al., 2002; Kato et al., 2007). A recent study performed a comprehensive genome-wide methylation analysis on all repetitive elements in human embryonic stem cells and fetal fibroblasts (Su et al., 2012). Among all classes of repetitive elements, LINE, LTR, DNA transposon, and also satellite tandem DNA repeats appear more susceptible to changes in DNA methylation, thus suggesting that they are specifically regulated and silenced during cellular differentiation. Importantly, transposon-free regions (TFRs) in the genome have been selectively conserved and are associated with regions including CpG islands, suggesting that in mammalian genomes there are fragments of DNA that are largely unable to tolerate transposon insertion (Simons et al., 2006).

Aberrant repetitive DNA methylation can be associated with diseases. For example, hypo-methylation of L1, Alu, LTR, but also of satellite repeats, is significantly associated with tumor progression in multiple cancers such as gastrointestinal stromal tumors, myeloma, and lung cancer (Rauch et al., 2008; Bollati et al., 2009; Igarashi et al., 2010).

Additionally, mobilization of L1 repeats has been associated with both physiological and pathological processes and is regulated by DNA methylation (Muotri et al., 2010). L1 mobilization has been associated with brain cell development, where the occurrence of L1 retrotransposition in adult cells has been suggested to contribute to neuronal somatic diversification (Muotri et al., 2005). This mechanism, so far assigned specifically to human neural progenitors and adult hippocampus, is modulated by the methyl-CpG-binding protein 2 (MeCP2) (Muotri et al., 2005, 2010; Coufal et al., 2009). Importantly, in RETT syndrome, a mental retardation disorder caused by mutation in the *MECP2* gene, an extensive de-regulation of L1 retrotransposition in neurons has been reported (Muotri et al., 2010; Solyom and Kazazian, 2012).

Besides DNA methylation, several repressive histone modifications, including H3K9me3, H3K27me3, and H4K20me3, are also enriched on interspersed repeats (Martens et al., 2005; Mikkelsen et al., 2007; Leeb et al., 2010). Importantly, a re-estimation of chromatin immunoprecipitation results on repetitive elements from high-throughput sequence data of human and mouse cells has been recently conducted (Day et al., 2010). According to this analysis, different members of the murine ERV family of repeats appear to assume distinct patterns of histone modifications, which are representative of a specific pattern of heterochromatin formation. While transposable elements belonging to ERV-K and ERV1 subfamilies are enriched for histone marks typical of constitutive heterochromatin such as H3K9me3 and H4K20me3 in mouse ES cells, ERV-L and MaLR families are characterized by the hallmark of Polycomb-mediated silencing H3K27me3 (Mikkelsen et al., 2007; Dong et al., 2008; Day et al., 2010).

A remarkable finding from these studies is that silencing of repetitive elements can be redundant and flexible. This has been shown by independent groups and within independent

silencing pathways. For example, during the stages of global DNA de-methylation in early embryonic mouse development, the RNA-interference guardian machinery become responsible for controlling the expression of intracisternal A particle (IAP), ERV-K, and ERV-L retrotransposons, thus preserving genome integrity (Svoboda et al., 2004). Additionally, studies of mouse ES cells deficient for the H3K9 histone methyltransferases Suv39h showed that decreased H3K9me3 levels in the repetitive elements were compensated by increases in H3K27me3 enrichment (Peters et al., 2003). Thus, different and largely independent repression pathways can converge and compensate each other's function. Most likely, this has to do with the necessity of the cells to guarantee multiple levels of protection from aberrant activation of mobile elements.

Overall, the epigenetic repression of repetitive elements on one hand prevents dramatic nuclear effects such as genotoxicity, but on the other hand allows the specific regulation of such elements occurring in the germ line (Peaston et al., 2004), embryonic cells (Kano et al., 2009) and, perhaps to a lesser extent, during later developmental phases (Muotri et al., 2005, 2010).

TANDEM REPEATS

Tandem repeats constitute a large portion of the human genome, and account for a significant amount of its copy number variation (Warburton et al., 2008). Besides their role in evolution (Warburton et al., 1996; Rudd et al., 2006; McLaughlin and Chadwick, 2011), they have been found to be critical in several other processes, including heterochromatin formation, chromosome segregation, (Morris and Moazed, 2007) and X-chromosome inactivation (XCI) (Chadwick, 2008). Moreover, repeat instability is at the basis of a number of diseases (Lopez Castel et al., 2010).

Tandem DNA repeats in the human genome show a wide range of unit sizes, spanning from a few base pairs in microsatellites, to several kilobases in megasatellites (Gelfand et al., 2007; Ames et al., 2008; Warburton et al., 2008). At a given locus, the tandem repeat copy number is usually polymorphic among individuals, and for this reason they are more commonly known as variable number tandem repeats (VNTRs).

One of the principal families of DNA tandem repeats in the genome is represented by the satellite DNA of chromosome centromeres. Indeed, maintenance of the structural integrity of centromeres and telomeres is one of the most important functions of tandem repeats (Blackburn, 1984). Centromeres have the fundamental role to ensure proper chromosome segregation during cell division. In the human genome, they consist of several Mb of alpha-satellite DNA, which is composed of a 171 bp repeat unit. Chromosome-specific higher-order repeat structures are typical of this type of repeat, as they are important for centromere function (Schueler et al., 2001). Forms of higher-order organization have also unexpectedly been characterized in "simple satellite" sequences such as GAATGn and VNTRs (Warburton et al., 2008), but whether this bears functional relevance has yet to be determined.

For their function, centromeres of higher eukaryotes require an epigenetic specification, rather than a defined DNA sequence. Indeed, centromeric regions localize in the pericentric heterochromatic domain of the interphase nucleus, and they are enriched in H3K9me3, H4K20me3, H3K27me1 histone marks (Peters et al., 2001, 2003; Guenatri et al., 2004; Martens et al., 2005; Mikkelsen et al., 2007; Dong et al., 2008) and in proteins like the centromere-specific H3 variant Centromere protein A (CENP-A) (Yoda et al., 2000; Lo et al., 2001; Blower et al., 2002). As already described for the epigenetic regulation of interspersed repeats, loss of the H3K9 histone methyltransferases (HMTases) Suv39h, which are responsible for the tri-methylation of H3K9 (Peters et al., 2003), activates a compensatory mechanism leading to increase in H3K27me3 (a hallmark of Polycomb-mediated silencing). This underscores an unexpected plasticity between the H3K9 and H3K27 methylation systems (Peters et al., 2003).

In mice, where two different types of repetitive DNA sequences are associated with centromeres, major satellite repeats (6 megabases of 234 bp units) in the pericentromeric region, and minor satellite repeats (600 kb of 120 bp units) in the centromeric region (Choo, 1997), two distinct heterochromatic domains are distinguishable, which became important signatures of mouse interphase nuclei (Guenatri et al., 2004). Pericentromeric satellite DNA of different chromosomes forms large heterochromatic clusters, which upon DAPI staining result in DAPI-dense structures called chromocenters. These formations are typically enriched for the heterochromatin protein 1 alpha (HP1α). The minor satellite DNA, instead, forms individual heterochromatin structures containing the CENP proteins (Guenatri et al., 2004).

In the human genome, the main groups of tandem repeats are the micro-, mini- or macro-satellites (Warburton et al., 2008). They are highly polymorphic in the general population and for this reason they are widely used as genetic markers. Macrosatellites consist of arrays of 1–12 kb repeat units, with a number of repeats ranging from a few to over one hundred (Warburton et al., 2008; Moseley et al., 2012). They can be either chromosome specific, as DXZ4 at chromosome Xq23 (Giacalone et al., 1992) and ZAV at chromosome 9q32 (Tremblay et al., 2010) or they can be associated with two or more chromosomal locations, such as D4Z4, on chromosomes 4q35 and 10q26; (Deidda et al., 1995; Winokur et al., 1996) and RS447, on 4p15 and 18p23; (Gondo et al., 1998).

DXZ4 and D4Z4 macrosatellites are both extensively regulated at the epigenetic level, and they have been described as being associated with either euchromatic or heterochromatic states. Contraction of the 3.3 kb polymorphic D4Z4 tandem repeat array on chromosome 4q35 is associated with facioscapulohumeral muscular dystrophy (FSHD) where a shortening below the threshold of 11 repeat units generates an epigenetic and topologic remodeling of the locus, thus leading to the pathology (Cabianca and Gabellini, 2010). The X-linked DXZ4 macrosatellite locus, instead, has an opposing conformation to that of the surrounding chromosome, constituting a euchromatic dot in the inactive X chromosome, and vice versa (Chadwick, 2008). For their very peculiar epigenetic features and for their involvement in fundamental biological and pathological processes, D4Z4 and DXZ4 could emerge as paradigms for understanding the epigenetic regulation of tandem DNA.

D4Z4 AND DXZ4

Two of the most extensively investigated macrosatellites are the X-linked DXZ4 and the chromosome 4-linked D4Z4. Despite lacking sequence similarity, D4Z4 and DXZ4 macrosatellites share several common aspects (Chadwick, 2009). DXZ4 and D4Z4 are extremely GC rich and belong to a family of human macrosatellites that are noncentromerically located (Giacalone et al., 1992; Kogi et al., 1997; Chadwick, 2009; Tremblay et al., 2010).

Each DXZ4 unit is 3.0 kb long and organized in a tandem array containing 12 to more than 100 copies, localized at Xq23 (Giacalone et al., 1992). As typical for an X-linked locus, DXZ4 is hemizygous in males and subject to XCI in females. However, DXZ4 adopts an opposite chromatin conformation compared to that of the surrounding X chromosome. In males and on the active X-chromosome (Xa), DXZ4 displays features of constitutive heterochromatin, like enrichments in the repressive histone mark H3K9me3, high levels of DNA methylation and association with heterochromatin protein 1 gamma (HP1γ). On the contrary, in the inactive X (Xi), DXZ4 is characterized by euchromatic histone marks such as H3K4me2 and H3K9Ac, a low level of DNA methylation, and is bound by the chromatin regulators CTCF and YY1 (Chadwick, 2008; Filippova, 2008; Moseley et al., 2012). Notably, these features of DXZ4 are remarkably similar to those of the mouse X-inactivation center (Xic), a region of the X chromosome required for XCI (Courtier et al., 1995; Chao et al., 2002; Boumil et al., 2006; Donohoe et al., 2007). Finally, DXZ4 resides at the distal edge of a heterochromatic region targeted by PcG epigenetic repressors (Chadwick and Willard, 2004; McLaughlin and Chadwick, 2011).

The D4Z4 macrosatellite maps to the subtelomeric region of the chromosome 4 long arm, in 4q35. Each unit is 3.3 kb and is present in 11 to 100–150 copies in the general population. Interestingly, reduction of D4Z4 copy number below 11 units is associated with FSHD, one of the most important forms of muscular dystrophy (Wijmenga et al., 1992; Van Deutekom et al., 1993). D4Z4 belongs to a family of repeats with high sequence identity present also in human chromosomes 10q26, 1p12, and the p-arm of acrocentric chromosomes (Lyle et al., 1995; Winokur et al., 1996). This results in frequent exchanges between the 4q35 and 10q26 arrays, which share the highest identity (Van Deutekom et al., 1993). Like DXZ4, D4Z4 is bound by the epigenetic factor YY1 (Gabellini et al., 2002) and displays alternative epigenetic states that parallel the ones of DXZ4 in Xa versus Xi. For D4Z4, the epigenetic make-up is copy number-dependent. The non-contracted array, which retains more than 11 D4Z4 units, displays heterochromatic features like the repressive histone marks H3K9me3 (Zeng et al., 2009) and H3K27me3 (Bodega et al., 2009; Cabianca et al., 2012), histone hypoacetylation (Jiang et al., 2003), as well as a high level of DNA methylation (Van Overveld et al., 2003). Reduction of D4Z4 copy number below 11 units is associated with reduced levels of repressive histone marks (Bodega et al., 2009; Zeng et al., 2009; Cabianca et al., 2012), acquisition of the activating histone marks H3K4me3 and H3K36me2 (Cabianca et al., 2012), DNA hypomethylation (Van Overveld et al., 2003), binding of CTCF (Ottaviani et al., 2009) and loss of Polycomb silencing (Cabianca et al., 2012).

Like DXZ4, D4Z4 is bi-directionally transcribed to generate non-protein-coding RNAs (ncRNAs) (Chadwick, 2008; Snider et al., 2009; Tremblay et al., 2011; Block et al., 2012; Cabianca et al., 2012). In particular, D4Z4 generates a long, chromatin-associated ncRNA (*DBE*-T) selectively in FSHD patients. *DBE-T* functions *in cis* by recruiting the Trithorax protein ASH1L to the FSHD locus leading to chromatin remodeling and de-repression of 4q35 genes (Cabianca et al., 2012). Hence, similarly to the dichotomous behavior observed for DXZ4 on Xi and Xa chromosomes, for D4Z4 the FSHD pathogenesis underlies a major epigenetic switch from a Polycomb repressed state to a Trithorax de-repressed state.

The last, most telomeric D4Z4 unit at 4q35 encodes for a protein called DUX4 (double homeobox 4), which represents one of the major candidates for FSHD (Lemmers et al., 2010). The *DUX4* gene itself originates from a repetitive element, as it is a processed pseudogene of the ancestral *DUXC* gene. Interestingly, *DUX4* and not *DUXC* has been selectively retained in the primate lineage (Clapp et al., 2007; Leidenroth and Hewitt, 2010). In healthy subjects *DUX4* is expressed only in the germ line, while it is epigenetically silenced in somatic tissues (Snider et al., 2010). In FSHD, *DUX4* is aberrantly expressed in skeletal muscle (Dixit et al., 2007; Snider et al., 2010).

DUX4 protein is a transcriptional activator able to bind and activate transcription of MaLR repetitive elements (Geng et al., 2012). Interestingly, MaLR retrotransposons are known Polycomb targets (Day et al., 2010). Hence, DUX4 could have the physiological role of collaborating with Polycomb for the regulation of repetitive elements during early developmental stages and in the germ line.

POLYCOMB

PcG proteins and their functional counterpart, the Trithorax Group (TrxG) proteins, are evolutionary-conserved chromatin regulatory factors that were originally identified in *Drosophila* (Schuettengruber et al., 2007, 2011; Morey and Helin, 2010). PcG and TrxG are essential for cellular identity and differentiation in multicellular organisms. Their activity is required to maintain an "epigenetic memory" of specific gene expression patterns. This is at the basis of the establishment of the correct spatio-temporal regulation of gene expression and, more importantly, of its transmission throughout cell division and cell fate choices. In general, PcG collaborates with transcriptional repressors to maintain gene silencing while TrxG works by counteracting PcG activity allowing, if the appropriate transcriptional activators are available, for gene activation (Schuettengruber et al., 2007). In vertebrates, PcG and TrxG play a central role in stem-cell plasticity and renewal, proliferation, genomic imprinting, X-inactivation, and cancer (Schuettengruber et al., 2007).

In *Drosophila*, where the Polycomb system was first described, PcG and TrxG are specifically recruited on so-called Polycomb Response Element (PRE)/Trithorax Response Element (TRE) sequences, which are switchable memory DNA modules, with PcG or TrxG as their effectors (Schuettengruber et al., 2011). The mechanisms underlying PcG recruitment in mammals are still controversial, though some vertebrate PRE-like elements have recently been described. Interestingly, these retain features of

Drosophila PREs including binding sites for DNA-binding of factors involved in PcG recruitment to PREs in *Drosophila* (Sing et al., 2009; Woo et al., 2010; Cuddapah et al., 2012). However, a defined role for mammalian homologs of PcG recruiters has not been established. Accordingly, additional mechanisms for PcG recruitment in mammals have been proposed. Several examples for a role of short and long ncRNAs in PcG recruitment in ES cells are available (Rinn et al., 2007; Zhao et al., 2008; Khalil et al., 2009; Gupta et al., 2010; Kanhere et al., 2010; Guil et al., 2012). Moreover, in mammals there is a strong correlation between PcG binding and CpG islands (Tanay et al., 2007; Ku et al., 2008; Mendenhall et al., 2010). In particular, non-methylated GC-rich sequences depleted of activating motifs have been shown to be sufficient for Polycomb recruitment in mammalian embryonic stem cells (Mendenhall et al., 2010).

Polycomb proteins form two major multiprotein complexes, Polycomb Repressive Complex 1 (PRC1) and 2 (PRC2). *Drosophila* PRC1 displays four core subunits: Polycomb (Pc), Polyhomeotic (Ph), Posterior sex combs (Psc), and Sex combs extra (Sce, also called dRing). PRC2 core subunits are Enhancer of zeste, E(z), Extra sex combs (Esc), Suppressor of zeste 12, Su(z)12, and the nucleosome-remodeling factor 55 (Nurf-55). In vertebrates, PRC1 and PRC2 are conserved in overall organization, but display a higher complexity in terms of subunits and interactions, so that their composition is cell type- and developmental stage-dependent (Kuzmichev et al., 2002, 2004, 2005; Gao et al., 2012).

Both PRC1 and PRC2 complexes retain an enzymatic activity. In PRC1, the RING domain containing protein dRing (Ring1B in vertebrates) is an E3 ubiquitin ligase mediating the ubiquitination of lysine 119 on histone H2A, which has been suggested to induce chromatin compaction and inhibit transcription elongation (De Napoles et al., 2004; Wang et al., 2004). Nevertheless, in the case of Ring1B the requirement of the enzymatic activity for chromatin compaction was recently challenged (Eskeland et al., 2010). PRC2 catalyzes the di-methylation and tri-methylation of histone H3 at lysine 27 (H3K27me2/me3). The catalytic subunit of PRC2, E(z) in flies, Enhancer of zeste homologs 1/2 (Ezh1/Ezh2) in vertebrates, contains the SET histone methyltransferase domain (Morey and Helin, 2010). Importantly, for its activity, E(z) requires the binding of two other PRC2 core components, Su(z)12/suppressor of zeste 12 (Suz12), and Esc/embryonic ectoderm development (Eed) (Morey and Helin, 2010).

H3K27me3 is a fundamental histone mark (hallmark) of Polycomb binding. Frequently, H3K27me3 is spread out to broad regions marking large PcG domains allowing for PREs-mediated repression several tens of kilobases away from target genes (Schuettengruber et al., 2007; Morey and Helin, 2010). H3K27me3 also represents a docking site recognized by PC (Cbx in vertebrates) contained in the PRC1 complex. Based on this, a sequential PRC2, PRC1 recruitment has been proposed (Cao et al., 2002). Nevertheless, it was recently shown that PRC1 recruitment to target genes in mammals can be also independent from PRC2 (Gao et al., 2012; Tavares et al., 2012).

POLYCOMB AND REPEATS

Polycomb-associated histone marks are prevalent in the mammalian genome. Quantitative mass spectrometry studies reported that up to 70% of histone H3 carries the PRC2 histone marks H3K27me2 or me3 (Peters et al., 2003; Schoeftner et al., 2006). However, genes and known functional elements comprise only up to 10% of the mammalian genome (Pheasant and Mattick, 2007), while over two-thirds of the remaining part is composed of repetitive elements (De Koning et al., 2011). Hence, this simple observation raises interesting questions about the possible acquirement of novel functions by the PcG proteins along with evolution, involving the non-coding fraction of the mammalian DNA.

Several reports show the presence of Polycomb repressive histone marks on repetitive elements. Initially, PcG silencing on repeats was described as a compensatory mechanism upon loss of H3K9me3 repression in pericentric DNA, where H3K27me1 was converted into H3K27me3 (Peters et al., 2003). More recently, the characterization of the epigenetic pattern of ERV-L and MaLR retrotransposons revealed that they are marked by H3K27me3 (Day et al., 2010), and importantly, a crucial role for ERV-L retrotransposons in embryo totipotency and development has been described (Macfarlan et al., 2011, 2012). At the very early two-cell stage, the murine endogenous retroviral elements ERVL (MuERV-L) are transiently de-repressed (Kigami et al., 2003). Their expression is significant, as it represents 3% of the total transcriptional output, and it is very sharply regulated in time, as it is specific for the developmental stage of the embryo where blastomeres are still totipotent (Svoboda et al., 2004).

Importantly, ERVL transcripts represent a source of regulatory elements which is co-opted by cellular genes to co-regulate their cell stage-specific expression (Macfarlan et al., 2012). In this process, more than 25% of MuERV-L copies are activated and 307 protein-coding genes generate 626 different chimeric transcripts with MuERV-L elements. Among the genes that use alternative MuERV-L-LTR promoters to initiate their transcription, there are transcription factors like Gata-4, which is involved in lineage determination and embryo development (Soudais et al., 1995) and is a known PcG target (Tiwari et al., 2008). Remarkably, MuERV-L expression is regulated by histone modifications like H3K4me3, the active histone mark typical of TrxG proteins (Schuettengruber et al., 2011). In fact, in the absence of the H3K4me3 demethylase LSD1/KDM1A, which is critical for the H3K4/H3K27 methylation balance in human ES cells (Adamo et al., 2011), MuERV-L/MERVL becomes overexpressed and embryonic development arrests at gastrulation (Macfarlan et al., 2011). Given that Polycomb and Trithorax are the major players in development and ERV-L is repressed via PcG mediated-silencing (Day et al., 2010), it is tempting to speculate that MuERV-L retrotransposons undergo a Polycomb/Trithorax regulation, with Polycomb mediating their repression and Trithorax their spatiotemporal-specific up-regulation in order to drive cell-fate specification.

A direct link between Polycomb and repeats-mediated silencing has been recently reported (Leeb et al., 2010). This work, in fact, not only identified both murine leukemia virus (MLV) and IAP retroelements as targets of Polycomb complexes, but also performed the first PcG loss-of-function study in a genomic repeat contest. Indeed, upon double knock out of key PRC1 and PRC2 components, Leeb et al. observed a strong increase

in expression of LTR retrotransposons, which in turn provoked their active mobilization (Leeb et al., 2010). In particular, both MLV and IAP elements were found strongly de-repressed in ES cells double null for the Polycomb proteins Eed and Ring1B when compared to both wild type and single KO cells. Importantly, loss of binding of Polycomb complexes on MLV and the subsequent de-repression of these elements was associated with a considerable increase in MLV mobilization (Leeb et al., 2010). Similarly, Eed$^{-/-}$ Ring1B$^{-/-}$ ES cells showed IAP de-repression, that was associated with reduced levels of DNA methylation on IAP repeats in the double KO and Eed$^{-/-}$ ES cells, in agreement with a previous report about repressive function on IAP retroelements of DNA methylation (Walsh et al., 1998). Hence, this work once again showed a redundancy in the mechanisms of repeat silencing, similarly to that previously reported for other repressive histone marks (Peters et al., 2003; Svoboda et al., 2004). Both PRC1 and PRC2 complexes, in fact, are recruited in parallel for LTR PcG-mediated silencing, as the single KO produced only a partial effect of de-repression (Leeb et al., 2010), thus suggesting that mechanisms of retrotransposon repression act redundantly even when mediated by the Polycomb machinery. Based on these results, it was suggested that genomic repeats, for their intrinsic feature of being present in several copies in the genome, could constitute binding platforms for mammalian PcG complexes (Leeb et al., 2010). Notably, epigenetic silencing of transgenes present in multiple copies has been already described in mice (Garrick et al., 1998; Festenstein et al., 1999; Hiragami and Festenstein, 2005) and it is well-established that proximity of DNA binding sites encourages cooperation among transcription factors (Amouyal et al., 1998; Amouyal, 2007).

Since the greatest proportion of Polycomb-mediated chromatin modifications is located in non-genic regions, a loss of PcG activity would need to be considered not only for its specific effect on Polycomb targets, but also for its possible effects on genome stability.

ncRNAs IN A POLYCOMB AND REPEAT LANDSCAPE

Repeats can be specifically transcribed. Around 6–30% of the total amount of transcripts in mammalian cells initiates within repetitive elements and their expression is frequently tissue-specific (Faulkner et al., 2009). Recent studies show that repeats play central roles in regulating gene expression at multiple levels (Norris et al., 1995; Speek, 2001; Faulkner and Carninci, 2009; Kaneko et al., 2011; Shen et al., 2011). Repetitive elements may regulate the expression of nearby protein-coding genes by providing tissue-specific promoters or enhancers (Speek, 2001; Conley et al., 2008; Faulkner et al., 2009); they can be co-opted to generate alternative exons (Zhang and Chasin, 2006); they can modulate the abundance of gene products, for example through generation of ncRNAs, working *in trans* or *in cis*, either enhancing (by anti-silencing) or reducing (by transcriptional interference) their expression (Allen et al., 2004); or they can produce short ncRNAs exploited by RNAi machinery (Ghildiyal et al., 2008; Watanabe et al., 2008; Faulkner and Carninci, 2009) (**Figure 1**).

If the regulatory functions are combined with the ability of retrotransposons to mobilize upon de-repression or of tandem repeats to rearrange during meiotic division, the scenario becomes even more complex, as novel insertions of mobile repetitive elements or change in tandem repeat copy number may

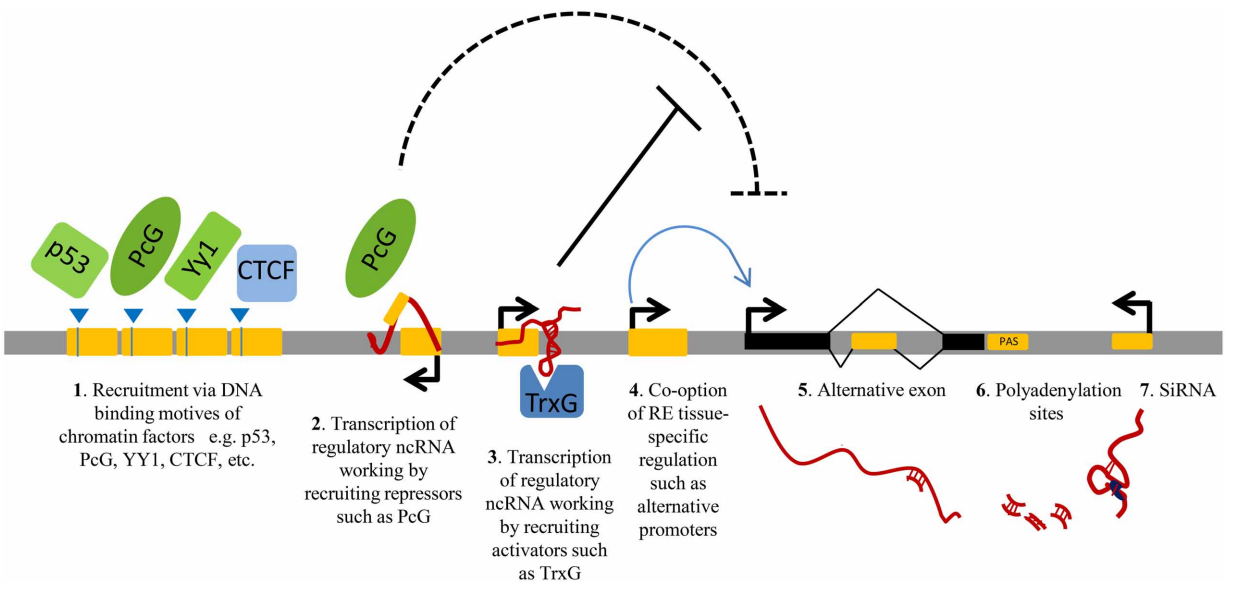


FIGURE 1 | A schematic view of the principal gene-regulatory functions of repeats. Repetitive elements (RE, represented as yellow modules) have an impact on gene expression by providing DNA binding sites for transcription factors or chromatin regulators **(1)**; upon transcription, by generating regulatory non-protein-coding RNAs (ncRNAs) involved in gene silencing **(2)** or gene activation **(3)**, for example via direct recruitment of repressors/activators such as Polycomb (PcG) and Trithorax (TrxG) **(2–3)**. Repeat sequences can also contribute to gene transcription by providing alternative promoters **(4)**; alternative exons **(5)**; polyadenylation sites (PAS, **6**) or they can influence the stability of other transcripts via RNA interference (RNAi) by producing short double-stranded RNAs (dsRNAs) **(7)**.

modify the chromatin structure (Lunyak et al., 2007) and the gene regulation of nearby genes (Cabianca et al., 2012).

Important examples of the interplay between Polycomb, repeats and ncRNAs in normal physiology and in disease are illustrated in the following sections.

X-INACTIVATION

X-inactivation, the process that leads to the silencing of one X chromosome in mammalian female cells, represents one of the most striking examples of long-range chromosomal regulation involving ncRNAs, Polycomb-mediating silencing and DNA repeats (Hall and Lawrence, 2010). In mammals, a large non-coding RNA named *Xist* "paints" the X-chromosome *in cis* (Brockdorff et al., 1992; Brown et al., 1992; Clemson et al., 2006; Chow et al., 2007) and induces a silencing cascade repressing the whole chromosome territory (Hall and Lawrence, 2003; Heard and Disteche, 2006). *Xist* works by recruiting PRC1, PRC2 and their respective histone marks (Leeb and Wutz, 2007) to the core of the inactive X chromosome, which contains genomic repeats (Chaumeil et al., 2006; Clemson et al., 2006). Besides local changes, a higher-order remodeling of the chromatin architecture takes place, thus producing the well-known silent core corresponding to the DAPI-dense Barr Body, which resides in the heterochromatic compartment at the nuclear or nucleolar periphery (Clemson et al., 2006).

Different classes of repeats play their roles in X-inactivation. Common repeats, like LINE-1 and Alu, participate structurally in the formation of the heterochromatic inner core of the Xi DNA territory (Hall and Lawrence, 2010), whereas a role for the euchromatic DXZ4 macrosatellite locus in Xi chromosome has been suggested (Chadwick, 2008). Moreover, the *Xist* ncRNA contains several tandem repeats termed A, B, C, D, E, and F (Hendrich et al., 1997; Nesterova et al., 2001; Yen et al., 2007; Horvath et al., 2011). Repeat A, with its conserved sequence and tetra-loop structure (Duszczyk et al., 2011), is essential for Polycomb-mediated silencing of X-linked genes (Wutz et al., 2002; Zhao et al., 2008). In fact, in the future Xi chromosome, PRC2 is initially recruited by the 1.6 kb *RepA* ncRNA, which is directly bound by the PRC2 subunit Ezh2. The *RepA*/PRC2 interaction enables the full-length *Xist* induction and thus the spreading of the *Xist* ncRNA and PcG silencing on the whole Xi chromosome (Zhao et al., 2008). The *RepA* region is the primary target of PcG binding also within the 17 kb full-length ncRNA *Xist* (Zhao et al., 2008), and indeed in *RepA* mutants, *Xist* recruits 80–90% less PRC2 (Kohlmaier et al., 2004).

The antisense 40 kb *Tsix* ncRNA is able to inhibit the *RepA*/Ezh2 interaction, probably by competing with *Xist* for PRC2 binding (Zhao et al., 2008). In pre-XCI cells, *Tsix* keeps in check the state of both X chromosomes and only a few molecules of *Xist* are transcribed (Zhao et al., 2008). When cell differentiation triggers dosage compensation, another regulatory ncRNA named *Jpx* becomes actively transcribed from the *Xist* loci of both X chromosomes, thus supplying the required activator for high-level *Xist* expression (Tian et al., 2010). In the future Xi, *Tsix* is now down-regulated, hence producing a permissive state for *Xist* induction, whereas, in the future Xa, the levels of *Tsix* continue titrating away PcG from *RepA*, thus maintaining blocked the repressive cascade (Zhao et al., 2008).

Another important *Xist* repeat is Repeat C, a C-rich sequence, specific of *Xist* and highly conserved, which is important for *Xist* localization on the inactive X chromosome (Memili et al., 2001; Sarma et al., 2010). A recent report provided an important role for another repeat of the *Xist* locus, Repeat F (Jeon and Lee, 2011). This region, characterized by the presence of CTCF and YY1 binding sites, is bound by YY1, which with its multiple zinc fingers is able to bind both DNA and RNA at the same time. YY1 bridges the *Xist* ncRNA via Repeat C (Sarma et al., 2010), and the X chromosome, via the Repeat F region. Overall, X inactivation provides a strong argument for an important physiological interplay between repeats, Polycomb, and ncRNAs (**Figure 2**).

FSHD MUSCULAR DYSTROPHY

FSHD (OMIM 158900) is a genetic disorder of particular interest for the atypical interactions between genetic and epigenetic players, which both contribute to the etiology of the disease (Neguembor and Gabellini, 2010). FSHD is an autosomal dominant disease and for more than 20 years it has been known to be associated with reduction in copy number of a macrosatellite repeat (called D4Z4) mapping to the subtelomeric 4q35 region (Wijmenga et al., 1990, 1991, 1992; Van Deutekom et al., 1993). Also, it has been known for a decade that D4Z4 deletions cause de-repression of genes located nearby (Gabellini et al., 2002). Nevertheless, the molecular understanding of the D4Z4 repeat mechanism of action was only recently provided (Cabianca et al., 2012).

Each D4Z4 unit is extremely GC rich, containing a sequence nearly identical to the consensus motif of *Drosophila* PREs and several putative DNA binding sites for factors which are Polycomb recruiters in *Drosophila*, such as YY1 and GAGA factor (Mihaly et al., 1998; Busturia et al., 2001; Mishra et al., 2001; Gabellini et al., 2002; Cabianca et al., 2012). Accordingly, in healthy subjects the D4Z4 tandem array is extensively bound by PRC1 and PRC2 and displays enrichment for the typical PcG-associated repressive histone marks H2AK119Ub and H3K27me3. The region is also bound by proteins associated to Polycomb recruitment in mammals like Jarid2 (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010) or homologs of PcG recruiters in *Drosophila* (YY1, HMGB2, c-Krox/Th-POK; vertebrate fly homologs Pho, Dsp1, GAGA factor, respectively) (Busturia et al., 2001; Mishra et al., 2001; Gabellini et al., 2002; Dejardin et al., 2005; Matharu et al., 2010). Finally, the repeats array also shows enrichment for the Polycomb-associated histone variant macroH2A (Buschbeck et al., 2009).

Importantly, D4Z4 is able to initiate PcG recruitment to ectopic sites and mediate copy number-dependent repression of gene expression, typical features of *Drosophila* PREs (Gabellini et al., 2002; Cabianca et al., 2012). In FSHD patients, the reduction in D4Z4 copy number is associated with a reduction in PcG silencing. This allows for the production of a long, chromatin-associated ncRNA: *DBE-T*. *DBE-T* works *in cis* by directly recruiting the TrxG protein ASH1L to the 4q35 locus. This leads to a structural and epigenetic remodeling of the FSHD locus, toward a more active chromatin state, which is responsible

for the de-repression of 4q35 genes. Altogether, FSHD constitutes an important example of the relevance of DNA repeats, Polycomb and ncRNAs in human genetic diseases (**Figure 3**).

REPETITIVE ELEMENTS AND CHROMATIN ORGANIZATION IN THE 3D NUCLEAR SPACE

In general, nuclear organization of chromatin reflects its active or inactive state. Euchromatin occupies the internal nucleoplasm, whereas heterochromatin preferentially localizes at the nuclear and nucleolar periphery (Kosak et al., 2002; Shopland et al., 2003). Accordingly, repetitive elements can also localize differently. For example, pericentromeric satellite repeats are usually confined to the heterochromatic domains of the nuclear periphery whereas telomeres of human chromosomes usually reside in the internal compartment (Tam et al., 2004). There are important exceptions; the FSHD-associated 4q35 telomere behaves differentially, being usually associated to the nuclear periphery (Masny et al., 2004; Tam et al., 2004).

The nuclear machineries are not uniformly distributed in the nucleoplasm, but are organized in functional sub-compartments, so-called "factories" or "hubs" (Lamond and Spector, 2003; Hall et al., 2006; Meaburn and Misteli, 2007). In fact, by staining for

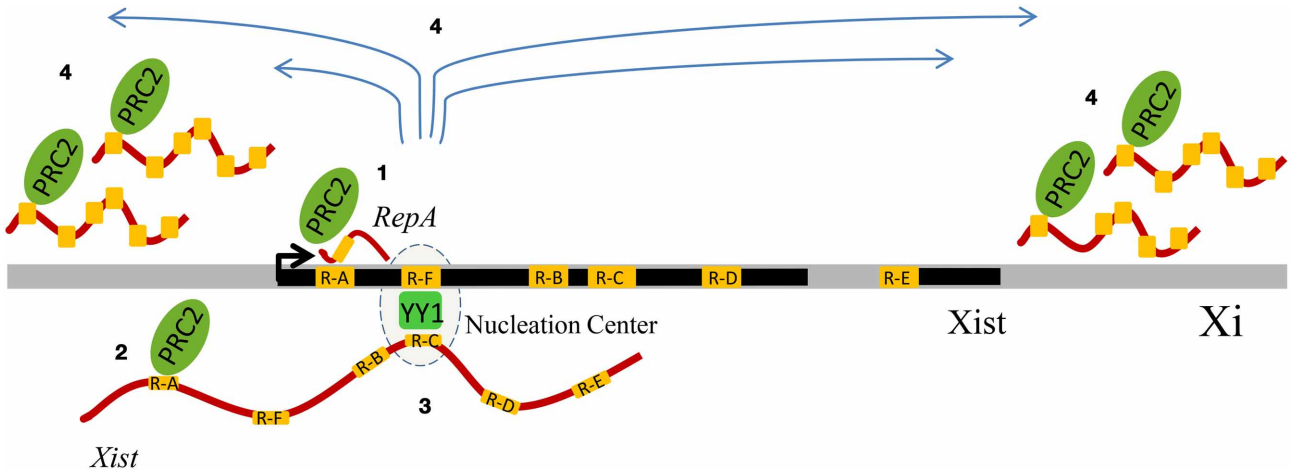


FIGURE 2 | Schematic summary of the influence of genomic repeats on X-chromosome inactivation (XCI). The *Xist* DNA locus displays tandem repeats (visualized in yellow) and generates multiple transcripts (such as *RepA* and *Xist*), whose contribution to XCI involves binding to Polycomb Repressive Complex 2 (PRC2) and YY1, which has also been associated to Polycomb. Four sequential events of XCI are represented. During the initiation phase of XCI, the Repeat A (R-A) region of the ncRNA *RepA* recruits PRC2, creating the conditions for the production of the full-length Xist RNA **(1)**. *Xist* co-transcriptionally binds PRC2 via its R-A region, and it is loaded onto chromatin **(2)**. YY1 functions as a bridge and anchors *Xist in cis*, by binding both *Xist* RNA and DNA, respectively via their Repeat C (R-C) and Repeat F (R-F) regions **(3)**. *Xist* RNA, first bound only on the nucleation center, spreads *in cis* and recruits PRC2, thus mediating the X-chromosome inactivation **(4)**.

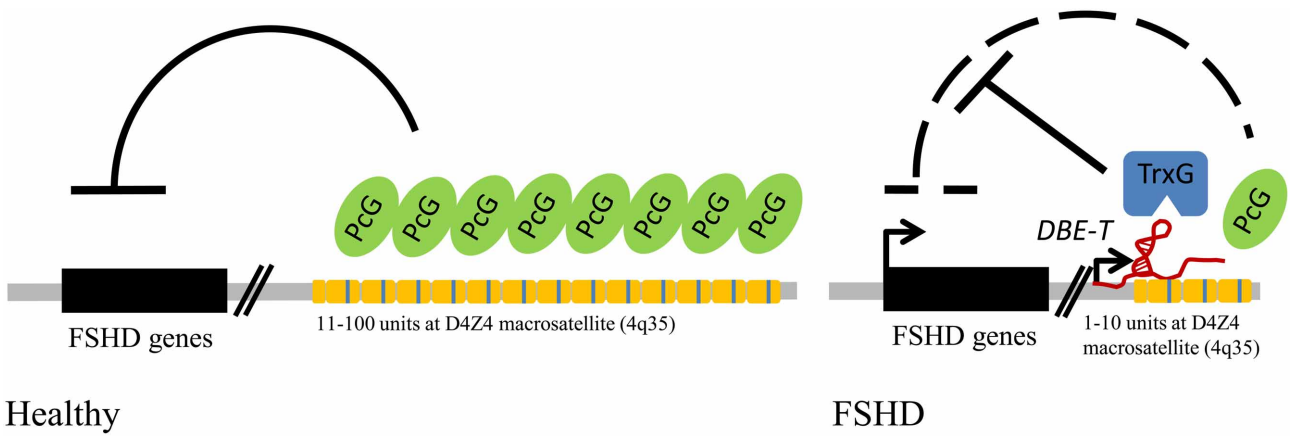


FIGURE 3 | FSHD muscular dystrophy links repetitive elements, Polycomb proteins, and ncRNAs in a human genetic disease. Model for FSHD molecular pathogenesis: in healthy individuals the repetitive elements (yellow modules) of the D4Z4 macrosatellite at 4q35 are bound by Polycomb (PcG) proteins, which mediate gene repression; in FSHD patients the shortening below the threshold of 11 copies generates an epigenetic remodeling of the locus, sustained by a long non-coding RNA (*DBE-T*), and the recruitment of Trithorax (TrxG) proteins, which leads to up-regulation of disease candidate genes.

a particular key factor of important nuclear processes (like transcription, RNA processing, replication, or DNA repair), a number of discrete structures appear in the nucleus, which result from the local concentration of proteins involved in specific nuclear processes. For example, "transcription factories" have been described and different genes, localized on distant chromosomal loci, can associate to the same active foci to be co-transcribed (Osborne et al., 2004). For nuclear compartments, patterns of distribution in the nucleus, characteristic of the different cell type or differentiation state, can be recognized (Lanctot et al., 2007). However, it is still an open question whether a fragment of DNA needs to be primarily attracted to one of these nuclear compartments in order to be functionally processed, or if the specific machinery can also activate elsewhere in the nucleus but needs to reach these structures for a higher efficiency.

Polycomb proteins and associated histone marks reside in discrete nuclear structures called Polycomb bodies, co-localizing with stably repressed homeotic genes (Messmer et al., 1992; Buchenau et al., 1998; Grimaud et al., 2006; Ferraiuolo et al., 2010; Bantignies et al., 2011). These repressive chromatin hubs are composed of chromatin loops involving PcG-bound regulatory elements and promoters of PcG target genes (Cleard et al., 2006; Comet et al., 2011). Hence not only events associated with gene activation, but also those associated with gene repression, including the ones involving Polycomb proteins, can localize on discrete foci, where long-range interactions take place.

The organization of these structures in *Drosophila* starts at the level of PREs, the DNA modules recruiting Polycomb complexes (Muller and Kassis, 2006; Schuettengruber et al., 2007). As already discussed, the histone-methylation activity of the PRC2 complex spreads out on neighboring regions, marking large PcG domains. Hence, PcG silencing reaches target genes that are tens of kilobases distant from a PRE. Moreover, PREs tend to cluster in larger domains (Bantignies and Cavalli, 2011) (**Figure 4**).

As characterized by chromosome conformation capture experiments, long distance intra- and even inter-chromosomal interactions among PcG targets are established, thus producing a major level of chromatin organization in the 3D nuclear space (Lanzuolo et al., 2007; Terranova et al., 2008; Tiwari et al., 2008; Eskeland et al., 2010; Comet et al., 2011; Tolhuis et al., 2011). It has been proposed that these long-range contacts are mediated by ncRNAs (Rinn et al., 2007), insulators DNA element (Li et al., 2011) and RNAi machinery (Grimaud et al., 2006). On top of such a hierarchal organization of PcG domains are found the PcG bodies. PcG bodies differ in size and Polycomb intensity. In particular, PcG domains with a larger linear size display a higher content of Polycomb and generate bigger and more intense PcG bodies (Cheutin and Cavalli, 2012).

The discovery of PcG bodies raised questions about their function: are they merely the result of the accumulation of PcG proteins to clustered Polycomb domains, or is the formation of these "hubs" required for PcG silencing (Buchenau et al., 1998)? The fact that PcG proteins organize in such PcG bodies instead of being uniformly distributed in the nucleus is already an indication toward a functional role for these structures. Indeed, PcG-mediated gene silencing occurs within PcG bodies

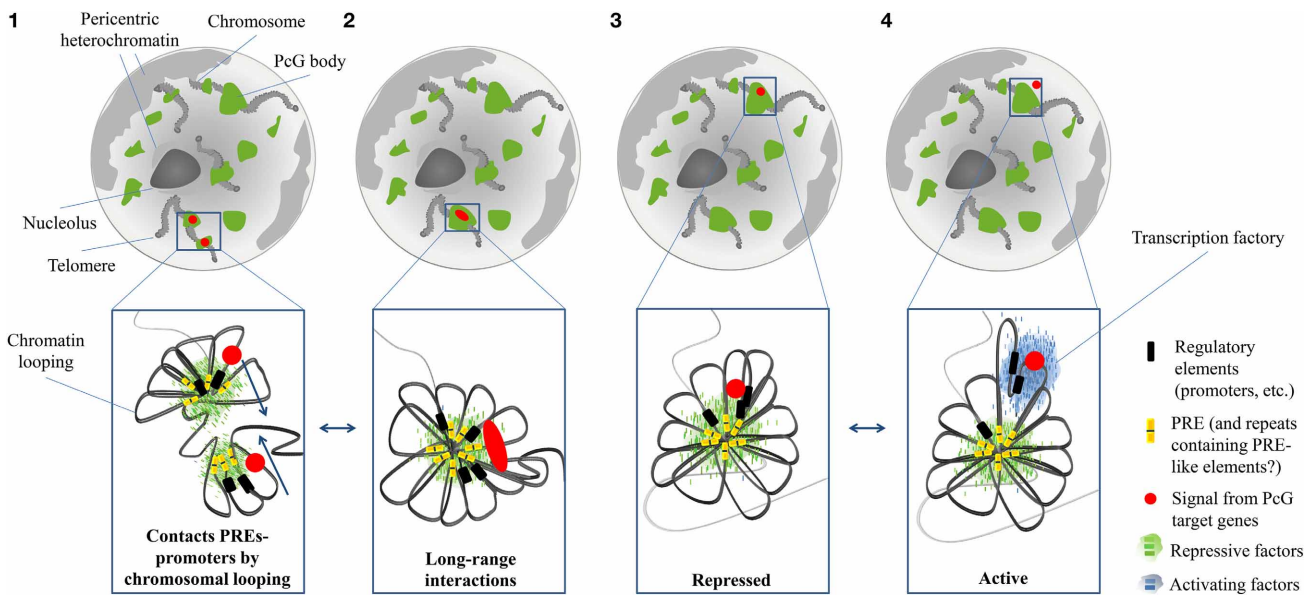


FIGURE 4 | Schematic representation of the distribution and dynamics of Polycomb (PcG) bodies within the nucleus. PcG bodies (green) are hubs where, by chromatin looping, Polycomb Response Elements (PREs; yellow bars) closely interact with promoters (black bars) of PcG target genes (red circles), and where PcG proteins and other repressive factors (small green bars) accumulate, thanks to the binding to PREs **(1)**. PcG-bound elements and promoters are able to engage long-range chromatin interactions, so that two different PcG bodies cluster into the same structure. Red oval symbolizes co-localization of independent signals from remote PcG target genes **(2)**. Chromatin loops can adopt different spatial conformations, so that PcG target genes can be retained or displaced from PcG bodies, depending on their transcriptional state. In the repressed state, a condensed structure tightens the interactions among all PcG-bound elements **(3)**. When a stimulus activates the transcription of a PcG target gene, its promoter loses the interaction with PREs, and co-localizes with activators (small blue bars), within transcription factories (blue cloud) **(4)**.

(Grimaud et al., 2006) and it has been proposed that the local concentration of PcG components and their target genes in PcG bodies may produce chromatin condensation (Terranova et al., 2008; Eskeland et al., 2010). Indeed, a correlation between repression of PcG targets and their localization in PcG bodies has been reported. For example, Fab-7, the PRE-containing region controlling the expression of the gene Abd-B, is found within PcG bodies when Abd-B is repressed, whereas it is outside the PcG bodies when Abd-B is expressed (Lanzuolo et al., 2007; Bantignies and Cavalli, 2011; Bantignies et al., 2011). Active genes are displaced from these repressive chromatin hubs not only in *Drosophila* but also in mammals. For example, the human GATA-4 locus, involving several PcG bound regions, shows a similar chromatin structure depending on its transcriptional state (Tiwari et al., 2008) (**Figure 4**).

As it primarily functions as a marker regulator of development, Polycomb accumulation, and thus the presence of PcG bodies are regulated during cell differentiation. Experiments of fluorescence recovery after photobleaching (FRAP) in both *Drosophila* and mammalian embryonic stem cells, showed a dynamic exchange of PcG proteins between PcG bodies and nucleoplasm (Ficz et al., 2005; Ren et al., 2008). In *Drosophila*, Polycomb starts accumulating in the nucleus during the early stages of development (stage 5), progressively increases and gets recruited to PcG bodies (stages 5–11), until it becomes stably associated with PcG bodies during late embryogenesis (Cheutin and Cavalli, 2012). To address the question of whether the formation of PcG bodies is the direct result of PcG binding to their targets or, on the contrary, PcG targets need to associate with PcG bodies in order to be repressed, *in vivo* live imaging approaches have been used to characterize the motion of PcG targets and PcG bodies in the nucleus. Interestingly, a motion away from PcG bodies from the nuclear periphery toward the nuclear interior, regulated by actin and nuclear myosin I, was observed immediately after inducing transcription (Chuang et al., 2006). Similarly to other chromatin domains, Polycomb bodies' motion sensitively decreases upon differentiation, and shows similar kinetics, either fast but limited to volumes much smaller than chromosome territory occupancy, or slow but involving overall a higher level of nuclear structure (Cheutin and Cavalli, 2012).

Based on the fact that Polycomb is concentrated in PcG bodies by immunofluorescence and in repeats by chromatin immunoprecipitation, it could be hypothesized that genomic repeats which are Polycomb targets in mammals could functionally behave in a similar way to PREs and mediate association between Polycomb-regulated genes. In this view, Polycomb complexes and repetitive elements would play a role in the compartmentalization of the nucleus, establishing large chromatin domains where PcG target genes are efficiently repressed. Interestingly, it has been shown that the 3D organization of PcG target genes can influence PcG-mediated silencing. In *Drosophila* the deletion of Fab-7 perturbed the interaction between BX-C and ANT-C, producing mild effects on gene expression at distant Polycomb target genes. However, sensitized genetic backgrounds had to be used in order to observe homeotic phenotypes (Bantignies et al., 2011). Interestingly, in mammals structural alterations of repetitive sequences can affect long-range PcG-mediated silencing *in cis* (Cabianca et al., 2012). Moreover, deletions or mutations of genetic elements on one chromosome can affect expression of interacting genes *in trans* (Spilianakis et al., 2005; Ling et al., 2006).

Collectively, these considerations strongly indicate that investigation of the role of repetitive sequences in nuclear structural organization in mammals is an important topic for future research. This will require a significant operational and conceptual shift. Operationally, genome-wide approaches would have to be tailored to the analysis of repetitive sequences, which represents a serious bioinformatics challenge. Conceptually, investigators should take into consideration the biological relevance of the major component of the human genome, being aware that this could potentially change the understanding of how the nuclear processes work.

ACKNOWLEDGMENTS

This work is a partial fulfillment of Valentina Casa's PhD in Molecular Medicine, Program in Neuroscience, San Raffaele University, Milano, Italy. The Gabellini laboratory is supported from the European Research Council (ERC), the Italian Epigenomics Flagship Project, the Italian Ministry of Health and the FSHD Global Research Foundation. D. Gabellini is a Dulbecco Telethon Institute Assistant Scientist.

REFERENCES

Adamo, A., Sese, B., Boue, S., Castano, J., Paramonov, I., Barrero, M. J., et al. (2011). LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat. Cell Biol.* 13, 652–659.

Allen, T. A., Von Kaenel, S., Goodrich, J. A., and Kugel, J. F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* 11, 816–821.

Ames, D., Murphy, N., Helentjaris, T., Sun, N., and Chandler, V. (2008). Comparative analyses of human single- and multilocus tandem repeats. *Genetics* 179, 1693–1704.

Amouyal, M. (2007). Transition from DNA looping to simple binding or DNA pairing in gene regulation and replication: a matter of numbers for the cell. *Genes Genomes Genomics Glob. Sci. Books* 1, 104–112.

Amouyal, M., Perez, N., and Rolland, S. (1998). Adjacent cooperation of proteins on DNA are not representative of long-distance interactions. *C R Acad. Sci. III* 321, 877–881.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. *Trends Genet.* 27, 454–464.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., et al. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* 144, 214–226.

Batzer, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379.

Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S., and Devine, S. E. (2004). Natural genetic variation caused by transposable elements in humans. *Genetics* 168, 933–951.

Blackburn, E. H. (1984). The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.* 53, 163–194.

Block, G. J., Petek, L. M., Narayanan, D., Amell, A. M., Moore, J. M., Rabaia, N. A., et al. (2012). Asymmetric bidirectional transcription from the FSHD-causing D4Z4 array modulates DUX4 production. *PLoS ONE* 7:e35532. doi: 10.1371/journal.pone.0035532

Blower, M. D., Sullivan, B. A., and Karpen, G. H. (2002). Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* 2, 319–330.

Bodega, B., Ramirez, G. D., Grasser, F., Cheli, S., Brunelli, S., Mora,

M., et al. (2009). Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biol.* 7, 41.

Bollati, V., Fabris, S., Pegoraro, V., Ronchetti, D., Mosca, L., Deliliers, G. L., et al. (2009). Differential repetitive DNA methylation in multiple myeloma molecular subgroups. *Carcinogenesis* 30, 1330–1335.

Boumil, R. M., Ogawa, Y., Sun, B. K., Huynh, K. D., and Lee, J. T. (2006). Differential methylation of Xite and CTCF sites in Tsix mirrors the pattern of X-inactivation choice in mice. *Mol. Cell. Biol.* 26, 2109–2117.

Bourque, G., Leong, B., Vega, V. B., Chen, X., Lee, Y. L., Srinivasan, K. G., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762.

Brack, C., Hirama, M., Lenhard-Schuller, R., and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. *Cell* 15, 1–14.

Britten, R. J. (2010). Transposable element insertions have strongly affected human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19945–19948.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., et al. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71, 515–526.

Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., et al. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D. J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. *J. Cell Biol.* 141, 469–481.

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., et al. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat. Struct. Mol. Biol.* 16, 1074–1079.

Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development* 128, 2163–2173.

Cabianca, D. S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., et al. (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 149, 819–831.

Cabianca, D. S., and Gabellini, D. (2010). The cell biology of disease: FSHD: copy number variations on the theme of muscular dystrophy. *J. Cell Biol.* 191, 1049–1060.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039–1043.

Chadwick, B. P. (2008). DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* 18, 1259–1269.

Chadwick, B. P. (2009). Macrosatellite epigenetics: the two faces of DXZ4 and D4Z4. *Chromosoma* 118, 675–681.

Chadwick, B. P., and Willard, H. F. (2004). Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17450–17455.

Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S., and Lee, J. T. (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. *Science* 295, 345–347.

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.

Cheutin, T., and Cavalli, G. (2012). Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. *PLoS Genet.* 8:e1002465. doi: 10.1371/journal.pgen.1002465

Choo, K. H. (1997). Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am. J. Hum. Genet.* 61, 1225–1233.

Chow, J. C., Hall, L. L., Baldry, S. E., Thorogood, N. P., Lawrence, J. B., and Brown, C. J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10104–10109.

Chuang, C. H., Carpenter, A. E., Fuchsova, B., Johnson, T., De Lanerolle, P., and Belmont, A. S. (2006). Long-range directional movement of an interphase chromosome site. *Curr. Biol.* 16, 825–831.

Clapp, J., Mitchell, L. M., Bolland, D. J., Fantes, J., Corcoran, A. E., Scotting, P. J., et al. (2007). Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 81, 264–279.

Cleard, F., Moshkin, Y., Karch, F., and Maeda, R. K. (2006). Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nat. Genet.* 38, 931–935.

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.

Comet, I., Schuettengruber, B., Sexton, T., and Cavalli, G. (2011). A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2294–2299.

Conley, A. B., Miller, W. J., and Jordan, I. K. (2008). Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet.* 24, 53–56.

Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., et al. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* 460, 1127–1131.

Courtier, B., Heard, E., and Avner, P. (1995). Xce haplotypes show modified methylation in a region of the active X chromosome lying 3′ to Xist. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3531–3535.

Cuddapah, S., Roh, T. Y., Cui, K., Jose, C. C., Fuller, M. T., Zhao, K., et al. (2012). A novel human polycomb binding site acts as a functional polycomb response element in Drosophila. *PLoS ONE* 7:e36365. doi: 10.1371/journal.pone.0036365

Day, D. S., Luquette, L. J., Park, P. J., and Kharchenko, P. V. (2010). Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol.* 11, R69.

De Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384

De Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* 7, 663–676.

Deidda, G., Cacurri, S., Grisanti, P., Vigneti, E., Piazzo, N., and Felicetti, L. (1995). Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur. J. Hum. Genet.* 3, 155–167.

Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., et al. (2005). Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. *Nature* 434, 533–538.

Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., et al. (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18157–18162.

Dong, K. B., Maksakova, I. A., Mohn, F., Leung, D., Appanah, R., Lee, S., et al. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* 27, 2691–2701.

Donohoe, M. E., Zhang, L. F., Xu, N., Shi, Y., and Lee, J. T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol. Cell* 25, 43–56.

Duszczyk, M. M., Wutz, A., Rybin, V., and Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *RNA* 17, 1973–1982.

Ekram, M. B., Kang, K., Kim, H., and Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. *Epigenetics* 7, 370–382.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., et al. (2010). Ring1B compacts

chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* 38, 452–464.

Faulkner, G. J. (2011). Retrotransposons: mobile and mutagenic from conception to death. *FEBS Lett.* 585, 1589–1594.

Faulkner, G. J., and Carninci, P. (2009). Altruistic functions for selfish DNA. *Cell Cycle* 8, 2895–2900.

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571.

Ferraiuolo, M. A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X. Q., Nadler, M., et al. (2010). The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* 38, 7472–7484.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405.

Festenstein, R., Sharghi-Namini, S., Fox, M., Roderick, K., Tolaini, M., Norton, T., et al. (1999). Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal-context-dependent manner. *Nat. Genet.* 23, 457–461.

Ficz, G., Heintzmann, R., and Arndt-Jovin, D. J. (2005). Polycomb group protein complexes exchange rapidly in living Drosophila. *Development* 132, 3963–3976.

Filippova, G. N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. *Curr. Top. Dev. Biol.* 80, 337–360.

Gabellini, D., Green, M. R., and Tupler, R. (2002). Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* 110, 339–348.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., et al. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* 45, 344–356.

Garrick, D., Fiering, S., Martin, D. I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. *Nat. Genet.* 18, 56–59.

Gelfand, Y., Rodriguez, A., and Benson, G. (2007). TRDB – the Tandem Repeats Database. *Nucleic Acids Res.* 35, D80–D87.

Geng, L. N., Yao, Z., Snider, L., Fong, A. P., Cech, J. N., Young, J. M., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38–51.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., et al. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* 320, 1077–1081.

Giacalone, J., Friedes, J., and Francke, U. (1992). A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat. Genet.* 1, 137–143.

Gondo, Y., Okada, T., Matsuyama, N., Saitoh, Y., Yanagisawa, Y., and Ikeda, J. E. (1998). Human megasatellite DNA RS447, copy-number polymorphisms and interspecies conservation. *Genomics* 54, 39–49.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* 124, 957–971.

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

Guil, S., Soler, M., Portela, A., Carrere, J., Fonalleras, E., Gomez, A., et al. (2012). Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat. Struct. Mol. Biol.* 19, 664–670.

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.

Hall, L. L., and Lawrence, J. B. (2003). The cell biology of a novel chromosomal RNA: chromosome painting by XIST/Xist RNA initiates a remodeling cascade. *Semin. Cell Dev. Biol.* 14, 369–378.

Hall, L. L., and Lawrence, J. B. (2010). XIST RNA and architecture of the inactive X chromosome: implications for the repeat genome. *Cold Spring Harb. Symp. Quant. Biol.* 75, 345–356.

Hall, L. L., Smith, K. P., Byron, M., and Lawrence, J. B. (2006). Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288, 664–675.

Heard, E., and Disteche, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* 20, 1848–1867.

Hendrich, B. D., Plenge, R. M., and Willard, H. F. (1997). Identification and characterization of the human XIST gene promoter: implications for models of X chromosome inactivation. *Nucleic Acids Res.* 25, 2661–2671.

Hiragami, K., and Festenstein, R. (2005). Heterochromatin protein 1, a pervasive controlling influence. *Cell. Mol. Life Sci.* 62, 2711–2726.

Horvath, J. E., Sheedy, C. B., Merrett, S. L., Diallo, A. B., Swofford, D. L., Program, N. C. S., et al. (2011). Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus. *Genome Res.* 21, 850–862.

Hua-Van, A., Le Rouzic, A., Boutin, T. S., Filee, J., and Capy, P. (2011). The struggle for life of the genome's selfish architects. *Biol. Direct* 6, 19.

Igarashi, S., Suzuki, H., Niinuma, T., Shimizu, H., Nojima, M., Iwaki, H., et al. (2010). A novel correlation between LINE-1 hypomethylation and the malignancy of gastrointestinal stromal tumors. *Clin. Cancer Res.* 16, 5114–5123.

Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261.

Jeon, Y., and Lee, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119–133.

Jiang, G., Yang, F., Van Overveld, P. G., Vedanarayanan, V., Van Der Maarel, S., and Ehrlich, M. (2003). Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum. Mol. Genet.* 12, 2909–2921.

Jurka, J., Kapitonov, V. V., Kohany, O., and Jurka, M. V. (2007). Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8, 241–259.

Kaneko, H., Dridi, S., Tarallo, V., Gelfand, B. D., Fowler, B. J., Cho, W. G., et al. (2011). DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* 471, 325–330.

Kanhere, A., Viiri, K., Araujo, C. C., Rasaiyaah, J., Bouwman, R. D., Whyte, W. A., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* 38, 675–688.

Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., et al. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303–1312.

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.

Kazazian, H. H. Jr., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164–166.

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11667–11672.

Kidwell, M. G., and Lisch, D. R. (2000). Transposable elements and host genome evolution. *Trends Ecol. Evol.* 15, 95–99.

Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* 68, 651–654.

Kim, P. M., Lam, H. Y., Urban, A. E., Korbel, J. O., Affourtit, J., Grubert, F., et al. (2008). Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res.* 18, 1865–1874.

Kogi, M., Fukushige, S., Lefevre, C., Hadano, S., and Ikeda, J. E. (1997). A novel tandem repeat sequence located on human chromosome 4p: isolation and characterization. *Genomics* 42, 278–283.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* 2:e171. doi: 10.1371/journal.pbio.0020171

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., et al. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158–162.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.

PLoS Genet. 4:e1000242. doi: 10.1371/journal.pgen.1000242

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* 14, 183–193.

Kuzmichev, A., Margueron, R., Vaquero, A., Preissner, T. S., Scher, M., Kirmizis, A., et al. (2005). Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1859–1864.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 16, 2893–2905.

Lamond, A. I., and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4, 605–612.

Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H. F., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 9, 1167–1174.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971.

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

Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J. Cell Biol.* 178, 219–229.

Leidenroth, A., and Hewitt, J. E. (2010). A family history of DUX4, phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol. Biol.* 10, 364.

Lemmers, R. J., Van Der Vliet, P. J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J. G., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650–1653.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B. E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.

Li, H. B., Muller, M., Bahechar, I. A., Kyrchanova, O., Ohno, K., Georgiev, P., et al. (2011). Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in *Drosophila melanogaster*. *Mol. Cell. Biol.* 31, 616–625.

Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. A., Li, E., et al. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell. Biol.* 22, 480–491.

Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., et al. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 312, 269–272.

Lo, A. W., Craig, J. M., Saffery, R., Kalitsis, P., Irvine, D. V., Earle, E., et al. (2001). A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J.* 20, 2087–2096.

Lopez Castel, A., Cleary, J. D., and Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* 11, 165–170.

Lunyak, V. V., Prefontaine, G. G., Nunez, E., Cramer, T., Ju, B. G., Ohgi, K. A., et al. (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 317, 248–251.

Lyle, R., Wright, T. J., Clark, L. N., and Hewitt, J. E. (1995). The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* 28, 389–397.

Macfarlan, T. S., Gifford, W. D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., et al. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* 25, 594–607.

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

Maksakova, I. A., Mager, D. L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* 65, 3329–3347.

Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., et al. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* 24, 800–812.

Masny, P. S., Bengtsson, U., Chung, S. A., Martin, J. H., Van Engelen, B., Van Der Maarel, S. M., et al. (2004). Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum. Mol. Genet.* 13, 1857–1871.

Matharu, N. K., Hussain, T., Sankaranarayanan, R., and Mishra, R. K. (2010). Vertebrate homologue of Drosophila GAGA factor. *J. Mol. Biol.* 400, 434–447.

McLaughlin, C. R., and Chadwick, B. P. (2011). Characterization of DXZ4 conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. *Genome Biol.* 12, R37.

Meaburn, K. J., and Misteli, T. (2007). Cell biology: chromosome territories. *Nature* 445, 379–781.

Memili, E., Hong, Y. K., Kim, D. H., Ontiveros, S. D., and Strauss, W. M. (2001). Murine Xist RNA isoforms are different at their 3′ ends: a role for differential polyadenylation. *Gene* 266, 131–137.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., et al. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* 6:e1001244. doi: 10.1371/journal.pgen.1001244

Messmer, S., Franke, A., and Paro, R. (1992). Analysis of the functional role of the Polycomb chromo domain in *Drosophila melanogaster*. *Genes Dev.* 6, 1241–1254.

Mihaly, J., Mishra, R. K., and Karch, F. (1998). A conserved sequence motif in Polycomb-response elements. *Mol. Cell* 1, 1065–1066.

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

Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191.

Mishra, R. K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., et al. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol. Cell. Biol.* 21, 1311–1318.

Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends Biochem. Sci.* 35, 323–332.

Morris, C. A., and Moazed, D. (2007). Centromere assembly and propagation. *Cell* 128, 647–650.

Moseley, S. C., Rizkallah, R., Tremblay, D. C., Anderson, B. R., Hurt, M. M., and Chadwick, B. P. (2012). YY1 associates with the macrosatellite DXZ4 on the inactive X chromosome and binds with CTCF to a hypomethylated form in some male carcinomas. *Nucleic Acids Res.* 40, 1596–1608.

Muller, J., and Kassis, J. A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.* 16, 476–484.

Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910.

Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446.

Myers, R. M., Stamatoyannopoulos, J. S., Dunham, I., Hardison, R. C., Bernstein, B. E., Gingeras, T. R., et al. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9:e1001046. doi: 10.1371/journal.pbio.1001046

Neguembor, M. V., and Gabellini, D. (2010). In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy. *Epigenomics* 2, 271–287.

Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M.

E., et al. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* 11, 833–849.

Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., et al. (1995). Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* 270, 22777–22782.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36, 1065–1071.

Ottaviani, A., Rival-Gervier, S., Boussouar, A., Foerster, A. M., Rondier, D., Sacconi, S., et al. (2009). The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet.* 5:e1000394. doi: 10.1371/journal.pgen.1000394

Pasini, D., Cloos, P. A., Walfridsson, J., Olsson, L., Bukowski, J. P., Johansen, J. V., et al. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306–310.

Peaston, A. E., Evsikov, A. V., Graber, J. H., De Vries, W. N., Holbrook, A. E., Solter, D., et al. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606.

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., et al. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290–1302.

Peters, A. H., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A., Perez-Burgos, L., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* 12, 1577–1589.

Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107, 323–337.

Pheasant, M., and Mattick, J. S. (2007). Raising the estimate of functional human sequences. *Genome Res.* 17, 1245–1253.

Rauch, T. A., Zhong, X., Wu, X., Wang, M., Kernstine, K. H., Wang, Z., et al. (2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci. U.S.A.* 105, 252–257.

Ren, X., Vincenz, C., and Kerppola, T. K. (2008). Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell. Biol.* 28, 2884–2895.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.

Rudd, M. K., Wray, G. A., and Willard, H. F. (2006). The evolutionary dynamics of alpha-satellite. *Genome Res.* 16, 88–96.

Sarma, K., Levasseur, P., Aristarkhov, A., and Lee, J. T. (2010). Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22196–22201.

Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., et al. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* 25, 3110–3122.

Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K., and Willard, H. F. (2001). Genomic and genetic definition of a functional human centromere. *Science* 294, 109–115.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735–745.

Schuettengruber, B., Martinez, A. M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* 12, 799–814.

Shen, S., Lin, L., Cai, J. J., Jiang, P., Kenkel, E. J., Stroik, M. R., et al. (2011). Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2837–2842.

Shen, X., Kim, W., Fujiwara, Y., Simon, M. D., Liu, Y., Mysliwiec, M. R., et al. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 139, 1303–1314.

Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J., and Lawrence, J. B. (2003). Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* 162, 981–990.

Simeonova, I., Lejour, V., Bardot, B., Bouarich-Bourimi, R., Morin, A., Fang, M., et al. (2012). Fuzzy tandem repeats containing p53 response elements may define species-specific p53 target genes. *PLoS Genet.* 8:e1002731. doi: 10.1371/journal.pgen.1002731

Simons, C., Pheasant, M., Makunin, I. V., and Mattick, J. S. (2006). Transposon-free regions in mammalian genomes. *Genome Res.* 16, 164–172.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., et al. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* 138, 885–897.

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285.

Smit, A. F. A., Hubley, R., and Green, P. (1996–2004). *RepeatMasker Open-3.0*. Available online at: http://www.repeatmasker.org. (Accessed date: September 14, 2012).

Snider, L., Asawachaicharn, A., Tyler, A. E., Geng, L. N., Petek, L. M., Maves, L., et al. (2009). RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.* 18, 2414–2430.

Snider, L., Geng, L. N., Lemmers, R. J., Kyba, M., Ware, C. B., Nelson, A. M., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* 6:e1001181. doi: 10.1371/journal.pgen.1001181

Solyom, S., and Kazazian, H. H. Jr. (2012). Mobile elements in the human genome: implications for disease. *Genome Med.* 4, 12.

Soudais, C., Bielinska, M., Heikinheimo, M., Macarthur, C. A., Narita, N., Saffitz, J. E., et al. (1995). Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation *in vitro*. *Development* 121, 3877–3888.

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 21, 1973–1985.

Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R., and Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* 435, 637–645.

Su, J., Shao, X., Liu, H., Liu, S., Wu, Q., and Zhang, Y. (2012). Genome-wide dynamic changes of DNA methylation of repetitive elements in human embryonic stem cells and fetal fibroblasts. *Genomics* 99, 10–17.

Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G. J., and Schultz, R. M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* 269, 276–285.

Taft, R. J., Pheasant, M., and Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29, 288–299.

Tam, R., Smith, K. P., and Lawrence, J. B. (2004). The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J. Cell Biol.* 167, 269–279.

Tanay, A., O'Donnell, A. H., Damelin, M., and Bestor, T. H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5521–5526.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* 148, 664–678.

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.

Tian, D., Sun, S., and Lee, J. T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 143, 390–403.

Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Akhavanfard, S., Coffman, E. J., et al. (2011). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331, 593–596.

Tiwari, V. K., McGarvey, K. M., Licchesi, J. D., Ohm, J. E., Herman, J. G., Schubeler, D., et al. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* 6:e306. doi: 10.1371/journal.pbio.0060306

Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., et al. (2011). Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* 7:e1001343. doi: 10.1371/journal.pgen.1001343

Tremblay, D. C., Alexander, G. Jr., Moseley, S., and Chadwick, B. P. (2010). Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genomics* 11, 632.

Tremblay, D. C., Moseley, S., and Chadwick, B. P. (2011). Variation in array size, monomer composition and expression of the macrosatellite DXZ4. *PLoS ONE* 6: e18969. doi: 10.1371/journal.pone.0018969

Tyekucheva, S., Yolken, R. H., McCombie, W. R., Parla, J., Kramer, M., Wheelan, S. J., et al. (2011). Establishing the baseline level of repetitive element expression in the human cortex. *BMC Genomics* 12, 495.

Van Deutekom, J. C., Wijmenga, C., Van Tienhoven, E. A., Gruter, A. M., Hewitt, J. E., Padberg, G. W., et al. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037–2042.

Van Overveld, P. G., Lemmers, R. J., Sandkuijl, L. A., Enthoven, L., Winokur, S. T., Bakels, F., et al. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* 35, 315–317.

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

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.

Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., et al. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18613–18618.

Warburton, P. E., Haaf, T., Gosden, J., Lawson, D., and Willard, H. F. (1996). Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. *Genomics* 33, 220–228.

Warburton, P. E., Hasson, D., Guillem, F., Lescale, C., Jin, X., and Abrusan, G. (2008). Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics* 9, 533.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543.

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.

Wijmenga, C., Frants, R. R., Brouwer, O. F., Moerer, P., Weber, J. L., and Padberg, G. W. (1990). Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* 336, 651–653.

Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., Dauwerse, H. G., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* 2, 26–30.

Wijmenga, C., Padberg, G. W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O. F., et al. (1991). Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics* 9, 570–575.

Winokur, S. T., Bengtsson, U., Vargas, J. C., Wasmuth, J. J., Altherr, M. R., Weiffenbach, B., et al. (1996). The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* 5, 1567–1575.

Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., and Kingston, R. E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140, 99–110.

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist, R. *Nat. Genet.* 30, 167–174.

Yen, Z. C., Meyer, I. M., Karalic, S., and Brown, C. J. (2007). A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* 90, 453–463.

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., et al. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7266–7271.

Zeng, W., De Greef, J. C., Chen, Y. Y., Chien, R., Kong, X., Gregson, H. C., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet.* 5:e1000559. doi: 10.1371/journal.pgen.1000559

Zhang, X. H., and Chasin, L. A. (2006). Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13427–13432.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.

Zhu, Q., Pao, G. M., Huynh, A. M., Suh, H., Tonnu, N., Nederlof, P. M., et al. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* 477, 179–184.

Conflict of Interest Statement: 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.

Received: 06 July 2012; accepted: 17 September 2012; published online: 08 October 2012.

*Citation: Casa V and Gabellini D (2012) A repetitive elements perspective in Polycomb epigenetics. Front. Gene. **3**:199. doi: 10.3389/fgene.2012.00199*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

MINI REVIEW ARTICLE
published: 11 September 2012
doi: 10.3389/fgene.2012.00181

MeCP2 as a genome-wide modulator: the renewal of an old story

*Floriana Della Ragione[1,2], Stefania Filosa[1,2], Francesco Scalabri[2] and Maurizio D'Esposito[1,2]**

[1] Institute of Genetics and Biophysics "A. Buzzati-Traverso," Naples, Italy
[2] Istituto Di Ricovero e Cura a Carattere Scientifico Neuromed, Pozzilli, Italy

Edited by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France

Reviewed by:
Michèle Amouyal, Centre National de la Recherche Scientifique, France
John Strouboulis, Biomedical Sciences Research Center Alexander Fleming, Greece

****Correspondence:***
Maurizio D'Esposito, Institute of Genetics and Biophysics "A. Buzzati-Traverso," Via Castellino 111, Naples 80131, Italy.
e-mail: maurizio.desposito@igb.cnr.it

Since the discovery of MeCP2, its functions have attracted the interest of generations of molecular biologists. Its function as a transducer of DNA methylation, the major post-biosynthetic modification found throughout genomes, and its association with the neurodevelopmental disease Rett syndrome highlight its central role as a transcriptional regulator, and, at the same time, poses puzzling questions concerning its roles in physiology and pathology. The classical model of the MeCP2 function predicts its role in gene-specific repression through the binding of methylated DNA, via its interaction with the histone deacetylases and co-repressor complexes. This view has been questioned and, intriguingly, new roles for MeCP2 as a splicing modulator and as a transcriptional activator have been proposed. Recent data have demonstrated that MeCP2 is extremely abundant in the neurons, where it reaches the level of histone H1; it is widely distributed, tracking the methylated CpGs, and regulates repetitive elements expression. The role of MeCP2 in maintaining the global chromatin structure is further sustained by its involvement in other biologically relevant phenomena, such as the Line-1 repetitive sequences retrotransposition and the pericentromeric heterochromatin clustering during cellular differentiation. These new concepts renew the old view suggesting a role for DNA methylation in transcriptional noise reduction, pointing to a key role for MeCP2 in the modulation of the genome architecture.

Keywords: MECP2, Rett syndrome, chromatin, DNA methylation, epigenetics

MeCP2 AND DNA METHYLATION: IN LIMINE

In 2012, the twentieth anniversary of MeCP2 protein identification will be celebrated (Lewis et al., 1992). The impulse that this discovery gave to research in various, often apparently unrelated biological fields, from gene regulation to medical genetics, has been immense. Here we cannot describe the enormous weight of data produced, in 20 years, by an increasing number of teams. Rather, we wish to review current research on the MeCP2 biology starting from older scientific hypotheses.

MeCP2 was the second methyl-CpG-binding protein to be identified, although it was the first to be cloned. In fact, Boyes and Bird (1991) demonstrated that the methyl-CpG-binding protein MeCP1 can mediate the repression of transcription from densely methylated genes. MeCP1 is able to bind various methylated sequences "*in vitro*," if at least 12 symmetrically methylated CpGs are available. Like many important findings, MeCP2 was discovered "by accident" by Boyes and Bird (1991), who were attempting to identify the factors that bind unmethylated DNA to protect CpG islands from DNA methylation (Clouaire and Stancheva, 2008). Rat MeCP2 had been successfully isolated through its ability to bind methylated substrates. Then, after its purification, its cDNA had been cloned, thus enabling the knowledge of the nucleotide sequence of the first methyl CpG DNA gene (Lewis et al., 1992).

MeCP2 is able to bind at a genome-wide level, with the need of a single, methylated CpG. This weak discrimination is in agreement with its diffuse nuclear signal in rat cells. In mouse cells, given their peculiar heterochromatin organization, the staining is extremely evident in the pericentromeric heterochromatin, closely resembling the distribution of major satellite DNA (Lewis et al., 1992). Mouse satellite DNA is enriched of methylated CpGs, thus explaining the co-localization of MeCP2 with these genomic regions. MeCP2 was the first methyl-binding protein to be biochemically dissected, revealing the presence of a number of functional domains. The most noticeable domains are the methyl-binding domain (Nan et al., 1993), responsible for binding with the methylated cytosines and the transcriptional repression domain (Nan et al., 1997), which mediates the link with the histone modifications (Jones et al., 1998; Nan et al., 1998; Fuks et al., 2003) and the co-repressors. They play a fundamental role in modulating the functions of MeCP2, the main one being, without doubt, the transduction of DNA methylation. These functions fit with an earlier study reporting that the loss of the X-linked methyl-CpG-binding protein 2 (MeCP2) caused embryonic lethality in chimeric mice (Tate et al., 1996). Taken together, these data highlighted the role(s) of MeCP2 as a genome modulator, whose functions are indispensable for life.

DNA methylation is present, in various degrees, from bacteria to invertebrates and vertebrates. It plays a role in defending bacterial genomes from foreign DNA invasion (Hendrich and Tweedie, 2003). Vertebrate genomes are globally methylated, whereas in invertebrate genomes DNA methylation is patchy. DNA

methylation is involved in chromatin remodeling in vertebrates, whereas it is often located inside the genes in invertebrates, such as in *D. melanogaster* (Mandrioli, 2007). Its genome-wide pattern, in vertebrates, prompted Bird to hypothesize an association between a global repressive effect of DNA methylation and the increase in gene number, which is evident when switching from invertebrate to vertebrate genomes (Bird, 1995). In fact, a major change in the distribution of DNA methylation occurred at the invertebrate–vertebrate boundary (Tweedie et al., 1997; Hendrich and Tweedie, 2003). Following Bird's hypothesis, the global repressive effects of DNA methylation may act as an additional mechanism to suppress transcriptional noise together with the acquisition of a nuclear envelope and the arrangement of the chromatin, which mark the prokaryotes/eukaryotes boundary. This is clearly postulated: "*global improvements in the ability to suppress noise will permit an increase in the maximum gene number, allowing more genes to be tolerated*" (Bird, 1995).

Hendrich and Tweedie (2003) added further substance to this hypothesis suggesting that "*to increase the fidelity of DNA methylation-mediated silencing, and to protect against extensive mutation, there was also a coordinate increase in the number and diversity of methyl CpG binding proteins encoded in the proto-vertebrate genome*". Hendrich and Bird identified a family of methyl-binding protein genes, characterized, similarly to MECP2, by the presence of the methyl-DNA binding domain (MBD). These proteins, called MBD1, MBD2, MBD3, and MBD4 (Hendrich et al., 1999) were all (except for MBD3) characterized by their ability to bind methylated DNA. Only MBD2 and MBD3 were conserved in invertebrates: the ancestral MBD2/3 gene was encoded by a single gene in invertebrate genomes, in contrast to the two separate genes encoded by vertebrates (Hendrich and Tweedie, 2003).

Thus, if a global DNA methylation has been used, by vertebrate genomes, to reduce unscheduled transcription, thereby increasing the gene number, this would similarly provide an evolutionary pressure to increase the number and diversity of the protein(s) capable of repressing transcription through the binding of methylated DNA.

MECP2 AND RETT SYNDROME

Rett syndrome (RTT) is a sporadic post-natal progressive neurodevelopmental disorder occurring with a frequency of 1/10000–15000 live females births and is considered the second most common cause of mental retardation in females (Rett, 1966; Hagberg et al., 1983). The large majority of cases (99%) are sporadic. In 1999, Zoghbi and colleagues (Amir et al., 1999) were able to associate loss-of-function heterozygous mutations in the MECP2 gene to classical RTT patients. The discovery of the MECP2 mutations underlying RTT was a surprise because the large amount of data, summarized above, makes the association of MECP2 to a monogenic disease astonishing.

Besides the large number of studies on patients, the modeling of RTT in mice has been instrumental in order to elucidate the molecular basis of the disease. Mouse models have also been pivotal in the study of expression profiling alterations, necessary to identify putative MeCP2 target genes. They have helped in the elucidation of many questions of biomedical importance: is RTT a pure neuronal disease? Is MECP2 dosage important for the establishment of a pathogenic status? Is RTT reversible?

Two Mecp2 null mice obtained with Cre-LoxP technology and carrying an ubiquitous deletion, were viable but affected by severe neurological symptoms characteristic of RTT (Chen et al., 2001; Guy et al., 2001). The comparative analysis of knock out and brain selective deletions of Mecp2 suggested that the function of this gene is relevant for the central nervous system (Chen et al., 2001; Guy et al., 2001). Moreover, the deletion of MeCP2 in selected brain regions or neuronal sub-types revealed the presence of specific subsets of null phenotypes, allowing to ascribe to MeCP2 different neuronal-specific functions (Fyffe et al., 2008; Samaco et al., 2009; Chao et al., 2010).

MeCP2 dosage matters: a mouse over-expressing a transgene containing the human MECP2 locus that shows a near twofold MeCP2 expression, showed severe progressive neurological phenotypes (Collins et al., 2004). The effect of MECP2 over-expression has also been observed in humans, where a double dosage of MECP2 causes a severe developmental delay and mental retardation (Lubs et al., 1999). Such evidence suggests that MeCP2 levels must be fine regulated *in vivo* and even a mild over-expression of this gene can have a dramatic effect.

The concept of RTT as a pure neuronal disease has recently been challenged with results implicating the involvement of the glial cells in the pathogenesis of RTT (Ballas et al., 2009; Maezawa et al., 2009; Zoghbi, 2009). More recently, it has been suggested that the microglia may influence the onset and progression of RTT by releasing elevated doses of glutamate, exerting a toxic effect on neurons in a non-cell autonomous fashion (Maezawa and Jin, 2010). Very interestingly, null phenotypes in mouse models can be reversed by the re-insertion of the Mecp2 gene (Collins et al., 2004; Luikenhuis et al., 2004; Jugloff et al., 2008), while its over-expression by twofold is deleterious (Collins et al., 2004; Luikenhuis et al., 2004). An almost complete reversibility of the null phenotypes was obtained after the onset of the symptoms, by removing a stop cassette in the Mecp2 gene by a Cre-mediated excision induced by tamoxifen administration (Guy et al., 2007). These data suggest that the neurological defects caused by Mecp2 mutations can potentially be reversed.

MeCP2 AND TRANSCRIPTIONAL CONTROL

The apparent dichotomy of MeCP2 functions (genome-wide vs gene-specific regulator) has been widely debated. Transcriptional profiling studies comparing the total brains of RTT patients or mouse models with controls have revealed only subtle differences in gene expression dampening a role for MeCP2 as a global regulator of transcription (Chadwick and Wade, 2007). A number of reports highlighted BDNF as a bona fide target of MeCP2 in rodent systems (Chen et al., 2003; Martinowich et al., 2003). BDNF is a key signaling molecule involved in brain development and plasticity (Greenberg et al., 2009; Cohen-Cory et al., 2010). The mechanism of its transcriptional regulation is, therefore, quite controversial (Dani et al., 2005; Chang et al., 2006).

To simplify the expression analysis of a complex tissue such as the brain, Zoghbi and colleagues (Chahrour et al., 2008;

Ben-Shachar et al., 2009) performed microarray expression analyses, respectively, in the hypothalamus and cerebellum of Mecp2 null mice and of over-expressing mice (MECP2-Tg; Collins et al., 2004), comparing the results with wild type (WT) mice. Surprisingly, both reports revealed that MeCp2 is responsible for a subtle repression but also for an activation of many genes, and that some of them were similarly, deregulated in both hypothalamus and cerebellum of the Mecp2 null and MECP2-Tg mice (**Figures 1A,B**). Furthermore, it has been confirmed that MeCP2 directly binds the promoter region of the genes down-regulated in the Mecp2 null mice and up-regulated in the MECP2-Tg mice, while sequential ChIP assays have revealed that the promoter of the activated genes is simultaneously associated with both MeCP2 and the known transcriptional activator CREB1 (**Figure 1B**). These data suggest that MeCP2 regulates the expression of a wide range of genes in different brain sub-regions and point to a role for MeCP2 as a modulator of transcription that can both activate or repress target genes (Chahrour et al., 2008; Ben-Shachar et al., 2009). Moreover, the transcriptional alterations observed in the MECP2-Tg mice have confirmed the deleterious effect of the Mecp2 over-expression reported by different research groups (Collins et al., 2004; Luikenhuis et al., 2004).

A category of genes investigated as a putative target of MeCP2 is that of imprinted genes, whose expression is regulated by differential methylation. For example, several studies have focused on Ube3A, a gene imprinted in the brain (Rougeulle et al., 1998) and associated with Prader–Willi and Angelman syndromes. However, to date, the expression alteration of this gene in Mecp2-null mice has not been clearly understood (Guy et al., 2011). Another imprinted region bound by MeCP2 in the mouse brain includes the Dlx5 and Dlx6 genes, located in an imprinted gene cluster on chromosome 6. Its transcription is nearly two times greater in brains of Mecp2-null mice compared to those of WT mice and, in the same model, the chromatin loop in the Dlx5/6 locus enriched with methylated H3K9 present in the WT brain is absent (**Figure 1C**; Horike et al., 2005).

MeCP2 deficiency affects also Line-1 (L1) transcription and retrotransposition: these are, in fact, increased in the mouse brains from null mice and in the neural precursor cells obtained from iPSC and postmortem brains from RTT patients (Muotri et al., 2010).

The L1 elements are retrotransposons representing 20% of mammalian genomes that may induce genomic alterations, such as insertions and deletions (Kazazian, 1998; Perepelitsa-Belancio and Deininger, 2003; Han and Boeke, 2004). Moreover, a massive somatic L1 insertion can occur in adult brain tissues, a phenomenon that can alter the expression of the neuronal genes (Muotri et al., 2005; Coufal et al., 2009). These data were confirmed by another report which revealed an increased transcript level of the L1-elements, intracisternal A particles, and tandem repetitive units of the mouse major satellite in the Mecp2 null brains compared to WT mice (Skene et al., 2010).

MeCP2 AND ITS COFACTORS

As already described regarding the interaction between MeCP2 and CREB, proteins with which it interacts may modify the roles of MeCP2 (**Figures 1A,B**).

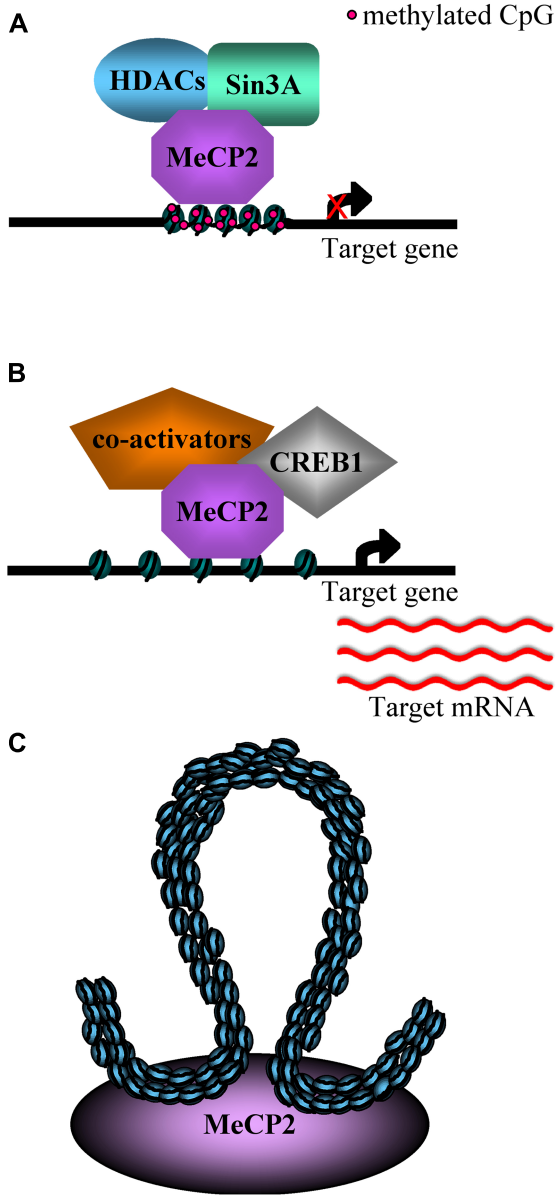


FIGURE 1 | (A) An example of MeCP2 mediated transcriptional repression: the methyl-DNA binding domain (MBD) binds methylated CpG sites and recruits histone deacetylases and co-repressors (Sin3A), inducing chromatin compaction and gene silencing (Nan et al., 1998; Chen et al., 2003; Klose and Bird, 2003). **(B)** MeCP2 is able to activate the transcription of some genes in hypothalamus, by binding unmethylated promoters and recruiting CREB1 and, potentially, co-activators (Chahrour et al., 2008). **(C)** MeCP2 is responsible for the silencing of Dlx5/6 imprinted locus by inducing the formation of a silent higher order chromatin loop (Horike et al., 2005).

The first potential connection between MeCP2 and chromatin came from the finding that MeCP2 copurifies with the Sin3-histone deacetylase complex (Jones et al., 1998; Nan et al., 1998). Based on this observation, most current models depict MeCP2 as a transcriptional repressor that facilitates repression through local histone deacetylation mediated by the passive recruitment of histone deacetylases (Bird and Wolffe, 1999). Klose and Bird

(2004) demonstrated that MeCP2 is a non-obligatory component of the Sin3a co-repressor complex. Moreover, MeCP2 exists as a monomeric protein in solution and does not stably associate with other proteins.

In addition to Sin3a, several other factors have been reported to bind mammalian MeCP2, including DNMT1, CoREST, Suv39H1, and c-SKI (Nan et al., 1998; Kokura et al., 2001; Lunyak et al., 2002; Kimura and Shiota, 2003) although the contribution of these factors to MeCP2-mediated repression is not known.

MeCP2 also interacts with ATRX, a SWI/SNF family ATPase. MeCP2 recruits ATRX to the heterochromatic foci, but this localization is disrupted in Mecp2 null neurons. ATRX localization is disrupted also by the A140V MECP2 mutation found in XLMR patients (Orrico et al., 2000; Nan et al., 2007). Unexpectedly, the complex MeCP2/ATRX with cohesin preferentially binds the unmethylated allele of the H19 gene. This may depend on the association of MeCP2 with this large complex or on regions of non-specific affinity present in MeCP2 (Guy et al., 2011).

A binding of MeCP2 to the trithorax-related protein Brahma (Brm) has also been reported. Brm and MeCP2 assemble on the methylated genes involved in cancer and on the FMR1 gene in fragile X syndrome (Harikrishnan et al., 2005). Therefore, this interaction is still controversial (Hu et al., 2006).

A physical interaction between the heterochromatin protein 1 (HP1) and MeCP2 has been demonstrated during the myogenic differentiation. In particular, this interaction leads to the re-localization of HP1γ to the heterochromatin, which correlates with the presence of MeCP2 (Agarwal et al., 2007). There is no doubt that works aimed at the dissection of the interactions of MeCP2 with other partners, in particular using the novel sequencing-based techniques (Skene et al., 2010), may open the way to a better understanding of the roles and functions of MeCP2.

MeCP2: GLOBAL REGULATORY ROLES

DNA methylation affects the nuclear architecture, as measured by the gene position alterations in the chromosome territories (Matarazzo et al., 2007). A direct role of MeCP2 in nuclear architecture rearrangements has not been reported. Rather, the role(s) of MeCP2 in genome-wide phenomena, such as pericentromeric heterochromatin clustering, has recently been analyzed (Brero et al., 2005; Agarwal et al., 2011; Singleton et al., 2011). During the myogenic differentiation of mouse C2C12 cells, the pericentric heterochromatin domains undergo a reorganization and cluster into a smaller number of larger chromocenters (**Figure 2**). These events are accompanied by an increase in the methylation of major satellite DNA and the accumulation of MeCP2 and MBD2 proteins in the nuclei of terminally differentiated muscle cells. Interestingly, the over-expression of MeCP2 and MBD2 in C2C12 myoblasts in the absence of differentiation also induces an aggregation of the chromocenters, indicating that these proteins may be directly involved in the reorganization of heterochromatin architecture. Moreover, studies in Mecp2 null mouse neurons have revealed significant differences in the number and size of the nucleoli and chromocenters compared to WT animals (Singleton et al., 2011). Already in 2002, it was shown that mice carrying a Mecp2 truncating mutation have a higher level of hyperacetylated histone H3

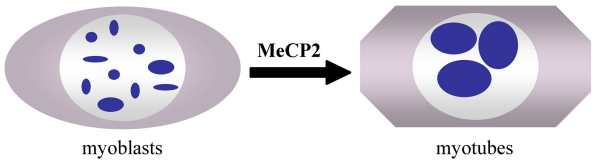


FIGURE 2 | Clustering of pericentromeric heterochromatin domains (chromocenters, blue spots) during myogenic differentiation of C2C12 myoblasts to myotubes (Brero et al., 2005).

compared with WT mice, emphasizing a generally altered chromatin architecture (Shahbazian et al., 2002). The development of techniques permitting genome-wide epigenomic studies are contributing to the assessment of MeCP2 functions in the chromatin architecture and genome organization.

In 2007, LaSalle and colleagues reported, by ChIP-chip analysis on a neuroblastoma cell line, that more than half of the MeCP2 binding sites are intergenic and that only a small number of them reside in the CpG islands. Moreover, among binding sites located in the CpG islands, many of them are associated with actively transcribed genes, supporting the view of a more complex function of MeCP2 (Yasui et al., 2007).

Different approaches, reagents, and technologies led, some years later, to the re-establishment of MECP2 as a protein with a global regulatory role (Skene et al., 2010). The utilization of next generation sequencing approaches in the neuronal nuclei from the mature mouse brain has revealed that the abundance of MeCP2 is similar to the number of nucleosomes (Skene et al., 2010). Moreover, as previously reported (Shahbazian et al., 2002), in the absence of Mecp2, the H3 acetylation levels are increased, while the H1 levels are doubled, pointing a role for MeCP2 in the global chromatin organization.

Furthermore, an analysis of binding sites around known regulated genes, such as BDNF and Dlx5/6, transcriptionally active in this cellular system, has revealed a MeCP2 binding across the entire locus, except for the CpG island regions, suggesting that these active promoters are unable to bind MeCP2 due to its hypomethylation state. Moreover, high-throughput data suggest that the MeCP2 binding *in vivo* tracks the density of methyl-CpG in the genome (Skene et al., 2010). These latter data have revealed that MeCP2 is one of the most abundant nuclear proteins in the mature neurons suggesting a crucial role for MeCP2 in neurons as a regulator of the entire genome.

The described data suggest that, in addition to the role of MeCP2 as a gene-specific transcriptional regulator, mediated by the association with specific cofactors, the global chromatin-binding function of MeCP2 is crucial for global chromatin dynamics especially during brain maturation. MeCP2 may thus be seen as a multifunctional and structural organizing factor. Furthermore, the interaction of MeCP2 with most regions of the genome, such as the intergenic DNA and repetitive elements, should contribute to keep the rate of somatic mutation and transcriptional noise in the brain low and allows to hypothesize further pathogenic roles for MeCP2 in RTT. This evidence recalls the concept we previously described, focusing on the role of MeCP2 as a key player in genome architecture and regulation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Maria Matarazzo and Dr. Maria Strazzullo for their critical reading of the manuscript, and Ms. Anna Aliperti for her expert secretarial assistance. Maurizio D'Esposito was supported by the UE Initial Training Network Project n°238242 "DISCHROM" and by the EPIGENOMICS FLAGSHIP PROJECT EPIGEN, MIUR-CNR. Francesco Scalabrì was supported by a Neuromed fellowship.

REFERENCES

Agarwal, N., Becker, A., Jost, K. L., Haase, S., Thakur, B. K., Brero, A., Hardt, T., Kudo, S., Leonhardt, H., and Cardoso, M. C. (2011). MeCP2 Rett mutations affect large scale chromatin organization. *Hum. Mol. Genet.* 20, 4187–4195.

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M. C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res.* 35, 5402–5408.

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188.

Ballas, N., Lioy, D. T., Grunseich, C., and Mandel, G. (2009). Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. *Nat. Neurosci.* 12, 311–317.

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A., and Zoghbi, H. Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum. Mol. Genet.* 18, 2431–2442.

Bird, A. P. (1995). Gene number, noise reduction and biological complexity. *Trends Genet.* 11, 94–100.

Bird, A. P., and Wolffe, A. P. (1999). Methylation-induced repression – belts, braces, and chromatin. *Cell* 99, 451–454.

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64, 1123–1134.

Brero, A., Easwaran, H. P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., and Cardoso, M. C. (2005). Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.* 169, 733–743.

Chadwick, L. H., and Wade, P. A. (2007). MeCP2 in Rett syndrome: transcriptional repressor or chromatin architectural protein? *Curr. Opin. Genet. Dev.* 17, 121–125.

Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J., and Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224–1229.

Chang, Q., Khare, G., Dani, V., Nelson, S., and Jaenisch, R. (2006). The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. *Neuron* 49, 341–348.

Chao, H. T., Chen, H., Samaco, R. C., Xue, M., Chahrour, M., Yoo, J., Neul, J. L., Gong, S., Lu, H. C., Heintz, N., Ekker, M., Rubenstein, J. L., Noebels, J. L., Rosenmund, C., and Zoghbi, H. Y. (2010). Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* 468, 263–269.

Chen, R. Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* 27, 327–331.

Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R., and Greenberg, M. E. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* 302, 885–889.

Clouaire, T., and Stancheva, I. (2008). Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell. Mol. Life Sci.* 65, 1509–1522.

Cohen-Cory, S., Kidane, A. H., Shirkey, N. J., and Marshak, S. (2010). Brain-derived neurotrophic factor and the development of structural neuronal connectivity. *Dev. Neurobiol.* 70, 271–288.

Collins, A. L., Levenson, J. M., Vilaythong, A. P., Richman, R., Armstrong, D. L., Noebels, J. L., David Sweatt, J., and Zoghbi, H. Y. (2004). Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* 13, 2679–2689.

Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V., and Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* 460, 1127–1131.

Dani, V. S., Chang, Q., Maffei, A., Turrigiano, G. G., Jaenisch, R., and Nelson, S. B. (2005). Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12560–12565.

Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., and Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* 278, 4035–4040.

Fyffe, S. L., Neul, J. L., Samaco, R. C., Chao, H. T., Ben-Shachar, S., Moretti, P., McGill, B. E., Goulding, E. H., Sullivan, E., Tecott, L. H., and Zoghbi, H. Y. (2008). Deletion of Mecp2 in Sim1-expressing neurons reveals a critical role for MeCP2 in feeding behavior, aggression, and the response to stress. *Neuron* 59, 947–958.

Greenberg, M. E., Xu, B., Lu, B., and Hempstead, B. L. (2009). New insights in the biology of BDNF synthesis and release: implications in CNS function. *J. Neurosci.* 29, 12764–12767.

Guy, J., Cheval, H., Selfridge, J., and Bird, A. (2011). The role of MeCP2 in the brain. *Annu. Rev. Cell Dev. Biol.* 27, 631–652.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315, 1143–1147.

Guy, J., Hendrich, B., Holmes, M., Martin, J. E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* 27, 322–326.

Hagberg, B., Aicardi, J., Dias, K., and Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann. Neurol.* 14, 471–479.

Han, J. S., and Boeke, J. D. (2004). A highly active synthetic mammalian retrotransposon. *Nature* 429, 314–318.

Harikrishnan, K. N., Chow, M. Z., Baker, E. K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J. M., Jones, P. L., Sif, S., and El-Osta, A. (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat. Genet.* 37, 254–264.

Hendrich, B., Abbott, C., McQueen, H., Chambers, D., Cross, S., and Bird, A. (1999). Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. *Mamm. Genome* 10, 906–912.

Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* 19, 269–277.

Horike, S., Cai, S., Miyano, M., Cheng, J. F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* 37, 31–40.

Hu, K., Nan, X., Bird, A., and Wang, W. (2006). Testing for association between MeCP2 and the brahma-associated SWI/SNF chromatin-remodeling complex. *Nat. Genet.* 38, 962–964; author reply 964–967.

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* 19, 187–191.

Jugloff, D. G., Vandamme, K., Logan, R., Visanji, N. P., Brotchie, J. M., and Eubanks, J. H. (2008). Targeted delivery of an Mecp2 transgene to forebrain neurons improves the behavior of female Mecp2-deficient mice. *Hum. Mol. Genet.* 17, 1386–1396.

Kazazian, H. H. Jr. (1998). Mobile elements and disease. *Curr. Opin. Genet. Dev.* 8, 343–350.

Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.* 278, 4806–4812.

Klose, R., and Bird, A. (2003). Molecular biology. MeCP2 repression goes nonglobal. *Science* 302, 793–795.

Klose, R. J., and Bird, A. P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J. Biol. Chem.* 279, 46490–46496.

Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. *J. Biol. Chem.* 276, 34115–34121.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69, 905–914.

Lubs, H., Abidi, F., Bier, J. A., Abuelo, D., Ouzts, L., Voeller, K., Fennell, E., Stevenson, R. E., Schwartz, C. E., and Arena, F. (1999). XLMR syndrome characterized by multiple

respiratory infections, hypertelorism, severe CNS deterioration and early death localizes to distal Xq28. *Am. J. Med. Genet.* 85, 243–248.

Luikenhuis, S., Giacometti, E., Beard, C. F., and Jaenisch, R. (2004). Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6033–6038.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* 298, 1747–1752.

Maezawa, I., and Jin, L. W. (2010). Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* 30, 5346–5356.

Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J. M., and Jin, L. W. (2009). Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J. Neurosci.* 29, 5051–5061.

Mandrioli, M. (2007). A new synthesis in epigenetics: towards a unified function of DNA methylation from invertebrates to vertebrates. *Cell. Mol. Life Sci.* 64, 2522–2524.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y. E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* 302, 890–893.

Matarazzo, M. R., Boyle, S., D'Esposito, M., and Bickmore, W. A. (2007). Chromosome territory reorganization in a human disease with altered DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16546–16551.

Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910.

Muotri, A. R., Marchetto, M. C., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F. H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446.

Nan, X., Campoy, F. J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 88, 471–481.

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M. J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2709–2714.

Nan, X., Meehan, R. R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* 21, 4886–4892.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386–389.

Orrico, A., Lam, C., Galli, L., Dotti, M. T., Hayek, G., Tong, S. F., Poon, P. M., Zappella, M., Federico, A., and Sorrentino, V. (2000). MECP2 mutation in male patients with non-specific X-linked mental retardation. *FEBS Lett.* 481, 285–288.

Perepelitsa-Belancio, V., and Deininger, P. (2003). RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* 35, 363–366.

Rett, A. (1966). On a unusual brain atrophy syndrome in hyperammonemia in childhood. *Wien. Med. Wochenschr.* 116, 723–726.

Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, L., and Lalande, M. (1998). An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat. Genet.* 19, 15–16.

Samaco, R. C., Mandel-Brehm, C., Chao, H. T., Ward, C. S., Fyffe-Maricich, S. L., Ren, J., Hyland, K., Thaller, C., Maricich, S. M., Humphreys, P., Greer, J. J., Percy, A., Glaze, D. G., Zoghbi, H. Y., and Neul, J. L. (2009). Loss of MeCP2 in aminergic neurons causes cell-autonomous defects in neurotransmitter synthesis and specific behavioral abnormalities. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21966–21971.

Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* 35, 243–254.

Singleton, M. K., Gonzales, M. L., Leung, K. N., Yasui, D. H., Schroeder, D. I., Dunaway, K., and LaSalle, J. M. (2011). MeCP2 is required for global heterochromatic and nucleolar changes during activity-dependent neuronal maturation. *Neurobiol. Dis.* 43, 190–200.

Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Andrews, R., and Bird, A. P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* 37, 457–468.

Tate, P., Skarnes, W., and Bird, A. (1996). The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nat. Genet.* 12, 205–208.

Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997). Methylation of genomes and genes at the invertebrate–vertebrate boundary. *Mol. Cell. Biol.* 17, 1469–1475.

Yasui, D. H., Peddada, S., Bieda, M. C., Vallero, R. O., Hogart, A., Nagarajan, R. P., Thatcher, K. N., Farnham, P. J., and LaSalle, J. M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19416–19421.

Zoghbi, H. Y. (2009). Rett syndrome: what do we know for sure? *Nat. Neurosci.* 12, 239–240.

Conflict of Interest Statement: 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.

Received: 13 July 2012; accepted: 28 August 2012; published online: 11 September 2012.

Citation: Della Ragione F, Filosa S, Scalabrì F and D'Esposito M (2012) MECP2 as a genome-wide modulator: the renewal of an old story. Front. Gene. 3:181. doi: 10.3389/fgene.2012.00181

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.

frontiers in
GENETICS

OPINION ARTICLE
published: 24 April 2012
doi: 10.3389/fgene.2012.00064

TFC6 (TFIIIC subunit): a bridge between prokaryotic and eukaryotic gene regulation

Michèle Amouyal*

Interactions à Distance, Centre National de la Recherche Scientifique, Paris, France
**Correspondence: michele.amouyal@club.fr*

TFIIIC is a multi-subunit protein first and best known as part of the transcriptional machinery required by RNA polymerase III for synthesis of transfer RNA (see Donze, 2012 for a recent review). Kleinschmidt et al. (2011) highlight a new aspect of this machinery, related to the sole TFIIIC, by demonstrating its capacity to directly regulate RNA polymerase II transcription – in fact, its own synthesis. The authors have dissected the role of TFIIIC and one of its subunits precisely, *in vivo*, providing an in-depth molecular picture, which is rather unusual when dealing with eukaryotic complexity.

THE MULTI-SUBUNIT TFIIIC COMPLEX AS A RECRUITING AGENT FOR THE RNA POLYMERASE III MACHINERY

In *Saccharomyces cerevisiae*, the polymerase III cannot initiate transcription before a certain number of components are assembled. At genes of transfer (tDNAs), TFIIIC first binds a sequence with two conserved regions, A and B. It then recruits TFIIIB, which in turn recruits the pol-III holoenzyme.

The B-box is sufficient for DNA-binding. TFIIIC is restricted to initiation and composed of six subunits. Only TFC3 and TFC6 bind the B-box.

THE TFC6 GENE

Kleinschmidt et al. (2011) were interested in the tfc6 gene which encodes the TFC6 subunit, because the promoter contains a B-box (ETC6 site) suggesting that the gene might be auto-regulated.

Mutations in ETC6 and a highly conserved adjacent region, severely affect transcription and cell growth. Impaired cell growth is restored by the sole TFC6 complementation, indicating a simple and direct correlation with the TFC6 defect, with no other origin.

TFIIIC, TFC6 specifically, indeed binds to the tfc6 B-box and conserved region, as shown by immunoprecipitation of the promoter-bound complexes. Its binding reduces the occupancy of the promoter by the TATA binding protein of the RNA polymerase II initiating complex, presumably interfering with a pol-II upstream factor (between −120 and −40 in yeast), as TFIIIC is newly found to bind a region larger upstream (from −150 to −90) than the previously specified ETC6 site.

Whether by mutational analysis of the DNA or with the appropriate protein mutants, TFIIIC binding to the promoter is prevented and the level of RNA transcripts (by Northern blot analysis) is increased in the same proportion, by twofold.

Similar data with respect to colony size, are obtained when the tfc6 gene is replaced by a reporter gene and when TFC6 is produced by an episomal plasmid. This effect is exclusively observed with the overexpressed TFC6, and with no other TFIIIC subunit.

TFIIIC AUTO-INHIBITION: WHAT FOR?

Auto-inhibition generally allows to strictly limit the level of a transcription factor in the cell. Pointed long ago in some model prokaryotic systems, as detailed in next section, these concentrations are determinant.

They can modify the mode of regulation. This finding allowed to unravel the contribution of non-proximal operator sequences, once thought to be cryptic, to repression of the *E. coli* lac operon (see Amouyal, 2006 for a review). In eukaryotes, over-expression of four key proteins is sufficient to re-program mice and human somatic cells into pluripotent stem cells like embryonic stem cells (Takahashi and Yamanaka, 2006). In *Caenorhabditis elegans*, it leads to terminal neuronal differentiation (Hobert, 2011).

Kleinschmidt et al. (2011) have focused on dissection of the etc6 gene and report for the first time that the level of TFC6 factor is restricted by tfc6 auto-regulation. In fact, over-expression affects cell growth. Is the production of other TFIIIC subunits also auto-regulated? TFIIIC is involved in a wide number of processes, from transcription to gene insulation and chromosomal organization (see Donze, 2012). Is its level critical for one of these processes?

This might be the case (Kleinschmidt et al., 2011). For instance, infection of human cells by the Epstein–Barr virus goes with an increase of TFIIIC concentration, which may be related to its carcinogenic potential.

AS SIMPLY AS A PROKARYOTIC FACTOR

Strikingly, it comes out from Kleinschmidt et al. (2011) work that TFC6 is much like a prokaryotic factor.

i The tDNAs and ETC sites are deprived of histones or covered with unstable histone variants, like several other insulators (Donze, 2012).
ii TFCIIIC binds strongly DNA, nearly like a prokaryotic factor, with an apparent dissociation constant of less than 10^{-10} M at tDNAs.
iii The mechanism of tfc6 auto-inhibition resembles that of prokaryotic operons or genes. TFC6 represses its own gene by simply interfering with the binding of the pol-II complex at the promoter.

The production of a wide number of *E. coli* transcriptional regulators is auto-repressed with help of an operator located on the promoter region or at the start of the gene. Thus, just in *E. coli*, over the 32 regulators reviewed by Collado-Vides et al. (1991), 19 are auto-repressed, and this number has probably expanded. The 107 listed promoters are mainly repressed through a direct interference with RNA polymerase II at the promoter, though in some instances, the repressor instead interferes with an upstream activator site or mRNA transcript elongation. The same regulator, depending upon its position with respect to the start

of transcription, is an activator or a repressor of its own synthesis, like TFIIIC. This is the case for 6 over the 32 listed regulators.

Since these proteins are only regulatory, they do not need to be extensively produced. Thus, auto-repression limits the number of AraC molecules to 20. Alternatively, the lac repressor is maintained at the low level of 10 copies per cell thanks to a weak promoter.

Like TFIIIC, the glnG product is an auto-regulated transcription factor that contributes to gene activation. In this case, auto-repression is linked to the modulation of environmental conditions. More precisely, the glnG product (NRI) is the regulator of the system for synthesis of many enzymes required for nitrogen assimilation in enterobacteria. The heart of the nitrogen control region is the glnALG operon, endowed with three promoters, glnAp1, glnAp2, and glnLp (p1p2-glnA-p-glnLG). It comprises the glnG regulatory gene, the glnL modulator gene, and the glnA structural gene for glutamine synthetase, the sensor of ammonia availability.

In cells growing in excess nitrogen, transcription from both glnAp1 and glnLG is repressed by NRI. Under these conditions, NRI limits the synthesis of glutamine synthetase as well as its own synthesis (five molecules in the cell). Ammonia deprivation results in phosphorylation of NRI by the modulator produced by the glnL gene, activation of the glnAp2 promoter by the phosphorylated NRI molecule and subsequent activation of a cascade of genes required under these new conditions. Activation of glnAp2 also requires a σ54 cofactor for RNA polymerase II, used in place of the common σ70 cofactor. Initiation of glnAp2 increases the intracellular level of glutamine synthetase and that of NRI (up to 70 molecules); This increased concentration is required for the activation of the other nitrogen regulated promoters, such as the 17 genes of the *Klebsiella pneumoniae* promoters of the nif regulon, or the hundred responsive genes in *E. coli*.

Chromosomal looping is associated with the modulation of NRI levels. It here allows to turn on the glnAp2 gene with an economy of means, since the NRI sites which were used for repression of the glnAp1 promoter under excess nitrogen, are also used for activation of the glnAp2 promoter, but are now 100 and 130 bp upstream of the promoter. It also contributes to convert a repressor into an activator by simply changing its position with respect to the promoter.

Restricting the number of regulatory proteins is also supposed to prevent non-specific binding and interference with other DNA–protein transactions in the cell, as prokaryotic proteins can easily bind DNA non-specifically.

Interestingly, several *E. coli* auto-repressed repressors such as GalS or deoR, contain an internal operator within part of the gene encoding the DNA-binding region of the protein. Thus, auto-regulation might have evolved from a few common ancestor DNA-binding motifs (Roy et al., 2002).

As for NRI, auto-repression is often associated with the coupling of distant loci by chromosomal looping with help of the corresponding regulator. This is the case for repression of the already mentioned ara and deo operons. The *E. coli* deo operon which encodes nucleoside and deoxynucleoside catabolizing enzymes is expressed from two promoters, deoP1 and deoP2, repressed by the deoR repressor (and CytR with different inducers). It is characterized by strong promoters and high affinity DNA-repressor interactions. DNA looping allows to lock very efficiently the two promoters controlled by the same repressor, 599 bp apart, in one operation.

Another example is provided by the bacteriophage λ (or 186) cI protein. λcI is the repressor of the functions needed for phage replication, gene assembly, and cell lysis, thereby maintaining the lysogenic state (integration into the *E. coli* chromosome). Though at this stage, the virus is dormant, it is important for its survival that it can switch to the lytic state to infect other cells. However, a simple increase by twofold of the cI repressor concentration, prevents the efficient switch from dormance to virulence, requiring self-repression.

As is often the case to secure a specific process, the maintenance of lysogeny in the present case, the cell makes use of several strategies, not just one. Thus, the same molecular process, through chromosomal looping between the P_L and P_{RM} promoters, 2800 bp apart, as well as repressor oligomerization, allows (i) to stringently control the intracellular level of repressor by two means, not a single one: directly, by auto-repression of the cI gene from P_{RM}, indirectly by repression of the gene synthesizing the N protein under the control of the P_L promoter, which activates repressor synthesis, (ii) to stop integrase production also controlled from the P_L promoter, as it is not anymore required once the phage has been integrated, (iii) to inhibit, through N repression at P_L, several functions required for phage lysis, replication, and assembly, also controlled by this promoter.

Chromosomal looping is not consistent with the extremely high concentrations of regulator that favor the non-cooperative occupancy of all available sites (see Amouyal, 2006), as well as non-specific binding. In this context, auto-regulation would also favor chromosomal looping for the coordination and organization of gene expression when it extends over distant loci.

The coupling of distant loci within a transcriptional unit by DNA looping and regulators in *E. coli*, presents some common features with the organization and coordination of gene expression by the so-called gene insulators, such as CTCF (Yang and Corces, 2012). TFIIIC would even be closer than CTCF to these prokaryotic repressors, as it can also be a transcriptional repressor and as it is involved in the clustering of distant genomic loci.

Since the frontier with prokaryotes when dealing with gene expression organization, is less clear than commonly thought, one expects that the eukaryotic systems will benefit from the comparison with their prokaryotic counterparts. Conversely, the prokaryotic transcription factors presenting common features with gene insulators, might also structure the genome and perform a similar function in eukaryotic cells.

REFERENCES

Amouyal, M. (2006). Transition from DNA looping to simple binding or DNA pairing in gene regulation and replication: a matter of numbers in the cell. *Gene Genomes Genomics* 1, 104–111.

Collado-Vides, J., Magasanik, B., and Gralla, J. D. (1991). Control site location and transcriptional regulation in *E. coli*. *Microbiol. Rev.* 55, 371–394.

Donze, D. (2012). Extra-chromosomal functions of the RNA polymerase III complexes: TFIIIC as a potential global chromatin bookmark. *Gene* 493, 169–175.

Hobert, O. (2011). Maintaining a memory by transcriptional auto-regulation. *Curr. Biol.* 21, R146–R147.

Kleinschmidt, R. A., Leblanc, K. E., and Donze, D. (2011). Auto-regulation of an RNA polymerase II promoter by the RNA polymerase III transcription factor IIIC

(TFIIIC) complex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8385–8389.

Roy, S., Sahu, A., and Adhya, A. (2002). Evolution of DNA binding motifs and operators. *Gene* 285, 169–173.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from adult human fibroblast by defined factors. *Cell* 126, 663–676.

Yang, J., and Corces, V. G. (2012). Insulators, long-range interactions, and genome function. *Curr. Opin. Genet. Dev.* 22, 1–7.

Received: 24 February 2012; accepted: 04 April 2012; published online: 24 April 2012.

Citation: Amouyal M (2012) TFC6 (TFIIIC subunit): a bridge between prokaryotic and eukaryotic gene regulation. Front. Gene. ***3****:64. doi: 10.3389/fgene.2012.00064*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.