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

In **??**, my co-authors optimized a low-input Hi-C method to interrogate genome organization in cryo-preserved prostate tissue slides. I then demonstrated that this could produce a high quality Hi-C library and helped produce that largest collection of genome organization data in prostate tumours to date. This technological step forward opens the door for profiling the three-dimensional genome in cancer patients without relying in cell lines or other models, and may be a critical step in moving personalized medicine forward. We add to existing evidence that SVs can, but rarely, alter 3D structure in disease [1–8]. Elucidating when and how SVs impact genome organization, then, is still an area that requires investigation. Developments in statistical methods, such as those discussed in **??**, may help identify the effects of individual, non-recurrent SVs. Subclonality of SVs

may interfere with the ability to detect rearranged domains in bulk Hi-C measurements. Thus, developments in high throughput sequencing and microscopy measurements in single cells, such as ORCA [9] and STORM [10], as well as organoid or explant models that recapitulate the chromatin state of the original tumour, may help in identifying the effect of such events []. This work also adds to our ability to detect chromatin interactions between promoters and enhancers in patient samples, allowing for better characterization of gene regulatory networks for each and every gene. Given the benefits of plexus-based approaches to interpreting aberrations in the chromatin architecture, this work serves as a foundation on which to integrate gene regulatory networks with chromatin aberrations in cancers more generally. This foundation can be extended to studying the evolution of these networks, their genome organization, and their resiliency between species or over time as tumours respond to therapeutic interventions [].

1.3 Implications of DNA methylation changes in relapse

?? identifies DNAme as an epigenetic marker that can mirror that mutational profile of B-ALL cells. The DNAme changes observed over the course of B-ALL relapse have the potential to become a biomarker predicting relapse, although variation in DNAme changes across patients necessitates larger sample sizes before recurrent events may be robustly identified. Our patient-oriented approach to identify recurrent changes to DNAme boosted our discovery of recurrent differentially methylated regions (DMRs) over a cohort-oriented approach, and thus may be a beneficial strategy for similar studies. Inter-tumour heterogeneity in tumours is a well-studied phenomenon [11–17], so patient-oriented discovery approaches may be advantageous when assessing molecular trajectories of relapse and therapeutic response, as well. Changes in DNAme can be detected through blood draws [18–20] and offers the potential for a non-invasive biomarker to predict relapse in B-ALL patients. Given the widespread hypermethylation observed at relapse in all patients, it is possible that B-ALL patients undergoing continuation/maintenance therapy may additionally benefit from demethylating agents such as 5-aza-cytidine and 5-aza-2'-deoxycytidine to prevent relapse. Finally, the widespread hypermethylation of B-ALL at relapse reflects revision of malignant blasts to a more stem-like phenotype, a characteristic typically observed in other leukemias such as acute myeloid leukemia (AML) [21–25]. This suggests that the role of leukemic stem cells and the impact that the cell-of-origin has on therapeutic response, should be prioritized in future research of B-ALL relapse.

1.4 Limitations

The main limitations in the studies presented here stem from cancer models, cohort sizes, and correlative measurements. In **??**, we used essentiality data from PCa cell lines that serve as models of primary prostate tumours. Further, our dissection of CREs relied on encyclopedias of DNA elements derived from cell lines, as did our validation experiments for assessing the impact of mutations seen in primary prostate tumours. Cell line models are not necessarily representative of primary tumours observed in patients []. Moreover, different cell lines may have different regulatory networks, as recent studies of chromatin organization and CRE contacts in PCa cell lines have highlighted [26]. This separation between the models we study and the patients we aim to serve may prohibit translation of discoveries from our models to the clinic. In **??**, we were able to profile the genome organization of 12 PCa patients and 5 benign prostate samples. Similarly, in **??**, we used patient-derived xenografts (PDXs) originating from 5 patients with B-ALL. In light of the large degree of inter-patient variation seen in tumours from the same tissue, and the intra-patient variation seen in PDXs (**??**), observations in these small sample sizes may not generalize to patients with these diseases as a whole. In both cases, we used patient-matched data to corroborate findings from other components of the chromatin architecture, where possible, but validation experiments for some observations remain difficult. Using CRISPR technologies to knock in SNVs or inactivate individual CREs is possible, but replicating complex SVs or selectively (de)methylating CREs to validate some findings remains technically challenging [27–29]. Finally, despite the small sample sizes in some of these studies, the amount of high throughput sequencing data generated was larger than many computational methods could handle efficiently, particularly with Hi-C data. Some methods used here had to be adapted from their original publications, and other methods that could have been used would have required extensive re-engineering to function properly, limiting some analyses.

1.5 Summary and concluding remarks

Diagnosis, treatment, and the foundational understanding of cancer has been revolutionized by high throughput sequencing technologies and the ability to detect and interpret genetic aberrations in tumours. When combined with information about CREs, the essentiality of genes in multiple cell types, and the three-dimensional genome organization, genetic aberrations can provide a significant amount of explanatory power for the hallmarks of cancer and what causes them. But genetic aberrations are not the only mechanism cancer cells can use to arrive at these hallmarks to drive

disease. Identifying aberrations across multiple components of the chromatin architecture can uncover complex mechanisms through which cancer cells turn off the expression of tumour suppressor genes, activate oncogenes, or activate pathways. Interrogating the role of mutations in a gene's set of CREs (**??**), identifying stable regions of genome organization and shifting chromatin interactions over oncogenesis (**??**), and comparing genetic and epigenetic states over disease progression (**??**) are all important approaches to better understand the aberrations, origins, and trajectories of cancers in patients. This multi-pronged approach requires computational, statistical, and molecular, methods, optimized for low-input samples and small cohort sizes (**????**), and methodological developments in these areas are still required. This thesis focuses on PCa and B-ALL, but the methodologies employed here are not restricted to a single disease. Proof-of-concept studies here can be extended to all cancers, or any other diseases caused by aberrations to the chromatin architecture.

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

References

1. Ghavi-Helm, Y. *et al.* Highly Rearranged Chromosomes Reveal Uncoupling between Genome Topology and Gene Expression. En. *Nature Genetics,* 1. ISSN: 1546-1718 (July 2019).

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

3. Despang, A. *et al.* Functional Dissection of the Sox9–Kcnj2 Locus Identifies Nonessential and Instructive Roles of TAD Architecture. en. *Nature Genetics* **51,** 1263–1271. ISSN: 1061-4036, 1546-1718 (Aug. 2019).

4. Williamson, I. *et al.* Developmentally Regulated *Shh* Expression Is Robust to TAD Perturbations. en. *Development* **146,** dev179523. ISSN: 0950-1991, 1477-9129 (Oct. 2019).

5. Dixon, J. R. *et al.* Integrative Detection and Analysis of Structural Variation in Cancer Genomes. En. *Nature Genetics* **50,** 1388. ISSN: 1546-1718 (Oct. 2018).

6. Akdemir, K. C. *et al.* Disruption of Chromatin Folding Domains by Somatic Genomic Rearrangements in Human Cancer. en. *Nature Genetics,* 1–12. ISSN: 1546-1718 (Feb. 2020).

7. Li, Y. *et al.* Patterns of Somatic Structural Variation in Human Cancer Genomes. en. *Nature* **578,** 112–121. ISSN: 1476-4687 (Feb. 2020).

8. Iyyanki, T. *et al.* Subtype-Associated Epigenomic Landscape and 3D Genome Structure in Bladder Cancer. en. *Genome Biology* **22,** 105. ISSN: 1474-760X (Dec. 2021).

9. Mateo, L. J. *et al.* Visualizing DNA Folding and RNA in Embryos at Single-Cell Resolution. En. *Nature* **568,** 49. ISSN: 1476-4687 (Apr. 2019).

10. Bates, M., Jones, S. A. & Zhuang, X. Stochastic Optical Reconstruction Microscopy (STORM): A Method for Superresolution Fluorescence Imaging. en. *Cold Spring Harbor Protocols* **2013,** pdb.top075143. ISSN: 1940-3402, 1559-6095 (June 2013).

11. Marusyk, A. & Polyak, K. Tumor Heterogeneity: Causes and Consequences. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1805,** 105–117. ISSN: 0006-3002 (Print)\r0006-3002 (Linking) (Jan. 2010).

12. Sottoriva, A., Barnes, C. P. & Graham, T. A. Catch My Drift? Making Sense of Genomic Intra-Tumour Heterogeneity. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. Evolutionary Principles - Heterogeneity in Cancer?* **1867,** 95–100. ISSN: 0304-419X (Apr. 2017).

13. Huang, S. Genetic and Non-Genetic Instability in Tumor Progression: Link between the Fitness Landscape and the Epigenetic Landscape of Cancer Cells. *Cancer and Metastasis Reviews* **32,** 423–448. ISSN: 0167-7659 (Dec. 2013).

14. McGranahan, N. & Swanton, C. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* **27,** 15–26 (Jan. 2015).

15. Landau, D. A. & Wu, C. J. Chronic Lymphocytic Leukemia: Molecular Heterogeneity Revealed by High-Throughput Genomics. en. *Genome Medicine* **5,** 47. ISSN: 1756-994X (May 2013).

16. Ben-David, U. *et al.* Genetic and Transcriptional Evolution Alters Cancer Cell Line Drug Response. en. *Nature,* 1. ISSN: 1476-4687 (Aug. 2018).

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

18. Peter, M. R. *et al.* Dynamics of the Cell-Free DNA Methylome of Metastatic Prostate Cancer during Androgen-Targeting Treatment. *Epigenomics* **12,** 1317–1332. ISSN: 1750-1911 (Aug. 2020).

19. Shen, S. Y. *et al.* Sensitive Tumour Detection and Classification Using Plasma Cell-Free DNA Methylomes. en. *Nature* **563,** 579–583. ISSN: 1476-4687 (Nov. 2018).

20. Nassiri, F. *et al.* Detection and Discrimination of Intracranial Tumors Using Plasma Cell-Free DNA Methylomes. en. *Nature Medicine* **26,** 1044–1047. ISSN: 1546-170X (July 2020).

21. Krivtsov, A. V. & Armstrong, S. A. MLL Translocations, Histone Modifications and Leukaemia Stem-Cell Development. en. *Nature Reviews Cancer* **7,** 823–833. ISSN: 1474-1768 (Nov. 2007).

22. Li, Y. *et al.* Clinical Implications of Genome-Wide DNA Methylation Studies in Acute Myeloid Leukemia. *Journal of Hematology & Oncology* **10,** 41–41 (Dec. 2017).

23. Kreso, A. & Dick, J. E. Evolution of the Cancer Stem Cell Model. *Cell Stem Cell* **14,** 275–291. ISSN: 1875-9777 (Electronic) (2014).

24. Shlush, L. I. *et al.* Identification of Pre-Leukaemic Haematopoietic Stem Cells in Acute Leukaemia. *Nature* **506,** 328–333. ISSN: 9788578110796 (Feb. 2014).

25. Shlush, L. I. *et al.* Tracing the Origins of Relapse in Acute Myeloid Leukaemia to Stem Cells. *Nature* **547,** 104–108 (June 2017).

26. Ahmed, M. *et al.* CRISPRi Screens Reveal a DNA Methylation-Mediated 3D Genome Dependent Causal Mechanism in Prostate Cancer. en. *Nature Communications* **12,** 1781. ISSN: 2041-1723 (Mar. 2021).

27. Nakamura, M., Gao, Y., Dominguez, A. A. & Qi, L. S. CRISPR Technologies for Precise Epigenome Editing. en. *Nature Cell Biology* **23,** 11–22. ISSN: 1476-4679 (Jan. 2021).

28. Pickar-Oliver, A. & Gersbach, C. A. The next Generation of CRISPR–Cas Technologies and Applications. en. *Nature Reviews Molecular Cell Biology* **20,** 490–507. ISSN: 1471-0080 (Aug. 2019).

29. Wang, H., Han, M. & Qi, L. S. Engineering 3D Genome Organization. en. *Nature Reviews Genetics* **22,** 343–360. ISSN: 1471-0064 (June 2021).