HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2017 November 17.

Published in final edited form as:

Cell. 2016 November 17; 167(5): 1188–1200. doi:10.1016/j.cell.2016.10.024.

Insulated neighborhoods: structural and functional units of mammalian gene control

Denes Hnisz[1,*,$], **Daniel S. Day**[1,*,$], and **Richard A. Young**[1,2,$]

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

Summary

Understanding how transcriptional enhancers control over 20,000 protein-coding genes to maintain cell type-specific gene expression programs in all human cells is a fundamental challenge in regulatory biology. Recent studies suggest that gene regulatory elements and their target genes generally occur within insulated neighborhoods, which are chromosomal loop structures formed by the interaction of two DNA sites bound by the CTCF protein and occupied by the cohesin complex. We review here evidence that insulated neighborhoods provide for specific enhancer-gene interactions, are essential for both normal gene activation and repression, form a chromosome scaffold that is largely preserved throughout development, and are perturbed by genetic and epigenetic factors in disease. Insulated neighborhoods are a powerful paradigm for gene control that provides new insights into development and disease.

Preamble

Many recent reports describe evidence that specific chromosome structures play important roles in gene control. A core principle that has emerged from these studies is that genes and their regulatory elements typically occur together within specific DNA loop structures, which we have called “insulated neighborhoods”. Here we review evidence that insulated neighborhoods are structural and functional units of gene control, and explain how they are used during development to control the diverse cell identities that contribute to complex animals. We explain how insulated neighborhoods form the mechanistic basis of higher-order chromosome structures such as Topologically Associating Domains (TADs), discuss

[$]Corresponding Authors: Denes Hnisz, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, Tel: (617) 258-7181, Fax: (617) 258-0376, hnisz@wi.mit.edu. Daniel S. Day, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142,. Tel: (617) 258-6978, Fax: (617) 258-0376, dsday@wi.mit.edu. Richard A. Young, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, Tel: (617) 258-5218, Fax: (617) 258-0376, young@wi.mit.edu.

[*]Co-first authors

Note on data availability

Maps of insulated neighborhoods in human embryonic stem cells (ESC) are available in Supplementary Table 3 in (Ji et al., 2016). The dataset described for primed hESCs were used for the quantitative analyses described here. Maps and features of insulated neighborhoods in human and murine ESCs are also found online at: http://younglab.wi.mit.edu/insulatedneighborhoods.htm

how genetic and epigenetic perturbations of neighborhood boundaries contribute to disease, and outline how further study of neighborhood structure and function will lead to additional insights into development and disease. There are other excellent reviews that provide historical perspective and summarize key insights into chromosome structure (Bickmore and van Steensel, 2013; Cavalli and Misteli, 2013; de Laat and Duboule, 2013; Dekker and Heard, 2015; Dekker and Mirny, 2016; Gibcus and Dekker, 2013; Gorkin et al., 2014; Merkenschlager and Nora, 2016; Phillips and Corces, 2009; Phillips-Cremins and Corces, 2013); here we focus on the insulated neighborhood as a model for further exploration of the principles that underpin gene control in mammalian systems.

The enhancer-gene specificity conundrum

Cell type specific gene expression programs in humans are generally controlled by gene regulatory elements called enhancers (Buecker and Wysocka, 2012; Heinz et al., 2015; Levine et al., 2014; Ong and Corces, 2011; Ren and Yue, 2015). Enhancers, first described over 30 years ago (Banerji et al., 1981; Benoist and Chambon, 1981; Gruss et al., 1981), are segments of DNA that are typically a few hundred base pairs in length and are occupied by multiple transcription factors that recruit co-activators and RNA polymerase II to target genes (Bulger and Groudine, 2011; Spitz and Furlong, 2012; Tjian and Maniatis, 1994). Tens of thousands of enhancers are estimated to be active in any given human cell type (ENCODE Project Consortium et al., 2012; Roadmap Epigenomics et al., 2015). Enhancers and their associated factors can regulate expression of genes located far upstream or downstream by looping to the promoters of these genes, so the features that cause enhancers to regulate only specific genes, generally on their own chromosomes, has been something of a mystery for several decades (Figure 1A). This mystery, which we will call the enhancer-gene specificity conundrum, is important to solve because the majority of disease-associated non-coding variation occurs in the vicinity of enhancers and thus likely impacts these enhancers' target genes (Ernst et al., 2011; Farh et al., 2015; Hnisz et al., 2013; Maurano et al., 2012).

Some of the specificity of enhancer-gene interactions may be due to the interaction of DNA-binding transcription factors at enhancers with specific partner transcription factors at promoters (Butler and Kadonaga, 2001; Choi and Engel, 1988; Ohtsuki et al., 1998). Each cell type expresses hundreds of different transcription factors and these bind to DNA sequences in enhancers and in promoter-proximal regions. Diverse factors bound at these two sites interact with large cofactor complexes and could, in principle, interact with one another to produce some degree of enhancer-gene specificity (Zabidi et al., 2015). It is not clear to what extent this mechanism contributes to specific enhancer-gene interactions throughout the human genome. Another potential solution to the enhancer-gene specificity conundrum lies in insulators, which are regulatory elements that can block the ability of an enhancer to activate a gene when located between them (Chung et al., 1993; Geyer and Corces, 1992; Kellum and Schedl, 1991; Udvardy et al., 1985). Insulators are bound by the transcription factor CTCF (Bell et al., 1999), but only a minority of CTCF sites function as insulators (Liu et al., 2015). The features that distinguish the subset of CTCF sites that function as insulators are not understood, so the extent to which insulators provide a solution to the enhancer-gene specificity conundrum has not been clear (Figure 1B).

Chromosome structure constrains enhancer-gene interactions

The idea that chromosome structures can influence phenotypic traits is nearly as old as the chromosome theory of inheritance (Boveri, 1909), but only recently have studies of chromosome structure suggested how enhancers might be constrained to interact with specific genes (Figure 2A). In situ hybridization techniques and microscopy have revealed that individual interphase chromosomes tend to occupy small portions of the nucleus, called “chromosome territories”, rather than spreading throughout this organelle (Cremer and Cremer, 2010); interactions between chromosomes would be minimized in this manner. Furthermore, individual chromosomes are partitioned into megabase- sized topologically associating domains (TADs), regions with relatively high intradomain DNA interaction frequencies as measured by Hi-C chromosome conformation capture data (Dixon et al., 2012; Nora et al., 2012). These TADs, which have similar boundaries in all human cell types examined, have been proposed to constrain enhancer-gene interactions because most DNA contacts occur within the TADs (Dixon et al., 2015; Dixon et al., 2012). This structuring of the genome helps explain why enhancer-gene interactions rarely occur between chromosomes and tend to be constrained within megabase-sized domains. However, they provide only limited insight into the molecular mechanisms that engender specific enhancer-gene interactions within TADs, which contain, on average, about eight genes whose expression is weakly correlated.

Further understanding of the mechanisms that engender specific enhancer-gene interactions have come from genome-wide maps of the proteins that bind enhancers, promoters and insulators, together with knowledge of the physical contacts that occur between these elements (Chepelev et al., 2012; DeMare et al., 2013; Dowen et al., 2014; Fullwood et al., 2009; Handoko et al., 2011; Phillips-Cremins et al., 2013; Tang et al., 2015). In the models that emerge from these data, each chromosome contains thousands of DNA loops, formed by the interaction of two CTCF molecules bound to different sites, and reinforced by a cohesin molecule (Figure 2A). Enhancer-bound proteins are constrained such that they tend to interact only with genes within these CTCF-CTCF loops. As described below, the subset of CTCF sites that form these “loop anchors” thus function to insulate enhancers and genes within the loop from enhancers and genes outside the loop. For these and other reasons, these CTCF-CTCF DNA loops have been called “insulated neighborhoods”.

Insulated neighborhoods

Insulated neighborhoods have been defined as chromatin loops that are formed by a CTCF-CTCF homodimer, co-bound with cohesin, and contain at least one gene (Dowen et al., 2014; Ji et al., 2016). In human embryonic stem cells (ESCs), there are ~13,000 insulated neighborhoods, which range from 25 kb to 940 kb in size, and contain from 1 – 10 genes (Figure 2B) (Dowen et al., 2014; Ji et al., 2016). The median insulated neighborhood is ~190kb and contains 3 genes. These numbers will vary depending on assumptions made for filtering genomic data, as described below, but they provide an initial description of genomic loops useful for further analysis. We describe below evidence that insulated neighborhood loop anchors have insulating properties, that they are largely maintained during

development, and that the subset of CTCF sites that form neighborhood loop anchors are especially conserved in the human germline and in primates.

Evidence for Insulation

Three lines of evidence argue that insulated neighborhood structures have insulating boundaries. The majority of enhancer-gene interactions occur within the insulated neighborhoods (Figure 2C). Perturbation of insulated neighborhood anchor sequences leads to local gene dysregulation (Figure 2D). Somatic mutations in multiple tumor types alter insulated neighborhood anchor sequences in order to activate oncogenes (Figure 2E). These lines of evidence, described in more detail below, indicate that the insulating function of the neighborhood loop anchors is generally necessary for normal gene activation and repression.

The vast majority of enhancer-gene interactions occur within insulated neighborhoods (Dowen et al., 2014; Hnisz et al., 2016; Ji et al., 2016; Phillips-Cremins et al., 2013). For example, in human ESCs, ~97% of all enhancer or promoter-related DNA loops occur within insulated neighborhood boundaries (Figure 2C) (Ji et al., 2016). Similarly, in T cell leukemia cells, ~90% of all enhancer or promoter-related DNA loops occur within insulated neighborhood boundaries (Hnisz et al., 2016). It is also possible to estimate each neighborhood's insulation efficacy using an "insulation score." The insulation score of a neighborhood is calculated as the percentage of enhancer-promoter interactions that are fully contained within the neighborhood. In human ES cells, 59% of insulated neighborhoods have an insulation score of 100%.

Genetic perturbation of neighborhood anchor sequences has provided evidence for their structural and functional roles as insulators (Dowen et al., 2014; Flavahan et al., 2016; Hnisz et al., 2016; Ji et al., 2016; Narendra et al., 2015). In a dozen loci and in multiple cell types, CRISPR/Cas9 deletion of CTCF binding sites at the anchors of insulated neighborhoods have been shown to produce changes in the expression of genes within the neighborhoods and immediately adjacent to the deleted neighborhood boundary. For example, the miR-290–295 miRNA gene cluster, which plays important roles in ES cell pluripotency, occurs within an insulated neighborhood together with a super-enhancer; when a CTCF loop anchor site of this neighborhood was deleted, there was a reduction in expression of the miRNA precursor and activation of an adjacent gene outside of the neighborhood concomitant with looping of the super-enhancer to this outside gene (Figure 2D). Furthermore, when genes occur within multiple nested insulated neighborhoods, deletion of multiple boundary sites were required to observe changes in gene expression (Dowen et al., 2014). Thus, insulated neighborhood boundaries constrain the activity of enhancers to genes within the neighborhood. Insulated neighborhood boundaries are also necessary to maintain repression of genes within the neighborhood; deletion of a CTCF anchor of an insulated neighborhood containing a Polycomb repressed gene led to the activation of that gene (Dowen et al., 2014).

The finding that cancer cells can activate oncogenes through somatic mutations or epigenetic modifications that disrupt insulated neighborhood boundaries provides additional evidence that neighborhood loop anchors have functional insulating properties (Figure 2E) (Flavahan et al., 2016; Hnisz et al., 2016; Katainen et al., 2015). Silent proto-oncogenes typically occur

within insulated neighborhoods and genetic modification of the neighborhood loop anchors can cause activation of these oncogenes (Flavahan et al., 2016; Hnisz et al., 2016). Somatic mutations occur frequently and recurrently in the loop anchors of oncogene-containing Insulated neighborhoods in a variety of cancer cells (Figure 2E). Indeed, the CTCF DNA-binding motif in loop anchor regions is among the most altered human transcription factor-binding sequence in cancer cells (Ji et al., 2016). These observations are consistent with the idea that mutations that alter the loop anchor sites of oncogene-containing insulated neighborhoods make an important contribution to the misregulation of gene expression that is inherent to the cancer state (Flavahan et al., 2016; Hnisz et al., 2016; Katainen et al., 2015).

Maintenance of loop anchors during development

The majority of insulated neighborhoods that have been mapped in human ES cells appear to be maintained during development because the experimental evidence indicates that CTCF binding and CTCF-CTCF loop structures are very similar in many other human cells (Ji et al., 2016). This constitutive behavior is consistent with the observation that CTCF is expressed in all cell types examined (Phillips and Corces, 2009). While different cell types share very similar insulated neighborhood boundaries, the enhancer-gene interactions that occur within these neighborhoods are cell-type specific because enhancer activity is cell-type specific (Figure 3) (Ji et al., 2016; Smith et al., 2016).

Evolutionary conservation

The CTCF sites that form insulated neighborhood boundaries are evolutionarily conserved. Human germline variation is rare in CTCF binding sites at insulated neighborhood boundaries and few GWAS variants occur in these sites (Ji et al., 2016). Analysis of CTCF-binding sites across primates indicates that the DNA sequence in anchor regions of insulated neighborhoods is far more conserved in primates than in regions bound by CTCF that do not participate in neighborhood loops (~55% of CTCF binding sites in the human genome do not appear to participate in insulated neighborhood loops) (Ji et al., 2016).

A subset of CTCF-CTCF loops connect enhancers and promoters, while others contribute to recombination

Although most CTCF-CTCF loops form insulated neighborhoods (Figure 2C), a subset of CTCF-CTCF loops (~19% in hESCs) occur at enhancer-promoter interaction sites. We infer that these interactions facilitate gene activation; previous studies have noted that some genes interact with their enhancers by this mechanism (Ong and Corces, 2014).

CTCF- and cohesin-associated loops also play essential roles in V(D)J recombination of the immunoglobulin heavy-chain in developing lymphocytes. Recombinase-assisted rearrangements of DNA segments encoding regions of antigen binding receptors occur during the development of cells of the adaptive immune system. CTCF-CTCF looping has been implicated in bringing these segments into spatial proximity and also in constraining the off-target effects of the recombinase (Dong et al., 2015; Hu et al., 2015). Thus a subset of CTCF-CTCF loops have evolved to control DNA recombination. It is possible that these

CTCF-CTCF loops and the enhancer-promoter CTCF-CTCF loops may also act as insulated neighborhoods.

Insulated neighborhoods are the mechanistic basis of TADs

TADs are megabase- sized domains with relatively high DNA interaction frequencies and are identified using a Hidden Markov Model-based analysis of Hi-C chromosome conformation capture data (Dixon et al., 2012; Nora et al., 2012). Two observations argue that TADs are generally composed of, and likely structured by, insulated neighborhoods.

Cohesin ChIA-PET data was used to identify insulated neighborhoods and the enhancer-promoter interactions that occur within them because cohesin occupies both CTCF-CTCF insulators and enhancer-promoter interaction sites (Kagey et al., 2010). This ChIA-PET DNA interaction data is biased: it is enriched for interaction sites where cohesin is present. Hi-C interaction data does not have this bias: it identifies interactions that should be independent of the functions of any one protein. Nonetheless, the TADs identified by processing Hi-C data with the Hidden Markov algorithm (Dixon et al., 2012; Nora et al., 2012) can also be identified when this algorithm is used to process cohesin ChIA-PET data (Ji et al., 2016). Murine and human ESC ChIA-PET data, processed with the same Hidden Markov Model, captures most TAD boundaries derived from Hi-C data in murine and human ESCs (Figure 4A). These results suggest that insulated neighborhoods are a major structuring component of TADs.

ChIA-PET data revealed that at least 50% of TADs have TAD-spanning CTCF-CTCF loops (Ji et al., 2016) and are thus insulated neighborhoods. Because the existing ChIA-PET data is not saturating, this is a minimal estimate; it is possible that the majority of TADs have TAD-spanning CTCF-CTCF loops. Some TADs appear to be a single insulated neighborhood, while others consist of multiple nested or multiple independent insulated neighborhoods (Figure 4B–D). TADs were originally discovered using Hi-C data that had ~40kb resolution, and improvements of the experimental and analytical aspects of Hi-C methods revealed that many TADs are composed of smaller TAD-like domains at higher resolution (Schmitt et al., 2016). It is thus possible that all high-resolution TADs are insulated neighborhoods and vice-versa, and so the CTCF-CTCF loops that encompass insulated neighborhoods, together with the enhancer-promoter loops within them, likely form the mechanistic basis of most interphase chromosome structures.

Relationships between insulated neighborhoods and other DNA loop models

Mammalian chromosome loop structures have been reported in multiple studies, which have used different descriptors for these loops, including sub-TADs, loop domains, and CTCF-contact domains (Phillips-Cremins et al., 2013; Rao et al., 2014; Tang et al., 2015). An analysis of the structures described in these studies suggests that they generally represent the same structural unit as the insulated neighborhoods described here (Figure 5).

Pioneering studies using 5C-technology first described TAD subtopologies, termed “sub-TADs”, together with the structuring proteins CTCF and cohesin, at seven genomic loci in murine ESCs (Phillips-Cremins et al., 2013). Sub-TADs were found to be constitutively present in multiple cell types and cell type-specific enhancer-promoter contacts occurred within sub-TAD boundaries (Phillips-Cremins et al., 2013). Examination of several of the sub-TAD loop structures (e.g., at the *Nanog* and *Olig1-2* loci) reveals that they are among the insulated neighborhoods described for murine ESCs (Dowen et al., 2014; Phillips-Cremins et al., 2013). Although these early studies of sub-TAD structures did not test the insulating properties of the sub-TADs, this may be one of the earliest descriptions of what we now term insulated neighborhoods. Another study used high-resolution hi-C technology to identify ~5,000 chromatin loops whose boundaries are occupied by CTCF and cohesin in multiple human cell types; these were termed “loop domains” (Rao et al., 2014). This and other studies have noted that the DNA binding motif of CTCF is asymmetric and thus directional, and the CTCF anchors of >90% the loop domains occur in a convergent orientation (de Wit et al., 2015; Gomez-Marin et al., 2015; Guo et al., 2015; Ji et al., 2016; Vietri Rudan et al., 2015). The convergent orientation of CTCF motifs in the anchors is also a general feature of insulated neighborhoods (Ji et al., 2016). A recent study mapped CTCF-associated contacts genome-wide using CTCF ChIA-PET and revealed ~2,000 “CTCF-contact domains (CCDs)” in human cells (Tang et al., 2015). CTCF-contact domains are clusters of CTCF-associated chromatin loops that appear separated from other CTCF-associated loops. The vast majority of RNA Polymerase II-associated interactions (e.g. enhancer-promoter loops) were found to occur within the boundaries of CTCF-contact domains. The anchor sites of CTCF contact domains are bound by CTCF and cohesin, and the CTCF anchors of ~90% of the CTCF contact domains occur in a convergent orientation (Tang et al., 2015). These features suggest that CTCF contact domains are either insulated neighborhoods or clusters of insulated neighborhoods.

Comparison of insulated neighborhoods with loop domains and CTCF contact domains in the same cell type suggests extensive overlap between these structures. For example, 70% of loop domains, and 54% CTCF contact domains have the same boundaries as an insulated neighborhood in human lymphoblastoid cells (Rao et al., 2014; Tang et al., 2015). Differences in experimental and analytical methods can explain many of the differences in loop structures reported by various studies; indeed, similarities among loop structures are more evident when data are analyzed with increasing stringency (Figure 5).

Insulated neighborhoods, gene regulation and disease

Gene regulation

Studies of gene control at imprinted loci were among the first to reveal the importance of CTCF loops in gene control and the role of DNA methylation in control of CTCF-associated loops. Parent-of-origin specific gene activity at the imprinted *IGF2/H19* locus is controlled by allele-specific CTCF-CTCF interactions that constrain enhancer-gene contacts in a DNA methylation-dependent manner (Figure 6A) (Kurukuti et al., 2006; Murrell et al., 2004). An insulated neighborhood on the maternal allele allows an enhancer-promoter interaction that activates the *H19* gene but not the *IGF2* gene, which is excluded from the neighborhood. A

larger insulated neighborhood is formed on the paternal allele to allow an enhancer-promoter interaction that activates the *IGF2* gene. Paternal allele-specific DNA methylation of a CTCF site in the *H19* promoter region abrogates CTCF binding, thus causing differential CTCF-CTCF loop formation while silencing *H19* expression (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000; Szabo et al., 2000). Individuals who lose these allele-specific insulated neighborhoods develop Beckwith-Wiedemann syndrome (when both alleles have the paternal type of insulated neighborhood; Figure 6B) or Silver-Russell syndrome (when both alleles have the maternal type of insulated neighborhood; Figure 6C) (Nativio et al., 2011).

Early studies of gene control at the beta-globin locus also demonstrated the importance of CTCF and its looping interactions in developmental control (Hou et al., 2008; Splinter et al., 2006; Tolhuis et al., 2002). In vertebrates, the beta globin locus contains a cluster of fetal and adult globin genes, and the developmental control of these genes is exerted by an upstream regulatory element called the locus control region (LCR) (Figure 6D). In erythroid cells expressing globin genes, a large CTCF-CTCF loop encompasses the beta-globin genes and the locus control region (LCR), consistent with the organization of the locus in an insulated neighborhood. In fetal brain cells, CTCF binding to one of the beta-globin loop anchor regions is absent and CTCF-CTCF loop formation is not detected (Splinter et al., 2006; Tolhuis et al., 2002), which suggests that tissue-specific CTCF-CTCF loops participate in developmental gene control.

Recent studies further support the view that the loop anchors of insulated neighborhoods play key roles in gene control. As described above, the vast majority of enhancer-promoter interactions occur within insulated neighborhoods in embryonic stem (ES) cells, and genetic perturbation of insulated neighborhood anchors leads to misregulation of local genes (Dowen et al., 2014; Ji et al., 2016). Positional information in the developing embryo depends on the precise expression of Homeobox (Hox) genes, and CTCF sites located within a Hox gene cluster play a critical role in proper expression of Hox genes (Narendra et al., 2015); some of these critical CTCF sites form insulated neighborhood anchors in ES cells. Hereditary mutations that invert or delete a TAD boundary at the EPHA4 locus have recently been linked to limb malformations in humans (Lupianez et al., 2015), and this TAD has a TAD-spanning CTCF-CTCF loop in ES cells, indicating that it is an insulated neighborhood. Inversion of a CTCF anchor has also been shown to cause altered enhancer-promoter contacts at the protocadherin locus (Guo et al., 2015). The effect of insulated neighborhoods on signal-responsive gene expression also supports the concept of insulation; gene activation by NOTCH signaling in T-cells was found to be restricted to genes that occur within the same CTCF-CTCF loops as NOTCH-dependent enhancers (Wang et al., 2014).

Altered neighborhoods in cancer

Recent studies have revealed that mutations that alter the loop anchor sites of oncogene-containing insulated neighborhoods make an important contribution to the misregulation of gene expression that is inherent to the cancer state (Flavahan et al., 2016; Hnisz et al., 2016; Katainen et al., 2015). Somatic mutations occur frequently and recurrently in loop anchors of oncogene-containing insulated neighborhoods in a variety of cancer cells and the CTCF

DNA-binding motif in loop anchor regions is among the most altered human transcription factor-binding sequence in cancer cells (Flavahan et al., 2016; Hnisz et al., 2016; Katainen et al., 2015). DNA hypermethylation occurs in some cancer cells, and tumor-specific DNA methylation has recently been implicated in the disruption of CTCF binding, alteration of chromosome structure and dysregulation of oncogene expression (Flavahan et al., 2016). Furthermore, chromosomal rearrangements such as translocations or deletions which activate oncogenes also disrupt insulated neighborhoods around those genes without altering the sequence of the gene itself (Groschel et al., 2014; Hnisz et al., 2016). Cancer genome sequencing has revealed that somatic mutations occur in CTCF and cohesin coding sequences in various solid tumors and leukemias (Lawrence et al., 2014), and it seems likely that these mutations contribute to oncogenesis by altering insulated neighborhoods.

Disease-associated variation in loop anchors

Genetic variants occur rarely in insulated neighborhood anchors. However, allelic non-coding variants in CTCF loop anchors have been shown to correlate with allele-specific enhancer-promoter interactions (Tang et al., 2015). Among these, one variant, associated with asthma, disrupts CTCF binding and CTCF loop formation (Tang et al., 2015). A recent human population genetics study showed several genetic variants linked with an individual's lipid profile (e.g., LDL, HDL) and present within at least 1% of the population were found within CTCF binding sites (UK10K Consortium et al., 2015), and it is possible that these variants disrupt CTCF binding at insulated neighborhood boundaries. With the new knowledge of CTCF loop anchors in human cells, geneticists will likely identify additional genetic variants that contribute to non-cancer disease through disruption of insulators.

Target genes of disease-associated enhancer variation

Insulated neighborhood models provide a new approach to identify the target genes of disease-associated enhancer variation. Tens of thousands of non-coding genetic variants have been linked with various human diseases and traits in genome-wide association studies (GWAS), and the majority of these variants occur in enhancers (Ernst et al., 2011; Farh et al., 2015; Hnisz et al., 2013; Maurano et al., 2012). The identification of the target genes of these variants is challenging because proximity-based assignment has proven, in some cases, to be inaccurate. Mapping interactions between enhancers and promoters in disease-relevant cells improves the accuracy of the assignment (Grubert et al., 2015; McGeachie et al., 2016; Pomerantz et al., 2009), but this is not always feasible. Because insulated neighborhoods tend to be shared by different cell types, existing maps of insulated neighborhoods should allow investigators to develop a hypothesis regarding the potential target genes of enhancer-associated variation (Figure 7A). For example, a recent study revealed that a genetic variant associated with obesity and previously assigned to the *FTO* gene in fact has no impact on *FTO*, but affects the *IRX3* and *IRX5* genes (Claussnitzer et al., 2016). Although the variant is located in an intronic enhancer within *FTO*, both *IRX3* and *IRX5* are located in the same insulated neighborhood as the variant (Figure 7B). Similarly, functional investigation of a genetic variant associated with Type 2 diabetes, and previously assigned to the *CDC123* and *CAMK1D* genes based on proximity, revealed that the variant affects the distal *CAMK1D* gene and not *CDC123* (Fogarty et al., 2014; GTEx Consortium, 2015). Examination of insulated neighborhood structures reveals that *CAMK1D* is located in the same

neighborhood as the variant, whereas *CDC123* is not (Figure 7C). These examples suggest that insulated neighborhood maps can facilitate the identification of genes affected by non-coding genetic variants.

Epigenetic editing of insulated neighborhood structures

The CTCF binding site in insulated neighborhood loop anchors is hypomethylated (Ji et al., 2016); DNA methylation abrogates CTCF DNA binding (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000; Szabo et al., 2000). This suggests that site-specific methylation and demethylation of a neighborhood anchor can alter neighborhood structures. Indeed, targeted methylation of a neighborhood anchor site with a dCas9-DNA-metyltransferase-3 fusion protein has been shown to disrupt the neighborhood (Figure 8A) (Liu et al., 2016). Similarly, targeted de-methylation with a dCas9-TET fusion protein has been demonstrated (Amabile et al., 2016; Liu et al., 2016), and this strategy could be used to restore an insulated neighborhood whose anchor site is disrupted by aberrant DNA methylation (Figure 8B). These tools might evolve to be useful for therapeutic purposes.

Challenges

How dynamic and heterogeneous are insulated neighborhood loops?

The example of CTCF-CTCF loop and gene control at the imprinted *IGF2/H19* locus suggests that the neighborhoods are sufficiently stable to prevent development of the diseases associated with neighborhood dysregulation. Furthermore, the striking similarity of TAD boundaries across cell types (Dixon et al., 2015; Dixon et al., 2012), which we argue is produced largely by insulated neighborhood structures (Figure 4A), suggests that these neighborhoods are rather stable. However, the dynamics of the loop structures that form insulated neighborhoods, and the enhancer-promoter interactions within them, are not yet understood. Similarly, the cell-to-cell heterogeneity of DNA loop structures is not understood, and the extent to which allele-specific loops occur is not clear. The fraction of time that CTCF is bound to a loop anchor site, the fraction of time that it spends in a dimerized state, and the extent to which CTCF switches its dimeric partner are three of the elements that factor into a potential solution to these questions. The experimental approaches used thus far to identify CTCF loops generally depend on the study of populations of cells, and thus the present data are inadequate to address questions of dynamics. Improvements in single-cell technologies will be needed to reveal the dynamics of insulated neighborhoods and enhancer-promoter interactions.

Computational simulations of chromosome loops have led to the suggestion that only a subset of insulated neighborhoods occur in each cell within a population at any given time, and to the hypothesis that an extrusion model can facilitate enhancer-promoter interactions and lead to insulation in all cells (Doyle et al., 2014; Fudenberg et al., 2016; Giorgetti et al., 2014; Sanborn et al., 2015). This model postulates that chromosome loops are formed by the extrusion of chromatin by an “extrusion complex”, an entity that acts as a molecular motor to draw DNA through a cohesin complex (reviewed in (Dekker and Mirny, 2016)). In this model, the extruded DNA loop forms CTCF-CTCF loop when the cohesin-containing extrusion complex is blocked by a pair of convergently oriented CTCF molecules bound to

two sites in the extruded DNA. Because transcription initiation by RNA polymerase II includes cohesin loading at the enhancer-promoter junction (Kagey et al., 2010), and two RNA molecules can transcribe bi-directionally from promoters and enhancers (Core et al., 2008; Seila et al., 2008; Sigova et al., 2013), it is possible that RNA polymerase II plays a role in this postulated extrusion. Condensin II, which is loaded onto DNA at sites of transcription initiation together with cohesin (Dowen et al., 2013), is another candidate "extrusion complex" factor. These extrusion models have yet to be tested experimentally.

To what extent are insulated neighborhoods shared across cell types?

Most genomic data is inherently noisy and is filtered to provide an interpretation at some arbitrarily chosen confidence interval. The experimental approaches used to determine DNA interactions, which include Hi-C and ChIA-PET technologies, produce especially noisy data. Furthermore, DNA interaction data can be sparse – especially when using small numbers of cells. These features of the data makes it challenging to provide good estimates of the extent to which all the DNA loop structures of any one cell type are shared by another cell type, but comparisons can be made for the set of loops that meet high-confidence criteria in similar experimental data from two or more cell types. Studies on TADs have estimated that most TAD boundaries are shared by any two cell types (Dixon et al., 2015; Dixon et al., 2012). Studies on insulated neighborhoods have estimated that approximately 80% of neighborhood boundaries of shared by any two cell types (Hnisz et al., 2016; Ji et al., 2016). Given the sparsity of data, and the noise in these datasets, it is possible that the vast majority of TADs and insulated neighborhoods are shared by most cell types, although there is some evidence that CTCF binding and CTCF-CTCF loop formation can be cell type-specific (Narendra et al., 2015; Splinter et al., 2006; Tolhuis et al., 2002; Wang et al., 2012). A broader survey of cell types will be needed to determine the extent to which cell type-specific insulated neighborhoods exist in human cells.

How are insulated neighborhood loop anchors regulated?

DNA methylation plays a key role in CTCF-CTCF loop anchor control and gene control, as illustrated in the imprinted *IGF2/H19* locus (Figure 6), but the regulatory mechanisms that produce site-specific methylation of loop anchors are not well-understood. In *Drosophila*, a number of proteins have been identified that influence CTCF binding and insulator function (Phillips-Cremins and Corces, 2013), but it is not clear whether similar proteins might contribute to regulation of mammalian loop anchors. CTCF-binding to DNA can also be modulated by post-translational modifications such as poly-ADP ribosylation (Ong et al., 2013). Non-coding RNA has also been implicated in regulation of CTCF binding to DNA at certain loci (Saldana-Meyer et al., 2014). Future studies will need to address the extent to which CTCF modifications and RNA modulate neighborhood loop anchors.

How does a loop insulate?

With the 2-dimensional representations of loops shown here, it is reasonable to ask how an insulated neighborhood boundary suppresses enhancer-promoter loop formation across the boundaries. An element in a 2-D chromatin loop should be able to contact elements in other loops. Additional structuring of the neighborhood, such as condensing the looped chromatin into a compact ball, would reduce the opportunity to interact with other neighborhoods. One

candidate for such a factor is Condensin II, which is loaded onto active promoters together with cohesin (Dowen et al., 2013). Condensin is known to inhibit transvection in *Drosophila* polytene chromosomes, where it presumably prevents interactions between two alleles (Hartl et al., 2008). Other proteins, such as the Polycomb repressive complex (Francis et al., 2004), may contribute to effectively condense a silent insulated neighborhood.

Do loop anchors vary in insulation strength?

Early efforts to quantify insulation on a genome-wide scale suggest that differences in insulation strength at CTCF anchors might occur, and perhaps correlate with certain genomic features (Phillips-Cremins and Corces, 2013). Insulated neighborhoods can have boundaries with multiple CTCF binding sites and can be nested within larger insulated neighborhoods. These features appear to produce a higher “insulation score” for enhancers and genes within the neighborhood and may thus represent a safeguard against perturbation. For example, the β-globin genes are located in nested neighborhoods, and perturbation of the inside neighborhood anchors has little effect on globin gene expression (Bender et al., 2006). In ES cells, deletion of multiple boundary sites were required to observe changes in gene expression at certain loci (Dowen et al., 2014). Interestingly, a recent study found that a set of adjacent neighborhoods show evidence of “merging” together during the differentiation of germinal center B cells (Bunting et al., 2016), indicating that the insulating properties of some neighborhoods may be under developmental control. Additional study is necessary to understand the structural and mechanistic features that contribute to insulator strength.

What additional mechanisms contribute to enhancer-gene specificity?

The insulated neighborhood model can explain how enhancer-promoter specificity is obtained when a single gene occurs together with its regulatory elements within the neighborhood, but it does not fully explain enhancer-promoter specificity when multiple genes are present. We estimate that in neighborhoods with two genes, the activity of the two is coherent in 60% (both are active or both are silent). The tendency for these two-gene neighborhoods to have coherent on or off activities suggests that genes in these neighborhoods may often be co-regulated. Indeed, recent evidence in *Drosophila* suggests that an enhancer can target all genes within an insulated chromatin structure (Fukaya et al., 2016). Further regulation may occur post-transcriptionally (*e.g.,* microRNAs), which could account for differential transcript accumulation. As noted above, it is also possible that some degree of enhancer-gene specificity is obtained through the interaction of specific factors bound at enhancer and promoters (Zabidi et al., 2015).

Future perspective

Evidence that proper activation and repression of genes is dependent on the integrity of insulated neighborhoods argues that these are structural and functional units of mammalian gene control. insulated neighborhoods provide a new framework for investigating gene control and interpreting the effects of non-coding genetic variation. A fuller understanding of the normal and abnormal control of any gene will require consideration of the potential contribution of any regulatory elements within its neighborhood and the possibility of loop

anchor regulation. New insights into the role of genome structure in selective gene control in development and disease will be accelerated with improvements in technologies to map chromosome structures at improved resolution, ideally in an allele-specific fashion in single cells.

Acknowledgments

Many ideas discussed in this perspective emerged from conversations with Brian Abraham, Frederick Alt, Jay Bradner, Daniel Dadon, Eric Guo, Rudolf Jaenisch, Xiong Ji, Tony Lee, Bryan Lajoie, Charles H. Li, Stuart Levine, Thoru Pedersen, Ana Pombo, Robert Roeder, Ben Sabari, Jurian Schuijers, Anne-Laure Valton, Robert Weinberg, Abraham Weintraub, Alicia Zamudio, Len Zon, and Thomas Zwaka. We are particularly grateful to Brad Bernstein, Victor Corces, Job Dekker, Edith Heard, Danny Reinberg, Bing Ren, Yijun Ruan and Phillip Sharp for comments on the manuscript. We thank Jennifer Cook-Chrysos for helping with the graphical illustrations. The work was supported by an NIH Grant HG002668 (R.A.Y.), an Erwin Schrödinger Fellowship (J3490) from the Austrian Science Fund (D.H.), and an American Cancer Society – New England Division Postdoctoral Fellowship (D.S.D). R.A.Y. is a founder of Syros Pharmaceuticals.

References

Amabile A, Migliara A, Capasso P, Biffi M, Cittaro D, Naldini L, Lombardo A. Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell. 2016; 167:219–232. e214. [PubMed: 27662090]

Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell. 1981; 27:299–308. [PubMed: 6277502]

Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000; 405:482–485. [PubMed: 10839546]

Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 1999; 98:387–396. [PubMed: 10458613]

Bender MA, Byron R, Ragoczy T, Telling A, Bulger M, Groudine M. Flanking HS-62.5 and 3' HS1, and regions upstream of the LCR, are not required for beta-globin transcription. Blood. 2006; 108:1395–1401. [PubMed: 16645164]

Benoist C, Chambon P. In vivo sequence requirements of the SV40 early promotor region. Nature. 1981; 290:304–310. [PubMed: 6259538]

Bickmore WA, van Steensel B. Genome architecture: domain organization of interphase chromosomes. Cell. 2013; 152:1270–1284. [PubMed: 23498936]

Boveri T. Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. Arch Zellforsch. 1909:181–268.

Buecker C, Wysocka J. Enhancers as information integration hubs in development: lessons from genomics. Trends in genetics : TIG. 2012; 28:276–284. [PubMed: 22487374]

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011; 144:327–339. [PubMed: 21295696]

Bunting KL, Soong TD, Singh R, Jiang Y, Beguelin W, Poloway DW, Swed BL, Hatzi K, Reisacher W, Teater M, et al. Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. Immunity. 2016; 45:497–512. [PubMed: 27637145]

Butler JE, Kadonaga JT. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes & development. 2001; 15:2515–2519. [PubMed: 11581157]

Cavalli G, Misteli T. Functional implications of genome topology. Nature structural & molecular biology. 2013; 20:290–299.

Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell research. 2012; 22:490–503. [PubMed: 22270183]

Choi OR, Engel JD. Developmental regulation of beta-globin gene switching. Cell. 1988; 55:17–26. [PubMed: 3167976]

Chung JH, Whiteley M, Felsenfeld G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell. 1993; 74:505–514. [PubMed: 8348617]

Claussnitzer M, Hui CC, Kellis M. FTO Obesity Variant and Adipocyte Browning in Humans. The New England journal of medicine. 2016; 374:192–193. [PubMed: 26760096]

Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008; 322:1845–1848. [PubMed: 19056941]

Cremer T, Cremer M. Chromosome territories. Cold Spring Harbor perspectives in biology. 2010; 2:a003889. [PubMed: 20300217]

de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature. 2013; 502:499–506. [PubMed: 24153303]

de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF Binding Polarity Determines Chromatin Looping. Molecular cell. 2015; 60:676–684. [PubMed: 26527277]

Dekker J, Heard E. Structural and functional diversity of Topologically Associating Domains. FEBS letters. 2015; 589:2877–2884. [PubMed: 26348399]

Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 2016; 164:1110–1121. [PubMed: 26967279]

DeMare LE, Leng J, Cotney J, Reilly SK, Yin J, Sarro R, Noonan JP. The genomic landscape of cohesin-associated chromatin interactions. Genome research. 2013; 23:1224–1234. [PubMed: 23704192]

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015; 518:331–336. [PubMed: 25693564]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

Dong J, Panchakshari RA, Zhang T, Zhang Y, Hu J, Volpi SA, Meyers RM, Ho YJ, Du Z, Robbiani DF, et al. Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. Nature. 2015; 525:134–139. [PubMed: 26308889]

Dowen JM, Bilodeau S, Orlando DA, Hubner MR, Abraham BJ, Spector DL, Young RA. Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. Stem cell reports. 2013; 1:371–378. [PubMed: 24286025]

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, et al. Control of cell identity genes occurs in insulated neighborhoods in Mammalian chromosomes. Cell. 2014; 159:374–387. [PubMed: 25303531]

Doyle B, Fudenberg G, Imakaev M, Mirny LA. Chromatin loops as allosteric modulators of enhancer-promoter interactions. PLoS computational biology. 2014; 10:e1003867. [PubMed: 25340767]

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–49. [PubMed: 21441907]

Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015; 518:337–343. [PubMed: 25363779]

Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016; 529:110–114. [PubMed: 26700815]

Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. PLoS genetics. 2014; 10:e1004633. [PubMed: 25211022]

Francis NJ, Kingston RE, Woodcock CL. Chromatin compaction by a polycomb group protein complex. Science. 2004; 306:1574–1577. [PubMed: 15567868]

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell reports. 2016; 15:2038–2049. [PubMed: 27210764]

Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. Cell. 2016; 166:358–368. [PubMed: 27293191]

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009; 462:58–64. [PubMed: 19890323]

Geyer PK, Corces VG. DNA position-specific repression of transcription by a Drosophila zinc finger protein. Genes & development. 1992; 6:1865–1873. [PubMed: 1327958]

Gibcus JH, Dekker J. The hierarchy of the 3D genome. Molecular cell. 2013; 49:773–782. [PubMed: 23473598]

Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, Tiana G, Heard E. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell. 2014; 157:950–963. [PubMed: 24813616]

Gomez-Marin C, Tena JJ, Acemel RD, Lopez-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, et al. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:7542–7547. [PubMed: 26034287]

Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. Cell stem cell. 2014; 14:762–775. [PubMed: 24905166]

Groschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BA, Erpelinck C, van der Velden VH, Havermans M, Avellino R, van Lom K, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell. 2014; 157:369–381. [PubMed: 24703711]

Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, Greenside P, Srivas R, Phanstiel DH, Pekowska A, et al. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. Cell. 2015; 162:1051–1065. [PubMed: 26300125]

Gruss P, Dhar R, Khoury G. Simian virus 40 tandem repeated sequences as an element of the early promoter. Proceedings of the National Academy of Sciences of the United States of America. 1981; 78:943–947. [PubMed: 6262784]

GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–660. [PubMed: 25954001]

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell. 2015; 162:900–910. [PubMed: 26276636]

Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nature genetics. 2011; 43:630–638. [PubMed: 21685913]

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature. 2000; 405:486–489. [PubMed: 10839547]

Hartl TA, Smith HF, Bosco G. Chromosome alignment and transvection are antagonized by condensin II. Science. 2008; 322:1384–1387. [PubMed: 19039137]

Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, Zhang MQ, Snyder MP. Genome-wide map of regulatory interactions in the human genome. Genome research. 2014; 24:1905–1917. [PubMed: 25228660]

Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nature reviews Molecular cell biology. 2015; 16:144–154. [PubMed: 25650801]

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]

Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 2016; 351:1454–1458. [PubMed: 26940867]

Hou C, Zhao H, Tanimoto K, Dean A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:20398–20403. [PubMed: 19074263]

Hu J, Zhang Y, Zhao L, Frock RL, Du Z, Meyers RM, Meng FL, Schatz DG, Alt FW. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. Cell. 2015; 163:947–959. [PubMed: 26593423]

Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, Weintraub AS, Hnisz D, Pegoraro G, Lee TI, et al. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell stem cell. 2016; 18:262–275. [PubMed: 26686465]

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]

Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Current biology : CB. 2000; 10:853–856. [PubMed: 10899010]

Katainen R, Dave K, Pitkanen E, Palin K, Kivioja T, Valimaki N, Gylfe AE, Ristolainen H, Hanninen UA, Cajuso T, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nature genetics. 2015; 47:818–821. [PubMed: 26053496]

Kellum R, Schedl P. A position-effect assay for boundaries of higher order chromosomal domains. Cell. 1991; 64:941–950. [PubMed: 1848159]

Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:10684–10689. [PubMed: 16815976]

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014; 505:495–501. [PubMed: 24390350]

Levine M, Cattoglio C, Tjian R. Looping back to leap forward: transcription enters a new era. Cell. 2014; 157:13–25. [PubMed: 24679523]

Liu M, Maurano MT, Wang H, Qi H, Song CZ, Navas PA, Emery DW, Stamatoyannopoulos JA, Stamatoyannopoulos G. Genomic discovery of potent chromatin insulators for human gene therapy. Nature biotechnology. 2015; 33:198–203.

Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. Editing DNA Methylation in the Mammalian Genome. Cell. 2016; 167:233–247. e217. [PubMed: 27662091]

Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell. 2015

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337:1190–1195. [PubMed: 22955828]

McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, Wise RA, Szefler SJ, Sharma S, Kho AT, et al. Genetics and Genomics of Longitudinal Lung Function Patterns in Asthmatics. American journal of respiratory and critical care medicine. 2016

Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annual review of genomics and human genetics. 2016; 17:17–43.

Murrell A, Heeson S, Reik W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nature genetics. 2004; 36:889–893. [PubMed: 15273689]

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 2015; 347:1017–1021. [PubMed: 25722416]

Nativio R, Sparago A, Ito Y, Weksberg R, Riccio A, Murrell A. Disruption of genomic neighbourhood at the imprinted IGF2-H19 locus in Beckwith-Wiedemann syndrome and Silver-Russell syndrome. Human molecular genetics. 2011; 20:1363–1374. [PubMed: 21282187]

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–385. [PubMed: 22495304]

Ohtsuki S, Levine M, Cai HN. Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes & development. 1998; 12:547–556. [PubMed: 9472023]

Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature reviews Genetics. 2011; 12:283–293.

Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nature reviews Genetics. 2014; 15:234–246.

Ong CT, Van Bortle K, Ramos E, Corces VG. Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in Drosophila. Cell. 2013; 155:148–159. [PubMed: 24055367]

Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009; 137:1194–1211. [PubMed: 19563753]

Phillips-Cremins JE, Corces VG. Chromatin insulators: linking genome organization to cellular function. Molecular cell. 2013; 50:461–474. [PubMed: 23706817]

Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013; 153:1281–1295. [PubMed: 23706625]

Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nature genetics. 2009; 41:882–884. [PubMed: 19561607]

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–1680. [PubMed: 25497547]

Ren B, Yue F. Transcriptional Enhancers: Bridging the Genome and Phenome. Cold Spring Harbor symposia on quantitative biology. 2015; 80:17–26. [PubMed: 26582789]

Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Roadmap Epigenomics C. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]

Saldana-Meyer R, Gonzalez-Buendia E, Guerrero G, Narendra V, Bonasio R, Recillas-Targa F, Reinberg D. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes & development. 2014; 28:723–734. [PubMed: 24696455]

Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:E6456–6465. [PubMed: 26499245]

Schmitt AD, Hu M, Ren B. Genome-wide mapping and analysis of chromosome architecture. Nature reviews Molecular cell biology. 2016

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. Divergent transcription from active promoters. Science. 2008; 322:1849–1851. [PubMed: 19056940]

Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C, Sharp PA, Giallourakis CC, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:2876–2881. [PubMed: 23382218]

Smith EM, Lajoie BR, Jain G, Dekker J. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. American journal of human genetics. 2016; 98:185–201. [PubMed: 26748519]

Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics. 2012; 13:613–626.

Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes & development. 2006; 20:2349–2354. [PubMed: 16951251]

Szabo P, Tang SH, Rentsendorj A, Pfeifer GP, Mann JR. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Current biology : CB. 2000; 10:607–610. [PubMed: 10837224]

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell. 2015; 163:1611–1627. [PubMed: 26686651]

Tjian R, Maniatis T. Transcriptional activation: a complex puzzle with few easy pieces. Cell. 1994; 77:5–8. [PubMed: 8156597]

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell. 2002; 10:1453–1465. [PubMed: 12504019]

Udvardy A, Maine E, Schedl P. The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. Journal of molecular biology. 1985; 185:341–358. [PubMed: 2997449]

Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JR, Xu C, Futema M, et al. UK10K Consortium. The UK10K project identifies rare variants in health and disease. Nature. 2015; 526:82–90. [PubMed: 26367797]

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell reports. 2015; 10:1297–1309. [PubMed: 25732821]

Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome research. 2012; 22:1680–1688. [PubMed: 22955980]

Wang H, Zang C, Taing L, Arnett KL, Wong YJ, Pear WS, Blacklow SC, Liu XS, Aster JC. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:705–710. [PubMed: 24374627]

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature. 2015; 518:556–559. [PubMed: 25517091]

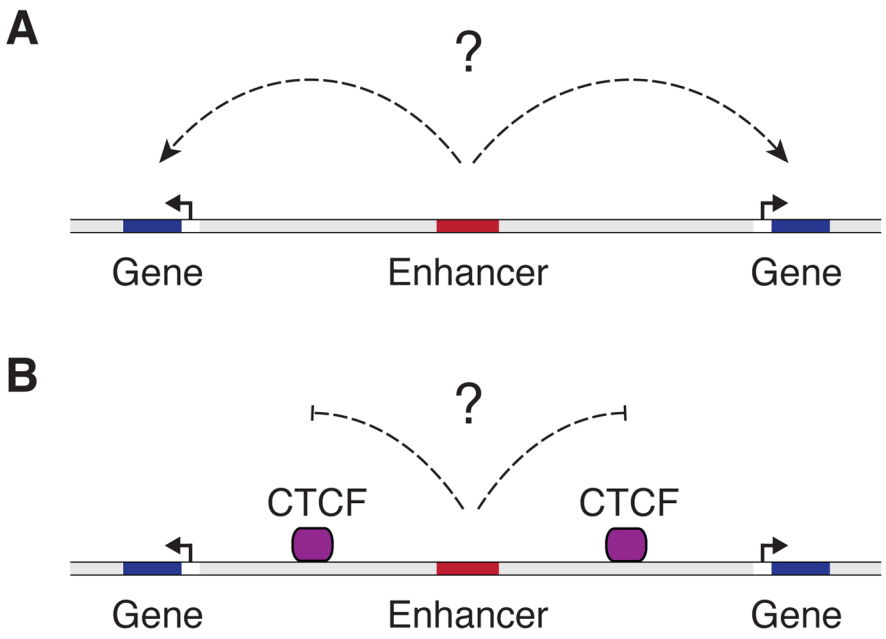

Figure 1. The enhancer-gene specificity conundrum

A. Model of a genomic region encompassing an enhancer and two genes. The features that cause an enhancer to regulate only specific genes are still not fully understood, which we refer to as the enhancer-gene specificity conundrum

B. Model of a genomic region encompassing an enhancer and two genes with the transcription factor CTCF bound in-between. CTCF is a component of enhancer-blocking insulators, but which CTCF-bound sites function as an insulator *in vivo* is still unclear.

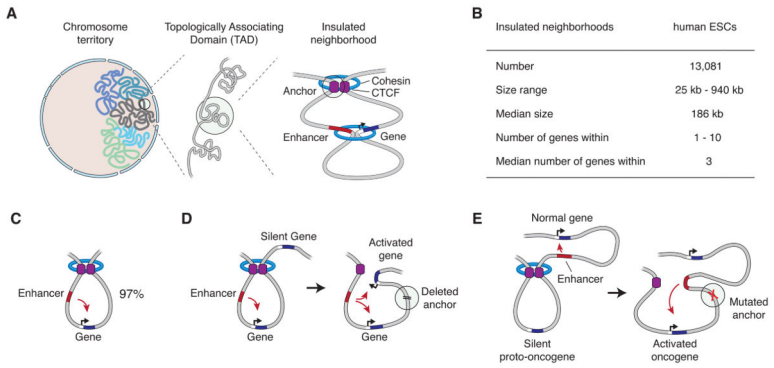

Figure 2. Insulated neighborhoods

A. Hierarchy of chromosome structures: Chromosome territories, Topologically Associating Domains and insulated neighborhoods. Anchor refers to the CTCF-bound site interacting with another CTCF-bound sites, both co—bound by a cohesin ring.

B. Features of insulated neighborhoods in human embryonic stem cells (ESCs). The values displayed for the size range and number of genes represent the middle 95% of the data range.

C. Evidence for insulation of insulated neighborhoods: 97% Enhancer-gene interactions occur within insulated neighborhoods in human ESCs.

D. Evidence for insulation of insulated neighborhoods: Deletion of insulated neighborhood anchors leads to gene misregulation.

E. Evidence for insulation of insulated neighborhoods: Mutations of insulated neighborhood anchors in tumor cells leads to oncogene activation.

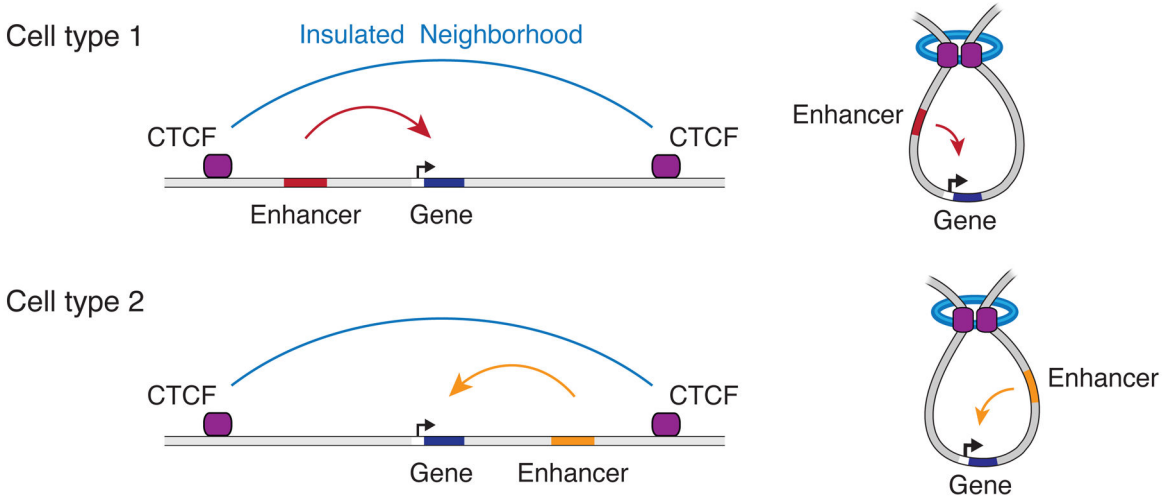

Figure 3. Insulated neighborhoods in development

Cell-specific enhancer-gene interactions occur within insulated neighborhoods that are generally maintained in different cell types. Left side displays a linear model of a genomic region encompassing a gene associated with cell type-specific enhancers, the right side displays the insulated neighborhood model of the locus.

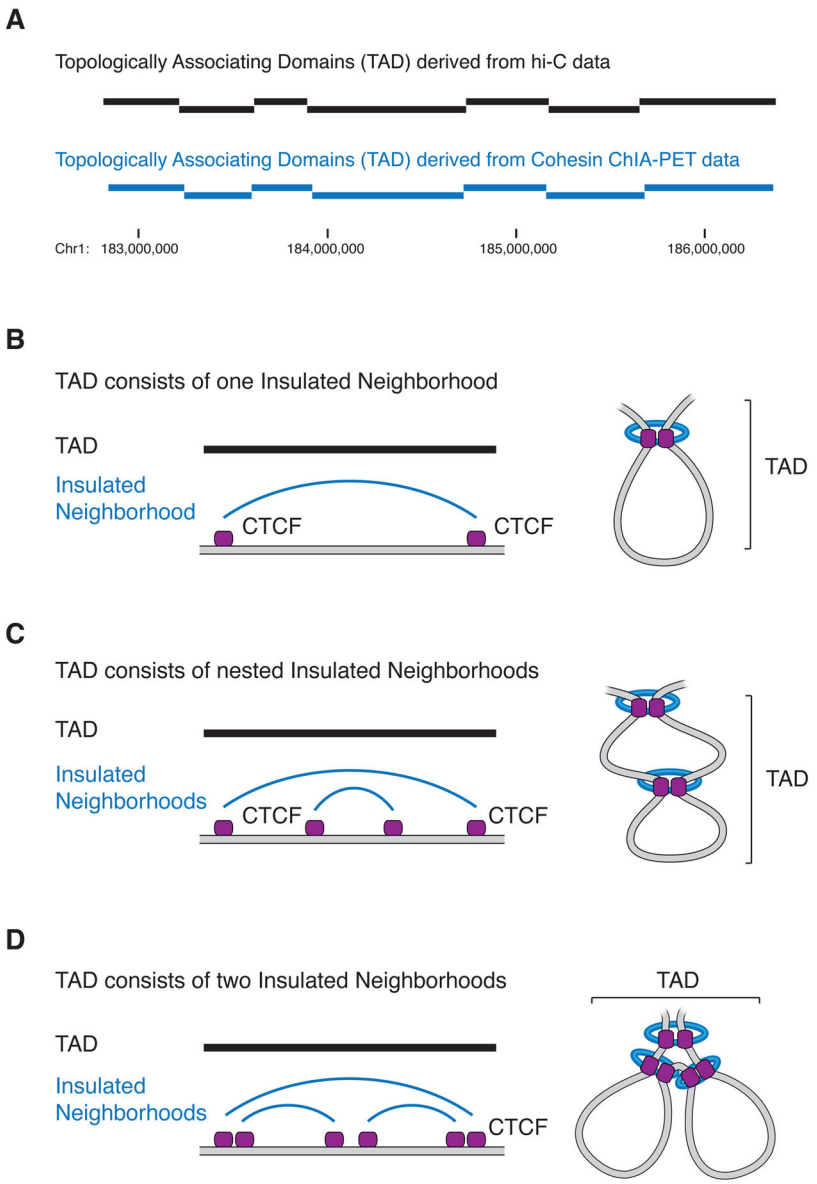

Figure 4. Insulated neighborhoods are the mechanistic basis of TADs

A. Hi-C and cohesin ChIA-PET identify similar Topologically Associating Domains (TAD). Bars indicate the TADs identified using Hi-C and ChIA-PET in human ESCs at the genomic region whose co-ordinates are indicated in the bottom.

B. Model of a TAD that consists of an insulated neighborhood

C. Model of a TAD that consists of nested insulated neighborhoods

D. Model of a TAD that consists of two insulated neighborhoods, nested within a TAD-spanning CTCF-CTCF loop

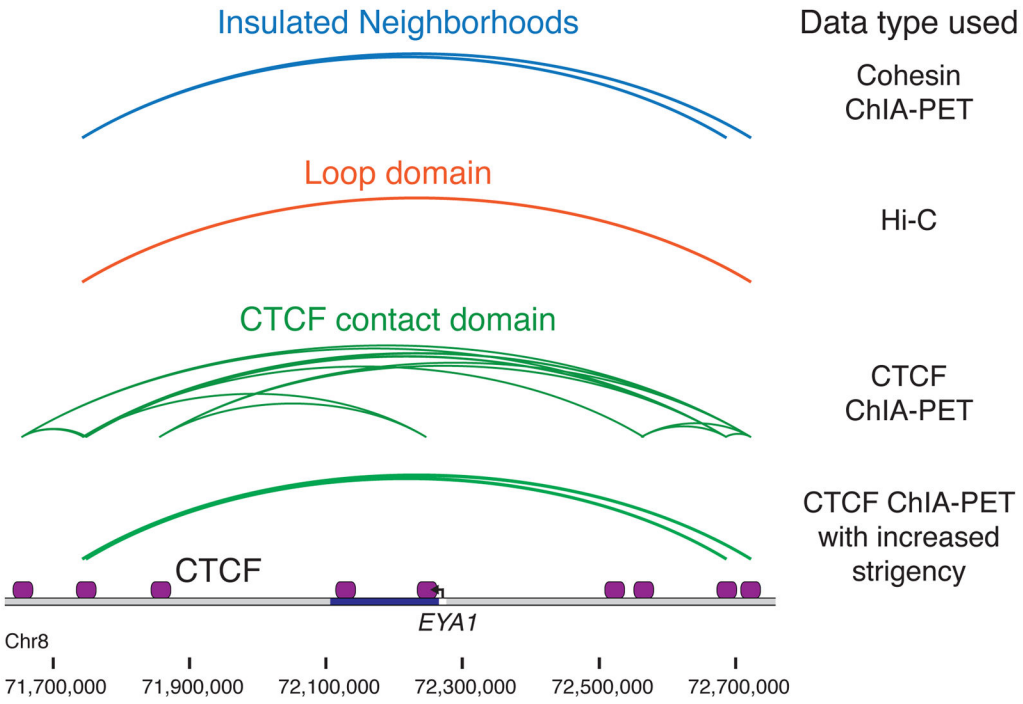

Figure 5. Relationships between insulated neighborhoods and other DNA loop models

DNA loops at the *EYA1* genomic locus generated using three different types of chromatin contact data in lymphoblastoid cells. Displayed are the cohesin ChIA-PET interactions (Heidari et al., 2014) used to identify insulated neighborhoods, Hi-C data (Rao et al., 2014) used to identify "Loop Domains", and CTCF ChIA-PET data (Tang et al., 2015) used to identify "CTCF contact domains". Increased stringency filtering of the CTCF ChIA-PET data reveals a chromosome structure similar to the insulated neighborhoods and Loop Domain. The coordinates of the genomic region are displayed at the bottom.

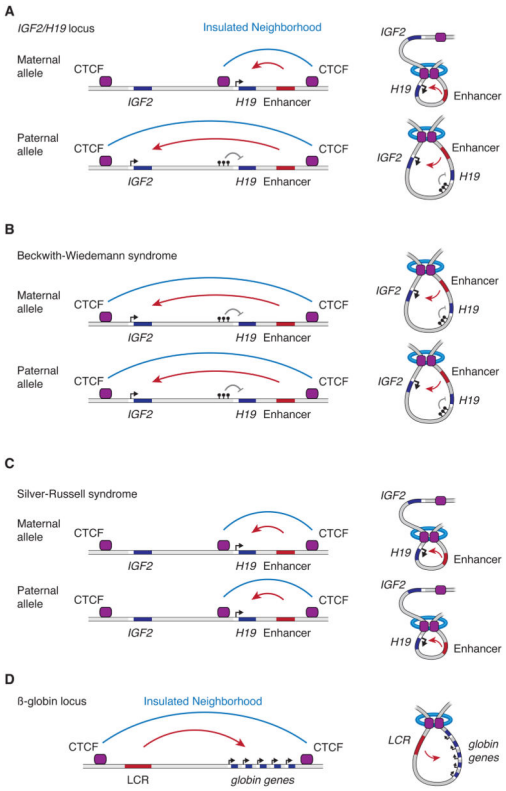

Figure 6. Insulated neighborhoods at the *IGF2/H19* and ß-globin locus

A. Insulated neighborhood model at the maternal and paternal alleles of the imprinted *IGF2/H19* locus. On the maternal allele, CTCF binding at the imprint control region upstream of the *H19* gene creates an insulated neighborhood around *H19* and an enhancer, which prevents the enhancer from activating the *IGF2* gene. On the paternal allele the imprint control region is methylated which leads to repression of the *H19* gene, and prevention of CTCF binding. On this allele, a large insulated neighborhood is formed allowing the downstream enhancer to activate the *IGF2* gene. Black lollipops indicate DNA methylation. The insulated neighborhood models are displayed on the right.

B. Lack of methylation at the imprint control region upstream of *H19* and the presence of the large insulated neighborhood on the maternal *IGF2/H19* allele occur in patients with Beckwith-Wiedemann syndrome.

C. Methylation at the imprint control region upstream of *H19* and the presence of the small insulated neighborhood on the paternal *IGF2/H19* allele occurs in patients with Silver-Russell syndrome.

D. Insulated neighborhood model at the β-globin locus containing a cluster of globin genes and an upstream locus control region (LCR).

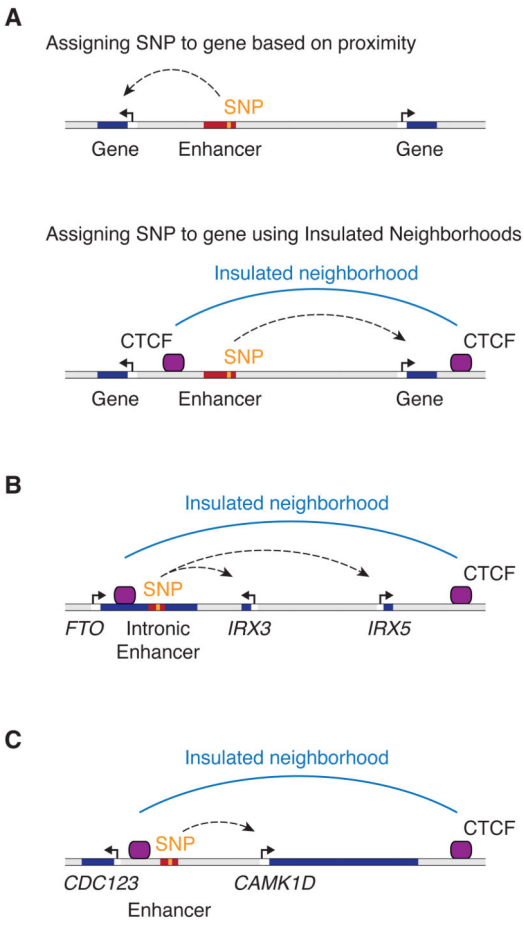

Figure 7. Insulated neighborhoods as a method to identify target genes of disease-associated enhancer variation

A. (Top) Assignment of an enhancer-associated Single Nucleotide Polymorphism (SNP) to a gene based on linear proximity. (Bottom) Assignment of a SNP to a gene based on the insulated neighborhood model.

B. Model of the insulated neighborhood organization at the *FTO-IRX3-IRX5* locus.

C. Model of the insulated neighborhood organization at the *CDC123-CAMK1D* locus.

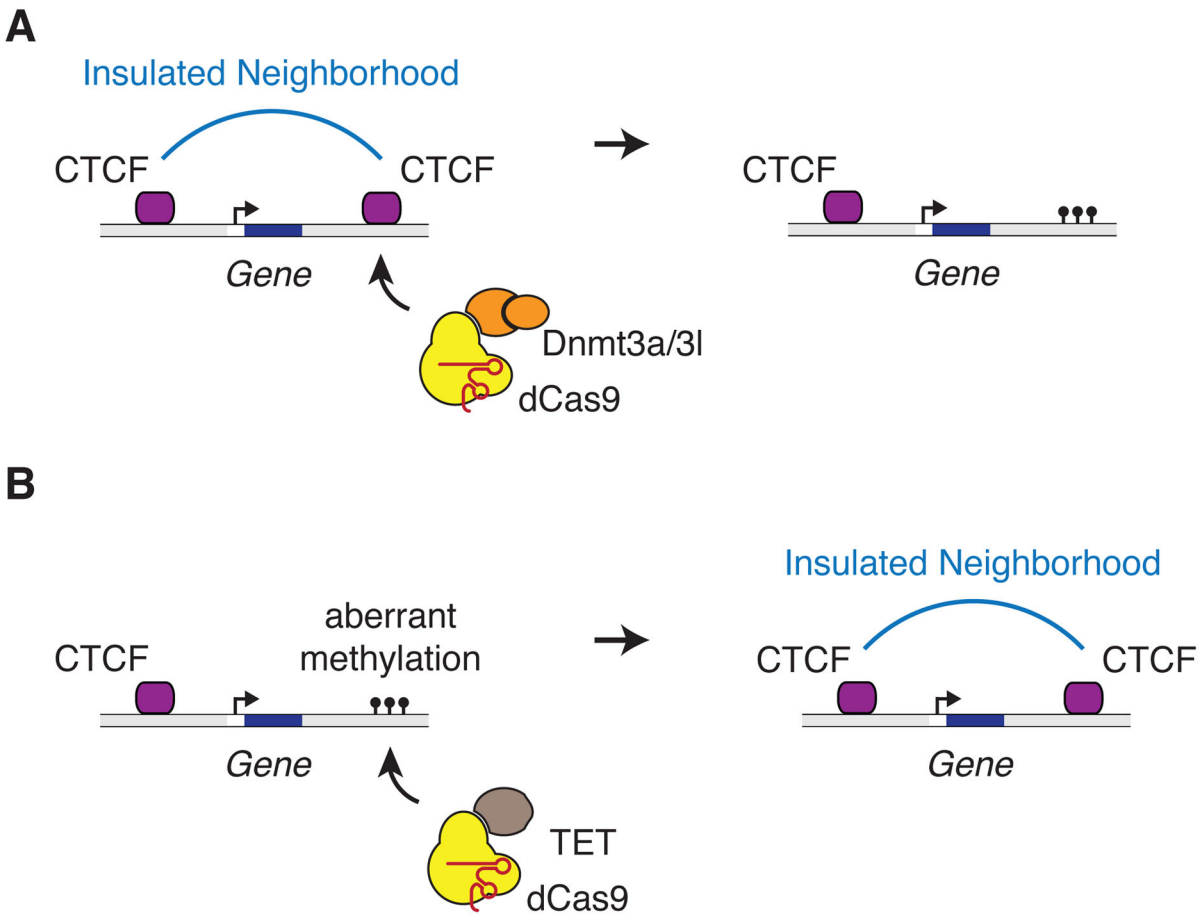

Figure 8. Neighborhood perturbation and repair through site-specific DNA methylation

A. Targeting a dCas9-DNA-methyltransferase 3a/3l (Dnmt3a/3l) fusion to an insulated neighborhood anchor leads to DNA methylation, abrogation of CTCF binding and loss of neighborhood integrity. Black lollipops indicate DNA methylation.

B. Targeting a dCas9-TET (Ten-eleven translocation) fusion to an aberrantly methylated insulated neighborhood anchor leads to DNA de-methylation, and restoration of neighborhood integrity.