Chromatin architecture aberrations in prostate cancer and acute lymphoblastic 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

Discussion & Future Directions

Each of the previous chapters have presented a story interrogating multiple components of the chromatin architecture, how they interact with each other, and the plethora of computational and experimental methods required to unravel this architecture. **??** identifies and validates *cis*-regulatory elements (CREs) of the *FOXA1* gene, a critical transcription factor (TF) that regulates prostate cancer (PCa) development and regulates androgen receptor (*AR*) expression to control disease progression. **??** expands on these ideas to investigate how the three-dimensional genome organization impacts gene regulatory networks and how genetic aberrations can alter this organization to promote oncogenesis. **??** develops a mathematical and computational framework to reduce uncertainty about how individual aberrations in chromatin architecture impact gene expression. Finally, **??** identifies the strong relationship between genetic and epigenetic profiles in B-cell acute lymphoblastic leukemia (B-ALL) relapse and investigates how DNA methylation (DNAme) changes and revision to a more stem-like chromatin state may underlie disease relapse. Together, the work presented in this thesis demonstrates that different components of the chromatin architecture, the genome, molecular chromatin modifications, and three-dimensional organization, can all individually contribute to cancer development and progression. Moreover, this thesis demonstrates that aberrations in these components work together to drive disease. These multiple components of the chromatin architecture need to be studied in tandem to understand the origins of cancer and how to develop curative treatments for it.

1.1 Implications of non-coding single nucleotide variants targeting a single gene

In **??**, I used gene essentiality screening data from multiple cell lines to prioritize the *FOXA1* TF as a critical factor across PCa cell lines. I also made use of the concept that single nucleotide variants (SNVs) converge on CREs of important genes in a given tumour type to predict how these mutations may impact candidate CREs for the *FOXA1* gene. *FOXA1* is also an important TF in breast cancers **??**. Similar investigations into the impact of SNVs in breast tumours may identify the impact of aberrations to the CREs of *FOXA1*. Identifying important genes in this manner is not limited to *FOXA1* and breast and prostate tumours. Critical genes may be identified in other cancer types using clustered regularly interspaced short palindromic repeat (CRISPR) screens or massively-parallel reporter assays (MRPAs). Similarly SNVs are not the only chromatin aberrations that can affect TF binding or gene regulation. Other chromatin aberrations may accumulate in CREs of important genes in a similar fashion. Complex structural variants (SVs), changes in DNAme, or histone modifications may only need to accumulate in the set of CREs for a given gene, rather than be recurrent in a single element, to affect its expression. Interpreting chromatin aberrations in cancer in light of this plexus-based approach may aid in identifying driver events for cancer by aggregating previously unrelated events together. These approaches are not limited to prostate tumours and can serve as a starting point to identify important genes in other cancers, more generally.

1.2 Implications of three-dimensional organization and enhancer hijacking in prostate cancer

- optimized low-input Hi-C method to interrogate genome organization in more tissues than previously possible
- largest collection of 3D organization data in prostate tumours to date
- add to evidence that SVs can, but rarely, alter 3D structure in disease, prompting further interrogation into how and why this appears to be the case
- motivates use of microscopy measurements in single cells to see how structure is altered and if sub-clonal mutations interfere with the ability to detect changes in structure
- statistical methods may make progress to help identify effect of individual, non-recurrent

events, such as in **??**, as well as computational libraries for developing better tooling in this area

- alternatively further push for development in organoid models that can replicate the chromatin state of tumour tissue from small input samples
- use genome organization methods to fully characterize GRNs for each individual gene (both first- and higher-order, depending on how much perturbations to these networks affect gene expression)
- consider evolutionary lens of GRN across species instead of just thinking of genes as single units

1.3 Implications of DNA methylation changes in relapse

- prioritize role of stem cells in disease relapse
- use increasing DNAme as a potential biomarker of relapse
- may be able to use blood-based DNAme detection to create a non-invasive test for this development
- may be able to treat B-ALL patients with de-methylating agents if gains in DNAme are observed to prevent relapse

1.4 Summary and concluding remarks

- work does not focus on a single disease, should extend this type of analysis to all cancers, since they all appear to harbour aberrations affecting multiple components of the chromatin architecture
- multi-pronged approach of computational, statistical, and molecular, and microscopy methods optimized for low-input samples targeting the set of DNA elements and their relationships to each other in individual patients to develop personalized medicines and treat cancer at its origins in the chromatin

Glossary

3C chromatin conformation capture

ALL acute lymphoblastic leukemia

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

ISUP International Society of Urological Pathology

JS James-Stein

KMT histone lysine methyltransferase

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

PSA prostate-specific antigen

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

T2E *TMPRSS2-ERG*

TAD topologically associated domain

TCGA The Cancer Genome Atlas

TET ten-eleven translocation

TSS transcription start site

TN true negative

TNM tumour node metastasis

TP true positive

TF transcription factor

tracrRNA trans-activating CPRISR RNA

UTR untranslated region

WES whole exome sequencing

WGBS whole genome bisulfite sequencing

WGS whole genome sequencing

WT wild-type