Demystifying cancer etiology via 3D genome mapping

Yuliang Feng and Siim Pauklin*

Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Old Road, University of Oxford, Oxford OX3 7LD
* correspondence siim.pauklin@ndorms.ox.ac.uk

Key words: prostate cancer, 3D genome organization, chromatin architecture

Abstract

Ramanand *et al*. perform the first high-resolution 3D genome mapping via ChIA-PET to capture RNAPII-associated chromatin interactions in normal prostate epithelial and prostate cancer cells. They describe how genetics, epigenetics, and the 3D genome architecture are coordinated in the aberrant gene expression that drives prostate cancer development.

Main text

Prostate cancer (PCa) is the most common cancer in men, with current strategies for diagnostics and treatment for PCa still suboptimal. Uncovering the molecular mechanisms that underpin PCa etiology is fundamental for its treatment. Large-scale omics profiling have revealed that PCa is associated with transcriptional dysregulation which drives cancer progression. Although thousands of enhancers seem to be aberrantly activated in PCa, their target genes remain largely unknown. This is because genome is organized in three-dimensional (3D) space within the nucleus, and enhancers in many cases, are located at long-range distance from their target genes on the linear DNA molecule. Due to the formation of dynamic chromatin loops, these distant regulatory elements can relocate close to their target genes in 3D space, and through direct contacts regulate gene transcription[1]. Hence, promoter-enhancer linkage cannot be assigned by using traditional 2D-epigenomic approaches. Protein-centric 3D genome mapping (e.g. ChIA-PET) enables capturing the chromatin interactions with specific transcription factors (e.g. RNAPII, the core component of transcription initiation) and thus can achieve high resolution insight into regulatory interactomes[2, 3]. Recently, Ramanand *et al.* perform RNAPII ChIA-PET mappings in normal and prostate cancer cell lines and uncover the distinct enhancer-promoter interactomes, which are the drivers for transcriptional aberration in PCa. They further show that lineage specific RNAPII-associated chromatin interactions are flanked by the cohesin subunit RAD21 associated chromatin interactions marked by CTCF. For example, RAD21 and CTCF restrict the communication of *KRT8-KRT18* and their enhancers by forming an insulated neighborhood to facilitate specific transcription in prostate luminal epithelial lineage, which is in line with the loop extrusion model [4].

In addition, Ramanand *et al.* identified an active enhancer cluster that interacts with the androgen receptor *(AR)* promoter (~700 kb distance) and regulates *AR* transcription in LNCaP cells. Interestingly, this enhancer cluster is bound by AR, suggesting an auto-regulation mechanism of the master regulator transcription in PCa. AR can act as a transcription factor and is a key driver of PCa. Additionally, they identify extensive RNAPII-associated chromatin interactions harboring

MYC and several *MYC* interacting anchors representing genes encoding non-coding RNAs. This phenomenon could be further explained from a recent study, which shows that RNAs derived from *MYC* enhancer and promoter interact with RNA-binding protein hnRNPK, whose oligomerization juxtaposes *MYC* enhancer *to MYC* promoter and thus promotes *MYC* transcription [5].

GWAS have identified a number of SNPs associated with PCa, but little is known about how these genetic variants contribute to PCa pathogenesis. The major challenges in inferring their functional consequences are: firstly, most of the associated SNPs are located outside of protein coding regions of genes and thus the precise molecular mechanism of disease etiology is not immediately apparent; secondly, the proximity based approach linking the SNP to its nearest gene is not reliable in some cases (e.g. rs1421085 in *FTO* intron [6]). To overcome these barriers, Ramanand *et al.* employed RNAPII ChIA-PET to identify the causal SNPs from PCa GWAS studies and infer their functionality. They found that *VPS53* promoter, which harbors the PCa risk SNP rs684232, interacted with *FAM57A* and *GEMIN4* genes. By stratifying the tumors from The Cancer Genome Atlas (TCGA) based on the genotypes, they found that the risk allele was associated with transcription repression of *VPS53*, *FAM57A* and *GEMIN4*. Mechanistically, this might be attributed to the presence of risk allele that is linked to decreased H3K27ac signal on *VPS53*, *FAM57A* and *GEMIN4* loci. This suggested that *VPS53* may function as an enhancer-like promoter [2, 7] to regulate *FAM57A* and *GEMIN4* expression. Thus, this approach highlighted that high-resolution 3D genome contact mapping is a powerful paradigm to pinpoint the causal SNP(s) and their target gene(s) for explaining disease susceptibility.

This study uses the ChIA-PET protocol employing in-solution ligation[3], which requires hundreds of millions cells. Recently, the development of *in-situ* ChIA-PET method dramatically reduced the material input need to several million cells, rendering its applicability in primary cells [8]. Moreover, CTCF is a key architectural protein in defining topological domains [9]. Hence, combined RNAPII and CTCF ChIA-PET profilings in the cells from PCa patients would provide a comprehensive understanding of the topological basis of RNAPII-associated chromatin interactions during PCa development. Furthermore, generation of a compendium of ChIA-PET maps from the patients with diverse clinical outcome will aid in identifying the structural codes as novel biomarkers for therapeutic sensitivity or therapeutic targeting.

This study identifies the interaction of *VPS53* promoter with *FAM57A* and *GEMIN4* promoters, and suggests this might be a regulatory hub for co-transcription. However, like other proximity-ligation based 3D genome mapping approaches, ChIA-PET can only detect binary interactions and cannot reflect whether the uncovered multigene clusters co-exist in individual cells or whether they are just an ensemble of diverse chromatin interactions from individual cells. Recently, the ligation-free approaches (such as ChIA-Drop) were developed [10], which enable interrogating multiplex chromatin interactions and examining the chromatin heterogeneity in individual cells. Thus, future work will be needed to further validate the co-transcription module (e.g. *VPS53*, *FAM57A* and *GEMIN4*) in PCa using a ChIA-Drop like approach to gain insights at a single-molecule level.

Linear chromatin map provides an incomplete view of the genome, which hampers unravelling the functionality of regulatory elements without a 3D-genome context. This study generates an unprecedented high-resolution 3D genome contact map in PCa and uncovers an interplay between genetics, epigenetics and 3D genome structure, and how they collectively drive the pathogenesis of PCa, by coordinating aberrant gene expression (**Figure 1**). Hence, this study not only sheds light on previously unrecognized mechanisms for PCa etiology but also provides a paradigm to study the impact of 3D genome organization on other cancers or diseases.

Acknowledgements

Work in S.P.'s laboratory is supported by a Cancer Research UK Career Development Fellowship, Grant ID C59392/A25064; Royal Society, UK; and The Clarendon Fund and St Edmund Hall Scholarship, UK, SFF1920_CB_MSD_759707.

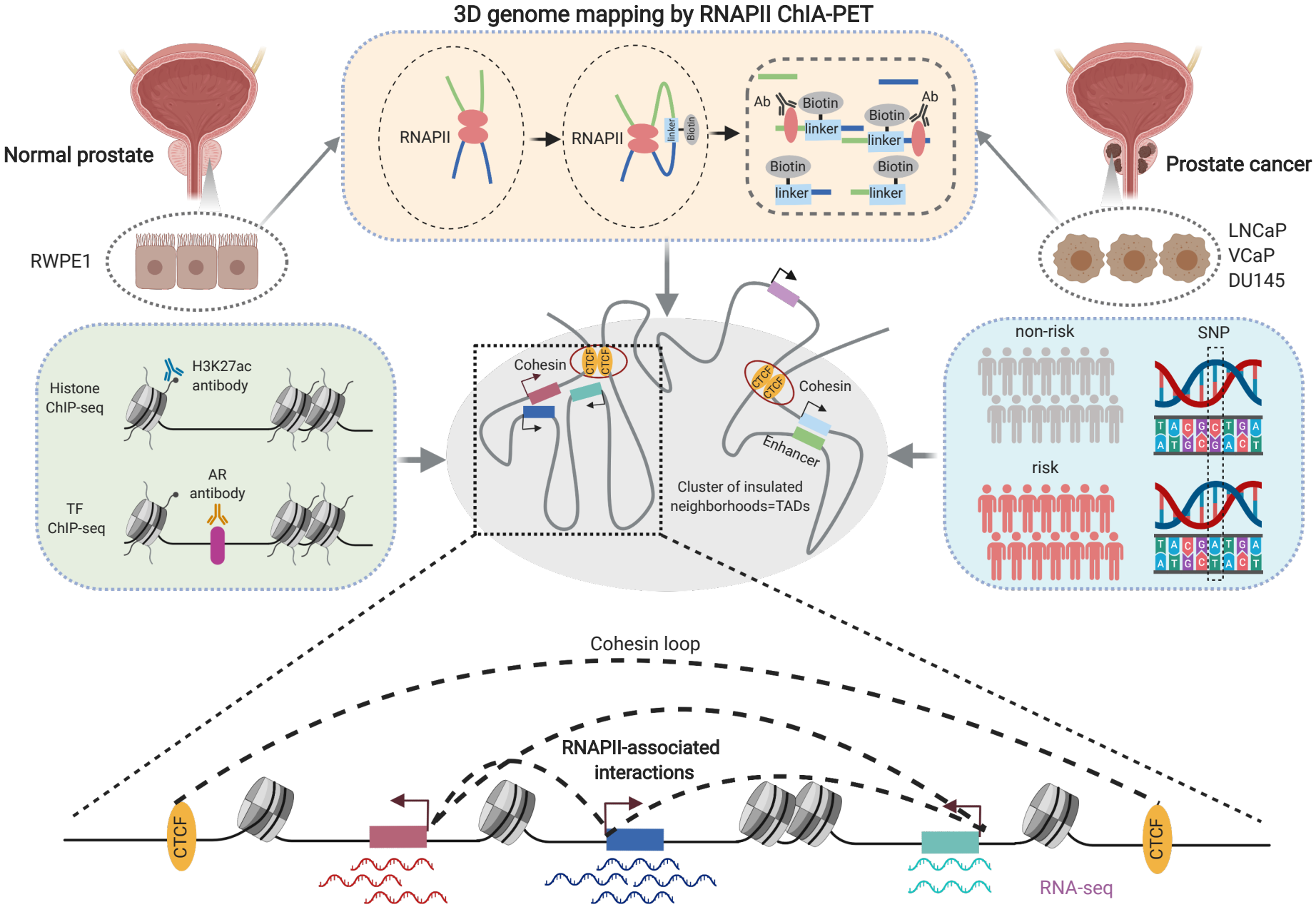
References

1. Schoenfelder, S. and Fraser, P. (2019) Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet 20 (8), 437-455.
2. Li, G. et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148 (1-2), 84-98.
3. Tang, Z. et al. (2015) CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 163 (7), 1611-27.
4. Kim, Y. et al. (2019) Human cohesin compacts DNA by loop extrusion. Science 366 (6471), 1345-1349.
5. Cai, Z. et al. (2020) RIC-seq for global in situ profiling of RNA–RNA spatial interactions. Nature, 1-6.
6. Claussnitzer, M. et al. (2015) FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med 373 (10), 895-907.
7. Jung, I. et al. (2019) A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat Genet 51 (10), 1442-1449.
8. Bertolini, J.A. et al. (2019) Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. Cell Stem Cell 24 (3), 462-476 e6.
9. Hnisz, D. et al. (2016) Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell 167 (5), 1188-1200.
10. Zheng, M. et al. (2019) Multiplex chromatin interactions with single-molecule precision. Nature 566 (7745), 558-562.

Figure legend

Figure 1. 3D genome mapping by RNAPII Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) in normal (RWPE1) and prostate cancer (LNCaP, VCaP, DU145) cell lines. Active histone mark (H3K27ac)/transcription factor (androgen receptor, AR) ChIP-seq, genome-wide association study (GWAS), RNA-seq and other datasets (cohesin ChIA-PET, Hi-C, CTCF ChIP-seq etc.) are integrated with RNAPII ChIA-PET data to reveal RNAPII-associated chromatin interactions (enhancer-promoter interactions, promoter-promoter interactions), their topological basis (insulated neighborhoods or topologically associating domain (TADs)) and how genetic variations modulate promoter/enhancer activity and thus affect enhancer-promoter interactions, leading to aberrant transcription in prostate cancer.

Figure 1


3D genome mapping by RNAPII ChIA-PET
Normal prostate
RWPE1
RNAPII
RNAPII
linker
Biotin
Ab
Biotin
linker
Prostate cancer
LNCaP
VCaP
DU145
Histone ChIP-seq
H3K27ac antibody
TF ChIP-seq
AR antibody
Cohesin
CTCF
Enhancer
Cluster of insulated neighborhoods=TADs
non-risk
SNP
risk
Cohesin loop
RNAPII-associated interactions
CTCF
RNA-seq