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

Chapter 1

Introduction

1.1 Cancer is a disease of the genome and epigenome

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 [**brayGlobalCancerStatistics2018**]. 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. 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 [**hanahanHallmarksCancer2000**, **hanahanHallmarksCancerNext2011**, **flavahanEpigeneticPlasticityHallmarks2017**, **pavlovaEmergingHallmarksCancer2016**] (**??**).

Many of these hallmarks of cancer can be achieved through aberrations to the molecular machinery that enables cells to function normally. For example, genome instability can be achieved by inhibiting DNA repair machinery, as is observed with abnormalities in *MLH1* and *MSH2* repair genes in colorectal cancers [**lengauerGeneticInstabilitiesHuman1998**] or mutations to *BRCA1*, *BRCA2*, and *ATM* genes in prostate cancer (PCa) [**abeshouseMolecularTaxonomyPrimary2015**]. Similarly, replicative immortality can be achieved through telomere elongation by over-expression of the *TERT* gene [**vinagreFrequencyTERTPromoter2013**]. Mutations to the *TERT* promoter, resulting in its over-expression, were first identified in melanomas [**huangHighlyRecurrentTERT2013**, **hornTERTPromoterMutations2013**], but have since been further identified in bladder, thyroid, and brain cancers [**vinagreFrequencyTERTPromoter2013**, **nagarajanRecurrentEpimutationsActivate2014**, **sternMutationTERTPromoter2015**]. But while cancer has long been viewed as a disease of the

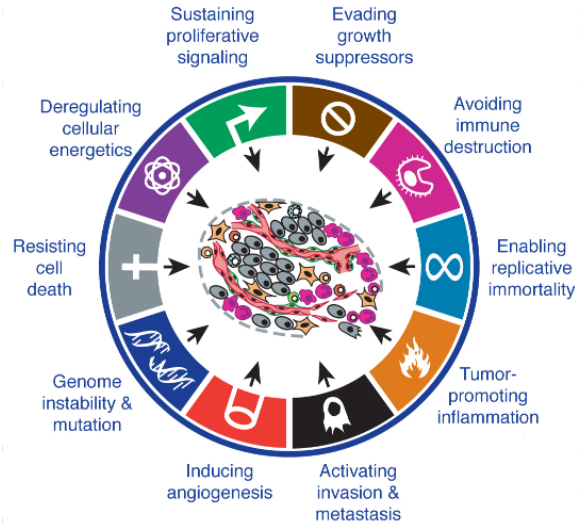

Figure 1.1: **The hallmarks of cancer.** Adapted from [**hanahanHallmarksCancerNext2011**].

genome [**hanahanHallmarksCancer2000**], there are many avenues cells can take to arrive at any of these hallmarks resulting from aberrations of how genes are expressed inside the cell nucleus.

Genes, encoded as DNA, are transcribed into messenger RNA (mRNA), which are then translated into proteins, in the process known as the central dogma of molecular biology [**albertsMolecularBiologyCell2015**] (**??**). The transcription of genes into mRNA requires RNA polymerase to bind at transcription start sites (TSSs) located within gene promoters [**goodrichUnexpectedRolesCore2010**]. 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 DNA elements termed enhancers, insulators, and silencers [**schoenfelderLongrangeEnhancerPromoter2019**, **spitzTranscriptionFactorsEnhancer2012**, **ongEnhancerFunctionNew2011**, **anderssonDeterminantsEnha** **gasznerInsulatorsExploitingTranscriptional2006**, **oudelaarRelationshipGenomeStructure2020**]. These different DNA elements can be tens to thousands of basepairs (bps) apart from each other, but the DNA polymer bends to take up a small space inside the nucleus, and this can bring distal DNA elements close together in three-dimensional space [**finnMolecularBasisBiological2019**, **zhouChartingHistoneModifications2011**].

The ability of TFs to bind at certain DNA elements is dependent on multiple features of the DNA inside the nucleus, including its sequence, its shape, and nearby chemical modifications, such as DNA methylation (DNAme) or modifications to the histone proteins that comprise the nucleosomes that DNA is wrapped around [**zhuTranscriptionFactorsReaders2016**, **fureyChIPSeqNew2012**, **carterEpigeneticBasisCellular2021**]. While the DNA sequence remains consistent across all

Figure 1.2: **The basics of gene expression inside the nucleus.. a.** The central dogma of molecular biology. **b.** Schematic of the transcription machinery to initiate transcription. Both panels are adapted from [**albertsMolecularBiologyCell2015**].

cells in an organism,

1.1.1

1.2 Dissertation structure

I begin with **??** by exploring the *cis*-regulatory landscape of PCa and delineating the *cis*-regulatory elements (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 single nucleotide variantss (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 chromatin conformation capture (3C) mapping in **??**. Using Hi-C, I characterize the three-dimensional chromatin organization of PCa and investigate the relationship between chromatin organization, structural variants (SVs), and the hijacking of *cis*-regulatory networks more generally.

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

CMP common myeloid progenitor

CPC-GENE Canadian Prostate Cancer Genome Network

CpG CG dinucleotide

crRNA CRISPR RNA

CRE *cis*-regulatory element

DEPMAP Cancer Dependency Map

DHS DNase I hypersensitive sites

DMR differentially methylated region

DNAme DNA methylation

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

IID independent and identically distributed

JS James-Stein

kbp kilobase

KO knockout

LDA limiting dilution assay

LMPP lymphoid-primed multi-potent progenitor

MeCapSeq DNA methylation capture sequencing

MEP megakaryocyte-erythrocyte progenitor

MSE mean square error

mCRPC metastatic castration-resistant prostate cancer

MDS myelodisplastic syndrome

MLP monocyte-lymphoid progenitor

MPP multi-potent progenitor

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

RNAi RNA interference

RNA-seq RNA sequencing

shRNA small hairpin RNA

siRNA small interfering RNA

SNV single nucleotide variants

SRA Seqeuence Read Archive

SNF similarity network fusion

SV structural variant

TAD topologically associated domain

TCGA The Cancer Genome Atlas

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