The three-dimensional cancer genome

M. Ryan Corces¹ and Victor G. Corces²

¹Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA 94305, USA
²Department of Biology, Emory University, 1510 Clifton Rd NE, Atlanta, GA 30322, USA

Abstract

The past decade of cancer research has ushered in a comprehensive understanding of the way that the sequence of the genome can be coopted during the process of tumorigenesis. However, only recently has the epigenome, and in particular the three-dimensional topology of chromatin, been implicated in cancer progression. Here we review recent findings of how the cancer genome is regulated and dysregulated to effect changes in 3D genome topology. We discuss the impact of the spatial organization of the genome on the frequency of tumorigenic chromosomal translocations and the effects of disruption of the proteins responsible for the establishment of chromatin loops. Alteration of the three-dimensional cancer genome is a rapidly emerging hallmark of multiple cancer subtypes.

Keywords

cancer; chromatin; epigenetics; 3D architecture; CTCF; cohesin

INTRODUCTION

Cancer is largely associated with the sequential acquisition of mutations in a single lineage of cells that ultimately leads to unrestrained proliferation. The foundations of cancer biology were laid through the discovery of oncogenes and tumor suppressors with canonical roles in proliferation and cell cycle control. Decades of research have elucidated the major drivers and the genetic mechanisms responsible for tumorigenesis, identifying point mutations and small-scale alterations that directly affect individual proteins in a one-dimensional fashion. The contribution of the epigenome to this process has become more apparent with the discovery of mutations in genes known to regulate DNA methylation and histone modification. These mutations are largely considered to affect gene expression in a two-dimensional fashion through the modulation of transcription factor recruitment. However, the precise mechanisms by which these alterations of the epigenome contribute to cancer progression has remained elusive. Only recently has the three-dimensional context of the
The genomes of higher eukaryotes are packaged into exquisitely organized hierarchical structures. Linear DNA is wrapped around histone proteins forming the 10 nm nucleosomal fiber which is subsequently folded in three-dimensions to create loops of DNA that form discrete neighborhoods of genes at the sub-megabase level [5,6]. These neighborhoods are formed through the action of multiple proteins including CCCTC-binding factor (CTCF) and the cohesin complex [7,8]. Groups of gene neighborhoods are further organized into large, isolated, megabase structures termed topologically associating domains (TADs) [9–13]. Each of these layers of organization have pronounced effects on gene expression and the control of cell identity and cell fate. The mechanisms by which these three-dimensional interactions are manipulated and coopted in the context of cancer are the subject of this review (Figure 1).

**Hijacking topology: the contribution of normal DNA architecture to carcinogenesis**

In the absence of alteration of the topological structure of the three-dimensional genome, recent work has demonstrated that the normal organization of the genome predisposes certain cell types to the acquisition of specific cancerous lesions. Chromosomal rearrangements, such as translocations, require the formation and incorrect resolution of DNA double strand breaks. Higher order chromatin structure has been shown to play a role in the formation of specific translocations through the spatial coordination of otherwise unrelated DNA sequences. Through the development of sequencing technologies to capture translocation junctions in B lymphocytes, multiple groups have shown that translocations between pairs of DSBs occurring on the same chromosome are strongly preferred over interchromosomal events [1,2]. The constrained physical proximity of intrachromosomal interactions implies that spatial organization of chromosomes influences the translocation process. In line with this hypothesis, modeling of somatic copy-number alterations and genome-wide chromosome conformation capture suggests that the distribution of chromosomal alterations is spatially related to three-dimensional genomic architecture [14]. These results were confirmed by direct comparisons of translocation frequency and spatial proximity of interchromosomal interactions [15,16]. Cumulatively, these studies show that the three-dimensional proximity between two loci is directly proportional to the likelihood of translocation. When combined with a requirement for positive selection in cancer cells, these observations explain the frequency of recurrent translocations such as BCR-ABL and MYC-IGH.

**Mutation or genetic alteration of genome organization components**

In addition to serving as a template for large-scale chromosomal aberrations, the three-dimensional genomic architecture is often perturbed in cancer through genetic alteration of the proteins involved in the establishment and maintenance of chromatin interactions. In particular, the cohesin complex, a multimeric ring structure involved in mediating looping interactions, has been found to be mutated in a wide variety of cancers. In addition to participating in three-dimensional looping, the cohesin complex is also involved in sister chromatid segregation during mitosis. Given the well-established role for aneuploidy in...
cancer, it has been hypothesized that mutation of the cohesin complex would contribute to carcinogenesis through the mis-segregation of chromosomes [17,18]. While this may be true in some cases, recent work has made it clear that genetic disruption of the cohesin complex plays a much more subtle role in cancer formation[19–22].

The first evidence that genetic mutation of the cohesin complex may not result in aneuploidy came from the genetic characterization of acute myeloid leukemia (AML), an aggressive malignancy of the bone marrow. AML is one of the most genetically stable adult cancers with minimal aneuploidy and an average of 10–15 coding mutations per patient [23,24]. Cohesin complex mutations were first described in AML in 2012, occurring in approximately 13% of patients [25–27]. These mutations occur in all four members of the cohesin complex (STAG2, SMC3, SMC1A, and RAD21) and are typically missense or truncating mutations. The spectrum of mutations observed implies a loss of function mechanism which is consistent with the finding that cohesin mutated leukemia cells have reduced levels of chromatin-bound cohesin components [3]. Recent work has demonstrated that mutations in the cohesin complex impair hematopoietic progenitor differentiation [19–21], suggesting a clear mechanism by which the cohesin complex may play a role in cancer progression in the absence of aneuploidy.

In addition to AML, the cohesin complex has been found to be mutated in multiple other types of cancer. Most prominently, mutations in STAG2 occur in 20–30% of urothelial bladder carcinoma and are not associated with aneuploidy [22,28–30]. Mutations in the cohesin complex have also been found in glioblastoma [31], medulloblastoma [32], breast cancer [33], pancreatic ductal adenocarcinoma [34], and Ewing sarcoma [35].

Beyond the cohesin complex, recurrent mutations have been observed in CTCF and other cohesin-interacting proteins. The first report of CTCF missense mutations in cancer identified multiple zinc finger domain mutations in breast, prostate, and Wilms’ tumors [36]. These missense mutations each selectively altered CTCF binding to a subset of target sites but did not completely abrogate DNA binding by CTCF. These results imply that selective alteration of chromatin architecture, perhaps in a cell type-specific manner, can play a causative role in cancer development. Notably, point mutations and copy number loss of CTCF are commonly observed in breast cancer [37], prostate cancer [37], and endometrial cancer [38], implicating a haploinsufficient phenotype for CTCF. Indeed, mouse models of CTCF haploinsufficiency indicate a strong predisposition to cancer with 80% of Ctcf heterozygous knockout mice succumbing to cancer by 100 weeks of age compared to only 40% of wildtype littermates. This 50% reduction in Ctcf gene dosage has profound effects on DNA methylation, suggesting a role for CTCF in maintaining the stability of global cytosine methylation [39].

Despite the abundance of cancerous mutations detected in the genes known to regulate chromatin topology, the precise mechanism of action of these mutations remains elusive. Many studies have addressed the consequence of loss of cohesin or CTCF via knockdown or knockout in post-mitotic cells showing widespread disruption of long-range interactions and concomitant changes in the expression of nearby genes [40–44]. However, cancer-associated mutations in these proteins are often heterozygous with the mutated allele expressed,
indicating that a reduction in wildtype protein levels by knockdown or knockout may not phenocopy a heterozygous mutation. Future work investigating the consequences of cohesin complex or CTCF mutation on three-dimensional chromatin architecture of cancer cells will provide key insights into the mechanism of action of these mutations. It remains unclear how these mutations confer a carcinogenic phenotype and whether all mutations in genes regulating the three-dimensional genome have the same mechanistic effect.

**Genetic and epigenetic dysregulation of chromatin architecture**

In the absence of direct disruption of chromatin organizers, genetic or epigenetic dysregulation of the non-coding genome can have profound effects on chromatin architecture. In particular, changes in the sequence or epigenetic milieu of transcription factor binding sites can lead to alterations in chromatin interactions which have broad-reaching effects on gene expression and cellular identity. Greater than 95% of genome-wide association study-identified SNPs are located in intergenic regions and more than 75% associated with DNase I-hypersensitive sites, indicating a strong link to regulatory elements [45]. Additional studies have linked these disease-associated polymorphisms in non-genic regions with regulatory elements involved in chromatin organization and looping [46–47]. Similarly, recent work studying colorectal cancer (CRC) has shown that certain genetic subtypes of CRC are characterized by a predominance of mutations at CTCF binding sites [48]. This enrichment for mutations at CTCF binding sites was only observed in the context of simultaneous cohesin binding, implicating a specific subset of CTCF binding sites in the pathogenesis of CRC. Moreover, these CTCF binding site mutations were highly associated with AT>GC mutations, previously identified as a unique mutational signature in cancer [49], and are enriched at specific positions in the CTCF consensus sequence. Importantly, as few as two SNPs in a CTCF binding site lead to complete abrogation of CTCF binding [50]. Across multiple patient samples, CTCF binding site mutations display a unimodal distribution whereby a small number of patients account for a majority of the mutations in CTCF binding sites. On a more global scale, CTCF binding site mutations are coupled to late replication timing domains and previous studies have shown that these CTCF/cohesin binding sites are not replicated by the leading strand DNA polymerase Pol ε but by another uncharacterized polymerase [44]. Taken together, these results imply that a subset of CRC patients may have global defects in the repair of mutations in CTCF binding sites. Whether these mutations are causative of or merely correlated with cancer progression remains to be shown. Importantly, this mutational signature was not unique to CRC and was observed in multiple other cancer types, suggesting a more universal role for dysregulation of CTCF binding in the pathogenesis of cancer.

CTCF binding patterns can also be influenced by epigenetic modification of its binding sites. DNA methylation of the CTCF consensus binding sequence has been shown to control cell type-specific CTCF binding [51–52], indicating that CTCF occupancy can be readily modulated by reversible epigenetic alterations. As disruption of DNA methylation is a hallmark of multiple types of cancer [53–55], it is possible that changes in DNA methylation directly or indirectly affect CTCF binding. One study of IDH mutant glioma has linked hypermethylation of CTCF binding sites to dissolution of important domain boundaries and aberrant expression of powerful oncogenes [4]. Future, more in-depth studies of CTCF
binding in cancer subtypes with altered DNA methylation will help to answer these questions and elucidate other potential mechanisms by which chromatin architecture is disrupted in cancer.

**How could changes in chromatin architecture mechanistically lead to cancer?**

While recent years of research have enumerated multiple examples of how chromatin architecture is dysregulated in cancer, no studies have provided direct mechanistic insight into how this dysregulation is translated into phenotypes associated with cancer. Multiple models can be posited but no direct evidence exists to distinguish these possibilities (Figure 2).

**Model 1 - Dysregulation of chromatin architecture prevents cell state changes**

—Mutation of components of the chromatin organization machinery such as CTCF and cohesin may not create new functional states but may, instead, prevent cells from changing states. In the context of cancer, acquisition of a CTCF or cohesin mutation in a stem cell would prevent that cell from differentiation which would increase the likelihood of acquiring additional mutations and potentially bestow a self-renewal phenotype to progeny cells. Indirect evidence supporting this model exists in AML [19–21] whereby mutations in the cohesin complex members lead to defects in differentiation. Mechanistically, this model may be possible through the action of pioneer factors. In the absence of proper CTCF or cohesin function, global chromatin accessibility decreases [44], implying that the only transcription factors capable of binding DNA may be pioneer factors that can bind condensed chromatin. In the context of a stem cell that acquires a mutation in CTCF or cohesin, the expressed pioneer factors would control a stem cell state and would perpetuate that state in the absence of additional changes to the chromatin architecture.

**Model 2 - Inappropriate 3D looping and insulation alter the cis-regulation of key genes and create neomorphic cell states**

—Disruption of the components maintaining chromatin architecture may lead to novel combinations of expressed and repressed genes and contribute to cancer development through the generation of neomorphic cell states. In depth studies of the binding sites and interactions of CTCF and cohesin in embryonic stem cells has shown that super enhancer-associated genes with important functions for cell identity exist in insulated neighborhoods created by looping interactions between two CTCF/cohesin binding sites [56]. Additionally, repressed lineage-specifying developmental regulators are also found in separate insulated neighborhoods. The integrity of these activational and repressive insulated neighborhoods is critical for the proper expression and repression of nearby genes that exist on the outskirts of these neighborhoods. Disruption of the neighborhood boundaries by mutation of CTCF binding sites or by dysregulation of cohesin or CTCF through mutational inactivation would lead to inappropriate expression and/or repression of key developmental genes [56] which could create unnatural cell states that have the potential to cause cancer. One example supporting this model has been demonstrated in glioblastoma whereby hypermethylation of CTCF binding sites leads to reduced CTCF binding at specific domain boundaries. In this study, the loss of one such boundary enables a constitutive enhancer to interact aberrantly with the PDGFRA gene, a prominent oncogene in glioma [4].
Model 3 - Disruption of chromatin architecture increases the epigenetic search space probed by cells and increases the likelihood of developing cancer—The loss of proper DNA repair leads to the acquisition of many more mutations, thus increasing the mutational search space of a given cell [57]. Similarly, disruption of the chromatin architecture of a cell may have pseudo-random effects on gene expression and repression. New chromatin loops may be established and old loops destroyed due to the random nature of whether a functional or non-functional CTCF or cohesin protein is recruited to a given chromosomal location. This increases the epigenetic search space of these cells as they probe pseudo-random chromatin configurations until acquiring an evolutionarily advantageous cell state that can be secured through positive selection. In this way, dysregulation of chromatin architecture may contribute to cancer by increasing epigenetic variability.

Ultimately, these models attribute the consequences of topological alterations to changes in gene expression. It has been well established that changes in gene expression can lead to increased proliferation and decreased differentiation, two key hallmarks of cancer [58]. Moreover, a causative link between changes in chromatin topology and changes in gene expression has been established through multiple lines of evidence [4,59].

DISCUSSION

The intricate interplay between chromatin architecture and cell identity has been extensively explored in the context of healthy cells. However, the involvement of three-dimensional chromatin organization in the pathogenesis of cancer has only recently been acknowledged. Work elucidating the interplay of spatial proximity and frequency of chromosomal translocations has enhanced our understanding of how the natural organization of chromatin can be co-opted to generate recurrent translocations directly responsible for cancer progression. Moreover, high-throughput sequencing efforts have identified mutations in the genes encoding for components of the chromatin organization machinery, such as CTCF and cohesin, as well as mutations in the sites bound by these factors. These genetic studies have shown that dysregulation of chromatin architecture may be a central hallmark of tumorigenesis in multiple cancer types. Nevertheless, it is still unclear precisely how, for example, a mutation in the cohesin complex affects chromatin organization and contributes to the pathogenesis of cancer. While it is known that programmed changes in genome topology occur during the normal process of differentiation [5,12,60,61], little is known about how genome organization changes in the setting of cancer. Recent work has shown that tumor cells exhibit a similar overall genomic architecture to their normal cell counterparts with TAD and sub-TAD compartments; however, characteristic and important local differences exist [4,13,62,63]. These minor differences may hold the key to understanding the epigenetics of cancer. As genome wide techniques for assaying chromatin conformation become more feasible and widely applied to the study of primary patient cancers, the answers to these questions will become clearer.

Acknowledgments

We would like to thank Brook Barajas and Adam Rubin for thoughtful discussion and critical review of this work. M.R.C. acknowledges funding from the National Institutes of Health training grant R25CA180993. V.G.C.
acknowledges support from U.S. Public Health Service Award R01GM035463 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES


4. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2015 Using integrative epigenomic profiling of IDH wild type and IDH mutant glioma samples, Flavahan et al. demonstrate that hypermethylation of specific CTCF sequences leads to loss of CTCF at a domain boundary, allowing for aberrant constitutive expression of PDGFRA, a known glioma oncogene. This is the first study to demonstrate a direct connection between a well-studied cancer mutation and disruption of chromatin architecture in a primary patient sample.


19. Mazumdar C, Shen Y, Xavy S, Zhao F, Reinish A, Li R, Corces MR, Lynn RA, Buenrostro JD, Chan SM, et al. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. Cell Stem Cell. 2015 In this study, Mazumdar et al. provide one of the first in-depth phenotypic characterizations of a cancerous mutation in a protein involved in chromatin organization. The authors demonstrate that mutation of various components of the cohesin complex in acute myeloid leukemia leads to defects in differentiation, linking chromatin architecture to a carcinogenic phenotype for the first time.
30. Taylor CF, Platt FM, Hurst CD, Thysgesen HH, Knowles MA. Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to

47. Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, Greenside PG, Srivas R, Phanstiel DH, Pekowska A, et al. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. Cell. 2015; 162:1–15. By integrating ChIP-seq data for three histone marks with Hi-C and ChIA-PET-based chromatin contact maps from lymphoblastoid cell lines, the authors discover a large number of histone quantitative trait loci. These loci are enriched for common sequence variants associated with autoimmune diseases, enabling the prediction of target genes of disease-associated variants from GWAS studies. Such techniques could be applied to cancer data in the future.


56. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, et al. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. Cell. 2014; 159:374–387. [PubMed: 25303531] Through the use of cohesin ChIA-PET in embryonic stem cells, the authors demonstrate that key cell identity genes and their enhancers are contained within insulated neighborhoods flanked by interacting CTCF sites. These results indicate that disruption of these insulated neighborhoods would have profound effects on local gene expression and serve as the basis for Model 2 of this review.


Through the study of embryonic stem cells and four separate embryonic stem cell-derived lineages, Dixon et al. demonstrate that 36% of the active and inactive chromosomal compartments change during differentiation. This work demonstrates that chromatin architecture is rapidly and coordinately changed during normal differentiation.

Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol Syst Biol. 2015; 11:1–14.


Chromatin architecture has been implicated in the pathogenesis of cancer through multiple lines of evidence. The normal topology of the genome has been shown to predispose certain cell types to the acquisition of certain chromosomal translocations such as the MYC/IgH translocations associated with certain types of lymphoma (left panel). The genetic mutation of components involved in chromatin looping, including the cohesin complex (red/green/blue ring) and CTCF (orange squares), has also been observed in many cancer types (center panel). These mutations (illustrated by *) likely cause quantitative changes in factor binding, illustrated here by a change in ChIP-seq signal. Lastly, disruption of regulatory regions that serve as the anchors for looping machinery has been identified in cancer and multiple other diseases (right panel).

Figure 1. Contribution of chromatin architecture to our understanding of cancer

Chromatin architecture has been implicated in the pathogenesis of cancer through multiple lines of evidence. The normal topology of the genome has been shown to predispose certain cell types to the acquisition of certain chromosomal translocations such as the MYC/IgH translocations associated with certain types of lymphoma (left panel). The genetic mutation of components involved in chromatin looping, including the cohesin complex (red/green/blue ring) and CTCF (orange squares), has also been observed in many cancer types (center panel). These mutations (illustrated by *) likely cause quantitative changes in factor binding, illustrated here by a change in ChIP-seq signal. Lastly, disruption of regulatory regions that serve as the anchors for looping machinery has been identified in cancer and multiple other diseases (right panel).
Figure 2. Models for the mechanistic underpinnings of how disruption of chromatin architecture can contribute to cancer progression

Disruption of the chromatin architecture machinery may abrogate a cell’s ability to establish new chromatin loops and thus inhibit it from changing cellular state (left panel). Alternatively, changes in chromatin domains caused by dysregulation of looping proteins (red/green/blue rings represent the cohesin complex; orange squares represent CTCF) could lead to neomorphic transcriptional states which could be tumorigenic (center panel). In this example, truncation of one member of the cohesin complex (*)& leads to dissolution of stabilized loops and inappropriate interaction of a super-enhancer with a previously silent gene. Lastly, pseudo-random cell states could be created through the inactivation of chromatin architecture machinery leading to an increased epigenetic search space. Some of these epigenetic configurations may be more advantageous or may facilitate transformation (right).