Regulation and Dysregulation of Chromosome Structure in Cancer

Regulation and dysregulation of chromosome structure in cancer

Denes Hnisz[1]*, Jurian Schuijers[1], Charles H. Li[1,2], Richard A. Young[1,2]*

[1] Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142, USA

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

* Corresponding authors

Corresponding Authors:

Denes Hnisz
Whitehead Institute for Biomedical Research
455 Main Street
Cambridge, MA 02142
Tel: (617) 258-7181
Fax: (617) 258-0376
hnisz@wi.mit.edu

Richard A. Young
Whitehead Institute for Biomedical Research
455 Main Street
Cambridge, MA 02142
Tel: (617) 258-5218
Fax: (617) 258-0376
young@wi.mit.edu

Summary

Cancer arises from genetic alterations that produce dysregulated gene expression programs. Normal gene regulation occurs in the context of chromosome loop structures called insulated neighborhoods, and recent studies have shown that these structures are altered and can contribute to oncogene dysregulation in various cancer cells. We review here the types of genetic and epigenetic alterations that influence neighborhood structures and contribute to gene dysregulation in cancer, present models for insulated neighborhoods associated with the most prominent human oncogenes, and discuss how such models may lead to further advances in cancer diagnosis and therapy.

Introduction

The idea that structural alterations of chromosomes may cause disease is nearly as old as the chromosome theory of inheritance (Boveri, 1914). The first discovery of a chromosomal translocation, the Philadelphia chromosome, in the blood cells of a leukemia patient (Nowell and Hungerford, 1960), stimulated further study of the potential roles of chromosome structural alterations in the neoplastic state of cancer cells. Such studies revealed that structural alterations of chromosomes often contribute to dysregulation of cellular gene expression programs in cancer cells (Rabbitts, 1994; Vogelstein and Kinzler, 2004). More recently, chromosome conformation capture technologies, which detect DNA interactions genome-wide, have led to important new insights into the roles that chromosome structures play in normal gene control and have revealed how various alterations in chromosome structure contribute to gene dysregulation in disease (Bickmore and van Steensel, 2013; Bonev and Cavalli, 2016; Corces and Corces, 2016; de Laat and Duboule, 2013; Dixon et al., 2016; Gibcus and Dekker, 2013; Gorkin et al., 2014; Groschel et al., 2014; Hnisz et al., 2016a; Krijger and de Laat, 2016; Lupianez et al., 2015; Merkenschlager and Nora, 2016; Ong and Corces, 2014; Phillips-Cremins and Corces, 2013; Valton and Dekker, 2016).

Recent studies have revealed that interphase chromosomes are organized into thousands of DNA loops, which are anchored, in part, through the interactions of CTCF proteins that bind to two separate sites in DNA and also bind one another, and these CTCF-CTCF interactions are reinforced by the cohesin complex. These loops generally contain one or more genes together with the regulatory elements that operate on the genes. The loop anchors constrain the regulatory elements to act predominantly on genes within the loop. In this manner, the anchors insulate genes and their regulatory elements from other regulatory elements located outside the neighborhood, and thus the CTCF-CTCF anchored loop structures have been called “insulated neighborhoods”.

Here we review recent evidence that genetic and epigenetic alterations can disrupt insulated neighborhoods in cancer cells and thereby contribute to the transformed phenotype. We present models for insulated neighborhoods associated with prominent human oncogenes, and identify neighborhoods that are altered based on cancer genome sequence data. Finally, we discuss how knowledge of insulated neighborhoods may lead to further advances in cancer diagnosis and therapy.

Chromosome structures

Interphase chromosomes are organized in a hierarchy of structures, and these can play important roles in transcriptional regulation (Figure 1). Detailed descriptions of the

various layers of genome organization and the history of this field can be found in other excellent reviews (Bickmore and van Steensel, 2013; Cavalli and Misteli, 2013; de Laat and Duboule, 2013; Dekker and Heard, 2015; Dixon et al., 2016; Gibcus and Dekker, 2013; Gorkin et al., 2014; Merkenschlager and Nora, 2016). We provide here a brief description of the layers of chromosome structural organization as background to the recent concept that chromosome loops are a structural and functional unit of gene control in mammalian cells.

Chromosomes in interphase nuclei tend not intermingle, but occupy distinct regions within the nuclear space (Figure 1). In situ hybridization and microscopy techniques revealed that these "chromosome territories" are a general feature in mammalian nuclei and that the territorial organization of chromosomes is maintained through cell division, although the positions of chromosome territories can be reshuffled (Cremer and Cremer, 2010). At present, the mechanisms that maintain chromosome territories are unknown.

Chromosome conformation capture technologies initially revealed that interphase chromosomes are partitioned into megabase sized folding entities that were termed "topologically associating domains" (TADs) (Figure 1) (Dixon et al., 2012; Nora et al., 2012). TADs are regions of DNA that show high frequency interactions relative to regions outside the TAD boundaries. Early studies reported about 2,000 TADs, which tend to have similar boundaries in all human cell types and contain on average 8 genes whose expression is weakly correlated (Dixon et al., 2015; Dixon et al., 2012). TADs were postulated to help constrain interactions between genes and their regulatory sequences (Dixon et al., 2012). The initial studies produced data at approximately 40kb resolution, which was not sufficient to determine the mechanistic basis of TAD formation and maintenance, although an abundance of CTCF bound sites was noted at TAD boundaries (Dixon et al., 2012).

Insights into the relationship between chromosome structure and gene regulation have emerged from studies that focused on the roles of chromosome-structuring proteins in DNA interactions, and used chromatin contact mapping technologies that provided a high resolution view of DNA contacts associated with those proteins (Figure 1) (DeMare et al., 2013; Dowen et al., 2014; Handoko et al., 2011; Ji et al., 2016; Phillips-Cremins et al., 2013; Splinter et al., 2006; Tang et al., 2015; Tolhuis et al., 2002). These studies showed that chromosomes are organized into thousands of DNA loops, formed by the interaction of DNA sites bound by the CTCF protein and occupied by the cohesin complex. The anchors of these CTCF-CTCF loops function to insulate enhancers and genes within the loop from enhancers and genes outside the loop. These CTCF-CTCF loops have thus been called insulated neighborhoods, but they have also been called sub-TADs, loop domains, and CTCF-contact domains (Dowen et al., 2014; Phillips-Cremins et al., 2013; Rao et al., 2014; Tang et al., 2015). For the purposes of this review, we will use the term "insulated neighborhoods" to describe these loops.

The Insulated Neighborhood model

Insulated neighborhoods are formed by an interaction between two DNA sites bound by the transcription factor CTCF and the cohesin complex (Figure 1) (Hnisz et al., 2016a). In human embryonic stem cells (ESCs), there are at least 13,000 insulated neighborhoods, which range from 25 kb to 940 kb in size and contain from 1–10 genes (Ji et al., 2016). The median insulated neighborhood is ~190kb and contains three genes. These numbers can vary depending on assumptions made when filtering

genomic data, but provide an initial description of genomic loops that is useful for further analysis.

Several lines of evidence argue that the CTCF-bound anchor sites of insulated neighborhoods insulate genes and regulatory elements within a neighborhood from those outside the neighborhood. Genome-wide analysis indicates that the majority (>90%) of enhancer-gene interactions occur within insulated neighborhoods (Dowen et al., 2014; Hnisz et al., 2016b; Ji et al., 2016). Perturbation of insulated neighborhood anchor sequences leads to changes in gene expression in the vicinity of the altered neighborhood boundary (Dowen et al., 2014; Ji et al., 2016; Narendra et al., 2015; Sanborn et al., 2015). Insulated neighborhood boundary elements are coincident with the endpoints of chromatin marks that spread over regions of transcriptional activity or repression (Dowen et al., 2014). These lines of evidence indicate that the insulating function of the neighborhood loop anchors contributes to normal gene regulation.

Insulated neighborhoods, and the CTCF-CTCF loops that form them, are largely maintained during development, and the subset of CTCF sites that form neighborhood loop anchors show little genetic variation in the germ-line (Hnisz et al., 2016a). However, allele-specific CTCF binding contributes to the formation of allele-specific insulated neighborhoods at imprinted genes (Hnisz et al., 2016a), and cell type-specific CTCF binding and neighborhoods appear to make some contribution to cell-specific transcriptional programs (Bunting et al., 2016; Narendra et al., 2015; Splinter et al., 2006; Tolhuis et al., 2002; Wang et al., 2012).

Although the descriptions of CTCF-CTCF loops and TADs to this point may imply to the reader that these are static structures, several lines of evidence suggest that they are dynamic. Both CTCF and cohesin dynamically interact with DNA, and as described below, their binding is influenced by a variety of different factors and post-translational modifications. Modeling studies suggest that chromatin contact mapping data represent an assembly of configurations that can differ between individual cells in the cell population, between time points within the same cell, and between alleles of a locus within the same cell (Figure 2) (Fudenberg et al., 2016; Giorgetti et al., 2014; Imakaev et al., 2012; Naumova et al., 2013). Consequently, the loop models displayed in this review and in other reports represent the predominant configurations deduced from cell population data or, in some cases, a combination of configurations that are inferred from the data.

Insulated neighborhoods cover the majority of the genome, and thus genes that play prominent rules in cancer biology are typically found within insulated neighborhoods. These genes include, but are not limited to, *KRAS*, *NRAS*, and *BRAF*, which are members of the RAS and RAF pathway (Figure 3A-C) (Bos, 1989; Downward, 2003); *MYC*, the most frequently overexpressed and amplified human oncogene (Figure 3D) (Beroukhim et al., 2010); *TP53,* which encodes the P53 protein and is the most frequently mutated gene in all cancers (Figure 3E) (Lawrence et al., 2014); EGFR , which encodes the epidermal growth factor receptor, a major drug target (Figure 3F) (Lynch et al., 2004); CD274, or Programmed death-ligand 1 (PD-L1), and the gene encoding its receptor PDCD1, which are immune checkpoint targets for cancer immunotherapy (Figure 3G-H) (Hamid et al., 2013; Pardoll, 2012). More detailed information on the structures of these loci is provided in Supplementary Figure 1. These models rely on data from a cell line, but provide the reader with one view of the structural features of these loci and a potential foundation for further exploration of these structures in primary cells of various cancer types.

Regulators of Insulated Neighborhood structure

The proteins that are best understood to contribute to insulated neighborhood anchor structures are CTCF and cohesin, as discussed in more detail below. There are additional factors that have been implicated in establishing, maintaining or modifying insulated neighborhood anchor structures (Figure 4). These include Structural Maintenance of Chromosomes (SMC) proteins such as condensin II, the CTCF-like protein BORIS, Poly (ADP-Ribose) polymerase (PARP), DNA methylation, noncoding RNA species, and the process of transcription by RNA Polymerase II.

CTCF

CTCF is a Zinc-finger transcription factor that was originally identified as a repressor of the c-MYC oncogene (Baniahmad et al., 1990; Lobanenkov et al., 1990). CTCF is conserved in eukaryotes from *Drosophila* to *Homo sapiens*, is essential for embryonic development in mammals, and is ubiquitously expressed in all cells (Ghirlando and Felsenfeld, 2016). CTCF has long been described as a component of insulators, which are DNA elements that can block the ability of enhancers to activate genes when placed between them (Bell et al., 1999). Several recent reviews provide more detailed information and historical perspective on CTCF (Ghirlando and Felsenfeld, 2016; Merkenschlager and Odom, 2013; Ong and Corces, 2014; Phillips and Corces, 2009).

Several lines of evidence suggest that CTCF contributes to the formation and maintenance of chromosome structures such as TADs and the insulated neighborhoods that comprise TADs. The majority of the boundary regions of topologically associating domains (TADs) are bound by CTCF (Dixon et al., 2015; Dixon et al., 2012; Nora et al., 2012), and global depletion of CTCF perturbs the insulating properties of TADs (Zuin et al., 2014). The CTCF protein is able to form homodimers and thus physical interactions between two CTCF molecules bound at two genomic locations can participate in the formation of DNA loops (Hou et al., 2008; Palstra et al., 2003; Splinter et al., 2006; Yusufzai et al., 2004).

Chromatin immunoprecipitation and sequencing (ChIP-Seq) experiments indicate that approximately 50,000-80,000 sites are bound by CTCF in mammalian genomes (Kim et al., 2007). However, functional assays of insulator function found that only a minority of these sites act as insulators (Liu et al., 2015) or participate in formation of insulated neighborhood boundaries (Ji et al., 2016). It is possible that two CTCF sites need to be in a specific orientation in order for the CTCF proteins to interact and have insulating function (Dekker and Mirny, 2016; Fudenberg et al., 2016; Sanborn et al., 2015).

The ability of CTCF to bind its DNA sequence motif and participate in insulator function is influenced by DNA methylation and protein modification (Figure 4A). CTCF binds to hypomethylated regions of the genome (Mukhopadhyay et al., 2004) and mechanistic studies of the H19/IGF2A imprinted locus revealed that methylation of DNA is sufficient to prevent CTCF binding to the methylated allele (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000; Szabo et al., 2000). CTCF can be poly(ADP-ribosyl)ated (PARylated), and at the imprinted *H19/IGF2A* locus, PARylation of CTCF regulates its insulator function (Figure 4A), which is associated with its ability to form DNA loops at the locus (Yu et al., 2004). Studies in *Drosophila* have identified additional proteins that interact with CTCF, including DNA helicases, nucleophosmin and topoisomerase (Phillips-Cremins and Corces, 2013), but whether such proteins associate with CTCF in human cells and modulate its function remains to be investigated.

Transcription by RNA polymerase II has been reported to evict CTCF from specific sites (Lefevre et al., 2008) and various RNA species can enhance or reduce CTCF binding at specific loci. The Tsix, Xite, and Xist RNAs produced during X chromosome inactivation can recruit CTCF to the X-inactivation center (Kung et al., 2015), whereas the Jpx RNA evicts CTCF from the Xist promoter (Sun et al., 2013)(Figure 4A). An antisense transcript (*Wrap53*) produced at the *TP53* locus was found to contribute to CTCF binding (Saldana-Meyer et al., 2014) (Figure 4A).

The CTCF gene has an ortholog in mammals called CTCFL or BORIS, which may also participate in DNA loops. While CTCF is ubiquitously expressed in all cell types, the expression of BORIS is thought to be restricted to male germ cells (Loukinov et al., 2002). BORIS appears to bind the same DNA sequence as CTCF and its expression is mutually exclusive with CTCF during germ cell development (Loukinov et al., 2002).

Cohesin

Cohesin is a multiprotein complex that belongs to the family of Structural Maintenance of Chromosome (SMC) family of proteins (Figure 4B), whose members are conserved both in prokaryotes and eukaryotes (Nasmyth and Haering, 2009). Cohesin consists of a tripartite ring of three subunits - SMC1, SMC3 and RAD21 - which in human cells is bound by accessory factors that include STAG1 or STAG2. Cohesin was initially studied for its role in sister chromatid cohesion, and later found to play important roles in gene regulation (Dorsett and Merkenschlager, 2013; Hirano, 2006; Merkenschlager and Odom, 2013; Nasmyth and Haering, 2009; Uhlmann, 2016).

Cohesin forms a ring whose internal dimensions are sufficient to entrap two DNA molecules, which provides a model to explain how it contributes to DNA loops, but it is also possible that two connected cohesin rings function in DNA loop formation (Nasmyth and Haering, 2009). Cohesin is loaded onto DNA by the SMC-loading factor NIPBL (Figure 4C), which is associated with the Mediator cofactor, which mediates interactions between enhancers and promoters at active genes (Kagey et al., 2010). Disruption of cohesin perturbs enhancer-promoter interactions and gene expression (Kagey et al., 2010; Seitan et al., 2011; Zuin et al., 2014).

Cohesin is also associated with CTCF-bound sites and contributes to insulation when two CTCF-bound sites interact to form the anchors of a DNA loop (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008). The SMC-loading factor NIPBL is not found at CTCF sites, so it is possible that cohesin is loaded at transcriptionally active sites and then migrates to CTCF bound sites, where further movement is inhibited. The STAG1/2 subunits of cohesin can engage in direct physical interaction with CTCF (Xiao et al., 2011), which may contribute to stable CTCF-cohesin association. DNA loop extrusion models have been proposed to account for the formation of DNA loops; these models posit that where cohesin is initially loaded, extrusion of DNA through the cohesin ring (or multiple connected cohesin rings) would drive cohesin migration to two CTCF-bound sites where, if the sites were properly oriented for CTCF-CTCF interaction, the DNA loop would be anchored (Dekker and Mirny, 2016; Fudenberg et al., 2016; Sanborn et al., 2015).

The regulation of cohesin has been studied primarily in the context of its role in sister chromatid cohesion, but these regulatory features may also contribute to cohesin regulation in enhancer-promoter and CTCF-CTCF interactions. For example, the SMC3 subunit of cohesin can be acetylated by the ESCO family of acetyltransferases and deacetylated by HDAC8, and the acetylation is important for normal retention of cohesin on DNA and sister chromatid cohesion (Figure 4C) (Deardorff et al., 2012). Furthermore,

cohesin is removed from chromatin in the mitotic prophase by the unloading factor Wapl, and depletion of Wapl leads to gross chromosome organization defects in interphase nuclei (Tedeschi et al., 2013).

Additional SMC complexes have been implicated in the control of chromosome organization. Vertebrate cells have two condensin complexes (Figure 4B). Although condensin I is excluded from interphase nuclei, condensin II is loaded, like cohesin, onto interphase chromatin by Nipbl at active enhancer-promoter interactions (Dowen et al., 2013). The contributions of condensin II to gene regulation, DNA looping, and larger chromosome structures are not yet understood.

Mutations in structuring components and neighborhood boundaries in cancer

Translocations of large portions of chromosome arms have been described for decades in tumor cells, but only recently were mutations in chromosome structure regulators and their binding sites described and appreciated for their potential impact on specific chromosome structures. In this section we describe the spectrum of mutations that have been described that impact neighborhood regulators and neighborhood boundary sites in tumor genomes, and review evidence suggesting that these mutations contribute to tumor development (Table 1, Supplementary Table 1).

CTCF mutations

Mutations in the CTCF gene have been reported in breast cancer, endometrial cancer (Lawrence et al., 2014; Walker et al., 2015), prostate cancer (Filippova et al., 1998), Wilms' tumor (Filippova et al., 2002), and head and neck carcinomas (Lawrence et al., 2014). These mutations are predominantly missense or nonsense and thus predicted to impair CTCF function (Lawrence et al., 2014). Some tumor cell mutations occur within the Zinc fingers of CTCF and may selectively perturb certain neighborhoods because they affect CTCF binding at only a subset of sites (Filippova et al., 2002). Loss of a CTCF allele can occur in some tumor types, suggesting that CTCF may act as a haplo-insufficient tumor suppressor (Filippova et al., 1998). Consistent with this notion, mice heterozygous for the CTCF gene display an increased susceptibility to develop tumors in various radiation and chemical- based cancer induction models (Kemp et al., 2014). Dysregulated expression of the germ line specific CTCF ortholog BORIS has been reported in several cancer types (Simpson et al., 2005), but it is not yet clear that this contributes to tumorigenesis.

Cohesin mutations

Mutations in the cohesin complex occur in acute myeloid leukemia (AML) (Cancer Genome Atlas Research et al., 2013), myeloid dysplastic syndrome (MDS) (Kon et al., 2013), bladder cancer (Guo et al., 2013), breast cancer (Stephens et al., 2012), colorectal cancer (Barber et al., 2008) and Ewing sarcoma (Crompton et al., 2014), among others (Table 1, Supplementary Table 1). In AML, mutations in all four cohesin subunits (SMC1A, SMC23, RAD21 and STAG2) have been reported, whereas in solid tumors mutations of the STAG2 subunit occur most frequently. The majority of mutations in the cohesin subunits are missense, nonsense or truncating (Lawrence et al., 2014), suggesting a loss-of-function effect, which is consistent with the reduced level of DNA-bound cohesin reported in cohesin-mutant AML cells (Kon et al., 2013).

Recent studies indicate that the tumor-promoting effect of at least a subset of cohesin alterations are linked to its roles in gene regulation and chromosome structure rather

than its roles in proper chromosome segregation. The classic role of cohesin in sister chromatid cohesion would predict that cohesin mutations in cancer contribute to the neoplastic state through defects in chromosome segregation and consequent aneuploidy. However, modeling of SMC3 mutations that occur in AML has revealed no association with chromosome segregation defects and aneuploidy, but rather with alteration of the gene expression program of the leukemia cells (Mazumdar et al., 2015; Mullenders et al., 2015; Viny et al., 2015). Furthermore, analysis of STAG2 mutant bladder cancer did not reveal any association of the STAG2 mutation with chromosome segregation defects and aneuploidy (Balbas-Martinez et al., 2013).

CTCF binding site mutations

Nucleotide substitutions in the DNA binding site of CTCF occur in the genomes of several cancer types (Supplementary Table 1), and such substitutions appear to be especially enriched in CTCF binding sites that form insulated neighborhood boundaries. Nucleotide substitutions in CTCF binding sites have been reported in colorectal cancer (Katainen et al., 2015), gastrointestinal cancer (Umer et al., 2016), esophageal cancer (Hnisz et al., 2016b), liver cancer (Hnisz et al., 2016b; Katainen et al., 2015), and melanoma (Poulos et al., 2016). Although the functional impact of these mutations has not been investigated in depth, the observation that insulated neighborhood boundary CTCF sites are conserved and show limited germ line variation (Ji et al., 2016), together with evidence that some of these mutations are recurrent, suggests that many of the somatic CTCF site mutations in tumor cells contribute to the neoplastic state by perturbing insulated neighborhoods.

Epigenetic alteration of CTCF binding

Because DNA hypermethylation is a feature of many cancer types and DNA methylation reduces CTCF binding, insulated neighborhood structures may be compromised in cells with hypermethylated DNA (Supplementary Table 1). Indeed, in a subset of gliomas that harbor mutations in the *IDH1* gene, tumor-specific hypermethylation is associated with the disruption of CTCF binding, alteration of chromosome structure, and dysregulation of oncogene expression (Flavahan et al., 2016).

Mutations in regulators of CTCF and cohesin

Several regulators of CTCF and cohesin have been recently implicated in cancer, and it is possible that mutations of some of these regulators contribute to the neoplastic state through alteration of insulated neighborhoods. For example, nucleophosmin, a direct physical interaction partner of CTCF is frequently mutated in AML (Cancer Genome Atlas Research et al., 2013). Many non-coding RNA species are implicated in cancer development (Lin and He, 2017), and CTCF interacts with thousands of RNAs, some of which impact its binding to DNA (Kung et al., 2015), so it is plausible that dysregulation of non-coding RNAs in tumor cells contributes to alterations of chromosome structures. ESCO1, the enzyme that acetylates cohesin is mutated in a subset of endometrial cancers (Price et al., 2014). In summary, defects in a diverse set of mechanisms may contribute to alterations of chromosome structure in cancer cells.

Impacts of neighborhood alterations in cancer

The structural and functional impact of mutations in chromosome structuring components and in neighborhood boundaries is only beginning to be studied in cancer, but it is useful nonetheless to consider models that explain how these mutations can

contribute to gene dysregulation in tumor cells (Figure 5). In some instances, these models are supported by experimental data and in others, they are predictive and await further study.

About half of T cell acute lymphoblastic leukemias contain mutations that activate the *TAL1* oncogene (Armstrong and Look, 2005; Van Vlierberghe and Ferrando, 2012). In a subset of these leukemias, the *TAL1* oncogene is activated by microdeletions that remove the boundary of an insulated neighborhood containing the *TAL1* gene. The disruption of the boundary leads to inappropriate contacts between the *TAL1* gene and regulatory elements normally located outside the *TAL1* neighborhood (Figure 5A) (Hnisz et al., 2016b). Similar microdeletions that disrupt an insulated neighborhood boundary also occur encompassing the *LMO2* oncogene in these T cell leukemias (Hnisz et al., 2016b).

Epigenetic alteration of CTCF binding sites at insulated neighborhood anchors can also lead to oncogene activation. A subset of gliomas harbor a mutation in the *IDH1* gene, and this mutation is associated with DNA hypermethylation (Cancer Genome Atlas Research et al., 2015; Dang et al., 2009). A recent study found that in *IDH1* mutant gliomas, methylation of an insulated neighborhood boundary encompassing the *PDGFRA* oncogene leads to a loss of the insulating property of the neighborhood, inappropriate contacts between *PDGFRA* and an upstream enhancer normally located outside the *PDGFRA* neighborhood, and elevated expression of *PDGFRA* (Figure 5A) (Flavahan et al., 2016).

Our understanding of insulated neighborhoods in normal gene control suggest additional models for the impact of genetic or epigenetic alterations of neighborhood structures in gene dysregulation in neoplastic cells; these have yet to be reported in cancer cells and thus serve merely as predictions. For example, alterations of neighborhood boundaries or their components may lead to the activation of silent oncogenes by enabling enhancers within the neighborhood to activate genes that are normally located outside the neighborhood (Figure 5B). Genetic or epigenetic perturbation of insulated neighborhood boundaries can lead to a down-regulation of genes found in neighborhoods (Figure 5C) (Dowen et al., 2014), and thus potentially loss of expression of a tumor suppressor. A minority of enhancers and promoters are bound by CTCF, and contacts between such enhancer-promoter pairs can be facilitated by CTCF-CTCF interactions (Guo et al., 2015), so disruption of these interactions can contribute to gene dysregulation (Figure 5D). Because the process of transcription and the presence of RNA species can affect CTCF binding, altered transcription in the cancer state may also be responsible for changes in neighborhood structure and function.

Future challenges: chromosome structures in cancer diagnostics and therapy

Genetic and epigenetic alterations of insulated neighborhoods can lead to dysregulation of prominent oncogenes that drive tumorigenesis. These new insights suggest new approaches to identify mechanisms associated with gene dysregulation in cancer and new approaches to target dysregulated expression of oncogenes and tumor suppressors.

Neighborhood alteration in cancer: identification of oncogenes and dependencies

Somatic mutations and epigenetic alterations that perturb insulated neighborhood boundaries may be useful to identify oncogenes and dependencies in cancers whose development and progression is not well understood. The finding that insulated

neighborhood boundary alterations occur at oncogene loci in leukemia and gliomas (Flavahan et al., 2016; Hnisz et al., 2013) suggests that neighborhoods with recurrently altered boundaries can identify new oncogenic drivers.

Cancer cells can become highly dependent on certain gene products during cancer progression. The disrupted neighborhood around *PDGFRA* in *IDH1* mutant glioma cells is associated with the sensitivity of these cells to PDGFRA inhibitors (Flavahan et al., 2016). This suggests that cancer dependencies engendered by similar mechanisms will be revealed through investigation of neighborhood alterations in cancer cells. Progress here will require improved understanding of the mutational landscape of non-coding regions of the genome where most neighborhood boundaries are located, and the epigenetic mechanisms that impact neighborhood boundary function.

Cancer susceptibility

Although cancer development occurs as a consequence of somatic alterations of the genome, DNA variants in the germ line contribute to the susceptibility of cancer development. Recent evidence that DNA polymorphisms in non-coding DNA are linked to cancer predisposition (Oldridge et al., 2015), that rare germ line variants occur in insulated neighborhood CTCF sites (Ji et al., 2016), and that such variants can impact neighborhoods (Tang et al., 2015) indicate that some of the genetic variation that contributes to cancer susceptibility may occur in insulated neighborhoods.

Epigenetic editing of CTCF anchors

Targeted disruption of CTCF binding and neighborhood integrity, with predictable effects on gene dysregulation, has been demonstrated through targeted methylation with a dCas9-DNA-metyltransferase-3 fusion protein (Liu et al., 2016). Targeted demethylation with a dCas9-TET fusion protein reversed this effect, allowing CTCF binding and insulated neighborhood formation (Liu et al., 2016). This suggests that targeted methylation and demethylation of CTCF binding sites could be used to alter CTCF-CTCF loops that form either insulated neighborhoods or enhancer-promoter interactions. Such epigenetic editing tools might evolve to be useful for therapeutic purposes in cancer and in other diseases where gene dysregulation is key to the disease state.

Acknowledgements

Supported by NIH Grant HG002668. D.H. was supported by an Erwin Schrödinger Fellowship (J3490) from the Austrian Science Fund (FWF), and a Margaret and Herman Sokol Postdoctoral Award. J.S. was supported by a Rubicon Fellowship for the Life Sciences from the Netherlands Organization for Scientific Research (NWO). R.A.Y. is a founder of Syros Pharmaceuticals and Marauder Therapeutics.

Figure legends

Figure 1. Chromosome structures

Hierarchy of chromosome structures: chromosome territories, Topologically Associating Domains (TADs), and Insulated Neighborhoods. The experimental methods typically used to identify these structures are listed on the right side.

Figure 2. Insulated neighborhood models

Dynamics and heterogeneity of insulated neighborhoods inferred from chromatin contact data. Displayed are schematic representations of the experimental data, and models of their interpretation. Insulated neighborhoods are thought to be dynamic, and alternative neighborhood configurations indicated by the experimental data may occur in different cells of the population or within the same cell at different times or on different alleles within the same cell.

Figure 3. Insulated neighborhoods containing genes with prominent roles in cancer

(A-H) Models of Insulated Neighborhoods identified from high confidence interactions detected in CTCF ChIA-PET data in GM12878 cells (Tang et al., 2015). Insulated neighborhoods are depicted as arcs, with those containing the gene of interest in red. The length of the largest such neighborhood is noted. CTCF binding profiles (ChIP-Seq) are displayed in gene tracks below the insulated neighborhood arcs. ChIP-Seq data is from (ENCODE Project Consortium et al., 2012), and read density is measured as reads per million mapped reads. The genes with prominent roles in cancer are depicted as black arrows and identified in black font. Only a subset of neighborhoods at each locus is shown for simplicity; more detailed information can be found in Supplementary Figure 1.

The genomic coordinates (hg19 genome assembly) of the displayed loci are:

(A) KRAS, chr12:23,328,472-26,234,964

(B) NRAS, chr1:114,471,740-116,063,184

(C) BRAF, chr7:140,149,898-141,280,727

(D) MYC, chr8:127,797,231-130,842,492

(E) TP53, chr17:7,398,136-7,751,726

(F) EGFR, chr7:54,800,278-56,193,912

(G) CD274, chr9:4,706,510-5,693,885

(H) PDCD1, chr2:241,704,545-243,199,373

Figure 4. Chromosome structure regulators

(A) Regulatory mechanism and their impact on the chromosome structure regulator CTCF.

(B) Schematic models of the composition of the SMC family members cohesin, condensin I, and condensin II.

(C) Regulatory mechanism and their impact on the chromosome structure regulator cohesin

(D) Model of DNA loop formation by loop extrusion

Figure 5. Insulated neighborhood models for gene regulation and dysregulation in cancer

(A) Disruption of an insulated neighborhood boundary leads to upregulation of gene that was in the neighborhood due to inappropriate contact with an enhancer that was outside the neighborhood.

(B) Disruption of an insulated neighborhood boundary leads to upregulation of gene that was outside the neighborhood due to inappropriate contact with an enhancer that was inside the neighborhood.

(C) Disruption of an insulated neighborhood boundary leads to downregulation of a gene that used to be inside the neighborhood.

(D) Disruption of a CTCF anchor that mediates enhancer-promoter interactions within a neighborhood leads to downregulation of a gene within the neighborhood.

Table 1. Mutations in structuring components and neighborhood boundaries in cancer

Listed are cancer types in which mutations in CTCF, subunits of cohesin, and cohesin regulators have been reported. Only studies that included at least 100 samples, and mutations that reached an at least 3% frequency are displayed. The complete list of mutations is displayed in Supplementary Table 1.

Supplementary Figure 1. Insulated neighborhoods around prominent human oncogenes

Displayed are chromatin interaction data and ChIP-seq, and their annotation in GM12878 cells. From top to bottom: (1) HiC interaction matrices, shown in brown color scale (data from (Rao et al., 2014), visualized with http://higlass.io). (2) Annotation of contact domains, which are higher resolution version of Topologically Associating Domains (TADs) derived from the Hi-C data, shown as black bars below the Hi-C interaction matrices. (3) Annotation of insulated neighborhoods identified using CTCF ChIA-PET data, displayed as thin bars. The neighborhoods that contain the genes highlighted in black are highlighted in green. (4) High confidence CTCF-CTCF interactions identified in CTCF ChIA-PET data (in purple), originally from (Tang et al., 2015). The high confidence interactions annotated as insulated neighborhoods that contain the genes highlighted in black are highlighted in green. The length of the largest such neighborhood is displayed for orientation. (5) CTCF ChIP-seq signal is shown in blue, measured in reads per million (data from (ENCODE Project Consortium et al., 2012)). (6) The orientation of the strongest CTCF motif under each CTCF ChIP-Seq peak is displayed as black bars. “+” indicates that the strongest motif is oriented from left to right on the top (i.e. “+” strand). “-” indicates that the strongest motif is oriented from

right to left on the bottom (i.e. “-” strand), (data from (Hnisz et al., 2016b)). (7) Simplified Refseq gene annotations. The oncogene is shown in black, other genes in grey.

The genomic coordinates (hg19 genome assembly) of the displayed loci are:

(A) KRAS, chr12:23,328,472-26,234,964

(B) NRAS, chr1:114,471,740-116,063,184

(C) BRAF, chr7:140,149,898-141,280,727

(D) MYC, chr8:127,797,231-130,842,492

(E) TP53, chr17:7,398,136-7,751,726

(F) EGFR, chr7:54,800,278-56,193,912

(G) CD274, chr9:4,706,510-5,693,885

(H) PDCD1, chr2:241,704,545-243,199,373

Supplementary Table 1. Mutations in structuring components and neighborhood boundaries in cancer

Listed are cancer types in which mutations in CTCF, subunits of cohesin, and cohesin regulators have been reported.

References

Armstrong, S.A., and Look, A.T. (2005). Molecular genetics of acute lymphoblastic leukemia. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *23*, 6306-6315.

Balbas-Martinez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Marquez, M., Vazquez, M., Lapi, E., Castro-Giner, F., Beltran, S., Bayes, M.*, et al.* (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nature genetics *45*, 1464-1469.

Baniahmad, A., Steiner, C., Kohne, A.C., and Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell *61*, 505-514.

Barber, T.D., McManus, K., Yuen, K.W., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S.*, et al.* (2008). Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America *105*, 3443-3448.

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.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M.*, et al.* (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899-905.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270-1284.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nature reviews Genetics *17*, 661-678.

Bos, J.L. (1989). ras oncogenes in human cancer: a review. Cancer research *49*, 4682-4689.

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren. Verlag von Gustav Fischer *Jena*

Bunting, K.L., Soong, T.D., Singh, R., Jiang, Y., Beguelin, W., Poloway, D.W., Swed, B.L., Hatzi, K., Reisacher, W., Teater, M.*, et al.* (2016). Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. Immunity *45*, 497-512.

Cancer Genome Atlas Research, N., Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M.*, et al.* (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. The New England journal of medicine *372*, 2481-2498.

Cancer Genome Atlas Research, N., Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A., Hoadley, K., Triche, T.J., Jr., Laird, P.W.*, et al.* (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. The New England journal of medicine *368*, 2059-2074.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nature structural & molecular biology *20*, 290-299.

Corces, M.R., and Corces, V.G. (2016). The three-dimensional cancer genome. Curr Opin Genet Dev *36*, 1-7.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harbor perspectives in biology *2*, a003889.

Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S.*, et al.* (2014). The genomic landscape of pediatric Ewing sarcoma. Cancer discovery *4*, 1326-1341.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C.*, et al.* (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739-744.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499-506.

Deardorff, M.A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D.*, et al.* (2012). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature *489*, 313-317.

Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. FEBS letters *589*, 2877-2884.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110-1121.

DeMare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. Genome research *23*, 1224-1234.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Molecular cell *62*, 668-680.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Dowen, J.M., Bilodeau, S., Orlando, D.A., Hubner, M.R., Abraham, B.J., Spector, D.L., and Young, R.A. (2013). Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. Stem cell reports *1*, 371-378.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K.*, et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in Mammalian chromosomes. Cell *159*, 374-387.

Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. Nat Rev Cancer *3*, 11-22.

ENCODE Project Consortium, Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Filippova, G.N., Lindblom, A., Meincke, L.J., Klenova, E.M., Neiman, P.E., Collins, S.J., Doggett, N.A., and Lobanenkov, V.V. (1998). A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. Genes, chromosomes & cancer *22*, 26-36.

Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E.*, et al.* (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer research *62*, 48-52.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110-114.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell reports *15*, 2038-2049.

Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. Genes & development *30*, 881-891.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Molecular cell *49*, 773-782.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell *157*, 950-963.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell stem cell *14*, 762-775.

Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A., Erpelinck, C., van der Velden, V.H., Havermans, M., Avellino, R., van Lom, K.*, et al.* (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell *157*, 369-381.

Guo, G., Sun, X., Chen, C., Wu, S., Huang, P., Li, Z., Dean, M., Huang, Y., Jia, W., Zhou, Q.*, et al.* (2013). Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nature genetics *45*, 1459-1463.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900-910.

Hamid, O., Robert, C., Daud, A., Hodi, F.S., Hwu, W.J., Kefford, R., Wolchok, J.D., Hersey, P., Joseph, R.W., Weber, J.S.*, et al.* (2013). Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. The New England journal of medicine *369*, 134-144.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F.*, et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nature genetics *43*, 630-638.

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.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Hnisz, D., Day, D.S., and Young, R.A. (2016a). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell *167*, 1188-1200.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A.*, et al.* (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454-1458.

Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proceedings of the National Academy of Sciences of the United States of America *105*, 20398-20403.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nature methods *9*, 999-1003.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I.*, et al.* (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell stem cell *18*, 262-275.

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.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C.F., Wolffe, A., Ohlsson, R., and Lobanenkov, V.V. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Current biology : CB *10*, 853-856.

Katainen, R., Dave, K., Pitkanen, E., Palin, K., Kivioja, T., Valimaki, N., Gylfe, A.E., Ristolainen, H., Hanninen, U.A., Cajuso, T.*, et al.* (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. Nature genetics *47*, 818-821.

Kemp, C.J., Moore, J.M., Moser, R., Bernard, B., Teater, M., Smith, L.E., Rabaia, N.A., Gurley, K.E., Guinney, J., Busch, S.E.*, et al.* (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell reports *7*, 1020-1029.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.

Kon, A., Shih, L.Y., Minamino, M., Sanada, M., Shiraishi, Y., Nagata, Y., Yoshida, K., Okuno, Y., Bando, M., Nakato, R.*, et al.* (2013). Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nature genetics *45*, 1232-1237.

Krijger, P.H., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. Nature reviews Molecular cell biology *17*, 771-782.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F.*, et al.* (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Molecular cell *57*, 361-375.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495-501.

Lefevre, P., Witham, J., Lacroix, C.E., Cockerill, P.N., and Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Molecular cell *32*, 129-139.

Lin, C.P., and He, L. (2017). Noncoding RNAs in Cancer Development. Annu Rev Cancer Biol, 163-184.

Liu, M., Maurano, M.T., Wang, H., Qi, H., Song, C.Z., Navas, P.A., Emery, D.W., Stamatoyannopoulos, J.A., and Stamatoyannopoulos, G. (2015). Genomic discovery of potent chromatin insulators for human gene therapy. Nature biotechnology *33*, 198-203.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247 e217.

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.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A*., et al.* (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proceedings of the National Academy of Sciences of the United States of America *99*, 6806-6811.

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

Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G*., et al.* (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. The New England journal of medicine *350*, 2129-2139.

Mazumdar, C., Shen, Y., Xavy, S., Zhao, F., Reinisch, A., Li, R., Corces, M.R., Flynn, R.A., Buenrostro, J.D., Chan, S.M*., et al.* (2015). Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. Cell stem cell *17*, 675-688.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annual review of genomics and human genetics *17*, 17-43.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285-1297.

Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A*., et al.* (2004). The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. Genome research *14*, 1594-1602.

Mullenders, J., Aranda-Orgilles, B., Lhoumaud, P., Keller, M., Pae, J., Wang, K., Kayembe, C., Rocha, P.P., Raviram, R., Gong, Y*., et al.* (2015). Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. The Journal of experimental medicine *212*, 1833-1850.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-1021.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annual review of genetics *43*, 525-558.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948-953.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Nowell, P.C., and Hungerford, D.A. (1960). Chromosome studies on normal and leukemic human leukocytes. Journal of the National Cancer Institute *25*, 85-109.

Oldridge, D.A., Wood, A.C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., McDaniel, L.D., Diamond, M., Hart, L.S., Zhu, S.*, et al.* (2015). Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature *528*, 418-421.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nature reviews Genetics *15*, 234-246.

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. Nature genetics *35*, 190-194.

Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer *12*, 252-264.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T.*, et al.* (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422-433.

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

Phillips-Cremins, J.E., and Corces, V.G. (2013). Chromatin insulators: linking genome organization to cellular function. Molecular cell *50*, 461-474.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Poulos, R.C., Thoms, J.A., Guan, Y.F., Unnikrishnan, A., Pimanda, J.E., and Wong, J.W. (2016). Functional Mutations Form at CTCF-Cohesin Binding Sites in Melanoma Due to Uneven Nucleotide Excision Repair across the Motif. Cell reports *17*, 2865-2872.

Price, J.C., Pollock, L.M., Rudd, M.L., Fogoros, S.K., Mohamed, H., Hanigan, C.L., Le Gallo, M., Program, N.I.H.I.S.C.C.S., Zhang, S., Cruz, P.*, et al.* (2014). Sequencing of candidate chromosome instability genes in endometrial cancers reveals somatic mutations in ESCO1, CHTF18, and MRE11A. PloS one *8*, e63313.

Rabbitts, T.H. (1994). Chromosomal translocations in human cancer. Nature *372*, 143-149.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

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. Proceedings of the National Academy of Sciences of the United States of America *105*, 8309-8314.

Saldana-Meyer, R., Gonzalez-Buendia, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes & development *28*, 723-734.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). 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 *112*, E6456-6465.

Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K.*, et al.* (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature *476*, 467-471.

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer *5*, 615-625.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes & development *20*, 2349-2354.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R.*, et al.* (2012). The landscape of cancer genes and mutational processes in breast cancer. Nature *486*, 400-404.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA activates Xist by evicting CTCF. Cell *153*, 1537-1551.

Szabo, 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. Current biology : CB *10*, 607-610.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B.*, et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611-1627.

Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V.*, et al.* (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564-568.

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. Molecular cell *10*, 1453-1465.

Umer, H.M., Cavalli, M., Dabrowski, M.J., Diamanti, K., Kruczyk, M., Pan, G., Komorowski, J., and Wadelius, C. (2016). A Significant Regulatory Mutation Burden at a High-Affinity Position of the CTCF Motif in Gastrointestinal Cancers. Human mutation *37*, 904-913.

Valton, A.L., and Dekker, J. (2016). TAD disruption as oncogenic driver. Curr Opin Genet Dev *36*, 34-40.

Van Vlierberghe, P., and Ferrando, A. (2012). The molecular basis of T cell acute lymphoblastic leukemia. The Journal of clinical investigation *122*, 3398-3406.

Viny, A.D., Ott, C.J., Spitzer, B., Rivas, M., Meydan, C., Papalexi, E., Yelin, D., Shank, K., Reyes, J., Chiu, A.*, et al.* (2015). Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. The Journal of experimental medicine *212*, 1819-1832.

Vogelstein, B., and Kinzler, K.W. (2004). Cancer genes and the pathways they control. Nature medicine *10*, 789-799.

Walker, C.J., Miranda, M.A., O'Hern, M.J., McElroy, J.P., Coombes, K.R., Bundschuh, R., Cohn, D.E., Mutch, D.G., and Goodfellow, P.J. (2015). Patterns of CTCF and ZFHX3 Mutation and Associated Outcomes in Endometrial Cancer. Journal of the National Cancer Institute *107*.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R.*, et al.* (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome research *22*, 1680-1688.

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

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. Molecular and cellular biology *31*, 2174-2183.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C.*, et al.* (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. Nature genetics *36*, 1105-1110.

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

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F.*, et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proceedings of the National Academy of Sciences of the United States of America *111*, 996-1001.

Figure 1

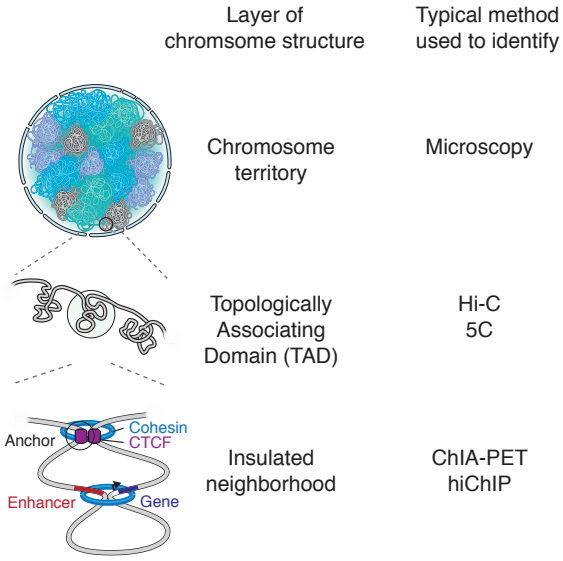

Layer of chromsome structure
Typical method used to identify
Chromosome territory
Microscopy
Topologically Associating Domain (TAD)
Hi-C
5C
Anchor
Cohesin
CTCF
Enhancer
Gene
Insulated neighborhood
ChIA-PET
hiChIP

Figure 2

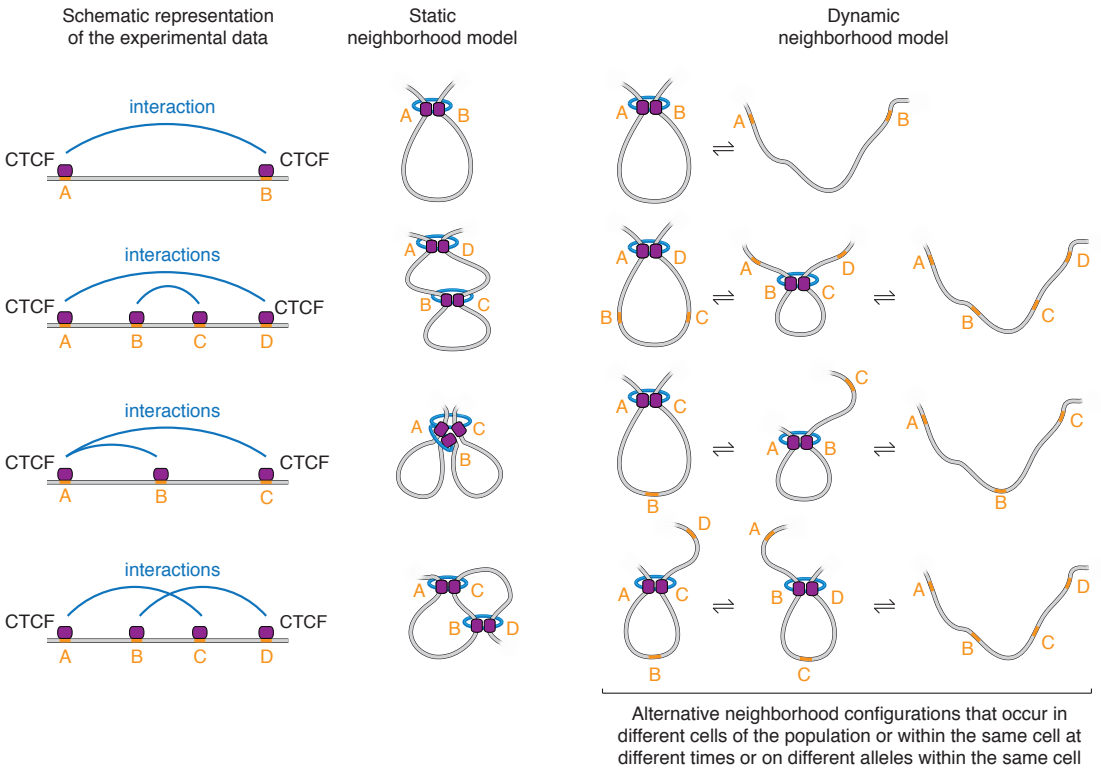

Schematic representation of the experimental data
Static neighborhood model
Dynamic neighborhood model
interaction
CTCF
CTCF
A
B
interactions
CTCF
CTCF
A
B
C
D
interactions
CTCF
CTCF
A
B
C
interactions
CTCF
CTCF
A
B
C
D
Alternative neighborhood configurations that occur in different cells of the population or within the same cell at different times or on different alleles within the same cell

Figure 3

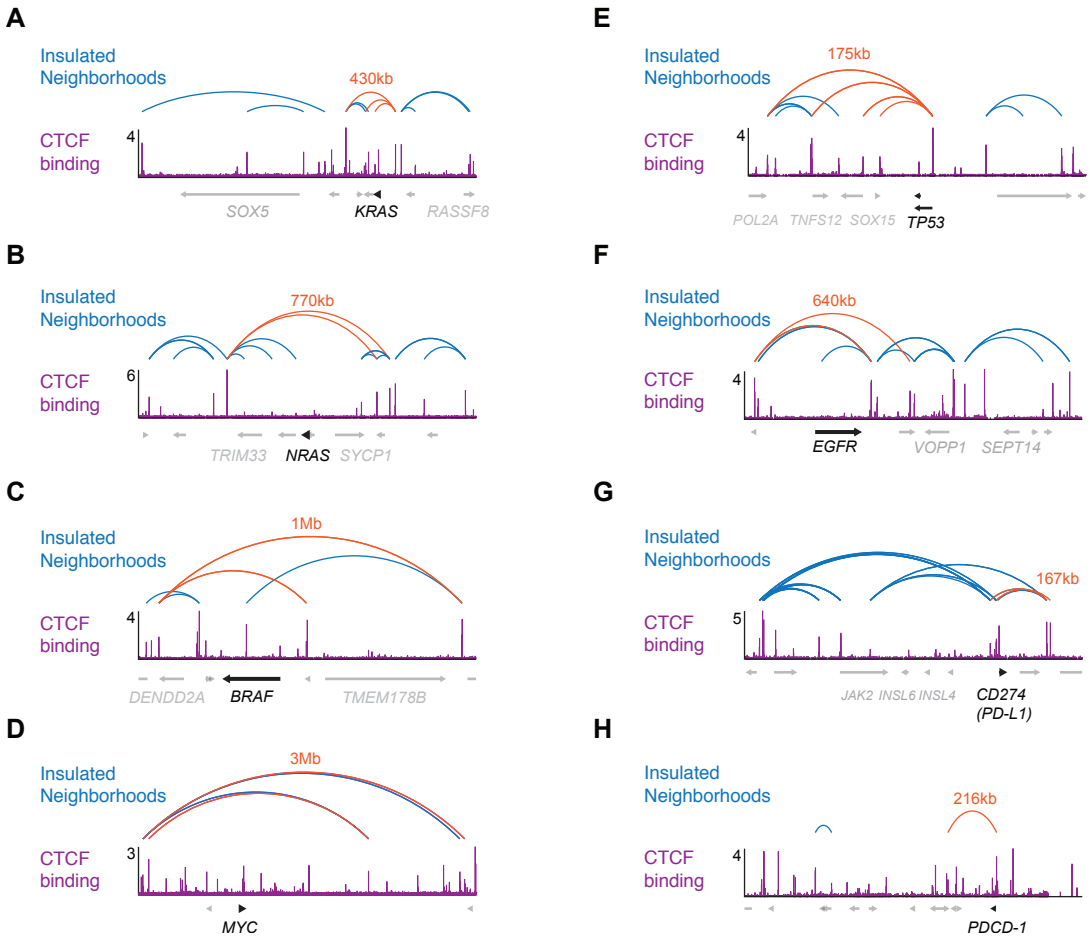

A
Insulated Neighborhoods
430kb
CTCF binding
4
SOX5
KRAS
RASSF8
B
Insulated Neighborhoods
770kb
CTCF binding
6
TRIM33
NRAS
SYCP1
C
Insulated Neighborhoods
1Mb
CTCF binding
4
DENDD2A
BRAF
TMEM178B
D
Insulated Neighborhoods
3Mb
CTCF binding
3
MYC
E
Insulated Neighborhoods
175kb
CTCF binding
4
POL2A
TNFS12
SOX15
TP53
F
Insulated Neighborhoods
640kb
CTCF binding
4
EGFR
VOPP1
SEPT14
G
Insulated Neighborhoods
167kb
CTCF binding
5
JAK2
INSL6
INSL4
CD274 (PD-L1)
H
Insulated Neighborhoods
216kb
CTCF binding
4
PDCD-1

Figure 4

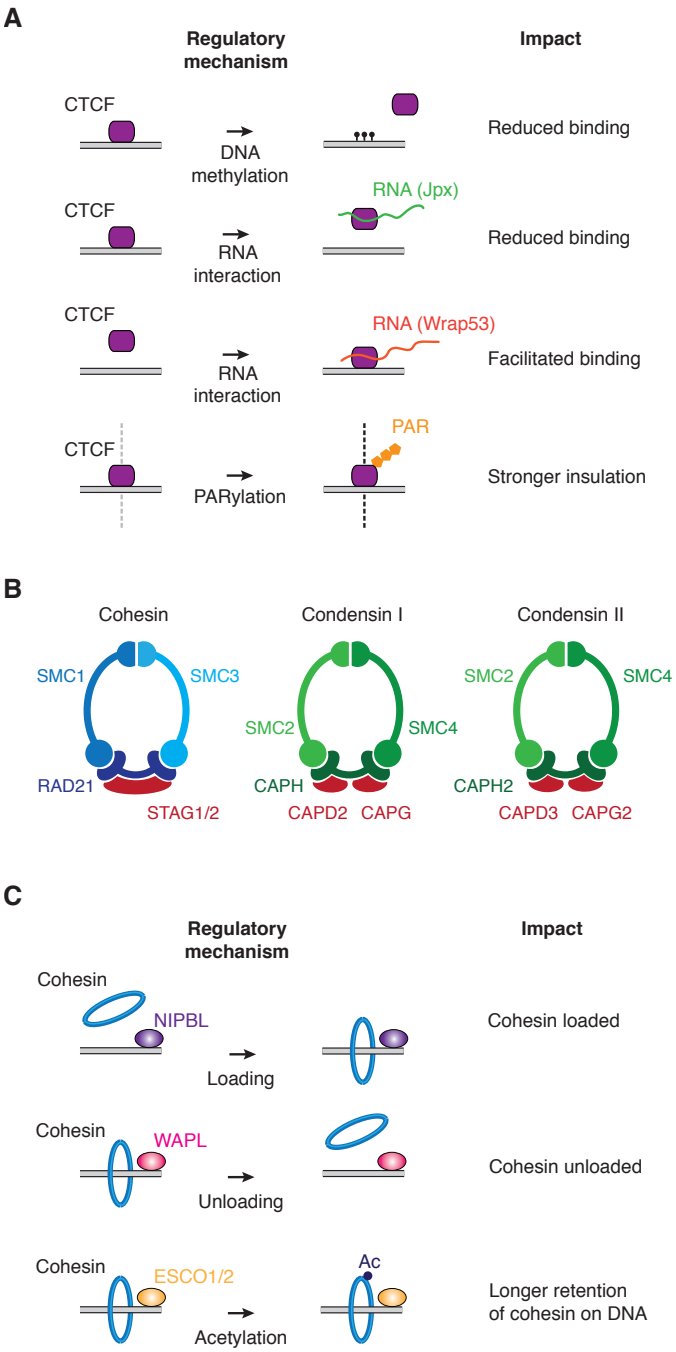
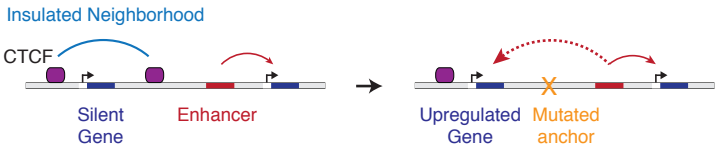

A
Regulatory mechanism
Impact
CTCF
DNA methylation
Reduced binding
RNA (Jpx)
CTCF
RNA interaction
Reduced binding
CTCF
RNA (Wrap53)
RNA interaction
Facilitated binding
CTCF
PAR
PARylation
Stronger insulation
B
Cohesin
Condensin I
Condensin II
SMC1
SMC3
RAD21
STAG1/2
SMC2
SMC4
CAPH
CAPD2
CAPG
SMC2
SMC4
CAPH2
CAPD3
CAPG2
C
Regulatory mechanism
Impact
Cohesin
NIPBL
Loading
Cohesin loaded
Cohesin
WAPL
Unloading
Cohesin unloaded
Cohesin
ESCO1/2
Acetylation
Ac
Longer retention of cohesin on DNA

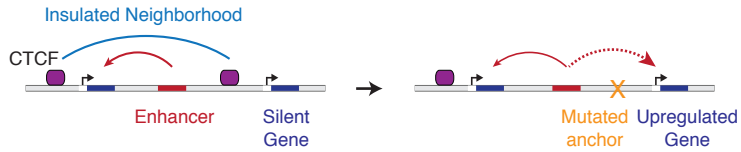
Figure 5

A



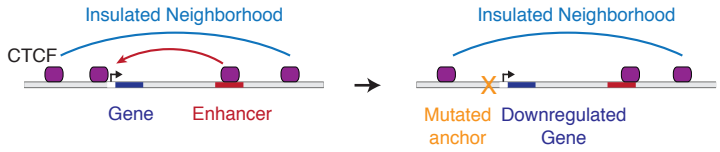
Mutation	Gene
CTCF site deletion (T-cell leukemia)	*TAL1* *LMO2*
DNA methylation of CTCF site (glioma)	*PDGFRA*

B



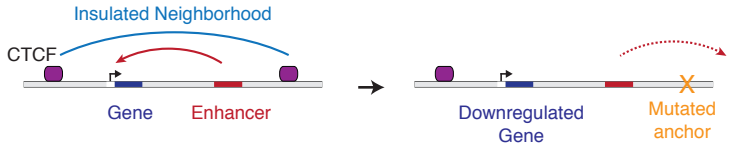
? ?

C



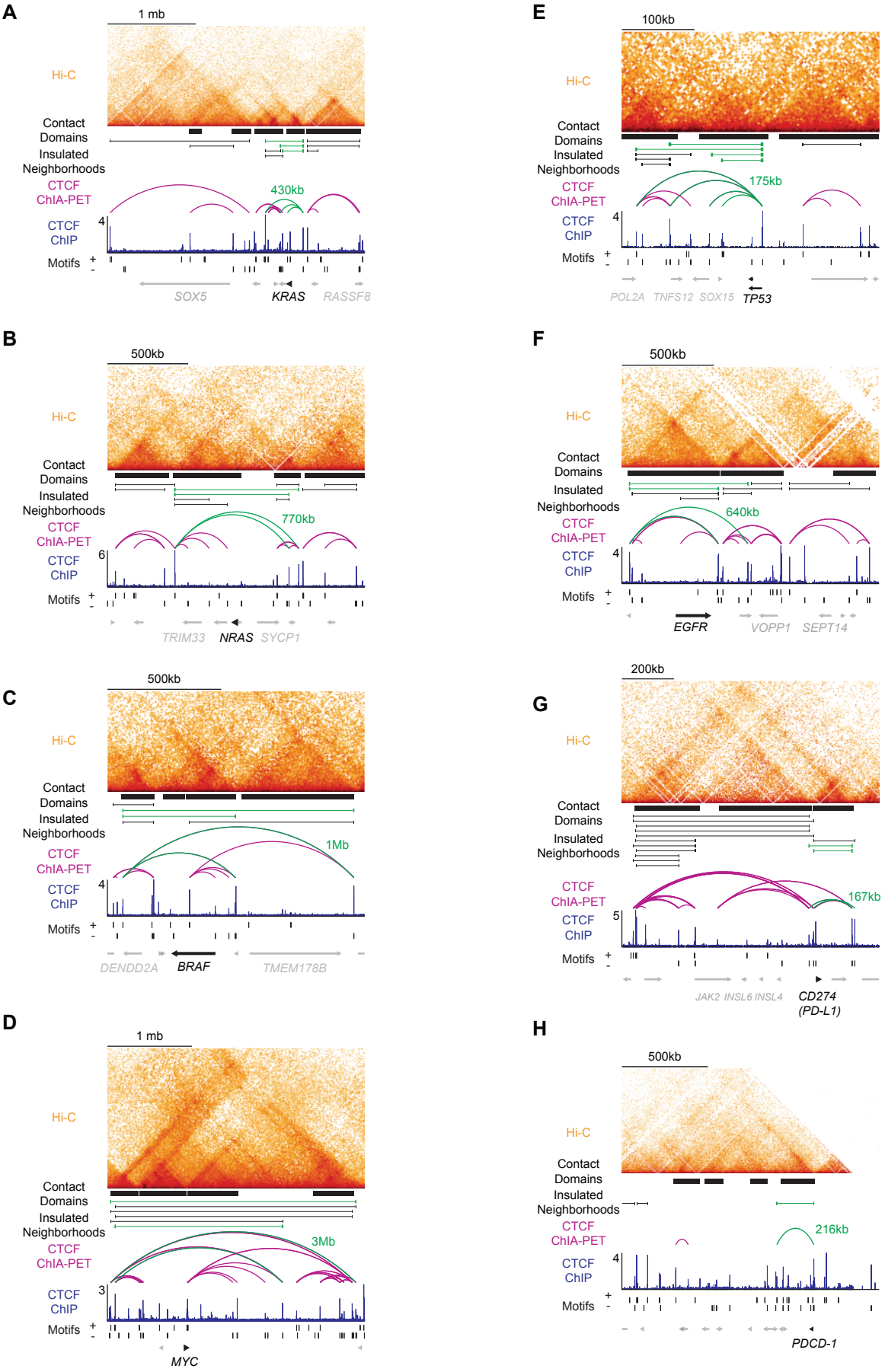
? ?

D



? ?

Supplementary Figure 1



Factor	Cancer Type	Type of mutation	Effect on gene	Hits	Sample Size	Frequency	References
CTCF	Endometrial cancer	unreported	unreported	46	248	18.5	Lawrence et al., Nature, 2014
CTCF	Endometrial cancer, endometrioid		frameshift, missense, splice, read through	136	538	25.3	Walker et al., Journal of the National Cancer Institute, 2015
CTCF	Head and neck cancer	unreported	unreported	12	384	3.1	Lawrence et al., Nature, 2014
CTCF	Uterine corpus endometrial carcinoma	substitution, indel	unreported	43	230	18.7	Kandoth et al., Nature, 2013
ESCO1	Endometrial cancer	substitution	nonsense, missense	4	107	3.7	Price et al., PLoS ONE, 2013
NIPBL	Colorectal cancer	substitution, indel	missense, frameshift	4	132	3	Barber et al., PNAS, 2008
NIPBL	Urothelial carcinoma	substitution	nonsense, missense, exon junction	5	131	3.8	Cancer Genome Atlas Research Network, Nature, 2014
RAD21	Acute myeloid leukemia	substitution, indel	frameshift, nonsense, splice	7	157	4.5	Kon et al., Nature Genetics, 2013
RAD21	Acute myeloid leukemia	unreported	unreported	6	196	3.1	Lawrence et al., Nature, 2014
RAD21	Urothelial carcinoma	substitution	missense, nonsense	5	131	3.8	Cancer Genome Atlas Research Network, Nature, 2014
SMC1A	Acute myeloid leukemia	substitution	missense, nonsense	7	200	3.5	Cancer Genome Atlas Research Network, NEJM, 2013
SMC1A	Acute myeloid leukemia	substitution	unreported	7	200	3.5	Kandoth et al., Nature, 2013
SMC1A	Acute myeloid leukemia	unreported	unreported	7	196	3.6	Lawrence et al., Nature, 2014
SMC1A	Colorectal cancer	substitution	missense	4	132	3	Barber et al., PNAS, 2008
SMC3	Acute myeloid leukemia	substitution	missense, nonsense	7	200	3.5	Cancer Genome Atlas Research Network, NEJM, 2013
SMC3	Acute myeloid leukemia	substitution, indel	missense, nonsense, splice, frameshift	14	450	3.1	Thota et al., Blood, 2014
SMC3	Acute myeloid leukemia	substitution	unreported	7	200	3.5	Kandoth et al., Nature, 2013
SMC3	Acute myeloid leukemia	unreported	unreported	7	196	3.6	Lawrence et al., Nature, 2014
STAG1	Urothelial carcinoma	substitution	missense, exon junction	4	131	3.1	Cancer Genome Atlas Research Network, Nature, 2014
STAG2	Acute myeloid leukemia	substitution, indel	frameshift, nonsense, splice	10	157	6.4	Kon et al., Nature Genetics, 2013
STAG2	Acute myeloid leukemia	substitution	nonsense	6	200	3	Cancer Genome Atlas Research Network, NEJM, 2013
STAG2	Acute myeloid leukemia	substitution, indel	nonsense, frameshift, splice, deletion	23	450	5.1	Thota et al., Blood, 2014
STAG2	Acute myeloid leukemia	substitution, indel	nonsense, frameshift, splice, missense, deletion	24	299	8	Lindsley et al., Blood, 2015
STAG2	Acute myeloid leukemia	substitution	unreported	6	200	3	Kandoth et al., Nature, 2013
STAG2	Acute myeloid leukemia	unreported	unreported	6	196	3.1	Lawrence et al., Nature, 2014
STAG2	Acute myeloid leukemia, de novo	unreported	frameshift, nonsense, missense	10	197	5.1	Kihara et al., Leukemia, 2014
STAG2	Bladder cancer	substitution, indel	missense, nonsense, exon junction, framshift	25	111	22.5	Solomon et al., Nature Genetics, 2013
STAG2	Bladder cancer	substitution, indel	missense, nonsense, frameshift, deletion, splice	67	307	21.8	Taylor et al., Human Molecular Genetics, 2014
STAG2	Ewing's sarcoma	substitution, indel, duplication	nonsense, missense, exon junction, frameshift, exon duplication	19	112	17	Tirode et al., Cancer Discovery, 2014
STAG2	Ewing's sarcoma	substitution, indel, duplication	nonsense, missense, exon junction, frameshift, exon duplication, in-frame deletion	41	199	20.6	Tirode et al., Cancer Discovery, 2014
STAG2	Gliobastoma multiforme	substitution, indel	unreported	12	290	4.1	Kandoth et al., Nature, 2013
STAG2	Gliobastoma multiforme	unreported	unreported	12	291	4.1	Lawrence et al., Nature, 2014
STAG2	Glioblastoma	unreported	unreported	12	291	4.1	Brennan et al., Cell, 2013
STAG2	Myelodysplastic syndromes	substitution, indel	frameshift, nonsense, splice	13	224	5.8	Kon et al., Nature Genetics, 2013
STAG2	Myelodysplastic syndromes	substitution, indel	missense, nonsense, frameshift, splice	30	386	7.8	Thota et al., Blood, 2014
STAG2	Myelodysplastic syndromes	substitution, indel	splice, nonsense, frameshift	9	150	6	Walter et al., Leukemia, 2013
STAG2	Myelodysplastic syndromes	unreported	unreported	71	944	7.5	Haferlach et al., Leukemiea, 2014
STAG2	Myelodysplastic syndromes/Myeloproliferati	substitution	missense, nonsense, splice	6	169	3.6	Thota et al., Blood, 2014
STAG2	Renal cell carcinoma, papillary	unreported	unreported	8	157	5.1	Cancer Genome Atlas Research et al., NEJM, 2016
STAG2	Urothelial carcinoma	substitution, indel	nonsense, missense, exon junction, frameshift	14	131	10.7	Cancer Genome Atlas Research Network, Nature, 2014

Supplementary Table 1

Factor	Cancer Type	Type of mutation	Effect on gene	Hits	Sample Size	Frequency	References
CTCF	Acute lymphoblastic leukemia	substitution	missense	1	23	4.3	Mullighan et al., Nature 2011
CTCF	Acute megakaryoblastic leukemia, Down syndrome-related		indel, deletion, splice, missense, nonsense	10	49	20.4	Yoshida et al., Nature Genetics, 2013
CTCF	Acute megakaryoblastic leukemia, non-Down syndrome-related		missense, deletion	4	19	21.1	Yoshida et al., Nature Genetics, 2013
CTCF	Acute myeloid leukemia	substitution	nonsense	1	8	12.5	Ding et al., Nature, 2012
CTCF	Acute myeloid leukemia	substitution	nonsense	1	200	0.5	Cancer Genome Atlas Research Network, NEJM, 2013
CTCF	Acute myeloid leukemia	indel	frameshift	1	50	2	Dolnik et al., Blood, 2012
CTCF	Bladder cancer, transitional cell carcinoma	substitution, indel	missense, frameshift	3	99	3	Guo et al., Nature Genetics, 2013
CTCF	Breast adenocarcinoma	substitution, indel	unreported	18	763	2.4	Kandoth et al., Nature, 2013
CTCF	Breast cancer	indel	frameshift	1	100	1	Stephens et al., Nature, 2012
CTCF	Breast cancer	unreported	unreported	20	892	2.2	Lawrence et al., Nature, 2014
CTCF	Breast cancer		rearrangement	1	4	25	Fillipova et al., Genes Chromosomes Cancer, 1998
CTCF	Breast cancer		missense	1	31	3.2	Fillipova et al., Cancer Research, 2002
CTCF	Breast cancer		unreported	13	507	2.6	Cancer Genome Atlas Research et al., Nature 2012
CTCF	Breast cancer, invasive ductal carcinoma	indel	nonsense	1	18	5.6	Aulmann et al., Breast Cancer Resarch and Treatment, 2003
CTCF	Clear cell renal cell carcinoma		unreported	2	417	0.5	Cancer Genome Atlas Research et al., Nature, 2013
CTCF	Colorectal cancer	substitution	missense	3	72	4.2	Seshagiri et al., Nature, 2012
CTCF	Endometrial cancer	unreported	unreported	46	248	18.5	Lawrence et al., Nature, 2014
CTCF	Endometrial cancer, endometrioid		frameshift, missense, splice, read through	136	538	25.3	Walker et al., Journal of the National Cancer Institute, 2015
CTCF	Endometrial cancer, MSI negative	substitution, indel	missense, nonsense, frameshift, splice	9	51	17.6	Zighelboim et al., Human Mutation, 2014
CTCF	Endometrial cancer, MSI positive	substitution, indel	missense, frameshift	3	8	37.5	Zighelboim et al., Human Mutation, 2014
CTCF	Endometrial cancer, MSI positive	substitution, indel	missense, nonsense, frameshift, splice	36	98	36.7	Zighelboim et al., Human Mutation, 2014
CTCF	Endometrial cancer, serous	substitution	missense	1	52	1.9	Le Gallo et al., Nature Genetics, 2012
CTCF	Ewing's sarcoma, pediatric	substitution	missense	1	92	1.1	Cromptom et al., Cancer Discovery, 2014
CTCF	Head and neck cancer	unreported	unreported	12	384	3.1	Lawrence et al., Nature, 2014
CTCF	Myelodysplastic syndromes	unreported	unreported	12	944	1.3	Haferlach et al., Leukemiea, 2014
CTCF	Prostate cancer		missense	1	41	2.4	Fillipova et al., Cancer Research, 2002
CTCF	Transient abnormal myelopoiesis		nonsense	1	41	2.4	Yoshida et al., Nature Genetics, 2013
CTCF	Urothelial carcinoma	substitution	missense, nonsense	3	131	2.3	Cancer Genome Atlas Research Network, Nature, 2014
CTCF	Uterine corpus endometrial carcinoma	substitution, indel	unreported	43	230	18.7	Kandoth et al., Nature, 2013
CTCF	Wilms' tumor		missense	2	59	3.4	Fillipova et al., Cancer Research, 2002
ESCO1	Chronic myelomonocytic leukemia	substitution	missense	1	88	1.1	Kon et al., Nature Genetics, 2013
ESCO1	Clear cell renal cell carcinoma		unreported	2	417	0.5	Cancer Genome Atlas Research et al., Nature, 2013
ESCO1	Colorectal cancer	substitution	missense	2	72	2.8	Seshagiri et al., Nature, 2012
ESCO1	Endometrial cancer	substitution	nonsense, missense	4	107	3.7	Price et al., PLoS ONE, 2013
ESCO1	Multiple myeloma	substitution	missense, nonsense	3	203	1.5	Lohr et al., Cancer Cell, 2014
ESCO1	Urothelial carcinoma	substitution	missense	1	131	0.8	Cancer Genome Atlas Research Network, Nature, 2014
ESCO2	Acute myeloid leukemia	substitution	missense	1	200	0.5	Cancer Genome Atlas Research Network, NEJM, 2013
ESCO2	Acute myeloid leukemia	substitution	missense	1	450	0.2	Thota et al., Blood, 2014
ESCO2	Clear cell renal cell carcinoma		unreported	1	417	0.2	Cancer Genome Atlas Research et al., Nature, 2013
ESCO2	Colorectal cancer	substitution	missense	1	72	1.4	Seshagiri et al., Nature, 2012
ESCO2	Myelodysplastic syndromes	substitution	missense	1	224	0.4	Kon et al., Nature Genetics, 2013
ESCO2	Myelodysplastic syndromes	unreported	unreported	1	944	0.1	Haferlach et al., Leukemiea, 2014
ESCO2	Urothelial bladder cancer	substitution	missense	1	77	1.3	Balbás-Martínez et al., Nature Genetics, 2013
ESCO2	Urothelial carcinoma	substitution	missense	1	131	0.8	Cancer Genome Atlas Research Network, Nature, 2014
HDAC8	Urothelial carcinoma	substitution	missense	1	131	0.8	Cancer Genome Atlas Research Network, Nature, 2014
MAU2	Urothelial bladder cancer	substitution, indel	missense, deletion	2	77	2.6	Balbás-Martínez et al., Nature Genetics, 2013
NIPBL	Acute megakaryoblastic leukemia, Down syndrome-related		nonsense, deletion	3	49	6.1	Yoshida et al., Nature Genetics, 2013
NIPBL	Acute myeloid leukemia	indel	frameshift	1	157	0.6	Kon et al., Nature Genetics, 2013
NIPBL	Acute myeloid leukemia	substitution	missense	1	200	0.5	Cancer Genome Atlas Research Network, NEJM, 2013
NIPBL	Bladder cancer, transitional cell carcinoma	substitution	missense	4	99	4	Guo et al., Nature Genetics, 2013
NIPBL	Breast cancer	substitution	missense, nonsense	2	100	2	Stephens et al., Nature, 2012
NIPBL	Chronic myelogenous leukemia	substitution	missense	1	64	1.6	Kon et al., Nature Genetics, 2013
NIPBL	Clear cell renal cell carcinoma		unreported	4	417	1	Cancer Genome Atlas Research et al., Nature, 2013
NIPBL	Colorectal cancer	substitution, indel	missense, frameshift	4	132	3	Barber et al., PNAS, 2008
NIPBL	Colorectal cancer	substitution	missense, nonsense	5	72	6.9	Seshagiri et al., Nature, 2012
NIPBL	Glioma		unreported	19	820	2.3	Ceccarelli et al., Cell, 2016
NIPBL	Myelodysplastic syndromes	substitution	missense	1	224	0.4	Kon et al., Nature Genetics, 2013
NIPBL	Myelodysplastic syndromes	unreported	unreported	4	944	0.4	Haferlach et al., Leukemiea, 2014
NIPBL	Urothelial bladder cancer	substitution	exon junction, missense	2	77	2.6	Balbás-Martínez et al., Nature Genetics, 2013
NIPBL	Urothelial carcinoma	substitution	nonsense, missense, exon junction	5	131	3.8	Cancer Genome Atlas Research Network, Nature, 2014
PDS5A	Breast cancer	substitution	missense	1	100	1	Stephens et al., Nature, 2012
PDS5A	Clear cell renal cell carcinoma		unreported	4	417	1	Cancer Genome Atlas Research et al., Nature, 2013
PDS5A	Colorectal cancer	substitution	missense	3	72	4.2	Seshagiri et al., Nature, 2012
PDS5A	Multiple myeloma	substitution	missense	3	203	1.5	Lohr et al., Cancer Cell, 2014
PDS5A	Urothelial carcinoma	substitution	missense, nonsense	2	131	1.5	Cancer Genome Atlas Research Network, Nature, 2014
PDS5B	Acute myeloid leukemia	indel	frameshift	1	157	0.6	Kon et al., Nature Genetics, 2013
PDS5B	Acute myeloid leukemia	substitution	missense	1	200	0.5	Cancer Genome Atlas Research Network, NEJM, 2013
PDS5B	Acute myeloid leukemia	substitution, indel	missense, frameshift	2	450	0.4	Thota et al., Blood, 2014
PDS5B	Acute myeloid leukemia	indel	frameshift, splice	2	299	0.7	Lindsley et al., Blood, 2015
PDS5B	Bladder cancer, transitional cell carcinoma	substitution	missense	1	99	1	Guo et al., Nature Genetics, 2013
PDS5B	Breast cancer	substitution, indel	missense, frameshift	2	100	2	Stephens et al., Nature, 2012
PDS5B	Chronic myelomonocytic leukemia	substitution	missense	1	88	1.1	Kon et al., Nature Genetics, 2013
PDS5B	Clear cell renal cell carcinoma		unreported	5	417	1.2	Cancer Genome Atlas Research et al., Nature, 2013
PDS5B	Colorectal cancer	substitution, splice	missense	3	72	4.2	Seshagiri et al., Nature, 2012
PDS5B	Ewing's sarcoma, pediatric	substitution	exon junction	1	92	1.1	Cromptom et al., Cancer Discovery, 2014
PDS5B	Myelodysplastic syndromes	substitution	splice	1	386	0.3	Thota et al., Blood, 2014
PDS5B	Myelodysplastic syndromes	unreported	unreported	7	944	0.7	Haferlach et al., Leukemiea, 2014
PDS5B	Myeloproliferative neoplasms	substitution	missense	1	77	1.3	Kon et al., Nature Genetics, 2013
PDS5B	Myeloproliferative neoplasms	substitution	missense	1	55	1.8	Thota et al., Blood, 2014
PDS5B	Urothelial bladder cancer	substitution	missense	1	77	1.3	Balbás-Martínez et al., Nature Genetics, 2013
PDS5B	Urothelial carcinoma	substitution	missense	1	131	0.8	Cancer Genome Atlas Research Network, Nature, 2014
RAD21	Acute megakaryoblastic leukemia, Down syndrome-related	unreported	missense, nonsense, indel, deletion	11	49	22.4	Yoshida et al., Nature Genetics, 2013
RAD21	Acute myeloid leukemia	substitution, indel	frameshift, nonsense, splice	7	157	4.5	Kon et al., Nature Genetics, 2013
RAD21	Acute myeloid leukemia	substitution, indel	nonsense, frameshift	5	200	2.5	Cancer Genome Atlas Research Network, NEJM, 2013
RAD21	Acute myeloid leukemia	substitution, indel	missense, frameshift, nonsense	4	389	1	Thol et al., Blood, 2014
RAD21	Acute myeloid leukemia	substitution, indel	missense, nonsense, frameshift, splice	11	450	2.4	Thota et al., Blood, 2014
RAD21	Acute myeloid leukemia	substitution, indel	nonsense, frameshift	8	299	2.7	Lindsley et al., Blood, 2015
RAD21	Acute myeloid leukemia	substitution, indel	nonsense, frameshift	4	50	8	Dolnik et al., Blood, 2012
RAD21	Acute myeloid leukemia	substitution, indel	unreported	5	200	2.5	Kandoth et al., Nature, 2013
RAD21	Acute myeloid leukemia	unreported	unreported	6	196	3.1	Lawrence et al., Nature, 2014
RAD21	Acute myeloid leukemia, de novo	unreported	missense, indel	4	197	2	Kihara et al., Leukemia, 2014
RAD21	Bladder cancer, transitional cell carcinoma	insertion	frameshift	1	99	1	Guo et al., Nature Genetics, 2013
RAD21	Chronic myelogenous leukemia	substitution	splice	1	64	1.6	Kon et al., Nature Genetics, 2013
RAD21	Chronic myelomonocytic leukemia		deletion	1	53	1.9	Rocquain et al, American Journal of Hematology, 2010
RAD21	Colorectal cancer	substitution	missense	2	72	2.8	Seshagiri et al., Nature, 2012
RAD21	Ewing's sarcoma, pediatric	substitution	nonsense	1	92	1.1	Cromptom et al., Cancer Discovery, 2014
RAD21	Myelodysplastic syndromes	indel	frameshift	2	224	0.9	Kon et al., Nature Genetics, 2013
RAD21	Myelodysplastic syndromes	substitution, indel	missense, frameshift, nonsense	7	386	1.8	Thota et al., Blood, 2014
RAD21	Myelodysplastic syndromes	unreported	unreported	13	944	1.4	Haferlach et al., Leukemiea, 2014
RAD21	Myelodysplastic syndromes/Myeloproliferative neoplasms	indel	frameshift	1	169	0.6	Thota et al., Blood, 2014
RAD21	Urothelial bladder cancer	substitution	missense	1	77	1.3	Balbás-Martínez et al., Nature Genetics, 2013
RAD21	Urothelial carcinoma	substitution	missense, nonsense	5	131	3.8	Cancer Genome Atlas Research Network, Nature, 2014
SMC1	Acute myeloid leukemia	substitution	missense	2	389	0.5	Thol et al., Blood, 2014
SMC1A	Acute megakaryoblastic leukemia, Down syndrome-related		missense, splice	2	49	4.1	Yoshida et al., Nature Genetics, 2013
SMC1A	Acute myeloid leukemia	substitution	missense	2	157	1.3	Kon et al., Nature Genetics, 2013
SMC1A	Acute myeloid leukemia	substitution	missense, nonsense	7	200	3.5	Cancer Genome Atlas Research Network, NEJM, 2013
SMC1A	Acute myeloid leukemia	substitution	missense, nonsense	5	450	1.1	Thota et al., Blood, 2014
SMC1A	Acute myeloid leukemia	substitution, indel	missense, splice, deletion	7	299	2.3	Lindsley et al., Blood, 2015
SMC1A	Acute myeloid leukemia	substitution	missense	1	24	4.2	Welch et al., Cell ,2012
SMC1A	Acute myeloid leukemia	substitution	unreported	7	200	3.5	Kandoth et al., Nature, 2013
SMC1A	Acute myeloid leukemia	unreported	unreported	7	196	3.6	Lawrence et al., Nature, 2014
SMC1A	Acute myeloid leukemia, core binding factor, pediatric	unreported	unreported	4	83	4.8	Huether et al., Nature Communications, 2014
SMC1A	Acute myeloid leukemia, de novo	unreported	missense, insertion	5	197	2.5	Kihara et al., Leukemia, 2014
SMC1A	Bladder cancer, transitional cell carcinoma	substitution	missense, exon junction	3	99	3	Guo et al., Nature Genetics, 2013
SMC1A	Brain tumor, ependymoma, pediatric	unreported	unreported	1	40	2.5	Huether et al., Nature Communications, 2014
SMC1A	Breast cancer	indel	deletion	1	101	1	Stephens et al., Nature, 2013
SMC1A	Chronic myelogenous leukemia	substitution	missense	2	64	3.1	Kon et al., Nature Genetics, 2013
SMC1A	Clear cell renal cell carcinoma		unreported	1	417	0.2	Cancer Genome Atlas Research et al., Nature, 2013
SMC1A	Colorectal cancer	substitution	missense	4	132	3	Barber et al., PNAS, 2008
SMC1A	Colorectal cancer	substitution	missense	1	72	1.4	Seshagiri et al., Nature, 2012
SMC1A	Ewing's sarcoma	substitution	missense	1	112	0.9	Tirode et al., Cancer Discovery, 2014
SMC1A	Ewing's sarcoma, pediatric	substitution	missense	1	92	1.1	Cromptom et al., Cancer Discovery, 2014
SMC1A	Medulloblastoma	substitution	missense	1	125	0.8	Jones et al., Nature, 2012
SMC1A	Myelodysplastic syndromes	unreported	unreported	10	944	1.1	Haferlach et al., Leukemiea, 2014
SMC1A	Urothelial bladder cancer	substitution	missense	1	77	1.3	Balbás-Martínez et al., Nature Genetics, 2013
SMC1A	Urothelial carcinoma	substitution	missense, nonsense	3	131	2.3	Cancer Genome Atlas Research Network, Nature, 2014
SMC1B	Breast cancer	substitution	missense	1	102	1	Stephens et al., Nature, 2014

SMC1B	Clear cell renal cell carcinoma		unreported	2	417	0.5	Cancer Genome Atlas Research et al., Nature, 2013
SMC1B	Colorectal cancer	substitution	missense, nonsense	3	72	4.2	Seshagiri et al., Nature, 2012
SMC1B	Multiple myeloma	substitution	missense	1	203	0.5	Lohr et al., Cancer Cell, 2014
SMC1B	Myelodysplastic syndromes, del(5q)		missense	1	2	50	Pellagatti et al., Leukemia, 2014
SMC1B	Urothelial bladder cancer	substitution	missense, exon junction	4	77	5.2	Balbás-Martínez et al., Nature Genetics, 2013
SMC1B	Urothelial carcinoma	substitution	missense	2	131	1.5	Cancer Genome Atlas Research Network, Nature, 2014
SMC3	Acute megakaryoblastic leukemia, Down syndrome-related		splice	1	49	2	Yoshida et al., Nature Genetics, 2013
SMC3	Acute myeloid leukemia	substitution	missense	1	8	12.5	Ding et al., Nature, 2012
SMC3	Acute myeloid leukemia	substitution	missense	1	157	0.6	Kon et al., Nature Genetics, 2013
SMC3	Acute myeloid leukemia	substitution	missense, nonsense	7	200	3.5	Cancer Genome Atlas Research Network, NEJM, 2013
SMC3	Acute myeloid leukemia	substitution	missense	5	389	1.3	Thol et al., Blood, 2014
SMC3	Acute myeloid leukemia	substitution, indel	missense, nonsense, splice, frameshift	14	450	3.1	Thota et al., Blood, 2014
SMC3	Acute myeloid leukemia	substitution	missense	4	299	1.3	Lindsley et al., Blood, 2015
SMC3	Acute myeloid leukemia	substitution	missense	1	24	4.2	Welch et al., Cell ,2012
SMC3	Acute myeloid leukemia	substitution	splice	1	7	14.3	Walter et al., NEJM, 2012
SMC3	Acute myeloid leukemia	substitution	unreported	7	200	3.5	Kandoth et al., Nature, 2013
SMC3	Acute myeloid leukemia	unreported	unreported	7	196	3.6	Lawrence et al., Nature, 2014
SMC3	Acute myeloid leukemia, de novo	unreported	missense, insertion	4	197	2	Kihara et al., Leukemia, 2014
SMC3	Bladder cancer, transitional cell carcinoma	substitution	nonsense, exon junction	2	99	2	Guo et al., Nature Genetics, 2013
SMC3	Clear cell renal cell carcinoma		unreported	5	417	1.2	Cancer Genome Atlas Research et al., Nature, 2013
SMC3	Colorectal cancer	substitution	missense	1	130	0.8	Barber et al., PNAS, 2008
SMC3	Colorectal cancer	substitution	missense	3	72	4.2	Seshagiri et al., Nature, 2012
SMC3	Medulloblastoma	substitution	splice	1	125	0.8	Jones et al., Nature, 2012
SMC3	Myelodysplastic syndromes	substitution	missense	3	224	1.3	Kon et al., Nature Genetics, 2013
SMC3	Myelodysplastic syndromes	substitution, indel	missense, splice, deletion	3	386	0.8	Thota et al., Blood, 2014
SMC3	Myelodysplastic syndromes	substitution	missense, splice	2	150	1.3	Walter et al., Leukemia, 2013
SMC3	Myelodysplastic syndromes	unreported	unreported	16	944	1.7	Haferlach et al., Leukemiea, 2014
SMC3	Myelodysplastic syndromes/Myeloproliferative neoplasms	substitution	splice	1	169	0.6	Thota et al., Blood, 2014
SMC3	Urothelial bladder cancer	substitution	missense	1	77	1.3	Balbás-Martínez et al., Nature Genetics, 2013
SMC3	Urothelial carcinoma	substitution	missense	2	131	1.5	Cancer Genome Atlas Research Network, Nature, 2014
STAG1	Acute myeloid leukemia	substitution, indel	missense, frameshift	7	389	1.8	Thol et al., Blood, 2014
STAG1	Bladder cancer, transitional cell carcinoma	substitution	missense	2	99	2	Guo et al., Nature Genetics, 2013
STAG1	Breast cancer	substitution	missense	1	103	1	Stephens et al., Nature, 2015
STAG1	Chronic myelomonocytic leukemia	substitution	missense	1	88	1.1	Kon et al., Nature Genetics, 2013
STAG1	Clear cell renal cell carcinoma		unreported	4	417	1	Cancer Genome Atlas Research et al., Nature, 2013
STAG1	Colorectal cancer	substitution	missense	4	72	5.6	Seshagiri et al., Nature, 2012
STAG1	Ewing's sarcoma	substitution	missense	1	112	0.9	Tirode et al., Cancer Discovery, 2014
STAG1	Multiple myeloma	substitution	missense	2	203	1	Lohr et al., Cancer Cell, 2014
STAG1	Myelodysplastic syndromes	substitution	missense	1	224	0.4	Kon et al., Nature Genetics, 2013
STAG1	Myelodysplastic syndromes	substitution	nonsense	1	386	0.3	Thota et al., Blood, 2014
STAG1	Myelodysplastic syndromes	unreported	unreported	2	944	0.2	Haferlach et al., Leukemiea, 2014
STAG1	Urothelial bladder cancer	substitution	missense	5	77	6.5	Balbás-Martínez et al., Nature Genetics, 2013
STAG1	Urothelial carcinoma	substitution	missense, exon junction	4	131	3.1	Cancer Genome Atlas Research Network, Nature, 2014
STAG2	Acute megakaryoblastic leukemia, Down syndrome-related		nonsense, splice, indel, deletion	9	49	18.4	Yoshida et al., Nature Genetics, 2013
STAG2	Acute megakaryoblastic leukemia, non-Down syndrome-related		missense, nonsense	2	19	10.5	Yoshida et al., Nature Genetics, 2013
STAG2	Acute myeloid leukemia	substitution, indel	frameshift, nonsense, splice	10	157	6.4	Kon et al., Nature Genetics, 2013
STAG2	Acute myeloid leukemia	substitution	nonsense	6	200	3	Cancer Genome Atlas Research Network, NEJM, 2013
STAG2	Acute myeloid leukemia	substitution, indel	missense, frameshift, nonsense	5	389	1.3	Thol et al., Blood, 2014
STAG2	Acute myeloid leukemia	substitution, indel	nonsense, frameshift, splice, deleetion	23	450	5.1	Thota et al., Blood, 2014
STAG2	Acute myeloid leukemia	substitution, indel	nonsense, frameshift, splice, missense, deletion	24	299	8	Lindsley et al., Blood, 2015
STAG2	Acute myeloid leukemia	substitution	nonsense	1	24	4.2	Welch et al., Cell ,2012
STAG2	Acute myeloid leukemia	indel	frameshift	1	7	14.3	Walter et al., NEJM, 2012
STAG2	Acute myeloid leukemia	substitution	splice	1	15	6.7	Walter et al., Leukemia, 2013
STAG2	Acute myeloid leukemia		deletion	1	51	2	Rocquain et al, American Journal of Hematology, 2010
STAG2	Acute myeloid leukemia	substitution	unreported	6	200	3	Kandoth et al., Nature, 2013
STAG2	Acute myeloid leukemia	unreported	unreported	6	196	3.1	Lawrence et al., Nature, 2014
STAG2	Acute myeloid leukemia, de novo	unreported	frameshift, nonsense, missense	10	197	5.1	Kihara et al., Leukemia, 2014
STAG2	Bladder cancer	substitution, indel	missense, nonsense, exon junction, framshift	25	111	22.5	Solomon et al., Nature Genetics, 2013
STAG2	Bladder cancer	substitution, indel	missense, nonsense, frameshift, deletion, splice	67	307	21.8	Taylor et al., Human Molecular Genetics, 2014
STAG2	Bladder cancer	unreported	unreported	10	99	10.1	Lawrence et al., Nature, 2014
STAG2	Bladder cancer, transitional cell carcinoma	substitution, indel	nonsense, missense, exon junction	11	99	11.1	Guo et al., Nature Genetics, 2013
STAG2	Bladder cancer, transitional cell carcinoma	CNV, deletion	deletion	5	99	5.1	Guo et al., Nature Genetics, 2013
STAG2	Bladder urothelial carcinoma	substitution, indel	unreported	10	98	10.2	Kandoth et al., Nature, 2013
STAG2	Breast cancer	substitution	nonsense	1	104	1	Stephens et al., Nature, 2016
STAG2	Breast cancer	unreported	unreported	11	892	1.2	Lawrence et al., Nature, 2014
STAG2	Chronic myelogenous leukemia	substitution	nonsense	2	64	3.1	Kon et al., Nature Genetics, 2013
STAG2	Chronic myelomonocytic leukemia	substitution, indel	frameshift, nonsense, splice	9	88	10.2	Kon et al., Nature Genetics, 2013
STAG2	Clear cell renal cell carcinoma		unreported	7	417	1.7	Cancer Genome Atlas Research et al., Nature, 2013
STAG2	Colorectal cancer	substitution	missense	4	72	5.6	Seshagiri et al., Nature, 2012
STAG2	Ewing's sarcoma	substitution	5' UTR	1	24	4.2	Solomon et al., Science, 2011
STAG2	Ewing's sarcoma	substitution, indel, duplication	nonsense, missense, exon junction, frameshift, exon duplication	19	112	17	Tirode et al., Cancer Discovery, 2014
STAG2	Ewing's sarcoma	substitution, indel, duplication	nonsense, missense, exon junction, frameshift, exon duplication, in-frame deletion	41	199	20.6	Tirode et al., Cancer Discovery, 2014
STAG2	Ewing's sarcoma	substitution, indel	exon junction, nonsense, frameshift, duplication, deletion, 3' UTR	14	65	21.5	Brohl et al., PLoS Genetics, 2014
STAG2	Ewing's sarcoma	substitution, indel	frameshift, nonsense	3	11	27.3	Agelopoulos et al., Clinical Cancer Research, 2015
STAG2	Ewing's sarcoma, pediatric	substitution, indel	nonsense, missense, exon junction, frameshift	8	96	8.3	Cromptom et al., Cancer Discovery, 2014
STAG2	Glioblastoma multiforme	substitution, indel	unreported	12	290	4.1	Kandoth et al., Nature, 2013
STAG2	Glioblastoma multiforme	unreported	unreported	12	291	4.1	Lawrence et al., Nature, 2014
STAG2	Glioblastoma	unreported	unreported	12	291	4.1	Brennan et al., Cell, 2013
STAG2	Glioma	unreported	unreported	15	820	1.8	Ceccarelli et al., Cell, 2016
STAG2	Medulloblastoma	substitution	nonsense	1	125	0.8	Jones et al., Nature, 2012
STAG2	Melanoma	indel	insertion	1	48	2.1	Solomon et al., Science, 2011
STAG2	Multiple myeloma	substitution	missense	2	203	1	Lohr et al., Cancer Cell, 2014
STAG2	Myelodysplastic syndromes	substitution, indel	frameshift, nonsense, splice	13	224	5.8	Kon et al., Nature Genetics, 2013
STAG2	Myelodysplastic syndromes	substitution, indel	missense, nonsense, frameshift, splice	30	386	7.8	Thota et al., Blood, 2014
STAG2	Myelodysplastic syndromes	substitution, indel	splice, nonsense, frameshift	9	150	6	Walter et al., Leukemia, 2013
STAG2	Myelodysplastic syndromes	unreported	unreported	71	944	7.5	Haferlach et al., Leukemiea, 2014
STAG2	Myelodysplastic syndromes/Myeloproliferative neoplasms	substitution	missense, nonsense, splice	6	169	3.6	Thota et al., Blood, 2014
STAG2	Myeloproliferative neoplasms	substitution	missense	1	77	1.3	Kon et al., Nature Genetics, 2013
STAG2	Pancreatic ductal adenocarcinoma		missense	1	50	2	Evers et al., Genome Medicine, 2014
STAG2	Renal cell carcinoma, papillary	unreported	unreported	8	157	5.1	Cancer Genome Atlas Research et al., NEJM, 2016
STAG2	Urothelial bladder cancer	substitution, indel	nonsense, missense, exon junction, frameshift	12	77	15.6	Balbás-Martínez et al., Nature Genetics, 2013
STAG2	Urothelial carcinoma	substitution, indel	nonsense, missense, exon junction, frameshift	14	131	10.7	Cancer Genome Atlas Research Network, Nature, 2014
STAG2	Urothelial carcinoma, upper urinary tract	substitution, indel	missense, nonsense, frameshift, splice	7	26	26.9	Hoang et al., Science Translational Medicine, 2013
STAG3	Colorectal cancer	substitution	missense	1	130	0.8	Barber et al., PNAS, 2008
WAPAL	Acute myeloid leukemia	substitution	missense	1	200	0.5	Cancer Genome Atlas Research Network, NEJM, 2013
WAPAL	Acute myeloid leukemia	substitution	missense	2	450	0.4	Thota et al., Blood, 2014
WAPAL	Breast cancer	substitution	missense, nonsense	2	105	1.9	Stephens et al., Nature, 2017
WAPAL	Myelodysplastic syndromes	unreported	unreported	1	944	0.1	Haferlach et al., Leukemiea, 2014
WAPL	Clear cell renal cell carcinoma		unreported	1	417	0.2	Cancer Genome Atlas Research et al., Nature, 2013
WAPL	Colorectal cancer	substitution	missense	2	72	2.8	Seshagiri et al., Nature, 2012
WAPL	Ewing's sarcoma	substitution	nonsense	1	112	0.9	Tirode et al., Cancer Discovery, 2014
WAPL	Multiple myeloma	substitution	missense	2	203	1	Lohr et al., Cancer Cell, 2014
WAPL	Urothelial bladder cancer	substitution	missense	1	77	1.3	Balbás-Martínez et al., Nature Genetics, 2013
WAPL	Urothelial carcinoma	substitution	missense	3	131	2.3	Cancer Genome Atlas Research Network, Nature, 2014