Chromatin architecture aberrations in prostate cancer and leukemia

by

James Hawley

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Graduate Department of Medical Biophysics
University of Toronto

Contents

Chapter 1

Introduction

Cancer is one of the largest causes of death worldwide, ranking in the top ten most frequent causes in over 150 countries and most frequent in over 40 [1]. Disease treatment is complicated by the fact that cancers are a myriad of diseases with unique origins, symptoms, and treatment options, often related to the cell of origin [2]. However, numerous hallmarks of cancers have emerged over the last 50 years to provide understanding about what biological aberrations cause tumours to initiate, how they develop over time, and how they respond to therapeutic interventions [3–6] (Figure 1.1).

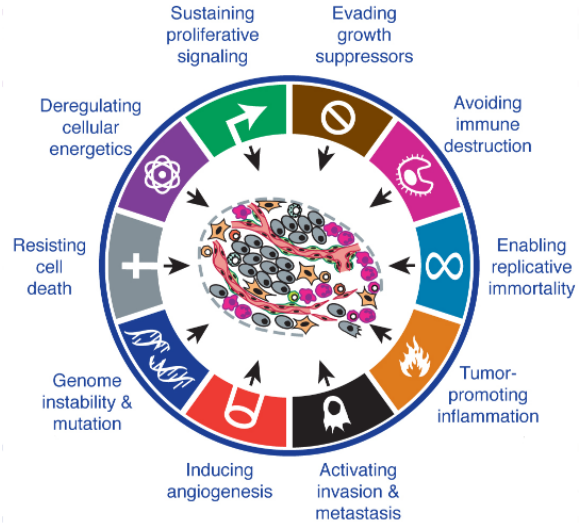

Figure 1.1: **The hallmarks of cancer.** Adapted from [REF 4].

Many of these hallmarks of cancer can be achieved through aberrations to the genome and the molecular machinery that enables cells to function normally [7]. For example, genome instability can be achieved by inhibiting deoxyribonucleic acid (DNA) repair machinery, as is observed with

abnormalities in *MLH1* and *MSH2* repair genes in colorectal cancers [8] or mutations to *BRCA1*, *BRCA2*, and *ATM* genes in prostate cancer (PCa) [9]. Similarly, replicative immortality can be achieved through telomere elongation by over-expression of the *TERT* gene [10]. Mutations to the *TERT* promoter, resulting in its over-expression, were first identified in melanomas [11, 12], but have since been further identified in bladder, thyroid, and brain cancers [10, 13, 14]. But while cancer has long been viewed as a disease of the genome [3, 7], there are many avenues cells can take to arrive these hallmarks resulting from aberrations of how genes are expressed inside the cell nucleus.

1.1 Normal chromatin architecture in mammalian cells

Genes, encoded as DNA, are expressed by being transcribed into ribonucleic acid (RNA) and subsequently translated into proteins in the process known as the Central Dogma of molecular biology [15] (Figure 1.2a). The transcription of genes into messenger RNA (mRNA) requires RNA polymerase to bind at transcription start sites (TSSs) within DNA elements found at the beginning of genes, termed promoters [16]. Promoters are one example of a class of DNA elements, termed *cis*-regulatory elements (CREs) because of their roles in regulating the expression of genes on the same strand of DNA. The recruitment of RNA polymerase is aided by a special class of proteins, termed transcription factors (TFs), that can bind at DNA sequences either close to a gene's promoter, or far from it at other CREs such as enhancers and insulators [17–22] (Figure 1.2b). Together, the binding of TFs to the DNA at specific CREs is fundamental for to initiating transcription and expressing genes.

1.1.1 DNA elements and features regulating transcription

The ability of TFs to bind at specific CREs is dependent on multiple features of the DNA. Many TFs bind to DNA at specific sequences, termed motifs [18, 23]. The structural protein CCCTC-binding factor (CTCF) has a well-defined motif and binds to this sequence at thousands of locations across the human genome [24, 25]. Mutations to the sequence motif can alter CTCF's binding affinity for DNA, as is the case with many TFs [26–28]. Relying on more than just the genetic sequence, CTCF is also an example of a TF that is sensitive epigenetic features such as DNA methylation (DNAme), the addition of a methyl group to DNA nucleotides [29–33], as are DNA methyltransferases DNMT1, DNMT3A, and DNMT3B [34, 35]. TF binding to DNA can also be affected by the presence of other proteins at binding sites. TFs can bind in a combinatorial manner at

the same location [18, 19, 23] or be blocked from binding altogether by the presence of nucleosomes, protein complexes that DNA winds around to make it compact in three-dimensional space [36, 37]. The collection of DNA, nucleosomes, DNA-bound transcription factors, and chemical modifications is defined as the chromatin, and the presence and density of nucleosomes, as well as DNA coiling, make certain segments of the chromatin more or less accessible for TF binding (euchromatin and heterochromatin, respectively). This can affect normal cellular behaviour such as cell-type-specific gene expression [38, 39] and DNA damage repair in inaccessible regions [40]. Thus, both genetic and epigenetic chromatin features affect how TFs can bind and regulate transcription.

In addition to TF binding, transcription regulation depends on the ability of CREs to localize together in three-dimensional space across large genomic distances [41–43] (Figure 1.2c). Localization of CREs tens to thousands of basepairs (bps) apart form focal interactions is aided by the formation of topologically associated domains (TADs), domains of chromatin whose boundaries are linked by structural proteins, including CTCF and cohesin [22, 44–46]. In addition to TADs which can range in size from $10^4 - 10^6$ bp, chromatin is also organized into active or inactive compartments (A and B compartments, respectively) that range in size from $10^5 - 10^6$ bp [22, 47–49]. These two modes of chromatin organization facilitate the proper localization of CREs and TFs at the right time. While TADs and compartments are largely conserved across cell types [25, 50, 51], focal chromatin interactions can differ up to 45 % between cell types, providing a further mechanism to change chromatin state [48]. Different chromatin states enable cells with the same DNA sequence to express genes differently [17, 19, 44, 52–54], and thus identifying the repertoire of CREs, chromatin interactions, TADs, and compartments are vital in determining the regulation of genes in various cell types.

1.1.2 Methods for identifying DNA elements and chromatin interactions

High throughput sequencing protocols have enabled the characterization of functional elements from across the genome and rely on a similar concept to do so. This concept is to take a molecular feature of interest, be it an RNA transcript or nucleosome position, associate it with a short fragment of DNA, sequence these DNA fragments, and map it to the reference genome to identify where the original molecules came from (Figure 1.3). RNA sequencing (RNA-seq) methods reverse transcribed RNA into DNA that map back to individual genes, with the abundance of fragments indicating how much the gene is expressed [55]. Protein binding sites and histone post-translational modifications can be identified by fragmenting DNA around antibodies that bind to these proteins with techniques like chromatin immunoprecipitation sequencing (ChIP-seq) and cleavage under targets and release

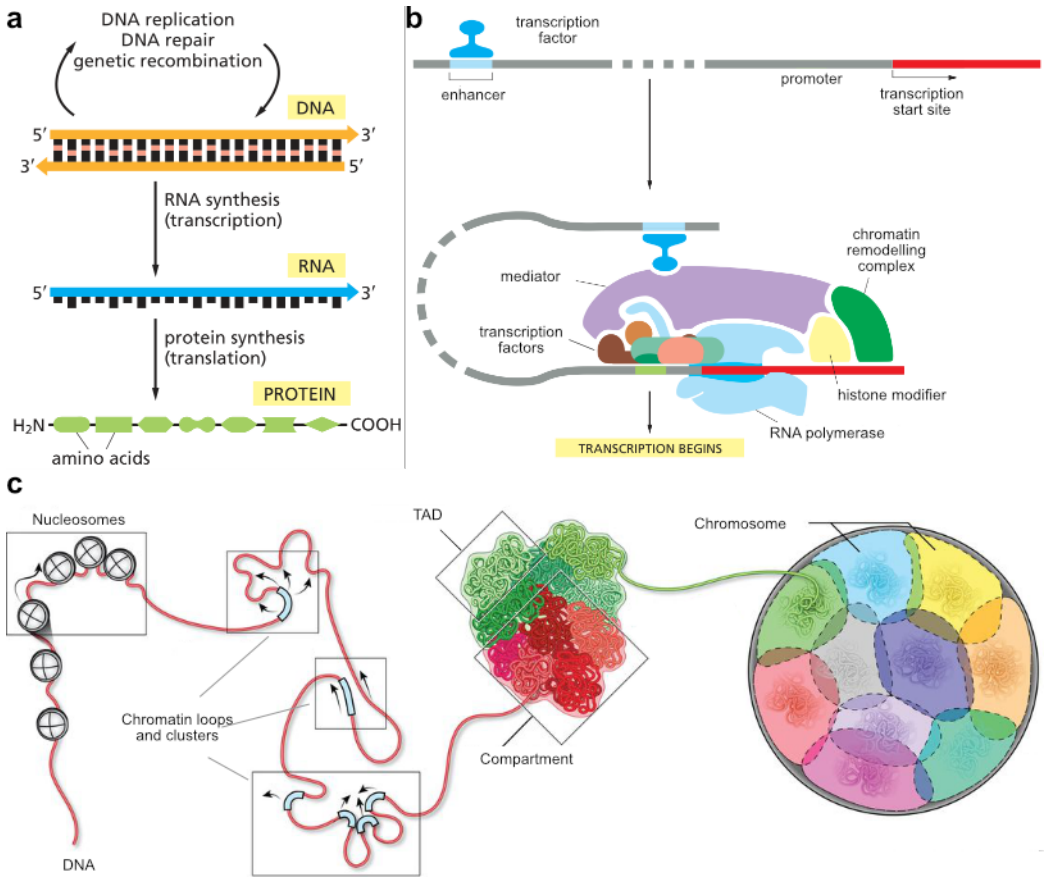

Figure 1.2: **The basics of gene expression inside the nucleus. a.** The central dogma of molecular biology. Adapted from [REF 15]. **b.** Schematic of the transcription machinery to initiate transcription. Adapted from [REF 15]. **c.** The scale of chromatin interactions across length scales. Adapted from [REF 46].

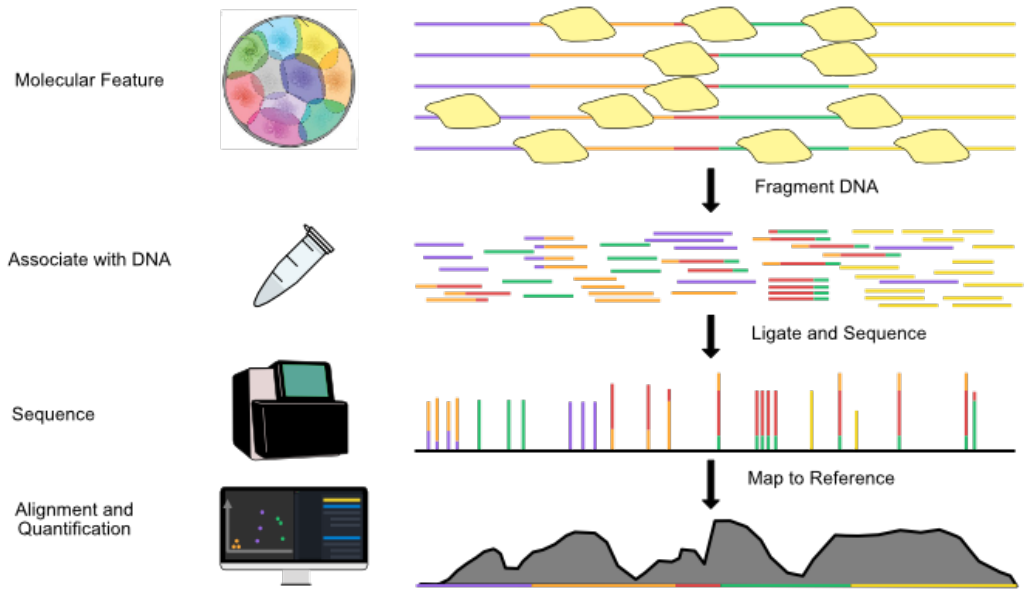

Figure 1.3: **Characterizing functional DNA elements with high throughput sequencing.**

using nuclease (CUT&RUN) [56–58]. Accessible and inaccessible chromatin can be assessed by the chromatin's propensity to be cut by enzymes like DNase I, Tn5 transposase, and micrococcal nuclease in DNase I hypersensitive sequencing (DNase-seq), assay for transposase-accessible chromatin sequencing (ATAC-seq), and micrococcal nuclease sequencing (MNase-seq) protocols, respectively [59–63]. DNAme can be measured with bisulfite-sequencing assays [64], and distal chromatin interactions can be identified with chromatin conformation capture (3C) and 3C-based methods such as Hi-C [25, 47, 48, 65, 66]. Yet while these measurements help in identifying candidate CREs and important regions of the genome, determining their function and which target genes they regulate is a further complicating problem.

Varying chromatin states across cell types means that multiple measurements across multiple cell types are necessary to understand the breadth of functions a single CRE may have. In 2007, the ENCODE Project aimed to catalogue all biochemically functional elements in the human genome to better understand all the ways genes are expressed and how they are regulated in different cell types [67, 68]. Using these genome-wide sequencing techniques across a variety of human cell lines and tissues, the ENCODE Project has since catalogued nearly 10^6 candidate CREs, comprising nearly 8 % of the human genome [68]. Interpreting this data requires computational methods to correlate and interpret measurements across samples. Genome segmentation methods such as ChromHMM [69] and Segway [70, 71] classify genomic regions according to their predicted function which can be validated with *in vitro* or *in vivo* experiments. Many techniques for experimental validation,

including clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9, small interfering RNA (siRNA), and small hairpin RNA (shRNA), can interfere with candidate CREs by deleting them from the genome, preventing TFs from binding to the chromatin, or preventing translation of mRNA transcripts into proteins [72, 73]. These same techniques can also be used to screen for candidate CREs themselves, through massively-parallel reporter assays (MRPAs) and CRISPR screens [73], necessitating their own suite of statistical and software tools for analyzing observations. Altogether, a collection of experimental and computational techniques enable the cataloguing and interpretation of thousands of CREs and chromatin interactions across many cell types. These catalogues facilitate understanding how genes are expressed within the complex chromatin architecture in normal cells and, importantly, how aberrations to this architecture can result in disease.

1.2 Aberrations to chromatin architecture in cancer

1.2.1 Genetic aberrations in cancer

Discovery of genetic mutations of oncogenes in tumours nearly 50 years ago spurred the widespread characterization of genetic aberrations in cancers [74–77]. These mutations occur within genic regions that code for proteins, but more than 98 % of somatic mutations acquired in tumours are found in non-coding regions [78]. Single nucleotide variants (SNVs), copy number variants (CNVs), and structural variants (SVs) are found throughout the genome, and interpreting the impact of these mutations on cancer is an active area of research [68, 77, 79, 80]. Analysis of recurrent somatic mutations in tumours led to the identification of *TP53* as a tumour suppressor gene [81], the frequently mutated *SPOP* gene to help define a molecular subtype of prostate tumours [82], and the interpretation of recurrent rearrangements of the proto-oncogene *MYC* in multiple cancers [83]. The impact of a mutation can also be predicted by identify overlapping regulatory elements or TF binding sites [27, 84, 85]. Grouping CREs by their putative target genes led to the identification of the *ESR1* gene as having its gene regulatory network recurrently mutated in ~10 % breast cancers, resulting in its over-expression, despite the gene itself being mutated in ~1 % of breast cancers [86]. Similarly, the binding sites of the *FOXA1*, *HOXB13*, *AR*, and *SOX9* TFs are enriched with mutations affecting their binding affinities [87] and recurrent amplifications of enhancers near the *AR* and *FOXA1* genes are associated with increased rates of metastasis [88, 89]. Furthermore, mutations that do not directly target gene bodies or CREs can lead to oncogene over-expression. Multiple non-coding SVs in pediatric medulloblastoma patients were found to bring the *GFI1* and *GFI1B* oncogenes proximal

to enhancer clusters, causing the oncogenes to become aberrantly regulated by this enhancer cluster [90]. This mechanism of enhancer hijacking has also been observed in developmental diseases [91, 92]. While this is not an exhaustive list, it is clear that genetic aberrations are abundant in cancers and that integrating genetic information with other components of the chromatin architecture can help identify driver events that promote oncogenesis or aggressive disease.

Mutations to DNA methyltransferases and chromatin remodelling proteins are common in cancers, and the impact of these mutations can be observed in their chromatin state. The isocitrate dehydrogenase (*IDH*) enzymes *IDH1*, *IDH2*, and the ten-eleven translocation (*TET*) enzymes *TET1* and *TET2* are frequently mutated in cancers, most often in leukemias and gliomas [93–97]. These mutations often affect the DNAme profiles of tumours and differentiation programs [93], such as loss of enhancer hydroxymethylation and germinal centre hyperplasia in diffuse large B-cell lymphoma (DLBCL) [98]. Similarly, mutations to the *EZH2* gene in leukemias can affect the ability of the EZH2 protein to write the H3K27me3 histone mark [99–102] and *EZH2* over-expression is associated with poor survival in PCa [103–106]. Together, these findings show that genetic aberrations to genes regulating other aspects of the chromatin architecture are abundant in multiple cancers and can drive specific programs in tumours. These programs can, in turn, affect progression of the disease and treatment strategies for patients. Importantly, the impact of these mutations is dependent on the function of the affected protein or CRE, which varies between different cancers. Thus, understanding how non-genetic aberrations affect tumours can be a vital step in understanding the impact of genetic aberrations.

1.2.2 Non-genetic aberrations in cancer

DNA methylation aberrations

Histone modifications

Chromatin accessibility and nucleosome positioning

Enhancer hijacking and chromatin organization aberrations

1.3 Interactions between chromatin architecture components in prostate cancer and leukemia

1.4 Dissertation structure

I begin with **??** by exploring the *cis*-regulatory landscape of PCa and delineating the CREs of the prostate oncogene *FOXA1*. I demonstrate the essentiality of *FOXA1* for prostate tumours, identify putative CREs based on integration of multiomic datasets in PCa cell lines, and assess the functional impact of recurrent PCa SNVs on *FOXA1* expression and TF binding.

With the *cis*-regulatory network of *FOXA1* established in PCa, I attempt to construct the *cis*-regulatory landscape genome-wide in PCa with 3C mapping in **??**. Using Hi-C, I characterize the three-dimensional chromatin organization of PCa and investigate changes to this structure over oncogenesis, and explore the relationship between chromatin organization, SVs, and CRE hijacking.

In assessing the impact of SVs on chromatin organization, I uncovered a statistical problem stemming from the lack of recurrent SVs across PCa patients, leading to unbalanced experimental comparisons. To address this problem, I developed a statistical method for reducing error in gene expression fold-change estimates from unbalanced experimental designs in **??** and characterize the method.

Given the shared importance of mutations to TFs and epigenetic enzymes in prostate cancer and leukemias, in **??** I explore the epigenetic landscape of B-cell acute lymphoblastic leukemia (B-ALL) and its relapse after treatment. I characterize molecular changes to B-ALL tumours over the course of disease relapse and identify important changes to DNAme that indicate the reversion to a stem-like phenotype, often present in a subpopulation of cells at diagnosis.

Together, this dissertation investigates the multiple layers of the chromatin architecture that contribute to oncogenesis and cancer progression. I demonstrate that aberrations to the genome,

epigenome, and three-dimensional organization of chromatin play important roles individually, and together, in the orchestration of the disease.

Glossary

3C chromatin conformation capture

AML acute myeloid leukemia

ANOVA Analysis of Variance

AR androgen receptor

ATAC-seq assay for transposase-accessible chromatin sequencing

B-ALL B-cell acute lymphoblastic leukemia

bp basepair

cDNA complementary DNA

ChIP-seq chromatin immunoprecipitation sequencing

CLL chronic lymphocytic leukemia

CML chronic myeloid leukemia

CMP common myeloid progenitor

CNV copy number variant

CPC-GENE Canadian Prostate Cancer Genome Network

CpG CG dinucleotide

crRNA CRISPR RNA

CRE *cis*-regulatory element

CRISPR clustered regularly interspaced short palindromic repeat

CTCF CCCTC-binding factor

CUT&RUN cleavage under targets and release using nuclease

DEPMAP Cancer Dependency Map

DHS DNase I hypersensitive sites

DLBCL diffuse large B-cell lymphoma

DMR differentially methylated region

DNA deoxyribonucleic acid

DNAme DNA methylation

DNase-seq DNase I hypersensitive sequencing

dRI disease relapse-initiating

Dx diagnosis

EarlyProB early progenitor B cell

FDR false discovery rate

FN false negative

FP false positive

FOX forkhead box

GLM generalized linear model

GMP granulocyte-macrophage progenitor

GO gene ontology

gRNA guide RNA

HSC hematopoietic stem cell

HSPC hematopoietic stem and progenitor cell

IDH isocitrate dehydrogenase

IID independent and identically distributed

JS James-Stein

KO knockout

LDA limiting dilution assay

LMPP lymphoid-primed multi-potent progenitor

MeCapSeq DNA methylation capture sequencing

MEP megakaryocyte-erythrocyte progenitor

MNase-seq microccocal nuclease sequencing

MSE mean square error

mCRPC metastatic castration-resistant prostate cancer

MDS myelodisplastic syndrome

MLP monocyte-lymphoid progenitor

MPP multi-potent progenitor

MRPA massively-parallel reporter assay

NSG NOD scid gamma

OLS ordinary least squares

mRNA messenger RNA

PCa prostate cancer

PDX patient-derived xenograft

PreProB pre-progenitor B cell

ProB progenitor B cell

Rel relapse

RNA ribonucleic acid

RNAi RNA interference

RNA-seq RNA sequencing

shRNA small hairpin RNA

siRNA small interfering RNA

SNV single nucleotide variant

SRA Seqeuence Read Archive

SNF similarity network fusion

SV structural variant

TAD topologically associated domain

TCGA The Cancer Genome Atlas

TET ten-eleven translocation

TSS transcription start site

TN true negative

TP true positive

TF transcription factor

tracrRNA trans-activating CPRISR RNA

UTR untranslated region

WES whole exome sequencing

WGS whole genome sequencing

WT wild-type

References

1. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. en. *CA: A Cancer Journal for Clinicians* **68,** 394–424. ISSN: 00079235 (Nov. 2018).

2. Gilbertson, R. J. Mapping Cancer Origins. English. *Cell* **145,** 25–29. ISSN: 0092-8674, 1097-4172 (Apr. 2011).

3. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100,** 57–70 (Jan. 2000).

4. Hanahan, D. & Weinberg, R. A. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674. ISSN: 1097-4172 (Electronic)\r0092-8674 (Linking) (Mar. 2011).

5. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic Plasticity and the Hallmarks of Cancer. *Science* **357,** eaal2380–eaal2380 (July 2017).

6. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. en. *Cell Metabolism* **23,** 27–47. ISSN: 1550-4131 (Jan. 2016).

7. Garraway, L. A. & Lander, E. S. Lessons from the Cancer Genome. English. *Cell* **153,** 17–37. ISSN: 0092-8674, 1097-4172 (Mar. 2013).

8. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic Instabilities in Human Cancers. *Nature* **396,** 643–649 (Dec. 1998).

9. Abeshouse, A. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. en. *Cell* **163,** 1011–1025. ISSN: 00928674 (Nov. 2015).

10. Vinagre, J. *et al.* Frequency of TERT Promoter Mutations in Human Cancers. en. *Nature Communications* **4,** 2185. ISSN: 2041-1723 (Oct. 2013).

11. Huang, F. W. *et al.* Highly Recurrent TERT Promoter Mutations in Human Melanoma. en. *Science* **339,** 957–959. ISSN: 0036-8075, 1095-9203 (Feb. 2013).

12. Horn, S. *et al.* TERT Promoter Mutations in Familial and Sporadic Melanoma. en. *Science* **339,** 959–961. ISSN: 0036-8075, 1095-9203 (Feb. 2013).

13. Nagarajan, R. P. *et al.* Recurrent Epimutations Activate Gene Body Promoters in Primary Glioblastoma. *Genome Research* **24,** 761–774 (May 2014).

14. Stern, J. L., Theodorescu, D., Vogelstein, B., Papadopoulos, N. & Cech, T. R. Mutation of the *TERT* Promoter, Switch to Active Chromatin, and Monoallelic *TERT* Expression in Multiple Cancers. en. *Genes & Development* **29,** 2219–2224. ISSN: 0890-9369, 1549-5477 (Nov. 2015).

15. Alberts, B. *Molecular Biology of the Cell* Sixth edition. ISBN: 978-0-8153-4432-2 978-0-8153-4464-3 978-0-8153-4524-4 (Garland Science, Taylor and Francis Group, New York, NY, 2015).

16. Goodrich, J. A. & Tjian, R. Unexpected Roles for Core Promoter Recognition Factors in Cell-Type-Specific Transcription and Gene Regulation. en. *Nature Reviews Genetics* **11,** 549–558. ISSN: 1471-0056, 1471-0064 (Aug. 2010).

17. Schoenfelder, S. & Fraser, P. Long-Range Enhancer–Promoter Contacts in Gene Expression Control. En. *Nature Reviews Genetics,* 1. ISSN: 1471-0064 (May 2019).

18. Spitz, F. & Furlong, E. E. M. Transcription Factors: From Enhancer Binding to Developmental Control. en. *Nature Reviews Genetics* **13,** 613–626. ISSN: 1471-0064 (Sept. 2012).

19. Ong, C.-T. & Corces, V. G. Enhancer Function: New Insights into the Regulation of Tissue-Specific Gene Expression. en. *Nature Reviews Genetics* **12,** 283–293. ISSN: 1471-0064 (Apr. 2011).

20. Andersson, R. & Sandelin, A. Determinants of Enhancer and Promoter Activities of Regulatory Elements. en. *Nature Reviews Genetics* **21,** 71–87. ISSN: 1471-0064 (Feb. 2020).

21. Gaszner, M. & Felsenfeld, G. Insulators: Exploiting Transcriptional and Epigenetic Mechanisms. en. *Nature Reviews Genetics* **7,** 703–713. ISSN: 1471-0064 (Sept. 2006).

22. Oudelaar, A. M. & Higgs, D. R. The Relationship between Genome Structure and Function. en. *Nature Reviews Genetics.* ISSN: 1471-0056, 1471-0064 (Nov. 2020).

23. Farnham, P. J. Insights from Genomic Profiling of Transcription Factors. en. *Nature Reviews Genetics* **10,** 605–616. ISSN: 1471-0064 (Sept. 2009).

24. Kim, T. H. *et al.* Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. en. *Cell* **128,** 1231–1245. ISSN: 0092-8674 (Mar. 2007).

25. Dixon, J. R. *et al.* Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. en. *Nature* **485,** 376–380. ISSN: 1476-4687 (May 2012).

26. Kasowski, M. *et al.* Variation in Transcription Factor Binding Among Humans. en. *Science* **328,** 232–235. ISSN: 0036-8075, 1095-9203 (Apr. 2010).

27. Maurano, M. T., Wang, H., Kutyavin, T. & Stamatoyannopoulos, J. A. Widespread Site-Dependent Buffering of Human Regulatory Polymorphism. *PLoS Genetics* **8.** ISSN: 1553-7404 (Electronic)\n1553-7390 (Linking) (2012).

28. Maurano, M. T. *et al.* Large-Scale Identification of Sequence Variants Influencing Human Transcription Factor Occupancy in Vivo. en. *Nature Genetics* **47,** 1393–1401. ISSN: 1061-4036, 1546-1718 (Dec. 2015).

29. Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. en. *Cell Reports* **12,** 1184–1195. ISSN: 22111247 (Aug. 2015).

30. Wang, H. *et al.* Widespread Plasticity in CTCF Occupancy Linked to DNA Methylation. *Genome Research* **22,** 1680–1688. ISSN: 1549-5469 (Electronic)\n1088-9051 (Linking) (Sept. 2012).

31. Wiehle, L. *et al.* DNA (de)Methylation in Embryonic Stem Cells Controls CTCF-Dependent Chromatin Boundaries. en. *Genome Research,* gr.239707.118. ISSN: 1088-9051, 1549-5469 (Apr. 2019).

32. Xu, C. & Corces, V. G. Nascent DNA Methylome Mapping Reveals Inheritance of Hemimethylation at CTCF/Cohesin Sites. *Science* **359,** 1166–1170 (2018).

33. Viner, C. *et al.* Modeling Methyl-Sensitive Transcription Factor Motifs with an Expanded Epigenetic Alphabet. en. *bioRxiv,* 043794 (Mar. 2016).

34. Goll, M. G. & Bestor, T. H. Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry* **74,** 481–514. ISSN: 0066-4154 (June 2005).

35. Lister, R. *et al.* Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* **462,** 315–322. ISSN: 1476-4687 (Electronic)\n0028-0836 (Linking) (Nov. 2009).

36. Henikoff, S. Nucleosome Destabilization in the Epigenetic Regulation of Gene Expression. en. *Nature Reviews Genetics* **9,** 15–26. ISSN: 1471-0064 (Jan. 2008).

37. Jiang, C. & Pugh, B. F. Nucleosome Positioning and Gene Regulation: Advances through Genomics. en. *Nature Reviews Genetics* **10,** 161–172. ISSN: 1471-0064 (Mar. 2009).

38. Vierstra, J. *et al.* Global Reference Mapping of Human Transcription Factor Footprints. en. *Nature* **583,** 729–736. ISSN: 1476-4687 (July 2020).

39. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. en. *Cell* **174,** 1309–1324.e18. ISSN: 0092-8674 (Aug. 2018).

40. Polak, P. *et al.* Cell-of-Origin Chromatin Organization Shapes the Mutational Landscape of Cancer. en. *Nature* **518,** 360–364. ISSN: 1476-4687 (Feb. 2015).

41. Zhu, H., Wang, G. & Qian, J. Transcription Factors as Readers and Effectors of DNA Methylation. *Nature Reviews Genetics* **17,** 551–565 (Aug. 2016).

42. Furey, T. S. ChIP–Seq and beyond: New and Improved Methodologies to Detect and Characterize Protein–DNA Interactions. en. *Nature Reviews Genetics* **13,** 840–852. ISSN: 1471-0064 (Dec. 2012).

43. Carter, B. & Zhao, K. The Epigenetic Basis of Cellular Heterogeneity. en. *Nature Reviews Genetics* **22,** 235–250. ISSN: 1471-0064 (Apr. 2021).

44. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting Histone Modifications and the Functional Organization of Mammalian Genomes. *Nature Reviews Genetics* **12,** 7–18 (Jan. 2011).

45. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. English. *Cell* **164,** 1110–1121. ISSN: 0092-8674, 1097-4172 (Mar. 2016).

46. Finn, E. H. & Misteli, T. Molecular Basis and Biological Function of Variability in Spatial Genome Organization. en. *Science* **365,** eaaw9498. ISSN: 0036-8075, 1095-9203 (Sept. 2019).

47. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. en. *Science* **326,** 289–293. ISSN: 0036-8075, 1095-9203 (Oct. 2009).

48. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. English. *Cell* **159,** 1665–1680. ISSN: 0092-8674, 1097-4172 (Dec. 2014).

49. Mirny, L. A., Imakaev, M. & Abdennur, N. Two Major Mechanisms of Chromosome Organization. en. *Current Opinion in Cell Biology. Cell Nucleus* **58,** 142–152. ISSN: 0955-0674 (June 2019).

50. Stergachis, A. B. *et al.* Conservation of Trans-Acting Circuitry during Mammalian Regulatory Evolution. *Nature* **515,** 365–370. ISSN: 1476-4687 (Electronic)\r0028-0836 (Linking) (2014).

51. Berthelot, C., Villar, D., Horvath, J., Odom, D. T. & Flicek, P. Complexity and Conservation of Regulatory Landscapes Underlie Evolutionary Resilience of Mammalian Gene Expression. *bioRxiv,* 1–31 (2017).

52. Spurrell, C. H., Dickel, D. E. & Visel, A. The Ties That Bind: Mapping the Dynamic Enhancer-Promoter Interactome. English. *Cell* **167,** 1163–1166. ISSN: 0092-8674, 1097-4172 (Nov. 2016).

53. Buenrostro, J. D. *et al.* Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation. en. *Nature* **523,** 486–490. ISSN: 0028-0836, 1476-4687 (June 2015).

54. Lee, T. I. & Young, R. A. Transcriptional Regulation and Its Misregulation in Disease. English. *Cell* **152,** 1237–1251. ISSN: 0092-8674, 1097-4172 (Mar. 2013).

55. Conesa, A. *et al.* A Survey of Best Practices for RNA-Seq Data Analysis. *Genome Biology* **17,** 13–13. ISSN: 1474-760X (Electronic)\r1474-7596 (Linking) (Dec. 2016).

56. Robertson, G. *et al.* Genome-Wide Profiles of STAT1 DNA Association Using Chromatin Immunoprecipitation and Massively Parallel Sequencing. en. *Nature Methods* **4,** 651–657. ISSN: 1548-7091, 1548-7105 (Aug. 2007).

57. Bailey, T. *et al.* Practical Guidelines for the Comprehensive Analysis of ChIP-Seq Data. en. *PLOS Computational Biology* **9,** e1003326. ISSN: 1553-7358 (Nov. 2013).

58. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers. en. *Nature Protocols* **13,** 1006–1019. ISSN: 1754-2189, 1750-2799 (May 2018).

59. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. English. *Cell* **132,** 311–322. ISSN: 0092-8674, 1097-4172 (Jan. 2008).

60. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position. *Nature Methods* **10,** 1213–8 (Dec. 2013).

61. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology* **6,** 356–372. ISSN: 6314442508 (2015).

62. Corces, M. R. *et al.* An Improved ATAC-Seq Protocol Reduces Background and Enables Interrogation of Frozen Tissues. en. *Nature Methods* **14,** 959–962. ISSN: 1548-7105 (Oct. 2017).

63. Schones, D. E. *et al.* Dynamic Regulation of Nucleosome Positioning in the Human Genome. English. *Cell* **132,** 887–898. ISSN: 0092-8674, 1097-4172 (Mar. 2008).

64. Laird, P. W. Principles and Challenges of Genome-Wide DNA Methylation Analysis. *Nature Reviews Genetics* **11,** 191–191. ISSN: 1471-0064 (Electronic)\r1471-0056 (Linking) (Mar. 2010).

65. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. en. *Science* **295,** 1306–1311. ISSN: 0036-8075, 1095-9203 (Feb. 2002).

66. Nora, E. P. *et al.* Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre. en. *Nature* **485,** 381–385. ISSN: 0028-0836, 1476-4687 (May 2012).

67. Birney, E. *et al.* Identification and Analysis of Functional Elements in 1% of the Human Genome by the ENCODE Pilot Project. en. *Nature* **447,** 799–816. ISSN: 1476-4687 (June 2007).

68. Moore, J. E. *et al.* Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes. en. *Nature* **583,** 699–710. ISSN: 1476-4687 (July 2020).

69. Ernst, J. & Kellis, M. ChromHMM: Automating Chromatin-State Discovery and Characterization. en. *Nature Methods* **9,** 215–216. ISSN: 1548-7105 (Mar. 2012).

70. Hoffman, M. M. *et al.* Unsupervised Pattern Discovery in Human Chromatin Structure through Genomic Segmentation. en. *Nature Methods* **9,** 473–476. ISSN: 1548-7091, 1548-7105 (May 2012).

71. Chan, R. C. W. *et al.* Segway 2.0: Gaussian Mixture Models and Minibatch Training. en. *Bioinformatics* **34** (ed Birol, I.) 669–671. ISSN: 1367-4803, 1460-2059 (Feb. 2018).

72. Zhou, S., Treloar, A. E. & Lupien, M. Emergence of the Noncoding Cancer Genome: A Target of Genetic and Epigenetic Alterations. *Cancer Discovery* **6,** 1215–1229 (Nov. 2016).

73. Gasperini, M., Tome, J. M. & Shendure, J. Towards a Comprehensive Catalogue of Validated and Target-Linked Human Enhancers. en. *Nature Reviews Genetics,* 1–19. ISSN: 1471-0064 (Jan. 2020).

74. Croce, C. M. Oncogenes and Cancer. *New England Journal of Medicine* **358,** 502–511. ISSN: 0028-4793 (Jan. 2008).

75. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. English. *Cell* **173,** 371–385.e18. ISSN: 0092-8674, 1097-4172 (Apr. 2018).

76. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer Analysis Project. en. *Nature Genetics* **45,** 1113–1120. ISSN: 1546-1718 (Oct. 2013).

77. Pan-Cancer Analysis of Whole Genomes. en. *Nature* **578,** 82–93. ISSN: 1476-4687 (Feb. 2020).

78. Khurana, E. *et al.* Role of Non-Coding Sequence Variants in Cancer. en. *Nature Reviews Genetics* **17,** 93–108. ISSN: 1471-0056, 1471-0064 (Feb. 2016).

79. Rheinbay, E. *et al.* Analyses of Non-Coding Somatic Drivers in 2,658 Cancer Whole Genomes. en. *Nature* **578,** 102–111. ISSN: 1476-4687 (Feb. 2020).

80. Zhang, Y. *et al.* High-Coverage Whole-Genome Analysis of 1220 Cancers Reveals Hundreds of Genes Deregulated by Rearrangement-Mediated Cis -Regulatory Alterations. en. *Nature Communications* **11,** 1–14. ISSN: 2041-1723 (Feb. 2020).

81. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. P53 Mutations in Human Cancers. en. *Science* **253,** 49–53. ISSN: 0036-8075, 1095-9203 (July 1991).

82. Barbieri, C. E. *et al.* Exome Sequencing Identifies Recurrent SPOP , FOXA1 and MED12 Mutations in Prostate Cancer. en. *Nature Genetics* **44,** 685–689. ISSN: 1546-1718 (June 2012).

83. Meyer, N. & Penn, L. Z. Reflecting on 25 Years with MYC. en. *Nature Reviews Cancer* **8,** 976–990. ISSN: 1474-1768 (Dec. 2008).

84. Cowper-Sal·lari, R. *et al.* Breast Cancer Risk–Associated SNPs Modulate the Affinity of Chromatin for FOXA1 and Alter Gene Expression. en. *Nature Genetics* **44,** 1191–1198. ISSN: 1061-4036, 1546-1718 (Nov. 2012).

85. Kron, K. J., Bailey, S. D. & Lupien, M. Enhancer Alterations in Cancer: A Source for a Cell Identity Crisis. en. *Genome Medicine* **6,** 1–12. ISSN: 1756-994X (Dec. 2014).

86. Bailey, S. D. *et al.* Noncoding Somatic and Inherited Single-Nucleotide Variants Converge to Promote ESR1 Expression in Breast Cancer. *Nature Genetics* **48,** 1260–1269 (2016).

87. Mazrooei, P. *et al.* Cistrome Partitioning Reveals Convergence of Somatic Mutations and Risk Variants on Master Transcription Regulators in Primary Prostate Tumors. English. *Cancer Cell* **36,** 674–689.e6. ISSN: 1535-6108, 1878-3686 (Dec. 2019).

88. Quigley, D. A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. English. *Cell* **174,** 758–769.e9. ISSN: 0092-8674, 1097-4172 (July 2018).

89. Parolia, A. *et al.* Distinct Structural Classes of Activating FOXA1 Alterations in Advanced Prostate Cancer. en. *Nature* **571,** 413–418. ISSN: 1476-4687 (July 2019).

90. Northcott, P. A. *et al.* Enhancer Hijacking Activates GFI1 Family Oncogenes in Medulloblastoma. en. *Nature* **511,** 428–434. ISSN: 1476-4687 (July 2014).

91. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. en. *Cell* **161,** 1012–1025. ISSN: 0092-8674 (May 2015).

92. Allou, L. *et al.* Non-Coding Deletions Identify Maenli lncRNA as a Limb-Specific En1 Regulator. en. *Nature.* ISSN: 0028-0836, 1476-4687 (Feb. 2021).

93. Pirozzi, C. J. & Yan, H. The Implications of IDH Mutations for Cancer Development and Therapy. en. *Nature Reviews Clinical Oncology,* 1–17. ISSN: 1759-4782 (June 2021).

94. Im, A. P. *et al.* DNMT3A and IDH Mutations in Acute Myeloid Leukemia and Other Myeloid Malignancies: Associations with Prognosis and Potential Treatment Strategies. en. *Leukemia* **28,** 1774–1783. ISSN: 1476-5551 (Sept. 2014).

95. Issa, G. C. & DiNardo, C. D. Acute Myeloid Leukemia with IDH1 and IDH2 Mutations: 2021 Treatment Algorithm. en. *Blood Cancer Journal* **11,** 1–7. ISSN: 2044-5385 (June 2021).

96. Molenaar, R. J., Maciejewski, J. P., Wilmink, J. W. & van Noorden, C. J. F. Wild-Type and Mutated IDH1/2 Enzymes and Therapy Responses. en. *Oncogene* **37,** 1949–1960. ISSN: 1476-5594 (Apr. 2018).

97. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The Role of Mutations in Epigenetic Regulators in Myeloid Malignancies. en. *Nature Reviews Cancer* **12,** 599–612. ISSN: 1474-1768 (Sept. 2012).

98. Dominguez, P. M. *et al.* TET2 Deficiency Causes Germinal Center Hyperplasia, Impairs Plasma Cell Differentiation, and Promotes B-Cell Lymphomagenesis. en. *Cancer Discovery* **8,** 1632–1653. ISSN: 2159-8274, 2159-8290 (Dec. 2018).

99. Plass, C. *et al.* Mutations in Regulators of the Epigenome and Their Connections to Global Chromatin Patterns in Cancer. en. *Nature Reviews Genetics* **14,** 765–780. ISSN: 1471-0064 (Nov. 2013).

100. Nikoloski, G. *et al.* Somatic Mutations of the Histone Methyltransferase Gene EZH2 in Myelodysplastic Syndromes. en. *Nature Genetics* **42,** 665–667. ISSN: 1546-1718 (Aug. 2010).

101. Ernst, T. *et al.* Inactivating Mutations of the Histone Methyltransferase Gene EZH2 in Myeloid Disorders. en. *Nature Genetics* **42,** 722–726. ISSN: 1546-1718 (Aug. 2010).

102. Morin, R. D. *et al.* Somatic Mutations Altering EZH2 (Tyr641) in Follicular and Diffuse Large B-Cell Lymphomas of Germinal-Center Origin. en. *Nature Genetics* **42,** 181–185. ISSN: 1546-1718 (Feb. 2010).

103. Varambally, S. *et al.* The Polycomb Group Protein EZH2 Is Involved in Progression of Prostate Cancer. en. *Nature* **419,** 624–629. ISSN: 1476-4687 (Oct. 2002).

104. Xu, K. *et al.* EZH2 Oncogenic Activity in Castration-Resistant Prostate Cancer Cells Is Polycomb-Independent. en. *Science* **338,** 1465–1469. ISSN: 0036-8075, 1095-9203 (Dec. 2012).

105. Min, J. *et al.* An Oncogene–Tumor Suppressor Cascade Drives Metastatic Prostate Cancer by Coordinately Activating Ras and Nuclear Factor-κB. en. *Nature Medicine* **16,** 286–294. ISSN: 1546-170X (Mar. 2010).

106. Kim, K. H. & Roberts, C. W. M. Targeting EZH2 in Cancer. en. *Nature Medicine* **22,** 128–134. ISSN: 1546-170X (Feb. 2016).