

Molecular mechanisms regulating the chromatin architecture and gene transcription in human diseases



Yuliang Feng

St Edmund Hall

A thesis submitted for the degree of
Doctor of Philosophy

Nuffield Department of Orthopaedics, Rheumatology and
Musculoskeletal Sciences

University of Oxford

Trinity Term 2022

Contents

Declaration	6
Acknowledgement	7
Abstract	8
List of abbreviations	11
List of figures	13
List of tables	15
1 Introduction	16
1.1 Overview	17
1.2 Enhancers in transcriptional regulation	17
1.3 Measuring the regulatory epigenome	19
1.4 The hierarchy of 3D chromatin architecture	23
1.5 Mapping of 3D chromatin architecture	26
1.6 3D chromatin architecture in cancer stem cells	32
1.7 Targeting 3D chromatin architecture in cancer stem cells	37
1.8 Research questions	39
2 Rewiring of 3D Chromatin Topology Orchestrates Transcriptional Reprogramming and the Development of Human Dilated Cardiomyopathy	40
2.1 Introduction	41
2.2 Results	44
2.2.1 Genome-Wide Identification of the Active CREs in DCM and NF Hearts	44
2.2.2 Characterization of the 3D Chromatin Organization in DCM and NF Hearts	50
2.2.3 Specific Enhancer-Promoter Connectome Regulates DCM-Specific Transcription	58
2.2.4 DCM-Enriched Enhancer/Promoter Connectome Largely Resides in Conserved Higher-Order Chromatin Architectures and Pre-Accessible Chromatin Sites Bound by Reactivated HAND1	66
2.2.5 HAND1 Rewires Enhancer/Promoter Connectome in Concordance with Transcriptional Changes Associated with DCM Etiology in Human iPSC-Derived Cardiomyocytes	76
2.2.6 HAND1 Drives DCM Pathology in Human iPSC-CM and Mouse Models	81
2.3 Discussion	90
2.4 Materials and Methods	94
2.4.1 Data Availability	94
2.4.2 Culture of hiPSC-CMs and 293 cells	94
2.4.3 ChIP-seq on heart tissues and hiPSC-CMs	95
2.4.4 H3K27ac HiChIP (human heart tissues and hiPSC-CMs) and in-situ HiC (human heart tissues) ..	96
2.4.5 RNA-seq on human heart tissues, hiPSC-CMs and adult mouse CMs	97

2.4.6 ATAC-seq on human heart tissues	99
2.4.7 Library QC metrics collectionHearts	101
2.4.8 ChIP-seq data analysis	101
2.4.9 Enhancer breadth analysis	102
2.4.10 RNA-seq data analysis	103
2.4.11 HiChIP and Hi-C data processing	103
2.4.12 Identification of chromatin interactions and differential loops	104
2.4.13 Validation of differential loops by 3C-qPCR.....	104
2.4.14 Luciferase reporter gene assays	105
2.4.15 Functional enrichment of specific active enhancers or differential loop anchors.....	106
2.4.16 Identification of A/B compartments and topologically associated domains (TADs)	106
2.4.17 Differential TAD analysis	107
2.4.18 ATAC-seq data analysis	107
2.4.19 Transcription factor motif analysis	107
2.4.20 Quantitative analysis of ATAC-seq and ChIP-seq	107
2.4.21 Sample variation analysis for sequencing data	108
2.4.22 Multiple test correction	108
2.4.23 Visualization.....	108
2.4.24 Statistical analysis for sequencing data	109
2.4.25 Plasmid production	109
2.4.26 AAV9 production and purification	109
2.4.27 AAV9 delivery into the heart	110
2.4.28 Adult mouse CM and fibroblast isolation	111
2.4.29 Measurement of contractility and calcium transients in adult mouse CMs	111
2.4.30 hiPSC-CM culture and transduction	112
2.4.31 Measurement of contractility and calcium transients in hiPSC-CMs	113
2.4.32 Western blot analysis.....	113
2.4.33 Quantitative real-time PCR (qPCR)	114
2.4.34 Analysis of cardiac fibrotic area	115
2.4.35 Wheat Germ Agglutinin Staining and Quantify Cell Size	116
2.4.36 Echocardiography.....	116
2.4.37 Animal studies	117
2.4.38 Human studies.....	117
 3 3D genome atlas in pancreatic cancer stem cells	 118
3.1 Introduction	119
3.2 Results	122

3.2.1 Identification of the interactions of cis-regulatory elements by H3K27ac HiChIP	122
3.2.2 Specific CRE interactions regulate metastasis-specific transcription	124
3.2.3 Chromatin accessibility is increased on specific H3K27ac loop anchors	129
3.2.4 FOSL1 is enriched on specific H3K27ac loop anchors in A13D_Org	130
3.2.5 Topologically associating domains are conserved in A13A_Org and A13D_Org	133
3.2.6 Specific CTCF binding on the promoters links a portion of specific CRE interactions.....	136
3.3 Discussion	137
3.4 Materials and Methods	138
3.4.1 HiChIP (H3K27ac)	138
3.4.2 In-situ ChIA-PET (CTCF)	139
3.4.3 ChIP-seq	140
3.4.4 ATAC-seq	142
3.4.5 RNA-seq	143
3.4.6 ChIP-seq data analysis	144
3.4.7 RNA-seq data analysis	145
3.4.8 HiChIP data processing	145
3.4.9 In-situ ChIA-PET data processing	145
3.4.10 Functional enrichment of genes with specific H3K27ac loops.....	146
3.4.11 Identification of topologically associated domains (TADs).....	146
3.4.12 ATAC-seq data analysis	147
3.4.13 Transcription factor motif analysis	147
3.4.14 Quantitative analysis of ATAC-seq and ChIP-seq	147
3.4.15 Visualization.....	148
 4 BRD9 orchestrates stemness and tumorigenesis in pancreatic ductal adenocarcinoma via enhancer-promoter looping	 149
4.1 Introduction	150
4.2 Results	153
4.2.1 Unbiased compound screening for epigenetic regulatory enzymes identifies BRD9 as an important factor for governing PDAC stemness and tumor growth	153
4.2.2 BRD9 inhibition eliminates CSCs from patient tumour	158
4.2.3 BRD9 inhibition disrupts enhancer-promoter connectome of stemness-related genes in pancreatic CSCs	165
4.3 Discussion	169
4.4 Materials and Methods	172
4.4.1 Cell lines and cell culture	172
4.4.2 Tumour sphere assays	172

4.4.3 The small molecule screening library	173
4.4.4 Screening of the chemical compounds.....	177
4.4.5 Establishment of xenograft models in nude mice	178
4.4.6 Tumor growth and mouse survival.....	179
4.4.7 In vivo PET imaging	179
4.4.8 Preparation and Sequencing of Illumina RNA libraries.....	180
4.4.9 RNA-sequencing analysis.....	180
4.4.10 Chromatin Immunoprecipitation (ChIP)	181
4.4.11 Chromatin Immunoprecipitation (ChIP) sequencing	182
4.4.12 ATAC-sequencing.....	183
4.4.13 ATAC-sequencing analysis.....	184
4.4.14 Single-cell gene expression analysis	185
4.4.15 ChIP-sequencing data analysis	186
4.4.16 In-situ ChIA-PET (H3K27ac).....	186
4.4.17 In-situ ChIA-PET data processing	187
4.4.18 Functional enrichment of downregulated genes with IBRD9 treatment.....	188
4.4.19 Visualization.....	189
4.4.20 ChIP-qPCR of anchor regions	189
 5 Conclusions and future direction.....	 191
 Appendix A Publications related to the thesis.....	 196
Appendix B Awards and Honors	198
References	199

Declaration

I confirm that the work presented in this thesis was performed by myself at the Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford between October 2019 and July 2022. The collaborative aspects of this work in which additional parties have given their support have been duly acknowledged within the main text of this thesis. This thesis has been written by myself and has not been previously submitted for the award of a degree or other qualification at this university, or any other institution.

Yuliang Feng



St Edmund Hall
Trinity Term 2022

Acknowledgement

First of all, I would like to express my gratitude and appreciation to my supervisor, Dr. Siim Pauklin. I am grateful for the opportunity to spend the past three years working in his lab, where I can develop and pursue my own research ideas and questions. I am also grateful for my collaborator Dr. Wei Huang (University of Cincinnati), for her continued support for the functional study of cardiac and pancreatic cancer projects, and for her comfort to my mind in the past three years.

I would like to also thank my lab mates, colleagues, friends, Shihong Wu, Chao-hui Chang, Rhiannon French, Lei Jiang for a cherished time spent together in the lab, and in social settings. My appreciation also goes out to my parents and wife for their support all through my studies. Most importantly, I would like to express my gratitude to my grandfather and grandmother who raised me up and gave me endless love but unfortunately passed away during the time of my DPhil study.

Lastly, this work is dedicated to my children, Deyi and Deyue. You have made me stronger, better and more fulfilled than I could have ever imagined. You are my inspiration to achieve greatness.

Abstract

In this thesis, I leveraged multi-omics, computational and functional analysis to shed light on the 3D genome organization and its impact on transcriptional regulation in pancreatic ductal adenocarcinoma (PDAC) and dilated cardiomyopathy (DCM), which are two of the world's deadliest diseases.

I generated a compendium of high-resolution 3D-chromatin contact maps ($\sim 1.2 \times 10^{10}$ reads) from human DCM/non-failing hearts and uncovered a DCM-specific 3D chromatin architecture. I discovered that DCM-specific H3K27ac HiChIP loops largely reside in pre-established high-order chromatin architectures and pre-accessible chromatin sites bound by reactivated *HAND1*. Forced overexpression of *HAND1* in human iPSC-derived cardiomyocytes (hiPSC-CM) induces a distinct gain of enhancer/promoter connectivity and increases the expression of their associated genes implicated for DCM etiology, recapitulating the transcriptional signature in human DCM hearts. In line with this, electrophysiology analysis demonstrated that overexpression of *HAND1* in hiPSC-CM induces defects in calcium handling. Furthermore, cardiomyocyte-specific overexpression of *Hand1* resulted in cardiac hypertrophy/dilated left ventricle and cardiac dysfunction in mouse. These data provided unique insights into topological mechanisms of DCM and demonstrated that *HAND1* rewires enhancer/promoter connectome and thereby drives DCM pathogenesis.

Recent studies demonstrated that the dismal prognosis of pancreatic ductal adenocarcinoma (PDAC) is linked to the presence of pancreatic cancer stem-like cells (CSCs) that respond poorly to current therapeutics. I performed a comprehensive 3D-epigenomic analyses in A13A_Org (primary) and A13D_Org (metastatic) PDAC CSC spheres from the same patient using integrative approaches: HiChIP /ChIA-PET for enhancer connectome/chromatin topology, ChIP-seq for annotation of regulatory elements, ATAC-seq for chromatin accessibility and RNA-seq for transcriptional output to reveal potential topological mechanisms that regulate metastasis-specific transcription in PDAC. By comparing A13D_Org to A13A_Org, I found that H3K27ac-associated enhancer connectome significantly contributes to A13D_Org-specific gene transcription, especially to those prognostic molecular markers in PDAC (e.g. *NRP1*) and cancer stem cell markers (e.g. *BMI1*). In addition, we demonstrated that chromatin accessibility is associated with the formation of A13D_Org-specific H3K27ac HiChIP loop anchors. Moreover, by transcription factor motif scan analysis, we showed that FOSL1 is enriched on A13D_Org-specific H3K27ac HiChIP loop anchors, the expression of which is upregulated in A13D_Org and associated with poor prognosis in PDAC patients. To further dissect the topological mechanism of enhancer-promoter interactions, we performed CTCF ChIA-PET analysis and uncovered that CTCF binding on the promoters but not alteration of topological associated domains (TADs) is linked to a specific enhancer-promoter connectome. In sum, the comprehensive 3D-epigenomic analyses suggest integrated epigenomic and topological mechanisms that regulate PDAC-specific transcription, and provide novel insights into the interplay of epigenetics, and 3D chromatin architecture in PDAC progression.

By performing small-molecule compound screening (142 epigenetic enzymes), my colleague and I uncovered that BRD9 is a key chromatin regulator to orchestrate the self-renewal/stemness of pancreatic CSCs. Perturbation of BRD9 (chemical inhibitor I-BRD9 or genetic ablation) can block the self-renewal and spheroid formation of pancreatic CSCs. Single-cell RNA-seq further revealed that perturbation of BRD9, particularly combining with gemcitabine, dramatically reduces the CSC population within patient tumor cells. In line with this, perturbation of BRD9 combined with gemcitabine, dramatically ameliorates tumorigenesis in mice. Mechanistically, Inhibition of BRD9 reprograms the transcriptional landscape related to stemness and TGF-beta pathway via disruption of enhancer-promoter connectome. Collectively, this study identified the important roles of BRD9 in orchestrating PDAC stemness and tumorigenesis, highlighting BRD9 perturbation as a potential avenue for PDAC treatment.

In summary, I provided reliable and extensive 3D-epigenomic data to interrogate the underlying 3D-chromatin mechanisms that control gene transcription in two human lethal diseases (i.e. DCM and PDAC). These datasets represent a significant advance in our understanding of the etiology of human diseases from a new perspective (chromatin architecture). In addition, the disease-relevant chromatin loops and their regulators (e.g. HAND1, BRD9) identified in these studies could be developed as novel biomarkers for disease diagnosis and be targeted for disease intervention.

List of abbreviations

AAV	adeno-associated virus
ATAC	assay for transposase-accessible chromatin using sequencing
CAGE	cap analysis of gene expression
CFs	cardiac fibroblasts
ChIA-PET	chromatin interaction analysis by paired-end tag sequencing
ChIP-seq	chromatin immunoprecipitation sequencing
CM	cardiomyocyte
CREs	cis-regulatory elements
CSCs	cancer stem cells
DCM	dilated cardiomyopathy
EF	ejection fraction
EMT	epithelial-to-mesenchymal transition
E-P interaction	enhancer-promoter interaction
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit
FREM2	FRAS1 related extracellular matrix 2
FS	fraction shortening
Fbxo25	F-box only protein 25
GAS1	growth arrest-specific 1
GWAS	genome-wide association study
HAND1	heart and neural crest derivatives expressed 1
H3K27ac	acetylation of lysine 27 on histone H3
hiPSC-CMs	human induced pluripotent stem cell-derived cardiomyocytes
Hi-C	high-throughput chromosome conformation capture
hPSCs	human pluripotent stem cells
IGF-1	insulin like growth factor 1
LVDd	left ventricular end-diastolic diameter
LVDs	left ventricular end-systolic diameter
MEF2D	myocyte enhancer factor 2d
MYH7	myosin heavy chain
MYH6	myosin heavy chain
MTSS1	MTSS I-BAR domain containing 1
MYPN	myopalladin
NF	non-failing
NPPA	natriuretic peptide A
NPPA-AS1	NPPA antisense RNA 1
NPPB	natriuretic peptide B
NOTCH1	notch receptor 1
PCA	principal component analysis
PDAC	pancreatic ductal adenocarcinoma
PDE1C	phosphodiesterase 1c
PDLIM5	PDZ and LIM domain 5
PLAC-seq	proximity ligation-assisted ChIP-Seq

SMYD2	SET and MYND domain containing 2
SOX4	SRY-Box Transcription Factor 4
TAC	transverse aortic constriction
TAD	topologically associating domains
TBX5	t-box transcription factor 5
TCGA	The Cancer Genome Atlas
TGF β	transforming growth factor beta
TTSs	transcription start sites
WGA	wheat germ agglutinin
WT	wild-type
TUBA3D	tubulin alpha 3d

List of figures

1 Introduction

Figure 1. 3D genome organization in Nucleus.....	25
Figure 2. Dissection of three-dimensional chromatin architecture	30
Figure 3. Formation of cancer stem cells and epigenetic reprogramming.	33
Figure 4. 3D chromatin architecture in CSCs.	35

2 Rewiring of 3D Chromatin Topology Orchestrates Transcriptional Reprogramming and the Development of Human Dilated Cardiomyopathy

Figure 1. Genome-Wide Identification of Active CREs in DCM and NF Hearts.....	45
Figure 2. Characterization of RNA-seq and H3K27ac ChIP-seq data in DCM and NF Hearts.....	47
Figure 3. Assessment of the Intra and Inter Group Similarity between Contact Matrices in DCM and NF HiChIP Data.	51
Figure 4. Characterization of H3K27ac HiChIP Data in DCM and NF Hearts.	53
Figure 5. Characterization of the 3D Chromatin Organization of the Active CREs by H3K27ac HiChIP in DCM and NF Hearts.....	55
Figure 6. DCM-Enriched Enhancer/Promoter Connectome Contributes to DCM-Specific Transcription.	59
Figure 7. Characterization of DCM/NF-specific enhancer/promoter connectome.	62
Figure 8. Assessment of the Intra and Inter Group Similarity between Contact Matrices in DCM and NF in-situ Hi-C Data.....	67
Figure 9. DCM-Enriched Enhancer/Promoter Connectome Largely Resides in Conserved Higher-Order Chromatin Architectures and Pre-accessible Chromatin Sites Bound by Reactivated HAND1.....	70
Figure 10. DCM-enriched H3K27ac HiChIP Loops Largely Reside in Pre-established High-Order Chromatin Architectures and Pre-accessible Chromatin Sites Bound by Reactivated HAND1.....	73
Figure 11. Overexpression of HAND1 Rewires Long-Range Enhancer/Promoter Connectome and Transcriptome in Human iPSC-Derived Cardiomyocytes.....	77
Figure 12. Overexpression of HAND1 Induces Genome-Wide Gain of Enhancer/Promoter Interactions.	80
Figure 13. HAND1 Induces DCM Phenotypes in Human iPSC-Derived Cardiomyocytes and Mouse Hearts	83
Figure 14. HAND1 Induces DCM Phenotypes in Human iPSC-Derived Cardiomyocytes and Mouse...	87
Figure 15. Model Illustrating 3D Epigenome Organization in Human DCM and How HAND1 Rewires Enhancer/Promoter Connectome to Contribute to Transcriptome Reprogramming and DCM	

Pathogenesis	91
--------------------	----

3 3D genome atlas in pancreatic cancer stem cells

Figure 1. 3D Genome Mapping by H3K27ac HiChIP and CTCF in-situ Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) in A13A (primary pancreatic adenocarcinoma) and A13D (lung metastasis) cell lines derived from the same patient.	121
Figure 2. Interrogation of cis-regulatory element (CRE) by H3K27ac HiChIP.	123
Figure 3. H3K27ac HiChIP identifies specific looping of the active CREs for A13A_Org and A13D_Org-specific transcription.	126
Figure 4. Examples of A13A_Org-specific H3K27ac loops.	127
Figure 5. Examples of A13D_Org-specific H3K27ac loops.	128
Figure 6. Co-transcription of MLLT10-DNAJC1-BMI1 in PDAC patients.	129
Figure 7. Chromatin accessibility on specific H3K27ac loop anchors in A13A_Org and A13D_Org....	130
Figure 8. FOSL1 is enriched on A13D_Org-specific H3K27ac loop anchors.	132
Figure 9. Mapping of chromatin topology by CTCF in-situ ChIA-PET.	134
Figure 10. Conserved TAD boundaries in A13A_Org and A13D_Org.....	135
Figure 11. A portion of CRE contacts is associated with specific CTCF binding on the promoters.....	136

4 BRD9 orchestrates stemness and tumorigenesis in pancreatic ductal adenocarcinoma via enhancer-promoter looping

Figure 1. Inhibition of BRD9 reduces CSCs formation and tumorigenesis in PDAC ..	157
Figure 2. Inhibition of BRD9 reduces CSCs in patient tumors and the 3D enhancer-promoter connectome in CSCs.	161
Figure 3. BRD9 inhibition eliminates CSCs in patient-derived primary PDAC tumor cells.	164
Figure 4. BRD9 inhibition changes the enhancer-promoter connectome of CSCs.....	168

List of tables

2 Rewiring of 3D Chromatin Topology Orchestrates Transcriptional Reprogramming and the Development of Human Dilated Cardiomyopathy

Table 1. Summary of heart sample donors.....	44
--	----

3 3D genome atlas in pancreatic cancer stem cells

Table 1. Overview of 3D epigenome mapping in A13A/A13D cells.....	121
---	-----

4 BRD9 orchestrates stemness and tumorigenesis in pancreatic ductal adenocarcinoma via enhancer-promoter looping

Table 1. List of small molecule compounds.....	173
Table 2. List of antibodies for flow cytometry.....	177
Table 3. List of ChIP-qPCR primers.	189

Chapter 1

Introduction

This chapter is adapted from **Yuliang Feng** and Siim Pauklin. *Nucleic Acids Res.* 2020 Nov 4;48(19):10632-10647; **Yuliang Feng**, Xinguo Liu and Siim Pauklin. *Protein Cell.* 2021 Jun;12(6):440-454.

1.1 Overview

The cell nucleus contains a mixture of macromolecules with chromatin. Although it is crowded, sophisticated and precise gene regulation take place in this environment for cellular homeostasis, which is essential for normal cell function. In recent years, transcriptional dysregulation has been recognized as a universal feature of human diseases, such as cancers. Uncovering the molecular mechanisms of transcriptional dysregulation will provide novel insights into disease development and progression and identify potential therapeutic targets. In a cell, transcriptional output is largely orchestrated by regulatory DNA elements, in particular enhancers. Dysregulated enhancers have been implicated in various diseases¹. In most cases, enhancers control gene expression through interaction with transcription factors (co-factors) and physically contact promoters through long-range interactions. In addition, enhancer-promoter interactions are framed by 3D genome architecture, such as topologically associated domains (TADs), which further aggregate into A/B compartments². Altered chromatin topology has been linked to aberrant enhancer-promoter interaction, which drives transcriptional activation of oncogenes and formation of cancer phenotype³. This has led to a wave of recent studies that investigate how 3D genome architecture regulates chromatin biology in human diseases. Despite this, the mechanisms adopted by cells to construct their unique chromatin architectures remain relatively obscure.

In this chapter, I will provide an overview of chromatin architectures relevant for human diseases and the state-of-the-art approaches to measure the regulatory elements in bulk or single cell level, as well as their applications in biomedical research. I will discuss whether the relation between enhancer-promoter interaction and transcription output is causal or correlative and examine previously unrecognized players regulating enhancer-promoter interactions. Moreover, I explore how 3D chromatin topology is altered in human diseases and lead to aberrant enhancer-promoter interactions and pathogenesis. In addition, I will describe the 3D genome mapping technology in biomedical research.

1.2 Enhancers in transcriptional regulation

Enhancers are key cis-regulatory elements that mediate transcription regulation by converging signals from oncogenic and developmental pathways⁴. The clustering of many enhancers on the same DNA molecule that span tens of kilobases has been termed as a super-enhancer^{5,6}. These large regulatory clusters integrate signals from multiple cell fate pathways, and provide strength and robustness to cell-specific gene transcription^{4,5} via tissue-specific transcription factors binding⁷. Aberrant enhancer activity orchestrates the dysregulated transcription program that dictate disease development and progression. The enhancer landscape of tumor cells is extensively rewired as compared to their normal counterparts. Structural variations (e.g. inversions⁸, translocations⁹, copy number alterations¹⁰), can tether enhancers close to the promoters of oncogenes and thereby drive their expression. For instance, increased copy number of the enhancer 650 kb centromeric to androgen receptor (AR) gene is the key driver for aberrant activation of AR, and drives the metastasis of prostate cancer¹¹. Apart from canonical enhancers, other non-coding elements are emerging as non-canonical enhancers implicated for cancer development.

H3K4me3 is a predominant histone mark for promoters, while H3K4me1 is a predominant histone mark for enhancers¹². The ratio of H3K4me3/H3K4me1 signal can be used as a quantitative measurement to assess the likelihood that a given regulatory element is a promoter or enhancer (i.e. a low ratio suggests enhancer-like while high ratio suggests promoter-like element^{13,14}). In breast cancer and leukemia cell lines, cloning enhancer-like and promoter-like transcription start sites (TSSs) fragments into downstream and upstream of a luciferase reporter gene has shown that enhancer-like TSSs can augment the transcriptional activity of promoter-like TSSs, while showing lower transcriptional activity than the promoter-like TSSs. This suggests that enhancer-like TSSs may function also as enhancers, in addition to being the sites of transcriptional initiation. Currently, Massively Parallel Reporter Assay (MPRA)¹⁵ and Self-Transcribing Active Regulatory Region sequencing (STARR-seq)¹⁶ are two most common genome-wide approaches to measure the enhancer activity. For instance, to measure genome-wide the enhancer activity of promoters, Salvatore Spicuglia lab performed STARR-seq in cervical cancer and leukemia cells and found that 3% of promoters displayed enhancer activity. In

addition, they revealed that the activity of those enhancer-like promoters did not correlate with the transcription of their proximal genes, but instead, they enhanced the expression of distal genes. How those enhancer-like promoters regulated distal but not proximal gene expression needs further studies. In addition, Bing Ren lab developed an approach called cis-regulatory element scan by tiling-deletion and sequencing (CREST-seq) by using high-throughput CRISPR-Cas9-mediated mutagenesis¹⁷ and discovered a large number of enhancer-like promoters regulating *POU5F1* expression. His lab recently further confirmed the genome-wide presence of enhancer-like promoters from 27 human cell/tissue types¹⁸. Future work is required to systemically investigate the contribution of enhancer-like promoter in human disease pathogenesis and whether they have unique properties compared to canonical enhancers in disease-specific gene regulation.

1.3 Measuring the regulatory epigenome

Surveying the exposed DNA regions accessible to transcription factors and their cognate's binding is an essential step to determine the activity of enhancers and promoters. DNase-seq and ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) are the two most common methods to measure chromatin accessibility. DNase-seq uses DNase to digest DNA whereas nucleosome-bound DNA is protected. The digested fragment is then subjected to next generation sequencing to examine open chromatin (50 million uniquely mapping reads) and transcription factor footprinting (150-200 million uniquely mapping reads) according to ENCODE standards (<https://www.encodeproject.org/data-standards/dnase-seq/>). ATAC-seq uses mutated Tn5 transposase preloaded with sequencing adaptors and can simultaneously tag the accessible DNA with sequencing adaptors (called tagmentation), followed by PCR amplification and sequencing¹⁹. ATAC-seq is a simple, bench-efficient method taking a few hours and only requires fresh 50,000 (or even 500) cells, enabling its wide applicability in chromatin accessibility mapping especially for samples with low cell number.

Chang and Greenleaf labs have further optimized their original ATAC-seq protocol and developed Omni-ATAC-seq²⁰ which substantially reduced the mitochondrial DNA

contamination by introducing digitonin in lysis buffer. Moreover, this protocol enabled the interrogation of frozen tissues. As such, they mapped the chromatin accessibility in 410 biobanked primary tumor tissues obtained from The Cancer Genome Atlas (TCGA) representing 23 different types of human cancers. Those samples were performed for whole genome sequencing and RNA-seq, enabling multi-omics data integration to interrogate how the interplay of genetic variation and regulome drives carcinogenesis. In addition, these are great reference regulome maps for cancer research because chromatin accessibility mapping has also been done for K562 (leukemia) and GM12878 (lymphoblastoid), HepG2 (hepatocellular carcinoma), HeLa-S3 (cervical carcinoma), MCF-7 (breast cancer), A549 (lung adenocarcinoma), SK-N-SH (neuroblastoma) cancer cell lines in ENCODE consortium²¹, which complement data from primary tumors. In the future, large regulome mapping to understand the multilayered transcriptional regulation of each type of cancer comprising hundreds to thousands of samples can be performed to identify novel subtypes of cancers. Integrating clinical data for patients with such multi-omics data would further uncover novel biomarkers associated with disease prognosis and guide the clinical strategy for cancer patients.

Recently, single cell ATAC-seq (scATAC-seq) was developed independently by William Greenleaf lab and Jay Shendure lab, enabling the identification of the open/accessible chromatin at single cell level^{22,23}. After that, Eileen Furlong lab and Jay Shendure lab further collaborated to adapt the scATAC-seq protocol to work with nuclei from formaldehyde-fixed tissues²⁴. Based on scATAC-seq approach, it is possible to identify cell subpopulations with different chromatin accessibility profiles within complex samples. For instance, scATAC-seq approach has been employed to interrogate the formation of chromatin accessibility in human hematopoiesis^{25,26}, human acute myeloid leukemia (AML)²⁶, embryonic development in *Drosophila*²⁴, and mouse cell atlas²⁴. By combining the Chromium platform (10x Genomics), a massive scale of single cell mapping can be achieved. This has been used in tumor biopsies with pre- and post- PD-1 inhibition which identified therapy responsive-T cell subtypes and the associated regulatory programs in a report by Howard Chang, William Greenleaf and Grace Zheng laboratories²⁷. In addition, William Greenleaf lab has performed scATAC-seq and CITE-seq (cellular

indexing of transcriptomes and epitopes by sequencing) profiling of mixed-phenotype acute leukemias (MPAL), showing that RUNX1 is a key transcription factor regulating the transcriptional signature in MPAL²⁸.

In the future, it will be attractive to apply this approach for decoding the regulome of various human cancers including tumor cell heterogeneity and their microenvironments (e.g. immune cells, tumor-associated fibroblast) in cancer development and progression. It will also allow analysing pre- and post- treatment with candidate compounds, providing novel insight into evolutionary trajectories for cancers as well as identifying the unique subpopulations which are implicated in therapeutic resistance. Moreover, applying this approach on circulating tumor cells (CTCs) even allows tracking the alteration of cancer cells in 'real time' without invasive biopsy and assess the therapeutic response/outcome. Although scATAC-seq method is appealing, some technical challenges need to be overcome in the future: 1) its sensitivity to tissue dissociation and how to reduce the Tagmentation bias created by different cells subjected to the same nuclear lysis conditions; 2) some computational methods (e.g. label transfer²⁴) has been developed to integrate scATAC-seq and single-cell RNA-seq (scRNA-seq) data but pairing of subpopulations of scATAC-seq and scRNA-seq is challenging; 3) The coverage for each single cell is still low due to capturing a small subset of open chromatin in one cell, and it is also computationally challenging to distinguish the uncaptured regions due to technical limitations and transcriptionally inactive regions. Hence, simultaneous measurements of chromatin accessibility and gene transcription in the same individual cell will be the ideal solution to solve this issue. In light of this, single-cell combinatorial indexing for jointly profiling of chromatin accessibility and mRNA (sci-CAR) has been developed²⁹. To confirm that the profiling is truly from the same cells, the authors mixed human and mouse cells and found that 99% of co-assayed cells were assigned to the same species, indicating its high accuracy to perform these two assays within the same cells.

It should be noted that chromatin accessibility alone is not sufficient to alter the activity of regulatory elements. A recent study revealed that during the progression of pancreatic cancer, enhancers are extensively reprogrammed but metastasis-specific enhancers

have equivalent ATAC-seq signal compared to primary tumors, indicating the chromatin accessibility for those sites is pre-established in primary tumor³⁰. In addition, in T cells, most of subtype-specific enhancer loops shared equivalent chromatin accessibility on the anchors³¹. This indicated that other parameters should be included to comprehensively evaluate the regulatory potential of the elements, for example using ChIP-seq for mapping active enhancer marks (e.g. H3K27ac). Traditionally, ChIP-seq requires millions of cells which limits its application in primary cancer samples. Recently, Cleavage Under Targets and Tagmentation (CUT&Tag)³² and CoBATCH³³ were developed as novel low cell number ChIP-seq approaches by requiring as few as 60 cells. Moreover, their basic protocols can be further adapted for single cell analysis. These two approaches used the same principle: 1) nuclear permeabilization similar to ATAC-seq; 2) incubating with primary antibody to bind the protein of interest (e.g. H3K27ac) in the nuclei; 3) secondary antibody incubation that binds to primary antibody; 4) addition of transposome consisting of protein A/G-Tn5 transposase (pA/G-Tn5) fusion protein loaded with sequencing adapters; 5) Activation of Tn5 by adding Mg^{++} for *in situ* cutting of DNA/protein complex; 6) PCR amplification. These two approaches dramatically reduce the required materials such as cells/tissues for ChIP-seq, increase signal-noise ratio, reduce sequence depth (about 10 fold) and can generate the sequence-ready libraries in the same day.

Although H3K27ac is the best marker for active enhancer, whether H3K27ac only correlates or is causal to enhancer activity, is a hotly debated question in the epigenetic field. To address this question, Bing Zhu lab made a point mutation of lysine 27 to arginine (an unmodifiable residue) on histone H3.3 (H3.3K27R) in mouse embryonic stem cells, which made the cells lose the capability of being acetylated. However, loss of H3K27ac at enhancer regions did not affect the chromatin accessibility, transcription and self-renewal of mouse embryonic stem cells (mESCs)³⁴. A previous study indicated that acetylation can neutralize positive charges on the lysine of histone, reduce the binding of histone and DNA and open the chromatin³⁵. However, this study suggested that in mESCs, single lysine residue acetylation alone is not sufficient to affect the enhancer activity and it might need to synergistically work with other lysine residue acetylations to make the chromatin loosen and maintain the active state of enhancers. The effect of

H3K27ac on the enhancer activity of other cell types such as cancer cells would still need to be examined by using the same approach.

It should also be emphasized that assessing the functionality of enhancers only based on epigenomic mapping might not be sufficient due to the presence of redundancy: one enhancer can compensate the inactivation of another enhancer to regulate the same gene, which is supposed to confer the robustness in transcriptional activity^{36,37} by buffering the genetic or epigenetic changes of individual enhancers. CRISPR-based *in vivo* genome-editing methods should be introduced to mutate or delete an enhancer, measure the resulting transcriptional output and thus confirm their functional contribution to transcription in cancer.

1.4 The hierarchy of 3D chromatin architecture

Only approximately 7% of enhancers in human cells have been estimated to control their closest promoters while other enhancers can bypass their closest genes and regulate the target promoters through long-range physical communications and thereby impact cellular plasticity and differentiation propensity. In most cases, enhancers interact with their target gene promoters within the same chromosome. However, increasing evidence suggests the existence of interchromosomal enhancer-promoter interactions³⁸. For example, a recent report identified that specific super-enhancers from different chromosomes come into proximity and regulate the transcription essential for identity of olfactory neurons³⁹.

Mechanistically, enhancer-promoter contact regulates gene expression via increasing transcriptional bursting fraction (more transcriptional events per time frame) but not bursting size (more RNA molecules per transcriptional event)⁴⁰. It should be noted that enhancer-promoter contact may not be necessary for active transcription as seen from a recent report showing decreased enhancer-promoter proximity in *Sonic hedgehog (Shh)* gene activation⁴¹, implying that other mechanisms, such as phase separation may be involved in transcription regulation⁴².

Loop extrusion model proposed that topological structures called insulated neighbourhoods can frame and facilitate enhancer-promoter interactions. In this model, cohesin-containing extrusion complex loads onto the DNA and extrudes the DNA loop until blocked by convergently oriented CTCF molecules. This loop extrusion process depends on the hydrolysis of ATP by cohesin's ATPase activity⁴³⁻⁴⁵. Inhibition of ATP production by oligomycin dramatically disrupted establishment of RAD21 (a cohesin subunit)-associated chromatin loops. In addition, Encyclopedia of DNA Elements (ENCODE) consortium generated RAD21-associated chromatin contact maps by ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) in 24 human cell lines and identified that 28% of all RAD21 loops are cell-type specific and overlapped with cell-type specific enhancers marked by H3K27ac⁴⁶, suggesting that specific enhancer-promoter contacts may be orchestrated by cell-specific insulated neighbourhoods.

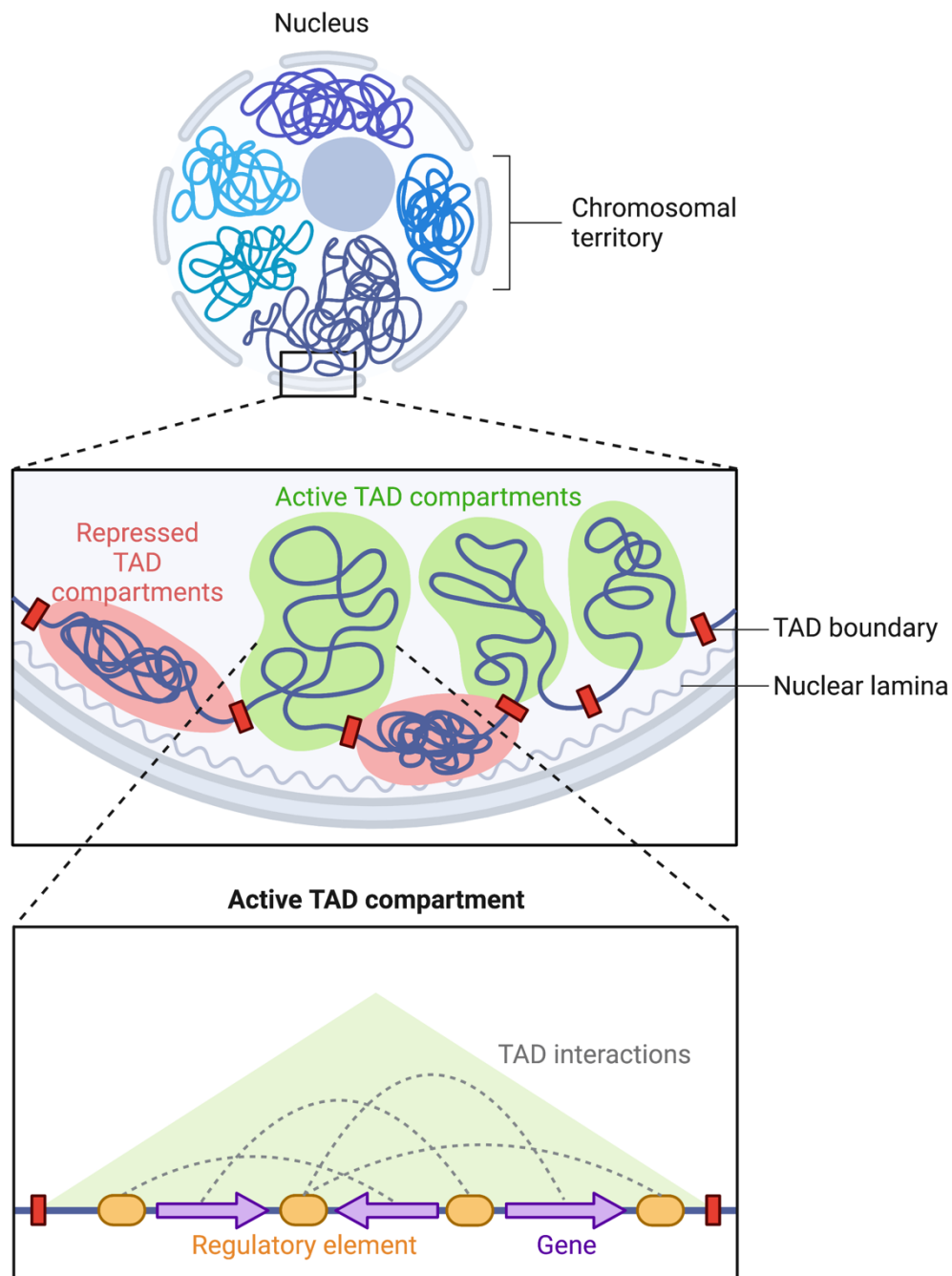


Figure 1. 3D genome organization in Nucleus. The human genome does not exist as a linear molecule but instead is hierarchically packaged into the tiny nucleus. This complex genome organization is essential for the transcriptional control in the genome. Basically the 3D genome includes multiscale structural units, like topologically associating domains (TADs), which frame enhancer-promoter interaction. TADs will

further aggregate into transcriptionally active compartment (A compartment) and transcriptionally inactive compartment (B compartment), which are organized into chromosome territories.

In addition, insulated neighbourhoods can further aggregate into topologically associating domains (TADs), whose boundaries are demarcated by CTCF binding^{47,48}. TADs are clustered into larger domains called A compartments (expression active) and B compartments (expression inactive)⁴⁹ (**Figure 1**). However, degradation of CTCF by auxin system only led to the loss of TADs but compartment structures remained largely unchanged, suggesting that the principles governing compartment organization are independent from TADs⁵⁰. A recent study found that compartment shift is correlated with suppression of the stemness program and tumor progression, and related to DNA hypomethylation⁵¹. Whether modulation of DNA methylation state can causally alter the compartment switch should be further examined in the future.

Based on cell population Hi-C data, previous studies have demonstrated that TADs remain largely unchanged during cellular specification^{52,53}. However, by single-cell Hi-C approach, a recent study showed that TADs structures vary substantially at single-cell level⁵⁴ and CTCF prevents the interactions of inter-TADs⁵⁵. Furthermore, by combining single-cell Hi-C and high-resolution microscopy, the authors revealed that TADs can be subdivided into chromatin nanodomains (CND), which depend on nucleosome-nucleosome interactions but are depleted of CTCF or cohesin⁵⁵. In trichostatin A (an inhibitor of histone deacetylases) treated cells, TADs remain unchanged but CND organization is largely disrupted, suggesting CND organization relies on the histone acetylation state⁵⁵.

1.5 Mapping of 3D chromatin architecture

Current Hi-C (in solution and in situ Hi-C) approaches are able to capture all-to-all interactions in the nucleus but are required to generate many replicates of libraries for ultra-deep sequencing⁵⁶. This requires much material for producing libraries and involves considerable costs for sequencing. Particularly, it is unrealistic for primary cancer

samples. Thus, in many current publications, the sequencing depth is not enough to examine enhancer-promoter interactions.

Amos Tanay lab developed a novel 4C approach, called unique molecular identifier (UMI)-4C, which can examine the chromatin interactions for one bait or multiplexed baits (up to 20-50) at high-resolution⁵⁷. In addition, Douglas Higgs/Jim Hughes labs further refined their previously developed Capture-C approach⁵⁸ and presented a new next-generation (NG) Capture-C protocol⁵⁹. By using a new oligonucleotide capture process, NG Capture-C mapping achieved unprecedented sensitivity and resolution of mammalian genomes. The initial Capture-C protocol required the design of a minimum of 40,000 probes (irrespective of the number of desired viewpoints). However, in NG Capture-C protocol, each set of capture oligos only targets one to many hundreds of regions and new capture oligos can be added to the existing pool for the experimental need⁵⁹, thus providing great flexibility and dramatically reducing the cost for an experiment with a small design.

In addition to Capture-C approach, protein-centric approaches (e.g. HiChIP⁶⁰, Proximity Ligation-Assisted ChIP-seq (PLAC-seq)⁶¹, in situ ChIA-PET⁶²) using an antibody of interest (e.g. H3K27ac, RNAPII, CTCF, SMC1a) can also examine chromatin looping at high-resolution with dramatically reduced sequencing depths and starting materials (e.g. 1 million cells), providing the technical basis for 3D genome mapping in cancer research using primary cells/tissues. It had been observed that many regulatory elements such as enhancers and promoters could be aggregated into multiplex chromatin structures, suggesting this might be the topological basis for gene co-regulation. However, the aforementioned 3D genome technologies are based on proximity-ligation and paired-end-tag (PET) sequencing and detect only binary interactions (**Figure 2a**). In addition, the multiplex chromatin structure mapping results derived from those approaches were mixed results derived from millions of cells. Thus, it cannot reflect true multiplex chromatin interactions in single chromosomes and do not reflect chromatin heterogeneity among individual cells. These limitations have created the biases in visualizing the bona fide picture of chromatin organization (**Figure 2b,c**). Considering these biases, three ligation-

free approaches ((genome architecture mapping) GAM⁶³, (split-pool recognition of interactions by tag extension) SPRITE⁶⁴ and chromatin-interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop)⁶⁵) were recently developed to examine genome-wide chromatin interactions. All these approaches can detect the genome-wide chromatin interactions between three or more genomic loci simultaneously. GAM uses ultra-thin cryosectioning with a 220 nm thickness to dissect the nuclei into slices, and then for each slice, the DNA is extracted, barcoded and amplified before all the slices are multiplexed for sequencing. For example, 400 slices (each slice=1 million sequencing reads) can achieve about 30 kb resolution. SPRITE and ChIA-Drop require crosslinking before chromatin preparation. These two methods need chromatin fragmentation (TurboDNase I for SPRITE, Mbol for short chromatin fragment and sonication for long chromatin fragment (~6 kb)) for ChIA-Drop) but do not require proximity ligation. For SPRITE, the fragmented chromatin is splitted into 96-well plate (each well has unique barcode), re-pooled and splitted into 96-well plate for couple rounds for barcoding and sequencing. For ChIA-Drop, the fragmented chromatin was subjected to chromatin immunoprecipitation to capture specific protein (e.g. RNAPII and CTCF) associated chromatin interactions. Then the immunoprecipitated DNA was loaded onto the microfluidic device to generate gel-bead-in-emulsion (GEM) droplets. Each droplet contains unique barcode and the reagents for DNA amplifications (**Figure 2d**). Data analysis can be done by a newly developed pipeline called ChIA-Dropbox⁶⁶. By employing RNAPII ChIA-Drop in *Drosophila* S2, Ruan lab found that 80% of RNAPII associated chromatin complex contains only one promoter. Although the remaining 20% of chromatin complexes contain multiple promoters and supported the notion of transcription factories⁶⁷, promoter-promoter interactions are less common than previously supposed¹³. It would be interesting to perform ChIA-Drop only using the sonicated chromatin without chromatin immunoprecipitation and ultra-deep sequencing (billions of reads) to generate “all-to-all” 3D contact map. This modification would potentially identify the chromatin interactions at single-molecule resolution, which are not associated with common histone marks (e.g. H3K27ac)/transcription factors (e.g. RNAPII) related to active transcription or architectural proteins (e.g. CTCF/Cohesin/YY1) related to chromatin topology. However, considering the sequencing cost, protein-centric ChIA-Drop would be a more “economic”

option for epigenome consortia (e.g. ENCODE, 4D Nucleome) to generate large-scale 3D chromatin maps at single-molecule resolution.

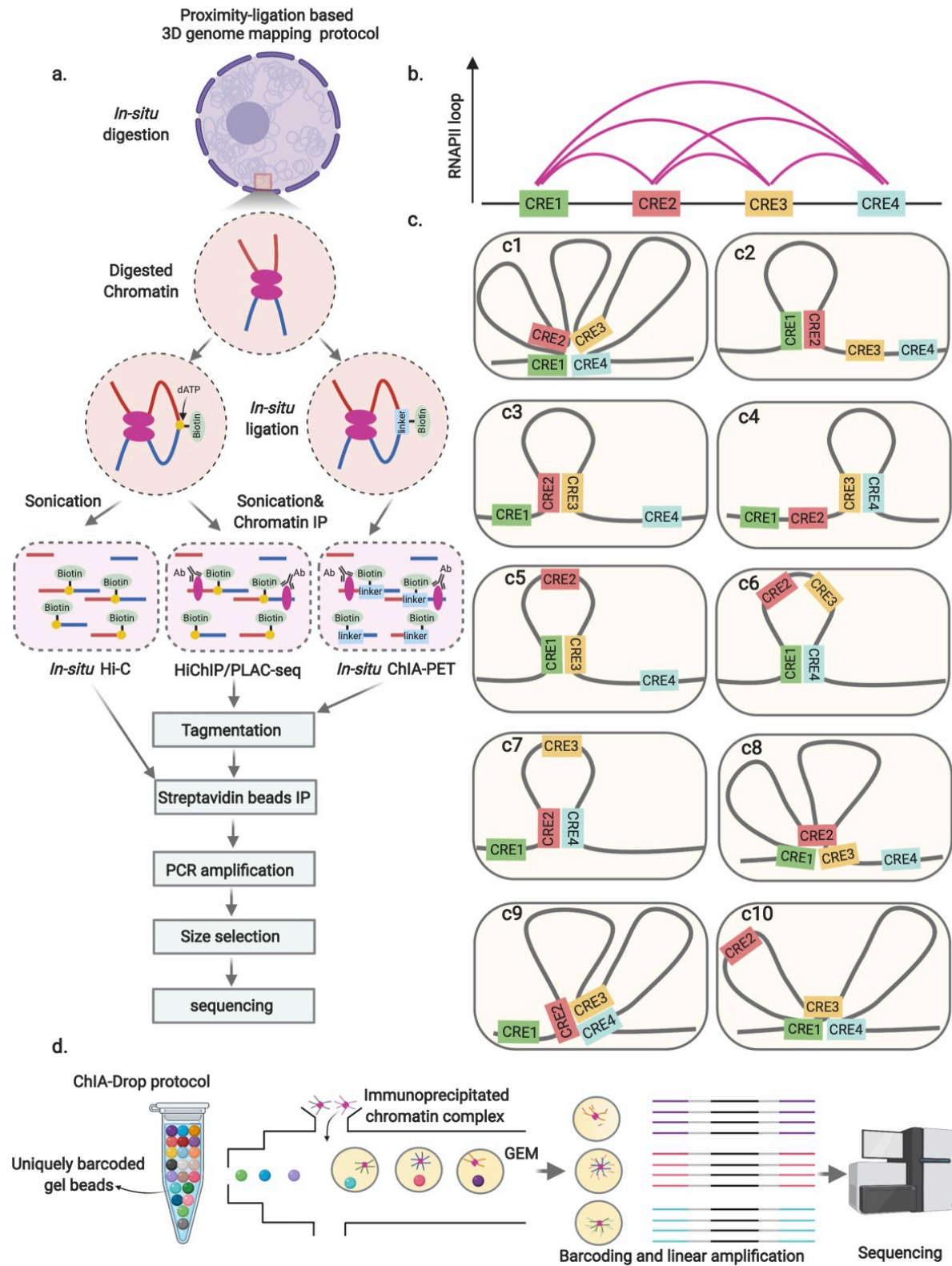


Figure 2. Dissection of three-dimensional chromatin architecture. *a.* Proximity-ligation based 3D genome mapping protocol: crosslinked cells are subjected to nuclei isolation. *In-situ* enzymatic digestion

of chromatin is performed, followed by in-situ ligation (in-situ Hi-C and HiChIP/PLAC-seq: biotin-labeled dATP, in-situ ChIA-PET: bio-labeled bridge linker) of two chromatin fragments which are proximal in space. Thereafter, the nuclei are subjected to sonication. For HiChIP/PLAC-seq and in-situ ChIA-PET, chromatin immunoprecipitation is required to capture the specific protein-centric chromatin interactions (e.g. RNAPII). Subsequently, the chromatin for all these assays is treated by decrosslinking and DNA purification. Then the purified DNA is tagged by transposon (Tn5) preloaded with sequencing adapters, followed by streptavidin bead immunoprecipitation (to capture the DNA fragments conjugated with biotin), PCR amplification, size selection and sequencing; **b.** Example of a chromatin interaction cluster (multiplex chromatin structure) that involves 4 cis-regulatory elements (CREs), as revealed by HiChIP/PLAC-seq or in-situ ChIA-PET using RNAPII antibody. **c.** 10 possible scenarios of interactions in individual cells speculated based on **3b**. Current proximity-ligation based 3D genome mapping technologies can only detect binary chromatin interactions, and cannot explore the existence of multiplex chromatin structures in single chromosome and its cellular heterogeneity, which hardly allow a clear view of the chromatin interactions and thus miss their implications in cancer. For example, it cannot be judged whether those chromatin interactions co-exist in individual cells (**3c1**) or are they just the ensemble of various chromatin interactions from individual cells (**3c2-c10**). **d.** illustration for the experimental procedures of ligation-free 3D genome mapping technology ChIA-Drop: crosslinked cells are subjected to sonication and chromatin immunoprecipitation. The precipitated chromatin is then loaded into the microfluidic device to generate gel-bead-in-emulsion (GEM). Each GEM from contains the reagents and unique barcodes for PCR amplification and sequencing. Amplicons from the same GEM have the same barcodes, enabling the inference of multiplex chromatin interactions. Figure adapted with permission from Feng et al., *Nucleic Acids Research*, 2020).

Visualization can further validate the findings from sequencing-based approaches and allow investigating the variability of individual cells. The most common method is 3D DNA FISH in fixed cells. For example, this approach was recently employed to visualize the 3D Cliques governing type 1 diabetes in individual T cells⁶⁸. Stochastic Optical Reconstruction Microscopy (STORM) and Photoactivated Localization Microscopy (PAML) has been widely used for super-resolution imaging of 3D genome organization⁶⁹. For example, by using a modified STORM approach, Xiaowei Zhuang lab identified TAD-like domains at the single-cell level in the absence of cohesin⁷⁰. CRISPR/Cas9 system has been employed to tag genomic loci in live cells using guide RNA conjugating catalytically inactive forms of the Cas9 (dCas9) with green fluorescent protein (GFP), red fluorescent protein (RFP), or blue fluorescent protein (BFP)⁷¹. A more robust approach termed Chimeric Array of gRNA Oligonucleotides (CARGO)-dCas9 was recently developed by delivering multiple RNAs to guide dCas9 and efficiently label the regulatory elements, and track their dynamics in living cells⁷². With the improvement of imaging techniques, more genomic loci can be probed simultaneously providing more comprehensive views of how the genome is reorganized during the progression of carcinogenesis.

1.6 3D chromatin architecture in cancer stem cells

Cancer stem cells or tumour initiating cells exhibit stem cell-like features such as self-renewal and differentiation capacity. CSCs are also resistant to anoikis, a form of programmed cell death when the cells are detached from the surrounding extracellular matrix (ECM). Therefore, the tumorigenic process with dedifferentiation of tissue cells to a stem cell-like or progenitor state seems to be the opposite process to normal development during early embryogenesis and organogenesis (**Figure 3**). Dedifferentiation has a similar developmentally reversed direction as the generation of inducible pluripotent stem cells by epigenetic reprogramming. During this reprogramming, exogenous expression of pluripotency factors induces extensive epigenetic remodelling that leads to the activation of an endogenous gene circuitry that maintains the pluripotent state of cells⁷³. The resemblance of these two processes – dedifferentiation and epigenetic reprogramming - is highlighted by the effects of tumour suppressors such as

p53 and cyclin dependent kinase inhibitors (e.g. p16) in blocking epigenetic reprogramming, while their inactivation increases the epigenetic reprogramming efficiency. p53 is also the most frequently mutated gene across all tumours. For instance, in pancreatic ductal adenocarcinoma it is, together with p16 loss of function, a hallmark mutation.

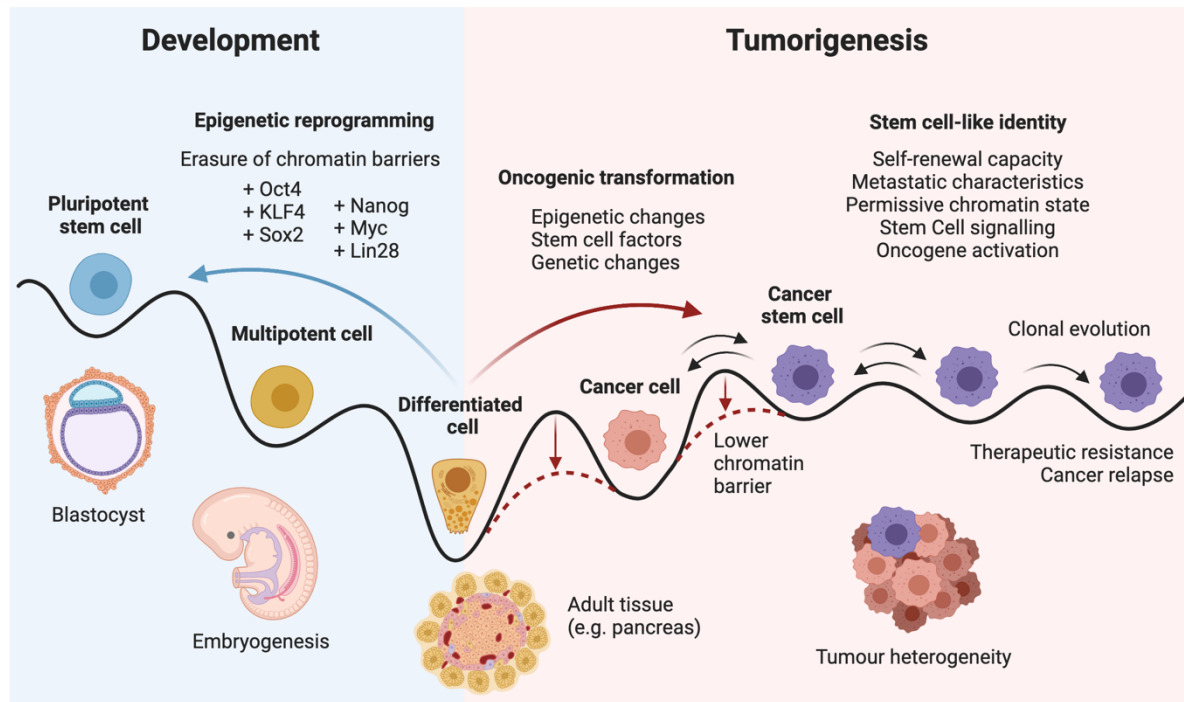


Figure 3. Formation of cancer stem cells and epigenetic reprogramming. Schematic depiction of cell state transitions during early development and tumorigenesis indicate dynamic changes according to cell types. Pluripotent stem cells are able to differentiate to all cell types and lead to fully differentiated cells in adult tissues. Epigenetic reprogramming by expression of various stem cell factors (e.g. Oct4, Sox2, KLF4, Myc, Nanog, Lin28) leads to the erasure of the epigenetic barriers and generation of induced pluripotent stem cells. Oncogenic transformation has the opposite direction to normal cell differentiation and the epigenetic changes, expression of stem cell factors and genetic mutations, can facilitate tumorigenesis by lowering the barriers that usually would prevent tumorigenesis and cell state changes. Tumorigenesis involves differentiation of the cell state to a dysregulated stem cell-like identity known as cancer stem cells. These cells are developmentally plastic and can self-renew but also differentiate to other cancer cells. Figure adapted with permission from Feng et al., *Protein and Cell*, 2021.

There is a further link between reprogramming and oncogenic transformation. Transient expression of reprogramming factors *in vivo* in mouse results in tumour development in

various tissues consisting of undifferentiated dysplastic cells exhibiting global changes in DNA methylation patterns. This indicates epigenetic regulation associated with reprogramming in the absence of irreversible genetic transformations may drive development of particular types of cancer⁷⁴. Furthermore, transient expression of reprogramming factors in *Kras* mutant mice is sufficient to induce the robust and persistent activation of ERK signaling in acinar cells and rapid formation of pancreatic ductal adenocarcinoma⁷⁵, indicating that reprogramming factors promote oncogenic transformation if the anti-tumorigenic barriers have been removed.

During cancer development and progression, we hypothesize that the dedifferentiation of cells with expression of some stem cell factors leads to the reorganisation of epigenetic signatures and 3D chromatin architecture in developmentally plastic cancer cells (**Figure 4**).

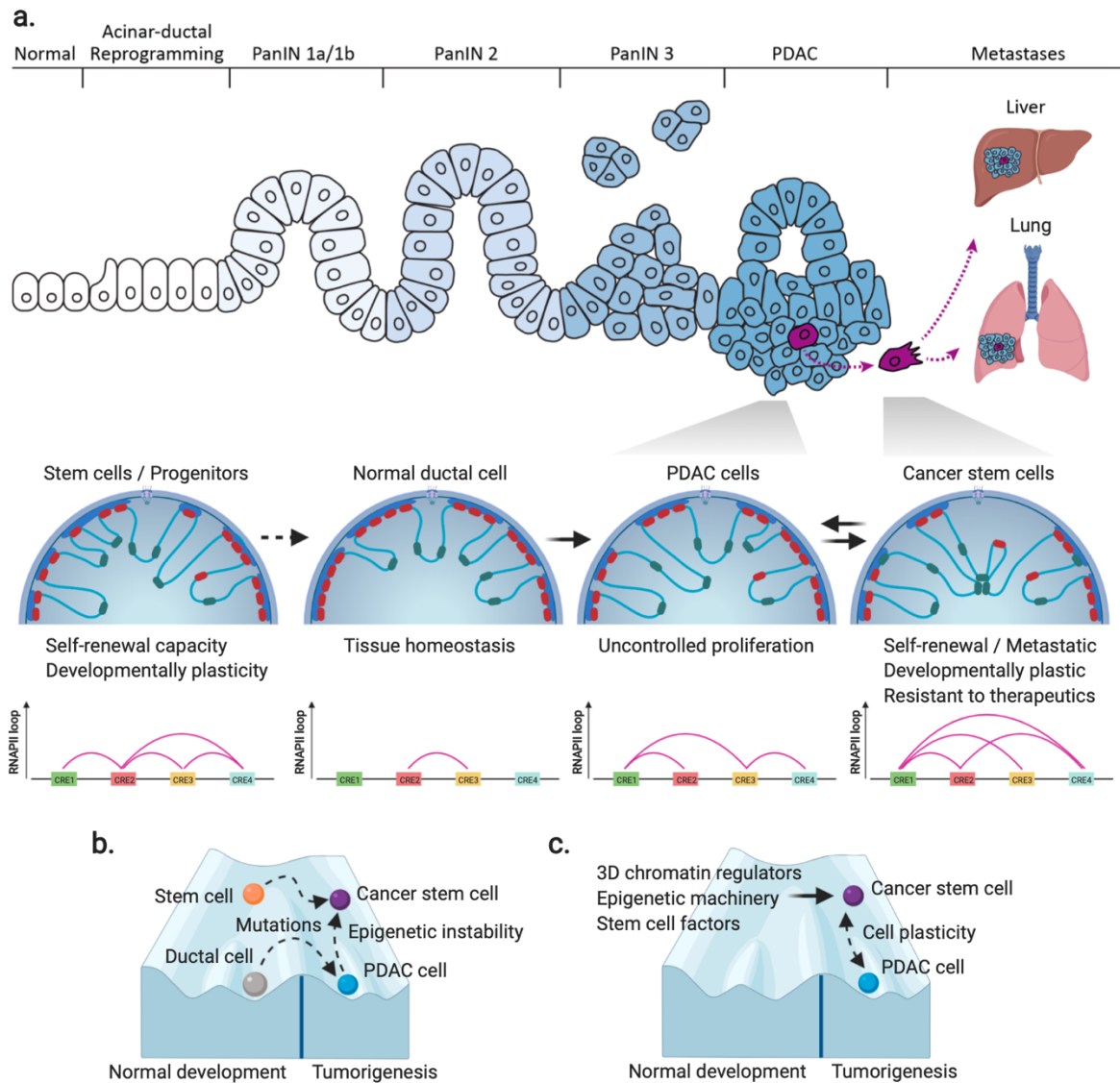


Figure 4. 3D chromatin architecture in CSCs. *a.* Rearrangement of the higher-order epigenome during tumorigenesis leads to changes in the cell state from differentiated cell to cancer cells and to cancer stem cells (CSCs). Pancreatic ductal adenocarcinoma (PDAC) formation occurs by gradual transitioning of normal ductal cells to Pancreatic intraepithelial neoplasia (PanIN) lesions that ultimately leads to PDAC formation. The PDAC is a heterogenous cell population containing cancer stem cells (CSCs) that can self-renew, are developmental plasticity, highly metastatic and more resistant to conventional therapeutics than other cancer cells. The simplified depiction of the higher-order epigenome inside the nuclei during tumorigenesis indicating rearrangements of active (green) and inactive (red) chromatin regions as well as lamina-associated domains (blue). The genome is organized into topologically associated domains that form gene expression domains with enhancer-promoter crosstalk. Different chromatin regions are postulated to be brought together in different cell states such as CSCs by transcription factors and

regulatory proteins that facilitate chromatin accessibility and gene expression regulation. The differences in cell cell states are postulated to manifest at the higher-order but also lower-order levels of chromatin that bind RNAPII and sequence-specific transcription factors (pink lines show interaction clusters), as depicted schematically by four cis-regulatory elements (CREs). **b.** The Waddington landscape of normal tissue development and tumorigenesis. In normal tissue formation, stem cells or progenitor cells differentiated to genetically and epigenetically stable cell types, such as ductal cells in pancreatic tissue. Genomic instability due to the accumulation of mutations in driver genes such as EGFR, KRAS and p53, and epigenomic instability leads to the formation of CSCs and non-CSCs either from normal stem cells or differentiated ductal cells, or through further dedifferentiation of non-CSCs to CSCs. **c.** The phenotypic plasticity of CSCs and non-CSCs depicted on the Waddington landscape. The plasticity of CSCs allows the cells to give rise to the whole tumour in its entire cellular heterogeneity while non-CSCs are less plastic. The plasticity of hypothetically regulated by the interplay between 3D epigenome, epigenetic machinery, stem cell factors and structural proteins that regulate gene expression. Figure adapted with permission from Feng et al., *Protein and Cell*, 2021.

The arising cells might share some or dysregulated features seen also in normal ESCs. Interestingly, the chromatin organisation of embryonic stem cells has several unique characteristics: they possess open chromatin characteristics and display chromatin hyperplasticity that distinguishes them from somatic cells.

Nuclei of hESCs have high physical plasticity that stiffens upon differentiation⁷⁶, a decondensed less compact chromatin^{77,78} with an enrichment of active histone modifications⁷⁹ and considerably decreased heterochromatin, including peripheral heterochromatin⁸⁰. The nuclear lamina in ESCs is largely devoid of Lamin A, less organized and more wrinkled⁸¹. hESCs have high levels of transcriptional activity and are considerably depleted for heterochromatin-associated histone modifications, particularly the Polycomb-associated H3K27me3, which spreads during differentiation and specification to distinct cell types^{82,83}. Structural proteins such as heterochromatin protein 1 (HP1), the linker histone H1, histone H3.3 and the core histones display highly dynamic kinetics in their association with chromatin in ESCs, further opening the chromatin structure⁸⁴⁻⁸⁶. Reprogrammed pluripotent stem cells also have increased dynamic association of a heterochromatin protein HP1b compared to differentiated cells^{85,87}. Conversely, the open chromatin of ESCs requires the involvement of chromatin

repressive complexes in protecting against inappropriate transcription of differentiation factors, and for the ordered repression of pluripotency-associated genes during differentiation. Developmental specification of hESCs is accompanied by progressive chromatin restriction from dynamic remodeling to generalized compaction with decreased transcriptional activity. This is accompanied by the reorganization of the histone variant H2A.Z from a broad distribution in hESCs to more restricted regions in promoters and distal regulatory elements in differentiated cells⁸², and formation of large H3K9me3-positive heterochromatic foci⁸⁵.

Deep sequencing of cancers has revealed the occurrence of mutations in epigenetic machinery, including regulators of DNA methylation, histone modification and chromatin organization, across a wide variety of cancer types. Recurrent somatic mutations in genes encoding chromatin-associated proteins (polycomb repressive complexes, HDAC1- and HDAC2-containing complexes) indicate that a deregulated chromatin environment can play a causal role in cancer^{88,89}. Cancers also often have aberrant promoter CpG island hypermethylation and transcriptional silencing of tumor suppressor genes and pro-differentiation factors. In ESCs, these genes are held in a “transcription-ready” state mediated by a “bivalent” promoter chromatin pattern consisting of H3K27me3 (repressive) and H3K4me3 (active) histone marks. Embryonic carcinoma cells have two additional repressive marks, H3K9me2 and H3K9me3, both associated with DNA hypermethylation in adult cancers^{90,91}. In human cancer, promoters with a high H3K27me3:H3K4me3 ratio are particularly susceptible to DNA hypermethylation⁹². These results suggest that tumor-specific targeting of *de novo* methylation is pre-programmed by an established epigenetic system that normally has a role in marking embryonic genes for repression. Altogether, transient silencing of regulatory genes in stem or progenitor cells with a bivalent mark may leave these genes vulnerable to aberrant DNA hypermethylation and heritable gene silencing during tumour initiation and progression.

1.7 Targeting 3D chromatin architecture in cancer stem cells

Cancer cells can establish *de novo* oncogenic super-enhancers that induce proliferation. In turn, disruption of super-enhancers can selectively inhibit tumour oncogenes⁵. Since

super-enhancers of stem cells receive signals from the extracellular milieu and this regulates stem cell/progenitor plasticity and lineage choice⁹³, it is possible that similar microenvironmental effects play important roles also for super-enhancer regulation in cancer cells, thus underlining the importance of the CSC niche. While targeting the tumour microenvironment or CSC niche is considered as a potential cancer therapeutic and alternative strategy could be to directly target the super enhancers and epigenetic machineries in CSCs that control the expression of key oncogenes or stemness regulators, thereby blocking CSC self-renewal capacity and developmental plasticity. This would reduce the metastatic spreading of CSCs and cancer relapse if this strategy would be combined with conventional therapeutic strategies that would aim to target the main hallmarks of cancers.

Several factors known to play a central role in embryonic stem cells and used in epigenetic reprogramming, such as pluripotency factors OCT4, NANOG, SOX2 and Klf4, have been reported to be expressed in subpopulations of cancer cells from different tumour type^{74,94-101}. This indicates an aberrant and partial reactivation of stem cell circuitries in cancer stem cells that usually would regulate the self-renewal and pluripotency of embryonic stem cells during early embryogenesis. Such stem cell factor expression has tumorigenic effects due to effects resembling their function in stem cells such as stem cell maintenance and self-renewal which block differentiation. Pluripotency gene loci form contacts with each other in *cis* and in *trans* in the nucleoplasm and utilize shared cell type-specific transcriptional machineries¹⁰². Pluripotency factors are involved in regulating the genomic architecture of pluripotent stem cells. Super enhancers bound by pluripotency factors OCT4, NANOG, SOX2 and KLF4 colocalize in the pluripotent cell nucleus¹⁰³ and establish the prevalent regulatory hubs¹⁰⁴. In line with this, recent study uncovered that KLF4 is a key driver for the organization of pluripotency-associated three-dimensional enhancer networks which are disassembled during differentiation^{105,106}.

The expression of pluripotency factors such as OCT4, NANOG and SOX2 in CSCs could similarly promote the emergence of stem-like cell states by reprogramming of 3D enhancer–promoter interaction that otherwise would prevent developmental plasticity and maintain differentiated cell states. Targeting the stem cell factors not only would reduce cancer risk, but also antagonize the growth of the primary tumour and metastatic

derivatives. Tryptophan derivatives and the AhR signalling pathway regulate the transcription of Oct4 and cancer cell stemness, opening a new therapeutic avenue to target stem-like cancer cells¹⁰⁷. Hence, these mechanisms are intriguing targets for small molecule compounds that would aim to reverse the developmental plasticity of CSCs and trap it to a more restricted cell fate that would re-sensitise these cancer cells to combined chemotherapeutic treatments. This would target the epigenetic instability and the emergence of CSCs during tumour evolution and formation of therapeutic resistance.

1.8 Research questions

The work presented in this thesis focusses on dissecting the 3D chromatin architectures in human diseases (Dilated cardiomyopathy (DCM) and pancreatic ductal adenocarcinoma), which are leading causes of death in the world. The specific aims of this thesis are:

1. To generate the 3D chromatin atlas in human DCM hearts and determine the main factors orchestrating the 3D chromatin architectures and pathological gene transcription in DCM.
2. To interrogate the specific 3D chromatin architectures that link to transcriptional signature in cancer stem cells of pancreatic ductal adenocarcinoma;
3. To identify the specific small molecule compounds that target the 3D chromatin architectures in cancer stem cells of pancreatic ductal adenocarcinoma;

Chapter 2

Rewiring of 3D Chromatin Topology Orchestrates Transcriptional Reprogramming and the Development of Human Dilated Cardiomyopathy

This chapter is adapted from Yuliang Feng *et al.* ***Circulation***. 2022 May 31;145(22):1663-1683.
(In vitro and in vivo functional study is in collaboration with Wei Huang at University of Cincinnati);
Computational analysis is in collaboration with Liuyang Cai at The Chinese University of Hong Kong)

2.1 Introduction

Heart failure is the leading cause of death worldwide and affects over 37.7 million individuals globally¹⁰⁸. In particular, dilated cardiomyopathy (DCM) is the most commonly diagnosed type of systolic heart failure¹⁰⁸. Currently, no effective treatment can prevent the progression of DCM to heart failure. Heart transplantation is the last resort for DCM patients, but it is hampered by limited donors and staggering medical and economic burdens¹⁰⁹. Thus, studies are urgently needed to demystify its etiology and develop effective management strategies.

DCM is characterized by pathogenic structural remodeling of the left ventricle (i.e., enlargement of the chamber and a thin ventricular wall) and poor contractility. Previous studies indicated that genetic variation in protein-coding genes (e.g., sarcomeric and cytoskeletal genes) contributed to the pathogenesis of DCM in a minority of cases¹¹⁰, suggesting that nongenetic (epigenetic) mechanisms may play a crucial role in DCM development¹¹¹. As cell identity is primarily determined by coordinated gene transcription, novel insights into the mechanisms by which the DCM transcriptome is controlled will enable us to develop novel strategies to diagnose and treat DCM.

Gene transcription is regulated by a variety of determinants, and *cis*-regulatory elements (CREs), such as enhancers and promoters, are being recognized as the key determinants that shape gene expression¹¹². Dysregulation of CREs has been implicated in the pathogenesis of human diseases, such as thalassemia, polydactyly and various types of

cancers¹¹³. During heart failure progression, *MYH6*/*MYH7* ratio is known to shift, as characterized by a decrease in *MYH6* expression and an increase in *MYH7* expression. Recently, a key enhancer has been identified to regulate the switch of *MYH6* expression to *MYH7* expression. Deletion of the predicted enhancer reduces *MYH7* and increases *MYH6* expression, leading to faster contraction in human engineered heart¹¹⁴. To identify the potential CREs in human failing hearts, Cap Analysis of Gene Expression (CAGE)-seq was performed in 3 healthy and 4 failed human left ventricles to map the initiation sites of both capped coding and noncoding RNAs¹¹⁵. This study identified the transcribing promoters and the first intronic enhancer in failed ventricles but most distal enhancers (e.g., intergenic enhancers) in heart failure were not explored yet. By performing ChIP-seq for H3K27ac (active enhancer and promoter mark¹¹⁶) in non-failing and end-stage failing human heart tissues, a recent study identified specific enhancers and promoters implicated for established molecular pathways in heart failure¹¹⁷. Nevertheless, the specific *cis*-regulome in DCM implicated for its pathogenesis remains unexplored.

Moreover, in many cases, enhancers regulate transcription via long-range interactions with target gene promoters, but not nearby gene promoters¹¹⁸. Thus, it is incorrect to assign the enhancer-promoter pair based on only the distance between the enhancer and promoter on the linear genome. Standard methods that map the enhancer-promoter interactome at high resolution (kilobase) require a large number of cells/tissue samples and an extremely deep sequencing depth (billion reads per sample). New approaches to protein-mediated chromatin interactions (e.g., HiChIP^{31,60,119}, PLAC-seq^{61,120}, and *in situ* ChIA-PET⁶²) have recently been developed to enable the generation of high-resolution

chromatin contact maps for CREs with substantially reduced cell numbers and sequencing depth and can thus be applied to clinical samples, such as human DCM tissues, with high resolution.

Furthermore, the enhancer-promoter interactome is embedded in topologically associated domains (TADs), which further aggregate into large-scale nuclear architectures called A/B compartments². In recent years, reorganization of higher-order chromatin architectures have been implicated in aberrant enhancer-promoter interactions in human diseases (e.g., abnormalities in limb development¹²¹ and gastrointestinal stromal tumors¹²²) and transcriptional derepression during cardiac lineage specification¹²³. However, the enhancer-promoter interactome and its link to higher higher-order chromatin architectures in DCM remains unexplored.

In this study, we aimed to develop the 3D epigenomic maps of DCM hearts and determine the main factors driving pathological gene transcription in DCM. We leveraged these datasets to a) map the differential active CREs in DCM and NF hearts; b) identify the differential active CREs interactome and determine how the components are linked to transcriptional changes in DCM; c) investigate whether the formation of DCM-enriched CREs interactome is related to higher-order chromatin structure (TADs and compartments) or chromatin accessibility; and d) uncover the key transcription factor that facilitates DCM-enriched CREs interactions and how it drives the pathogenesis of DCM.

2.2 Results

2.2.1 Genome-Wide Identification of the Active CREs in DCM and NF Hearts

We used 50 human DCM and 51 non-failing (NF) biobanked heart tissues (**Table 1**) to uncover novel mechanisms that regulate DCM pathogenesis by integrative analysis of multi-layered 3D epigenomic features: transcriptome, epigenome and 3D chromatin connectome.

Table 1. Summary of heart sample donors		
	Diagnosis	
	Non-failing (NF)	Failing (DCM)
No. of donors	51	50
Male	29	38
Female	22	12
Age (yrs)	43.5	57
Ejection fraction (%)	55	15

To accomplish this, we performed RNA-seq (n=101), ChIP-seq (H3K27ac) (n=20), HiChIP (H3K27ac) (n=20), *in situ* Hi-C (n=4) and ATAC-seq (n=20) mapping (**Figure 1A**). Principal component analysis (PCA) from RNA-seq data depicted a separation between DCM and NF hearts, suggesting that the transcriptome of DCM is globally reconfigured compared to NF (**Figure 2A**).

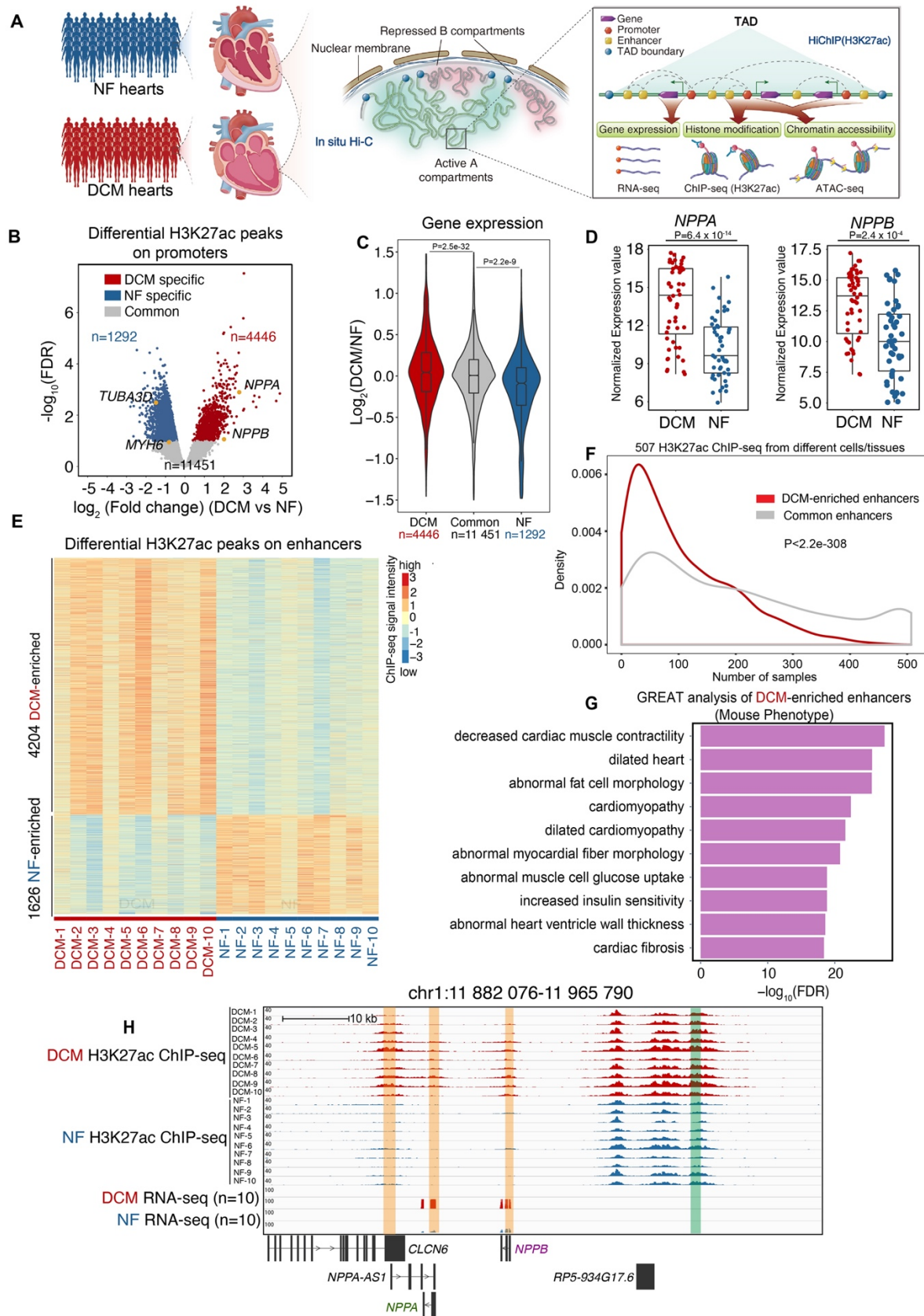


Figure 1. Genome-Wide Identification of Active CREs in DCM and NF Hearts. **A.** Schematic representation of 3D epigenome and transcriptome mapping for 50 DCM and 51 NF left ventricle heart tissue samples (left panel). In situ Hi-C was performed to examine the active A compartment and B compartment and topologically associating domains (TADs) (middle panel). HiChIP (H3K27ac) was performed to map the interactome of the active CREs, which was annotated by H3K27ac ChIP-seq. ATAC-seq was used to identify chromatin accessibility on the CREs. RNA-seq was used to measure transcription output (right panel). **B.** Volcano plot showing NF-enriched ($n=1,292$), DCM-enriched ($n=4,446$) and common ($n=11,451$) H3K27ac ChIP-seq peaks in promoter regions (within ± 2.5 kb of known transcription start sites (TSSs)) of DCM and NF hearts. Cardiac hypertrophy marker genes (NPPA, NPPB), normal cardiac marker gene (MYH6) and the cytoskeleton marker gene (TUBA3D) are labeled in yellow. **C.** Violin plot showing expression bias for genes exhibiting specific H3K27ac peaks in promoter regions ($n=4,446$ DCM-enriched H3K27ac peaks, $n=1,292$ NF heart-enriched H3K27ac peaks and $n=11,451$ common H3K27ac peaks). Data were analyzed by Wilcoxon rank sum test with continuity correction. **D.** Expression values ($\log_2(\text{normalized counts})$) for NPPA and NPPB in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **E.** Heatmap of DCM-enriched ($n=4,204$) and NF-enriched ($n=1,626$) H3K27ac ChIP-Seq peaks on enhancer regions in 10 DCM and 10 NF heart samples (rows represent peaks, and columns represent samples). **F.** Enhancer breadth (number of cells/tissues in which an enhancer is activated) of DCM-enriched and common enhancers. DCM-enriched enhancers (red) are significantly more represented as tissue-enriched than common enhancers (grey). Data were analyzed by Kolmogorov-Smirnov test. **G.** Bar chart showing the top 10 Ontology terms (Mouse Phenotype, from GREAT analysis) enriched for genes associated with DCM-enriched enhancers (ontology terms were ranked by negative $\log_{10}(\text{FDR})$ from smallest to largest). **H.** Genome browser view of the genes NPPA and NPPB and their associated promoters and enhancers in DCM and NF hearts. Enhancer is highlighted in the green box, while promoter is highlighted in the orange box. Figure adapted with permission from Feng et al., *Circulation*, 2022. The analysis of Figure 1F was designed by me and performed by Liuyang Cai (The Chinese University of Hong Kong).

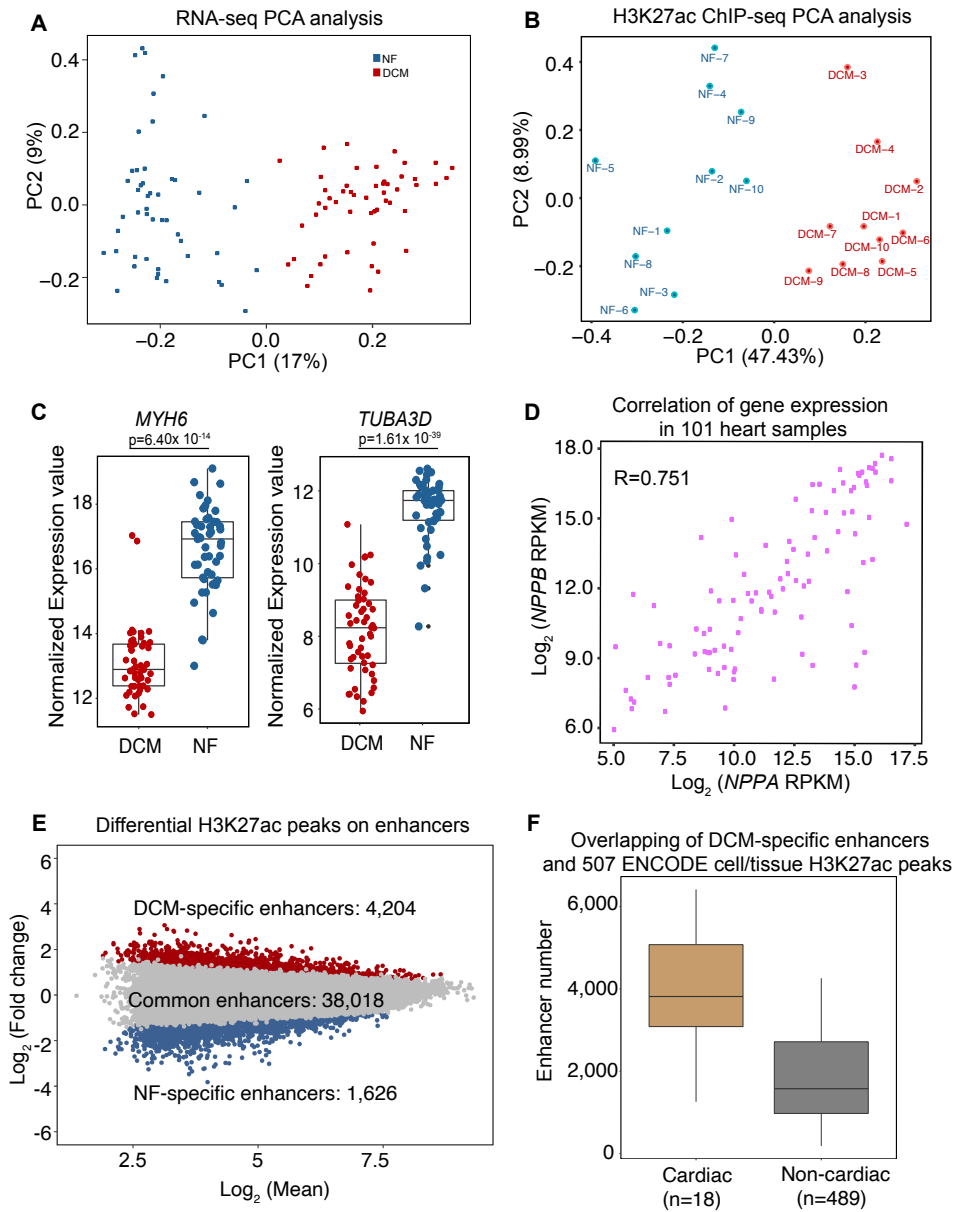


Figure 2. Characterization of RNA-seq and H3K27ac ChIP-seq data in DCM and NF Hearts. *A.* PCA plot of ensembled 50 DCM and 51 NF RNA-seq RNA-seq profiles using top 2000 variable genes. Red dots represent DCM heart samples. Blue dots represent NF heart samples. *B.* PCA plot of ensembled 10 DCM and 10 NF H3K27ac ChIP-seq profiles. Red dots represent DCM heart samples. Blue dots represent NF heart samples. *C.* Expression values ($\log_2(\text{normalized counts})$) for *MYH6* and *TUBA3D* in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. *D.* Correlation between the gene expression

of *NPPA* and *NPPB* in 101 DCM and NF heart tissues. **E.** MA plot showing DCM-enriched, NF-enriched, and common H3K27ac ChIP-seq peak signals in enhancer regions (outside of ± 2.5 kb to known TSSs) of DCM and NF hearts. **F.** Overlapping between DCM-specific enhancers identified from this study and ENCODE H3K27ac ChIP-seq peaks. Cardiac (left) and non-cardiac (right) cell/tissue-derived ENCODE data were analyzed separately. Figure adapted with permission from Feng et al., *Circulation*, 2022. The analysis of Figure 2F was designed by me and performed by Liuyang Cai (The Chinese University of Hong Kong).

As CREs have been recognized as essential sequences for cell-specific transcription, we performed H3K27ac ChIP-seq^{116,124} on 10 DCM and 10 NF human heart samples (left ventricle) to annotate active CREs. Similarly, PCA analysis showed a separation between the DCM and NF samples (**Figure 2B**), indicating that the active CREs were differentially used in DCM and NF hearts.

Since H3K27ac is associated with both promoters and distal enhancers¹¹⁶, we then separated H3K27ac peaks into promoter proximal peaks and enhancer peaks based on their distances (± 2.5 kb) to known transcription start sites (TSSs). Among the promoter regions, we identified 4,446 DCM-enriched H3K27ac peaks (e.g., promoters of the cardiac hypertrophy marker genes *NPPA* and *NPPB*¹²⁵), 1,292 NF-enriched H3K27ac peaks (e.g., promoters of the normal cardiac marker gene *MYH6*¹²⁶ and the cytoskeleton marker gene *TUBA3D*¹²⁷), and 11,451 H3K27ac peaks common to both DCM and NF hearts (**Figure 1B**).

By integrating the RNA-seq data for those samples, we found that the expression of genes with DCM-enriched H3K27ac peaks on their promoters was significantly upregulated in DCM hearts compared to NF hearts, while the expression of genes with NF-enriched H3K27ac peaks was significantly downregulated, and the expression of genes with common H3K27ac peaks was unchanged (**Figure 1C**). We further corroborated these results by analyzing the expression of *NPPA*, *NPPB*, *MYH6* and *TUBA3D* in human DCM and NF heart tissues. *NPPA* and *NPPB* were significantly upregulated (**Figure 1D**), whereas *MYH6* and *TUBA3D* were significantly downregulated in DCM compared to NF hearts (**Figure 2C**). These trends are consistent with the roles of H3K27ac in active transcription. Interestingly, the expression levels of *NPPA* and *NPPB* were highly correlated (Pearson correlation coefficient=0.751) (**Figure 2D**), suggesting that they are co-regulated at the transcriptional level and may share common regulatory elements.

Next, we identified 4,204 DCM-enriched enhancers, 1,626 NF-enriched enhancers (**Figure 1E**) and 38,018 common enhancers present in both DCM and NF hearts (**Figure 2E**). By integrating 507 H3K27ac ChIP-seq data derived from different cells/tissues from the ENCODE project, we further validated that the 4,204 DCM-enriched enhancers displayed high cell specificity compared to 38,018 common enhancers (**Figure 1F**). In addition, we split the H3K27ac ChIP-seq data into cardiac (n=18) and non-cardiac (n=489) cell types and found that H3K27ac peak regions in cardiac-cell types overlapped significantly more DCM-enriched enhancers compared to non-cardiac cell types (**Figure 2F**). These results suggested that DCM-enriched enhancers could play a role in orchestrating cell-type-specific regulatory activity in DCM hearts. In line with this, GREAT

analysis showed that DCM-enriched enhancers are located near genes essential for decreased muscle contractility, dilated cardiomyopathy, abnormal myocardial fiber morphology etc. (**Figure 1G**), suggesting the regulatory potential of these genes in DCM pathogenesis. DCM-enriched CREs can be exemplified by a genomic region on chromosome 1 (**Figure 1H**). This region contained a DCM-enriched enhancer near the coding gene *NPPB*. In addition, the *NPPA-AS1*, *NPPA*, and *NPPB* promoters had specific H3K27ac peaks. Previous study demonstrated a *cis*-regulatory role for *NPPA-AS1* transcript in the repression of *NPPA* expression in atria¹²⁸. However, *NPPA-AS1* was transcriptionally inactive, as shown from RNA-seq data, suggesting that this promoter may function in *trans* to regulate the transcription of other genes in DCM ventricle.

Taken together, the above results highlight that gene expression in DCM hearts is regulated by CREs and suggest that their interactions will bring to light the functions and target genes of these regulatory elements.

2.2.2 Characterization of the 3D Chromatin Organization in DCM and NF Hearts

The 3D connectome brings together regulatory chromatin regions and hence can potentially contribute to the regulation of gene expression. To map the high-resolution chromatin contact maps of active CREs, we performed H3K27ac HiChIP on 10 DCM and 10 NF samples. We obtained ~45 million high-quality paired-end reads (i.e., Unique valid interaction pairs) on average from the HiChIP libraries. For the HiChIP data, we found that the inter-group variation was greater than intra-group variation (**Figure 3**).

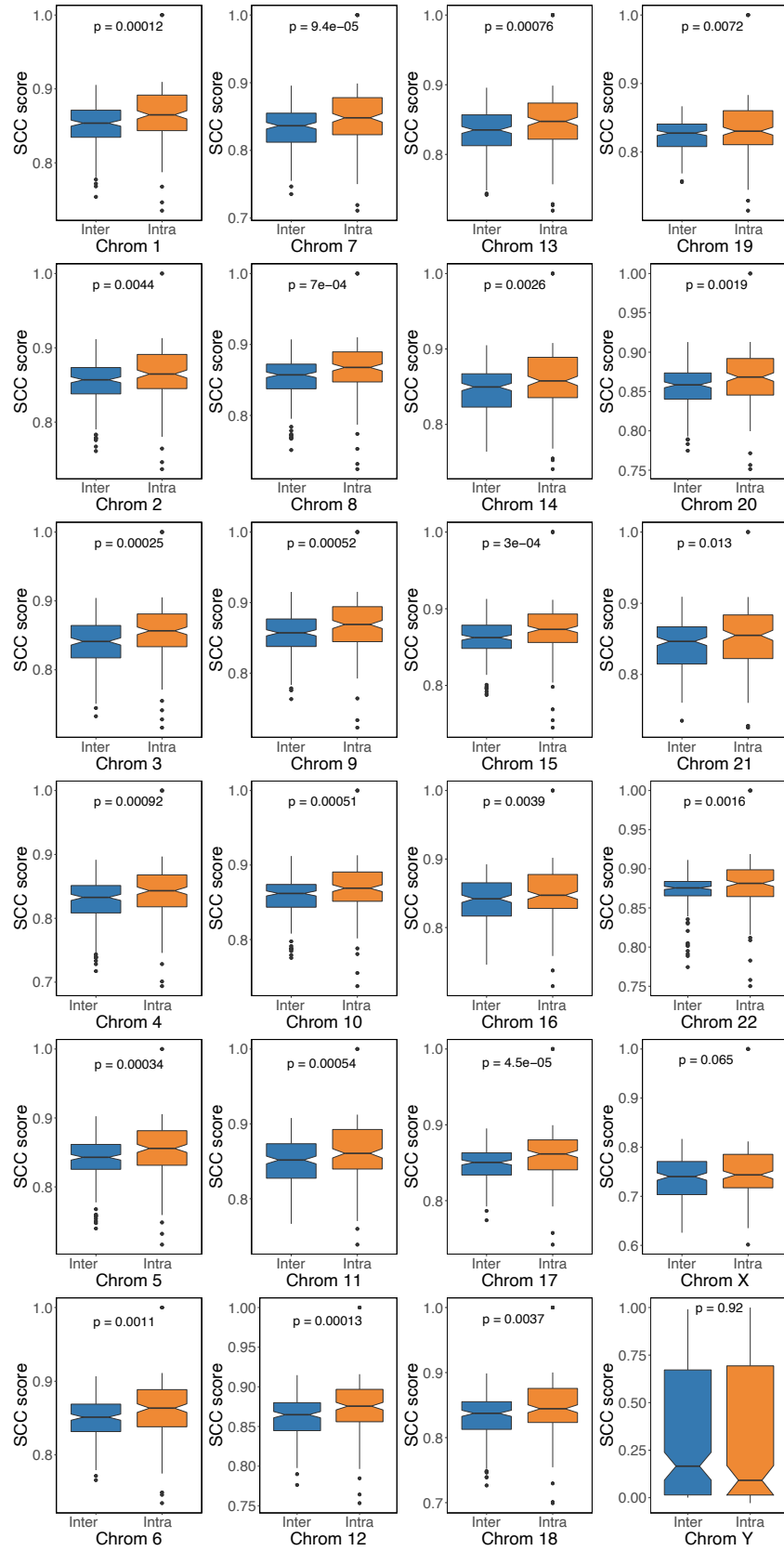


Figure 3. Assessment of the Intra and Inter Group Similarity between Contact Matrices in DCM and NF HiChIP Data. Stratum adjusted correlation coefficient (SCC) was computed separately for each chromosome. Figure adapted with permission from Feng et al., *Circulation*, 2022.

HiChIP data contain both 1D (binding sites of the targeted proteins) and 3D (chromatin interactions between binding sites) information^{31,60}. We found that the 1D signal of the HiChIP data was comparable to that of the H3K27ac ChIP-seq datasets (**Figure 4A**).

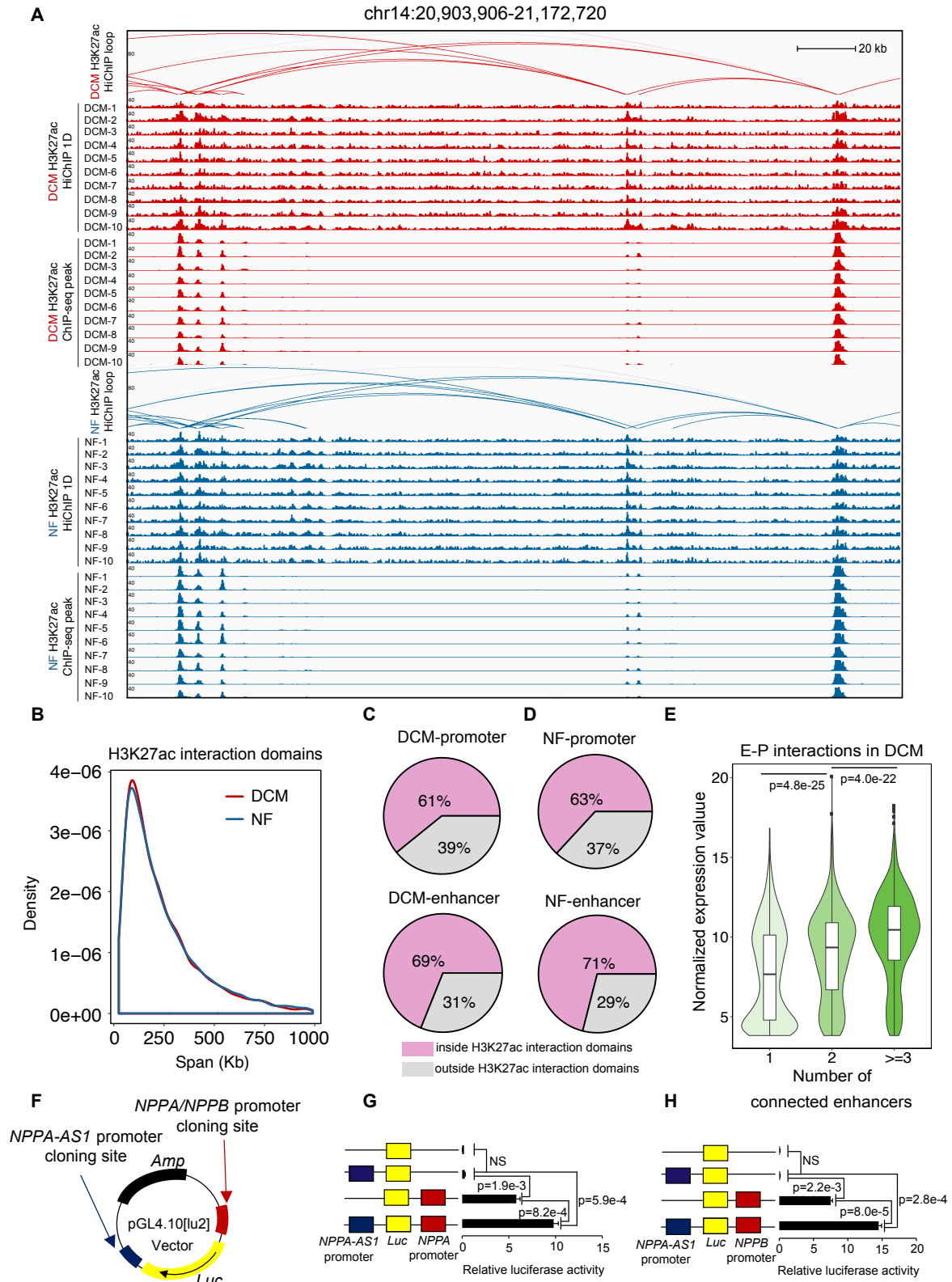


Figure 4. Characterization of H3K27ac HiChIP Data in DCM and NF Hearts. *A. Genome browser view of comparison of H3K27ac ChIP-seq and H3K27ac HiChIP 1D signals among 10 DCM and 10 NF heart tissue samples (chr14:20,903,906-21,172,720). B. Genomic span distribution of H3K27ac interaction domains in DCM and NF hearts by density. C. Distribution of all active promoters (upper panel) and active enhancers (lower panel) inside and outside of the H3K27ac interaction domains in DCM hearts. D. Distribution of all active promoters (upper panel) and active enhancers (lower panel) inside and outside of the H3K27ac interaction domains in NF hearts. E. Normalized expression values for genes with different numbers of enhancers connected to promoters in DCM hearts. Data were analyzed by Wilcoxon rank sum test with continuity correction. F-H. Luciferase reporter assays testing the enhancer function of the NPPA-AS1 promoter. Data were analyzed by One-way ANOVA. Figure adapted with permission from Feng et al., Circulation, 2022.*

Inspection of the combined interaction matrix at different resolutions showed typical compartments (500 kb resolution), TADs (25 kb resolution) and focal loops (5 kb resolution) for both DCM and NF hearts, similar to those reported in previous HiChIP data^{31,60} (**Figure 5A**). We then identified high-confidence chromatin interactions among gene promoters and distal enhancers in DCM and NF hearts (**Figure 5B**). We found that 13,562 promoters and 11,308 enhancers were involved in chromatin interactions in DCM and that anchor genes had significantly higher expression than that of non-anchor genes (**Figure 5C**), suggesting that chromatin interactions may play a role in transcriptional activation.

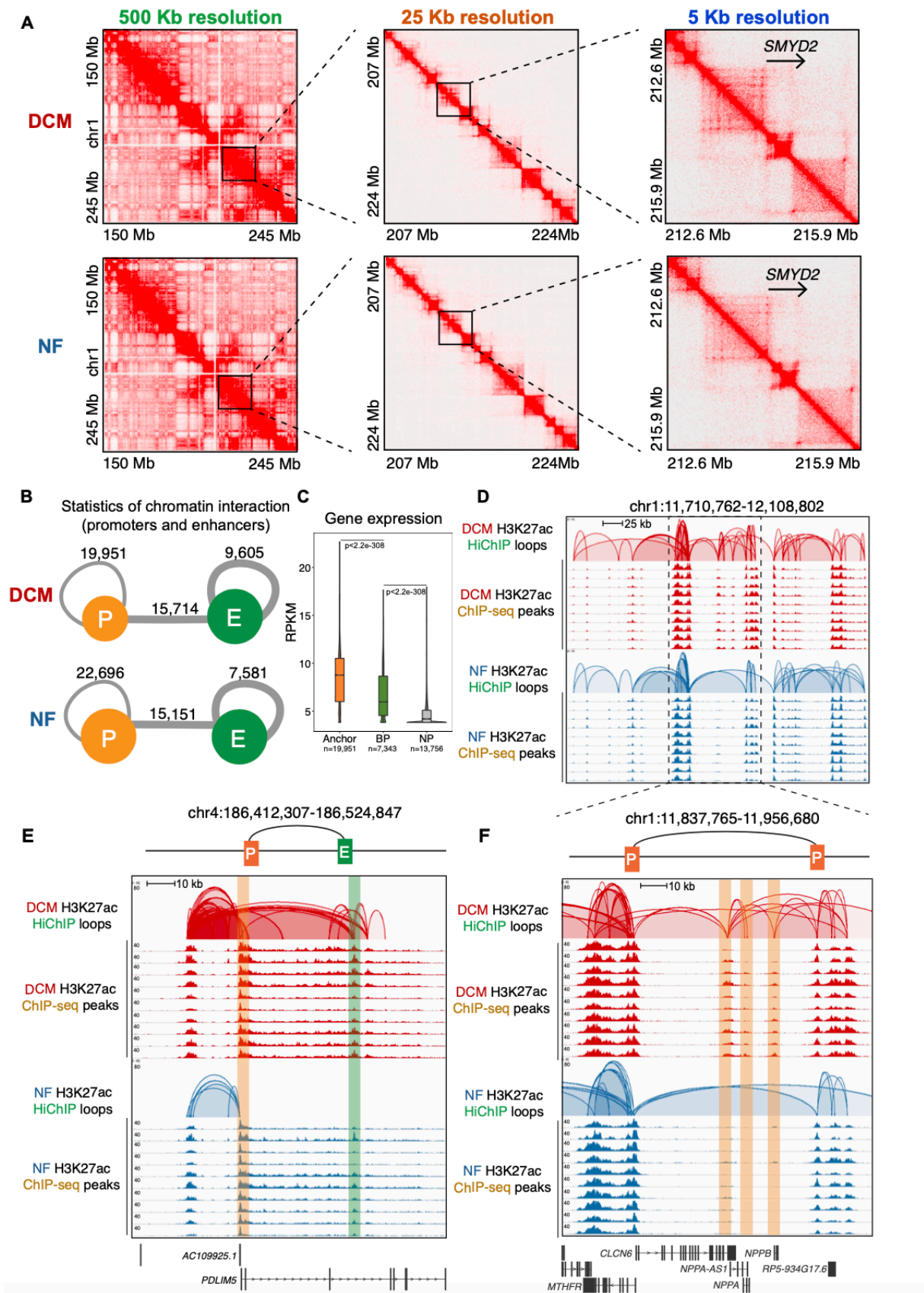


Figure 5. Characterization of the 3D Chromatin Organization of the Active CREs by H3K27ac HiChIP in DCM and NF Hearts. **A.** Knight–Ruiz (KR) matrix–balanced interaction maps of merged DCM (top) and NF (bottom) HiChIP valid interaction pairs of typical compartments at 500 kb resolution (left), TADs at 25 kb resolution (middle) and focal loops at 5 kb resolution (right). **B.** Summary for the number of chromatin interactions between promoters (P) and enhancers (E) in DCM (upper) and NF (lower) hearts. **C.** Violin plot of expression levels for anchor genes (n=19,951), basal promoter (BP) genes (n=7,343) and non-promoter (NP) genes (n=13,756) in DCM. Data were analyzed by Wilcoxon rank sum test with continuity correction. **D.** Genome browser view of CREs interactions in chr1:11,710,762-12,108,802. The top two rows represent the CREs interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. **E.** Genome browser view of enhancer-promoter interactions for PDLIM5 (chr4:186,412,307-186,524,847). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. The enhancer is highlighted in the green box, while the promoter is highlighted in the orange box. **F.** Genome browser view of promoter-promoter interaction for NPPA-AS1, NPPA and NPPB. The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. Enhancer is highlighted in the green box, while promoter is highlighted in the orange box. Figure adapted with permission from Feng et al., *Circulation*, 2022.

We then constructed chromatin interaction networks for DCM and NF hearts separately, as genome-wide chromatin interactions are organized into giant, modular interaction networks¹²⁹. Promoters and enhancers were organized into chromatin interaction domains (**Figure 5D**) across the genomes of both DCM and NF hearts; in addition, those interaction domains spanned similar genomic distances (**Figure 4B**) and contained most

of the active promoters and enhancers (**Figure 4C, D**). The numbers of connected enhancers were positively correlated with increase in expression levels of their connected genes (**Figure 4E**). Chromatin interactions within the domains can first be classified as promoter-enhancer interactions, i.e., gene promoters regulated by one or multiple enhancers. For example, the promoter of *PDLIM5*, a gene involved in actin cytoskeleton organization in heart development¹³⁰ and DCM¹³¹, was regulated by an enhancer embedded in its intron (**Figure 5E**). Second, these chromatin interactions may be classified as promoter-promoter interactions. For example, both the *NPPA* and *NPPB* promoters have strong interactions with the *NPPA-AS1* promoter as shown by paired-end tags (PETs) (**Figure 5F**), a finding that supports the conclusion that *NPPA* and *NPPB* may be co-regulated by *NPPA-AS1* promoter. *NPPA-AS1* transcript is shown to have an inhibitory effect on *NPPA* expression in atria¹²⁸, but our RNA-seq revealed that the *NPPA-AS1* gene is transcriptionally repressed in DCM ventricle. Interestingly, recent study showed that non-transcribing promoter can have enhancer activity to regulate other gene transcription via chromatin looping¹³², in line with previous identification of enhancer-like promoters^{13,17,133}. To test whether *NPPA-AS1* promoter can function as an enhancer to regulate *NPPA* and *NPPB* expression, we performed luciferase reporter assays, a commonly used method for promoter and enhancer characterization¹³⁴ (**Figure 4F**). *NPPA* or *NPPB* promoter fragment was cloned upstream of the luciferase gene as the driving promoter, and the *NPPA-AS1* promoter was cloned downstream of the luciferase gene as a presumed enhancer, as enhancers generally function in a manner independent of orientation. The analysis of transfection and luciferase activity was performed in an *in vitro* cardiac hypertrophy model induced by endothelin-1 (10 nM, 48 hours) in human

induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs). We found that the *NPPA-AS1* promoter alone was not able to activate luciferase gene expression. However, the *NPPA-AS1* promoter could act in concert with *NPPA/NPPB* promoter to significantly enhance the luciferase activity compared to *NPPA/NPPB* promoter alone (**Figure 4G, H**). This result indicated that the *NPPA-AS1* promoter could function as an enhancer to regulate *NPPA/NPPB* transcription. Overall, we identified the interactome of active CREs in DCM and NF in high resolution and uncovered that active promoters and enhancers were organized by multiple types of chromatin loops.

2.2.3 Specific Enhancer-Promoter Connectome Regulates DCM-Specific Transcription

To dissect how CRE interactions contribute to DCM-specific transcription, we performed differential analysis of HiChIP loops. Two classes of differential loops were identified based on H3K27ac signal on loop anchors: Class I (change of H3K27ac signal on loop anchors) and Class II (no change of H3K27ac signal on loop anchors). For Class I, we identified 670 DCM-enriched H3K27ac loops (DCM-enriched anchors) and 1,035 NF-enriched H3K27ac loops (NF-enriched anchors) (**Figure 6A**). We found that these DCM-enriched H3K27ac loop-connected genes were significantly upregulated in DCM hearts (**Figure 6B**), while the NF-enriched H3K27ac loop-connected genes were significantly downregulated in DCM hearts (**Figure 6C**). For Class II, we identified 4,883 DCM-enriched H3K27ac loops (Common anchors) and 3,712 NF-enriched H3K27ac loops (Common anchors) (**Figure 7A**).

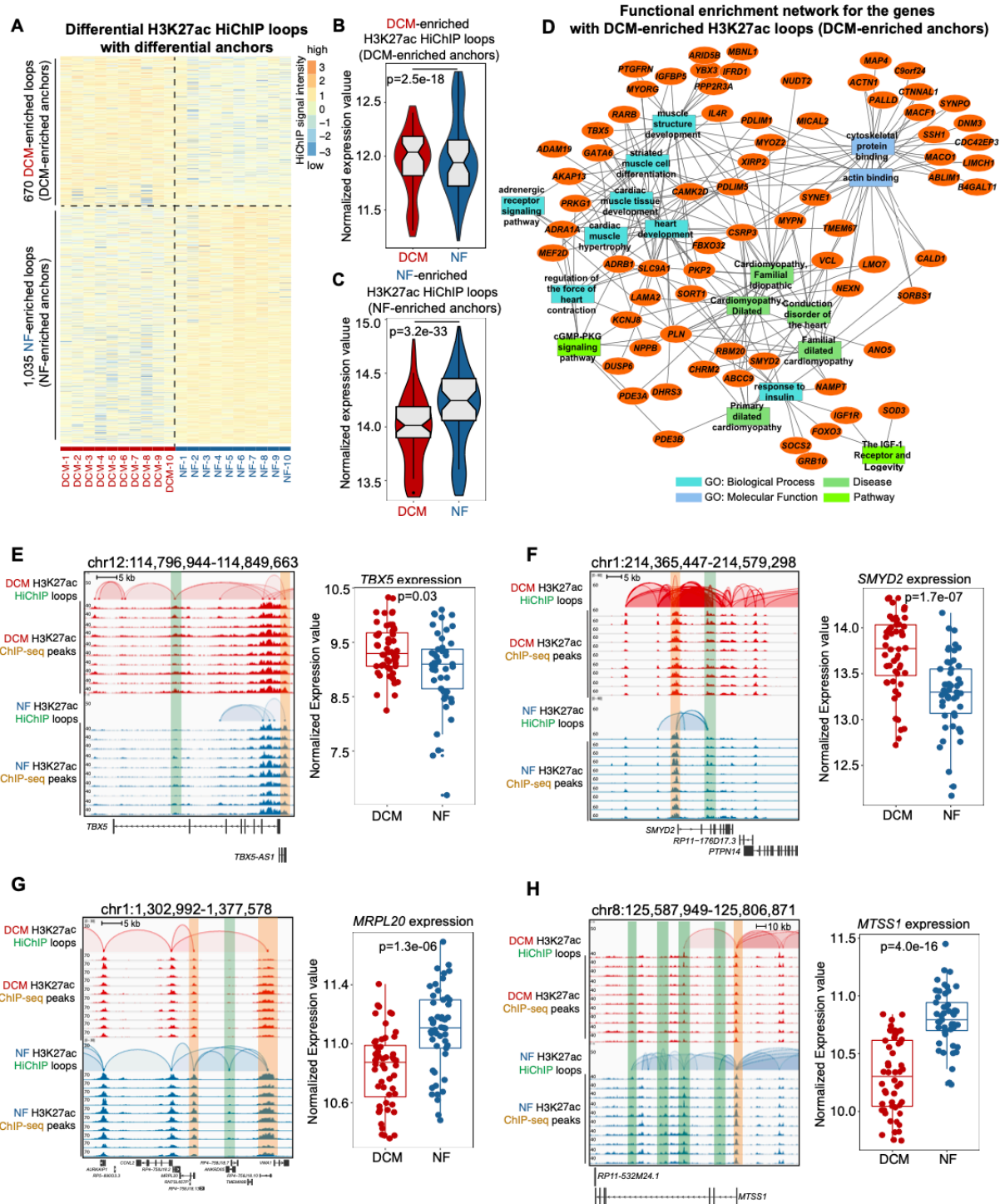


Figure 6. DCM-Enriched Enhancer/Promoter Connectome Contributes to DCM-Specific Transcription.

A. Heatmap showing the intensity of differential H3K27ac loops with differential loop anchors (H3K27ac signal) among 10 DCM and 10 NF heart samples. **B.** Normalized expression values of the genes with DCM-

enriched H3K27ac loops (DCM-enriched anchors). Data were analyzed by Wilcoxon rank sum test with continuity correction. **C.** Normalized expression values of the genes with NF-enriched H3K27ac loops (NF-enriched anchors). Data were analyzed by Wilcoxon rank sum test with continuity correction. **D.** Network representation of functional enrichment of the genes with DCM-enriched H3K27ac loops (DCM-enriched anchors). Functional enrichment was done using the ToppFun application. Orange nodes represent DCM-enriched genes, while the different colored rectangles represent enriched terms. Only select enriched terms are shown here. Network was generated using the Cytoscape application. **E.** Left panel: browser screenshot showing *TBX5* with DCM-enriched E-P interactions (chr12:114,796,944-114,849,663). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. Enhancer is highlighted in the green box, while promoter is highlighted in the orange box; Right panel: expression values ($\log_2(\text{normalized counts})$) for *TBX5* in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **F.** Left panel: browser screenshot showing *SMYD2* with DCM-enriched E-P interactions (chr1:214,365,447-214,579,298). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. Enhancer is highlighted in the green box, while promoter is highlighted in the orange box; Right panel: expression values ($\log_2(\text{normalized counts})$) for *SMYD2* in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **G.** Left panel: browser screenshot showing *MRPL20* with NF-enriched E-P interactions (chr1:1,302,992-1,377,578). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. Enhancer is highlighted in the green box, while promoter is highlighted in the orange box; Right panel: expression values ($\log_2(\text{normalized counts})$) for *MRPL20* in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **H.** Left panel: browser screenshot showing *MTSS1* with NF-enriched E-P interactions (chr8:125,587,949-125,806,871). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. Enhancer is highlighted in the green box, while promoter is highlighted in the orange

box; Right panel: expression values ($\log_2(\text{normalized counts})$) for MTSS1 in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. Figure adapted with permission from Feng et al., Circulation, 2022.

However, these differential loops have less interaction intensity (PET counts) compared to their counterparts in Class I (**Figure 7B**) and their connected genes are not significantly linked to transcription change (**Figure 7C, D**). Hence, in downstream analysis, we focus on Class I loops to explore the transcriptional regulation in DCM.

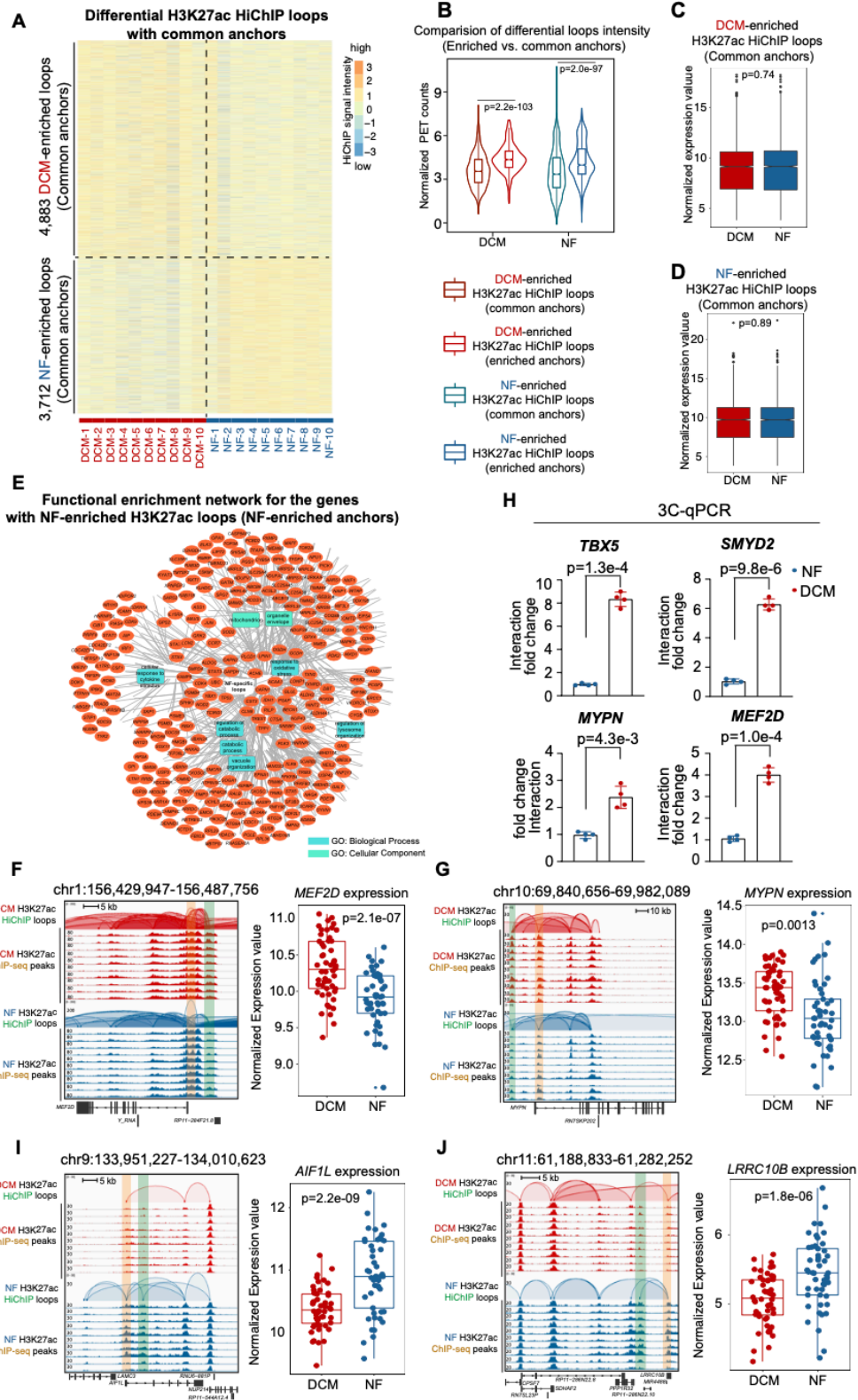


Figure 7. Characterization of DCM/NF-specific enhancer/promoter connectome. **A.** Heatmap showing intensity of differential H3K27ac loops with common loop anchors (H3K27ac signal) in 10 DCM and 10 NF heart samples. **B.** Normalized PET counts for differential loops (enriched anchors vs. common anchors). **C.** Normalized expression values of the genes with DCM-enriched H3K27ac loops (common anchors). Data were analyzed by Wilcoxon rank sum test with continuity correction. **D.** Normalized expression values of the genes with NF-enriched H3K27ac loops (common anchors). Data were analyzed by Wilcoxon rank sum test with continuity correction. **E.** Network representation of functional enrichment of the genes with NF-enriched H3K27ac loops (NF-enriched anchors). Functional enrichment was performed using the ToppFun application. Orange nodes represent NF-enriched genes, while the different colored rectangles represent enriched terms. Only select enriched terms are shown here. Network was generated using the Cytoscape application. **F.** Left panel: browser screenshot showing MEF2D with DCM-enriched E-P interactions (chr1:156,429,947-156,487,756). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. The enhancer is highlighted in the green box, while the promoter is highlighted in the orange box. Right panel: expression values ($\log_2(\text{normalized counts})$) for MEF2D in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **G.** Left panel: browser screenshot showing MYPN with DCM-enriched E-P interactions (chr10:69,840,656-69,982,089). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. The enhancer is highlighted in the green box, while the promoter is highlighted in the orange box. Right panel: expression values ($\log_2(\text{normalized counts})$) for MYPN in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **H.** Validation of E-P interactions with 3C-qPCR analysis in DCM and NF hearts (TBX5, SMYD2, MEF2D, MYPN, related to Figure 3E-H). Student t test was used for differential analysis. **I.** Left panel: browser screenshot showing AIF1L with NF-enriched E-P interactions (chr9:133,951,227-134,010,623). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. The enhancer is highlighted in the green box, while the promoter

is highlighted in the orange box. Right panel: expression values ($\log_2(\text{normalized counts})$) for *AIF1L* in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. **J.** Left panel: browser screenshot showing *LRRC10B* with NF-enriched E-P interactions (chr11:61,188,833-61,282,252). The top two rows represent the interactions in DCM hearts, and the bottom two rows represent those in NF hearts. The HiChIP loops and ChIP-seq peaks are shown for each. The enhancer is highlighted in the green box, while the promoter is highlighted in the orange box. Right panel: expression values ($\log_2(\text{normalized counts})$) for *LRRC10B* in 50 DCM and 51 NF heart RNA-seq datasets. Data were analyzed by Wald test. Figure adapted with permission from Feng et al., *Circulation*, 2022.

To understand the functional implications of the genes regulated by DCM/NF-enriched H3K27ac loops, we represent their functional enrichment network¹³⁵ (**Figure 6D, Figure 7E**). For DCM-enriched H3K27ac loops, the most highly represented GO terms included (1) molecular functions related to cytoskeletal protein binding and actin binding and (2) biological processes related to heart development, cardiac muscle development, adrenergic receptor signaling pathway, cardiac muscle hypertrophy, regulation of the force of heart contraction, response to insulin, muscle structure development, and striated muscle and differentiation. Pathway terms included cGMP-PKG signaling pathway, the IGF-1 receptor and longevity. Disease terms included, for example, cardiomyopathy (dilated), familial dilated cardiomyopathy, primary dilated cardiomyopathy, cardiomyopathy (familial idiopathic), and conduction disorder of the heart. These results suggested that specific enhancer/promoter connectome regulated specific transcription programs implicated in DCM development.

We uncovered specific examples of DCM-enriched E-P interactions. *TBX5* encodes a transcription factor crucial for proper cardiac development¹³⁶, and we found that its promoter interacted with an activated enhancer embedded in its 7th intron in DCM hearts (**Figure 6E**). *SMYD2* (histone lysine methyltransferase gene) is robustly expressed in neonatal heart and regulates cardiac development and regeneration^{137,138}. Similar to *TBX5*, we also found that its promoter interacted with an activated enhancer embedded in its 3rd intron in DCM hearts (**Figure 6F**). Other DCM-enriched E-P examples (*MEF2D*, regulator for myogenesis¹³⁹ and cardiac stress¹⁴⁰; *MYPN*, regulator for maintaining sarcomeric integrity¹⁴¹) are displayed in **Figure 7F, G** respectively. These specific examples of DCM-enriched E-P interactions were validated by 3C-qPCR (**Figure 7H**). In addition to DCM-enriched E-P interactions, we also examined NF-enriched E-P interactions. For example, *MRPL20* encodes the large 39S subunit of mitochondrial ribosomes for protein synthesis within the mitochondria. We identified NF-enriched enhancer-promoter and promoter-promoter interactions for *MRPL20*, which are lost in DCM, suggesting its potential impact on mitochondrial dysfunction in DCM¹⁴² (**Figure 6G**). In addition, a recent study employing integrated multiomics approaches like RNA-seq and methylation assays identified *MTSS1* as one of the key genes associated with left ventricular systolic function¹⁴³. We found that enhancer connectome associated with *MTSS1* promoter was lost in DCM as compared to NF (**Figure 6H**), thus indicating its possible association with cardiac dysfunction in DCM. Other NF-enriched E-P examples (*AIF1L*, regulator for actomyosin contractility¹⁴⁴; *LRRC10B*, whose paralog *LRRC10* was ablated to induce DCM in mice¹⁴⁵) are displayed in **Figure 7I, J** respectively. Collectively,

these findings indicate that DCM and NF-enriched E-P interactions regulate genes involved in cardiac homeostasis.

2.2.4 DCM-Enriched Enhancer/Promoter Connectome Largely Resides in Conserved Higher-Order Chromatin Architectures and Pre-Accessible Chromatin Sites Bound by Reactivated HAND1

Next, to investigate whether the altered H3K27ac looping in DCM is associated with changes in higher-order chromatin structures, we performed *in situ* Hi-C on DCM and NF heart tissue samples (left ventricles) (**Figure 8**).

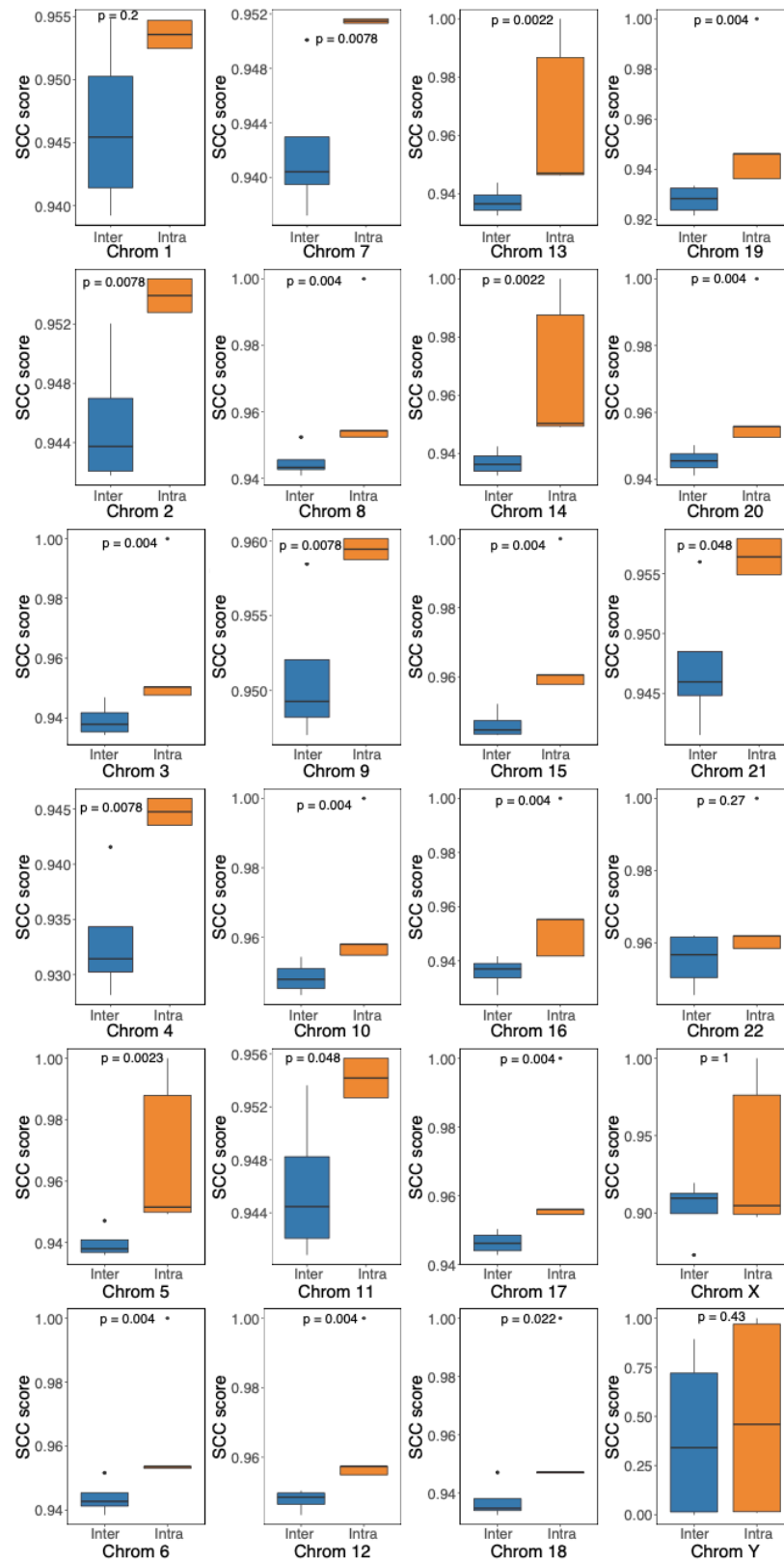


Figure 8. Assessment of the Intra and Inter Group Similarity between Contact Matrices in DCM and NF *in-situ* Hi-C Data. Stratum adjusted correlation coefficient (SCC) was computed separately for each chromosome. Figure adapted with permission from Feng et al., *Circulation*, 2022.

In situ Hi-C can capture the multilevel chromatin architecture, including the A and B compartments and TADs. We found that both the DCM and NF heart genomes were partitioned into A and B compartments (**Figure 9A**). We compared the A and B compartments in the genomes of NF and DCM hearts; overall, only a small fraction (A-to-B, 13.3%; B-to-A, 9.4%) of the genome switched compartments from NF to DCM hearts (**Figure 9B**). Gene expression was reduced for the genes that switched from the A-to-B compartment, while expression was increased for the genes that switched from the B-to-A compartment (**Figure 9C**); these results agree with previous findings showing that the A compartment corresponds to active transcription and that the B compartment corresponds to repressed transcription².

Then, we examined the TAD distribution and identified 1,428 DCM-specific TADs, 1,114 NF-specific TADs, and 4,803 common TADs (**Figure 9D**). Comparison of genome-wide insulation scores suggested that TADs were highly conserved between DCM and NF hearts (correlation coefficient = 0.97) (**Figure 9E**) and presented similar insulation profiles (**Figure 9F**). Hence, the overall higher-order chromatin structures remain largely invariant from NF to DCM hearts. Moreover, we found that the A and B compartments occupied similar portions of the genome, with TADs distributed evenly in the A and B compartments in both DCM and NF hearts, while H3K27ac loops located mostly in the A compartment,

suggesting that H3K27ac-associated chromatin interactions are involved in active transcription (**Figure 10A, B**).

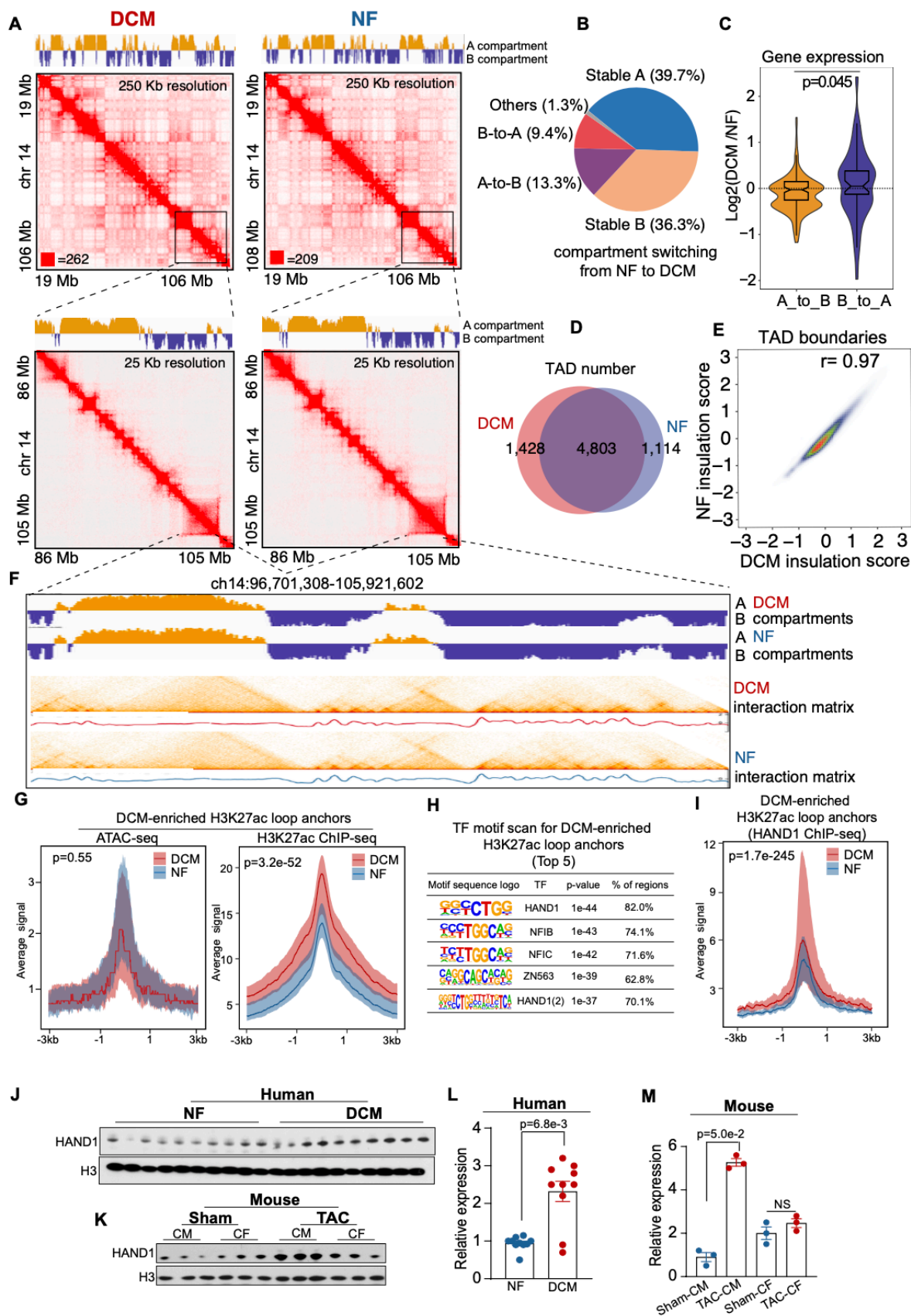


Figure 9. DCM-Enriched Enhancer/Promoter Connectome Largely Resides in Conserved Higher-Order Chromatin Architectures and Pre-accessible Chromatin Sites Bound by Reactivated HAND1. **A.** Knight–Ruiz (KR) matrix–balanced interaction maps of merged DCM (top) and NF (bottom) in situ Hi-C valid pairs for DCM and NF hearts at 250 kb and 25 kb resolution. The top tracks indicate the A (yellow) and B (blue) compartments. **B.** Proportion of genome switching between the A and B compartments (A-to-B (13.3%), B-to-A (9.4%), stable A (39.7%), and stable B (36.3%)) from NF to DCM hearts. **C.** Expression bias in A-to-B and B-to-A compartment-switched genomes in DCM hearts normalized to expression in NF hearts. Data were analyzed by Mann-Whitney U test. **D.** Venn diagram showing the number of TADs in DCM hearts (n=1,428) and NF hearts (n=1,114) and in both (n=4,803). **E.** Correlation of insulation scores for DCM and NF hearts; the Pearson correlation coefficient (R) is indicated (r=0.97). **F.** A/B compartment, Hi-C contact map (heatmap) and corresponding insulation profiles for DCM and NF hearts in the genomic region (ch14:96,701,308-105,921,602). **G.** Left panel: aggregation plot of DCM (n=20) and NF (n=20) normalized ATAC-seq signals $\pm$ 3 kb centered on the DCM-enriched H3K27ac loop anchors; Right panel: Left panel: aggregation plot of DCM (n=20) and NF (n=20) normalized H3K27ac ChIP-seq signals $\pm$ 3 kb centered on the DCM-enriched H3K27ac loop anchors. Data were analyzed by Mann-Whitney U test. **H.** Transcription factor motif scan for DCM-enriched loop anchors. The top 5 transcription factors are shown along with their motifs, p-values and percentage of regions. **I.** Aggregation plot of normalized HAND1 ChIP-seq signals $\pm$ 3 kb centered on the DCM-enriched H3K27ac loop anchors in DCM (n=4) and NF (n=4) hearts. Data were analyzed by Mann-Whitney U test. **J.** Evaluation of HAND1 expression (normalized to H3 expression) in human non-failing (NF) hearts and failing hearts (DCM) by Western blot analysis. **K.** Evaluation of HAND1 expression (normalized to H3 expression) in mouse cardiomyocytes (CMs) and cardiac fibroblasts (CF) obtained from non-failing hearts (Sham) and failing hearts (8 weeks post-transverse aortic constriction (TAC)) by Western blot analysis. **L.** Quantification of HAND1 expression in human NF and DCM hearts. Data were analyzed by Mann-Whitney U test. **M.** Quantification of HAND1 expression in mouse CMs and CFs isolated from Sham and TAC hearts. Data were analyzed by Mann-

Whitney U test. Figure adapted with permission from Feng et al., Circulation, 2022. The experiments for Figure 9 J, K, L, M were designed by me and performed by Wei Huang at University of Cincinnati.

Furthermore, we aimed to determine the distributions of enriched H3K27ac loops in altered compartments and TADs in DCM and NF hearts. We found that only 1.3% of DCM-enriched H3K27ac loops (**Figure 10C**) and 5.8% of NF heart-enriched H3K27ac loops (**Figure 10D**) were in switched compartments. In addition, 9.4% of DCM-enriched H3K27ac loops (**Figure 10E**) and 15.5% of NF-enriched H3K27ac loops (**Figure 10F**) were in changed TADs. The small percentage of specific loops within altered compartments and TADs suggested that factors other than higher-order chromatin structures might underlie the formation of specific H3K27ac loops in DCM.

As chromatin accessibility is essential for the function and interaction of CREs¹¹², we aimed to test whether DCM/NF-enriched H3K27ac loop formation could be explained solely by chromatin accessibility. To this end, we developed an optimized ATAC-seq protocol¹⁹ for frozen heart tissues and performed the experiments on the same 20 DCM/NF heart tissues described above (**Figure 10G**). Compared to NF hearts, DCM hearts had decreased chromatin accessibility for the NF-enriched H3K27ac loop anchors (**Figure 10H**). However, chromatin accessibility was equivalent for the DCM-enriched H3K27ac loop anchors in DCM and NF hearts (**Figure 9G**). These results suggested that DCM-enriched H3K27ac loops were formed on pre-accessible chromatin regions.

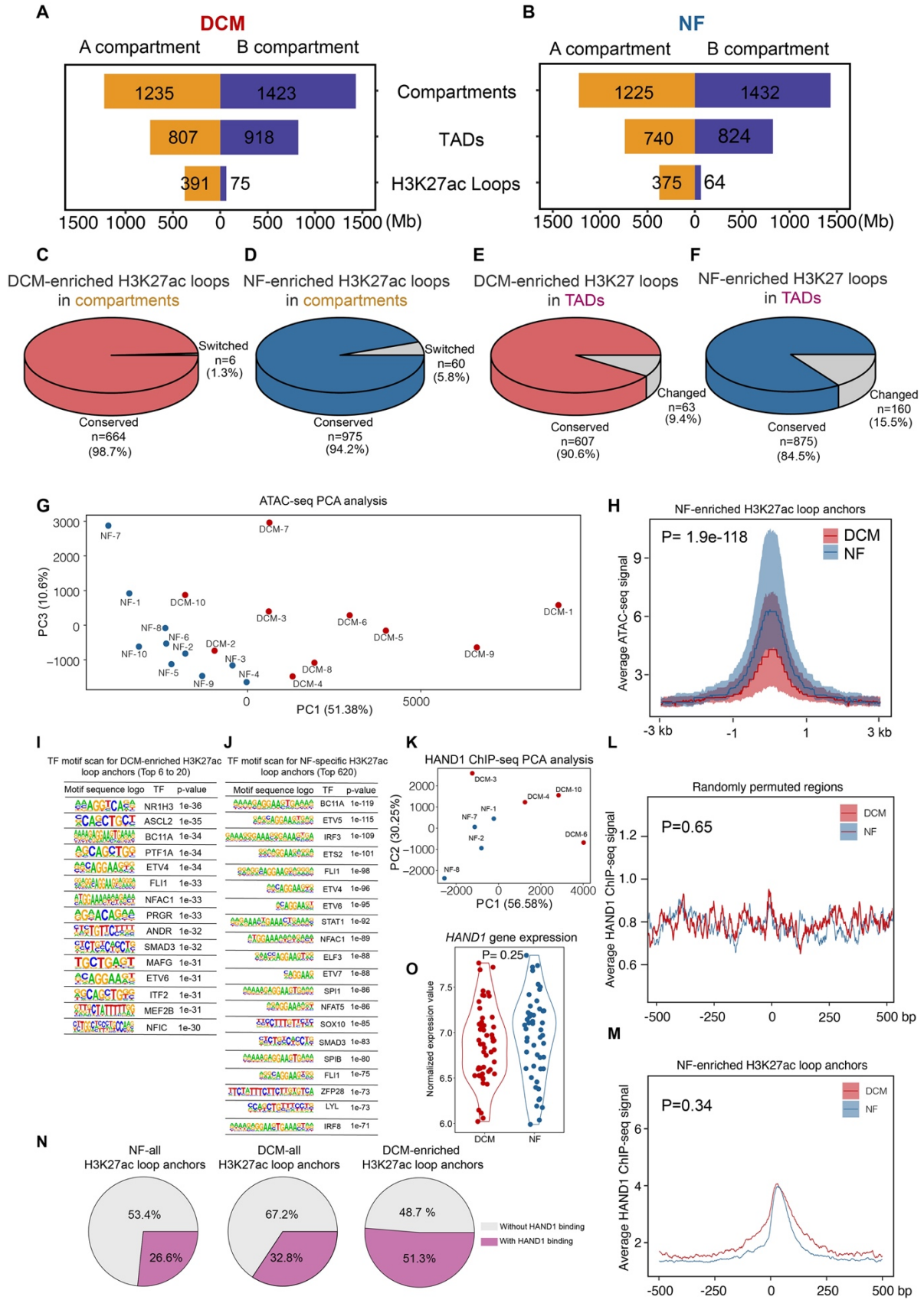


Figure 10. DCM-enriched H3K27ac HiChIP Loops Largely Reside in Pre-established High-Order Chromatin Architectures and Pre-accessible Chromatin Sites Bound by Reactivated HAND1. **A-B.** Bar chart of genomic coverage of A compartment (yellow) and B compartment (blue), TADs, and H3K27ac loops in DCM and NF hearts. Genomic coverage is shown in Mb. **C-D.** Distribution and number of DCM-enriched (C) and NF-enriched (D) H3K27ac loops in switched and conserved compartments. **E-F.** Distribution and number of DCM-enriched (E) and NF-enriched (F) H3K27ac loops in changed and conserved TADs. **G.** PCA plot of ensembled 10 DCM and 10 NF ATAC-seq profiles. Red dots represent DCM heart samples. Blue dots represent NF heart samples. **H.** Aggregation plot of NF (blue) and DCM (red) normalized ATAC-seq signals ± 3 kb centered on ATAC-seq peaks overlapped with NF-enriched H3K27ac loop anchors. Data were analyzed by Mann-Whitney U test. **I.** Transcription factor motif-enriched ATAC-seq peak overlapped with the DCM-enriched H3K27ac loop anchor (top 6 to top 20 transcription factors are shown). The p-values and the motif sequences are also shown. **J.** Transcription factor motif-enriched ATAC-seq peak overlapped with the NF-enriched H3K27ac loop anchor (Top 20 transcription factors are shown). The p-values and the motif sequences are also shown. **K.** PCA plot of ensembled 4 DCM and 4 NF HAND1 ChIP-seq profiles. Red dots represent DCM heart samples. Blue dots represent NF heart samples. **L-M.** Aggregation plot of normalized HAND1 ChIP-seq signals on randomly permuted regions (L) or centered on the ATAC-seq peaks overlapping with the NF-enriched H3K27ac loop anchors (M) in DCM (n=4) and NF (n=4) hearts. Data were analyzed by Mann-Whitney U test. **N.** Percentage of NF-all H3K27ac HiChIP loops, DCM-all H3K27ac HiChIP loops, DCM-enriched H3K27ac HiChIP loops, the anchors of which do or do not overlap with HAND1 ChIP-seq peaks. **O.** Comparison of HAND1 gene expression in DCM (n=50) and NF hearts (n=51) from the RNA-seq data in this study. Figure adapted with permission from Feng et al., *Circulation*, 2022.

Because transcription factors are key determinants shaping the 3D genome¹⁴⁶, we scanned the transcription factor motifs on the DCM-enriched H3K27ac loop anchors that overlapped with their ATAC-seq peaks. The motifs for the transcription factors SMAD3⁵³ and MEF¹⁴⁷, known to be involved in heart failure or DCM, were among the 20 most significantly enriched motifs (ranked No. 15 and No. 19, respectively) (**Figure 10I**). The most significantly enriched transcription factor motif on the DCM-enriched H3K27ac loop anchors was HAND1 (**Figure 9H**), which is not enriched in NF-enriched H3K27ac loop anchors (**Figure 10J**). HAND1 is a basic helix-loop-helix (bHLH) transcription factor that is highly expressed in the embryonic heart and essential for heart development and homeostasis. In humans and rodents, *HAND1* transcription in the adult heart is maintained at a much lower level than that in the embryonic heart¹⁴⁸. Recently, a report identified a heterozygous *HAND1* mutation in the blood DNA from some Chinese familial DCM patients¹⁴⁹. However, the functional roles of HAND1 in 3D epigenome reprogramming and transcription regulation within DCM hearts remains unexplored.

To further uncover the link of HAND1 with DCM-enriched H3K27ac looping, we performed HAND1 ChIP-seq (**Figure 10K**). We found that HAND1 binding was preferably enriched on DCM-enriched H3K27ac loop anchors (**Figure 9I**) in DCM hearts compared to randomly permuted regions (**Figure 10L**), NF-enriched H3K27ac loop anchors (**Figure 10M**) or NF/DCM-all loop anchors (**Figure 10N**). Although we demonstrated no significant difference of *HAND1* RNA expression between human DCM and NF hearts (**Figure 10O**), HAND1 protein expression in human DCM hearts was significantly upregulated compared to that in NF hearts (**Figure 9J, L**), suggesting post-transcriptional/post-translational

modifications of HAND1 in DCM. To further analyze HAND1 protein expression in mouse failing hearts, we constructed transverse aortic constriction (TAC) model and isolated cardiomyocytes (CMs) and cardiac fibroblasts (CFs) at 8 weeks after TAC. We found that HAND1 protein expression was significantly upregulated in TAC-CMs but not in TAC-CFs as compared to Sham controls (**Figure 9K, M**). These data suggested that HAND1 might be the potential transcription factor linked to the DCM-enriched CREs interactions in cardiomyocytes.

2.2.5 HAND1 Rewires Enhancer/Promoter Connectome in Concordance with Transcriptional Changes Associated with DCM Etiology in Human iPSC-Derived Cardiomyocytes

To dissect the functionality of HAND1 in organizing enhancer/promoter connectome of human cardiomyocytes, we transduced Ctrl and HAND1-overexpressing (*HAND1*^{OE}) adenovirus into human iPSC-derived cardiomyocytes (hiPSC-CM) for 72 hrs, which have emerged as invaluable platforms for disease modeling and therapeutic screening in DCM¹⁵⁰. Subsequently, we performed 3D epigenome mapping (H3K27ac HiChIP, H3K27ac ChIP-seq, HAND1 ChIP-seq and RNA-seq) (**Figure 11A, Figure 12A-C**) in these cells. HAND1-overexpression induced extensive rewiring of enhancer/promoter connectome with more gained H3K27ac HiChIP loops (N=5,456, 73%) compared to lost loops (N=1,968, 23%) (**Figure 11B**; global view of gained loops are demonstrated in **Figure 12D**). By integrating HAND1 ChIP-seq data, we found that 85% of the gained loops were bound by HAND1 on one or both anchors in *HAND1*^{OE} hiPSC-CM (**Figure 11C**). In addition, HAND1 binding intensity is significantly higher in those gained loop

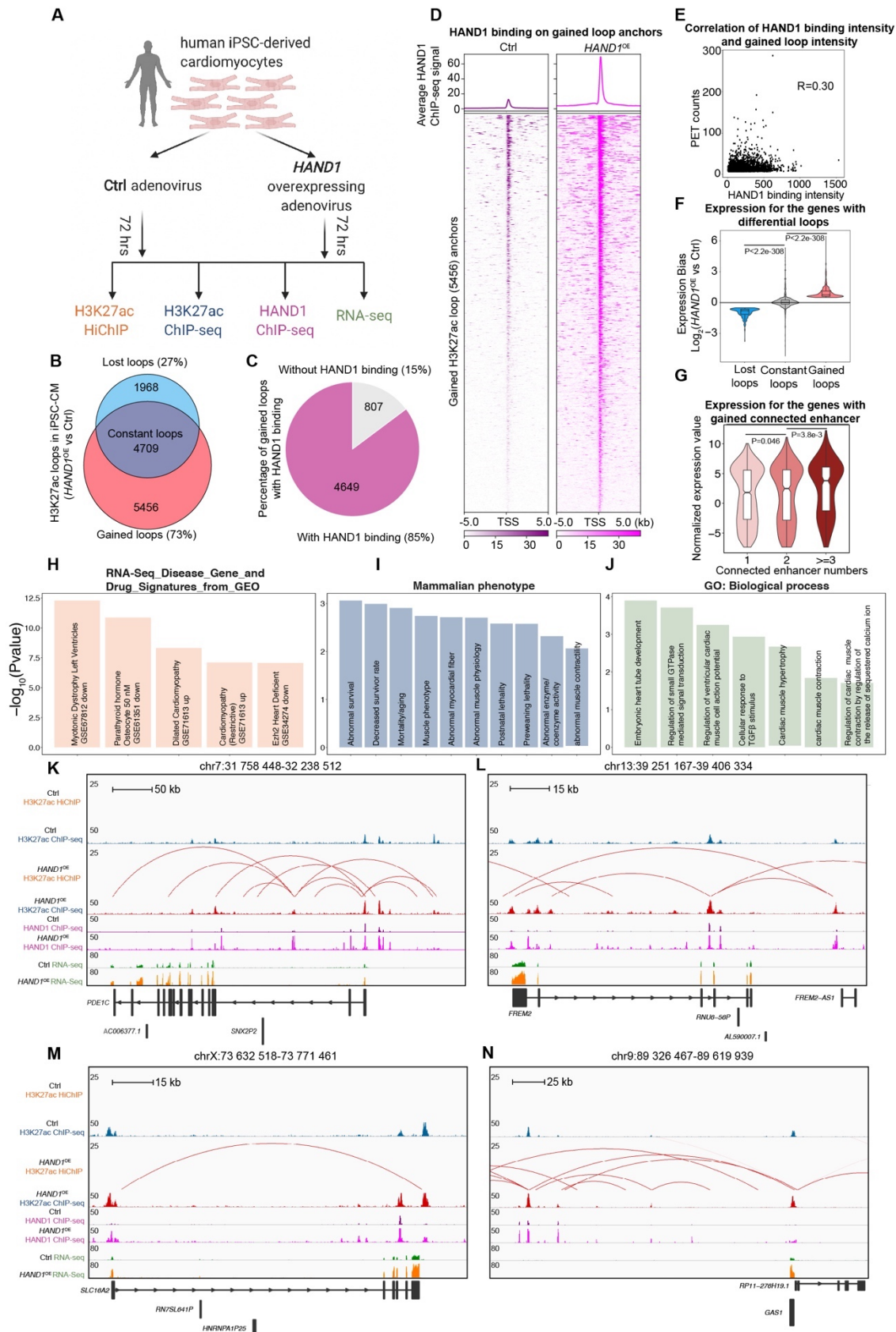


Figure 11. Overexpression of HAND1 Rewires Long-Range Enhancer/Promoter Connectome and Transcriptome in Human iPSC-Derived Cardiomyocytes. **A.** Schematic illustration of the experimental system and strategy to examine the roles of HAND1 in 3D chromatin organization and transcriptional control in human iPSC-derived cardiomyocytes (hiPSC-CM). **B.** Venn diagram showing the number of H3K27ac HiChIP loops gained, lost and constant in HAND1-overexpressing (HAND1^{OE}) versus Ctrl hiPSC-CM. **C.** Percentage of gained H3K27ac HiChIP loops, the anchors of which do or do not overlap with HAND1 ChIP-seq peaks in HAND1^{OE} hiPSC-CM. **D.** Aggregation plot and heatmap showing the intensity of HAND1 ChIP-seq peaks on the gained H3K27ac HiChIP loop anchors (HAND1^{OE} versus Ctrl hiPSC-CM). **E.** Correlation of the intensity of HAND1 ChIP-seq peaks and the intensity of gained H3K27ac HiChIP loops in HAND1^{OE} hiPSC-CM. **F.** Expression of genes with H3K27ac HiChIP loops gained, lost and constant in HAND1-overexpressing (HAND1^{OE}) versus Ctrl hiPSC-CM. Data were analyzed by Wilcoxon rank sum test with continuity correction. **G.** Normalized expression values for genes with gained H3K27ac HiChIP loops in HAND1^{OE} hiPSC-CM, the promoter anchors of which connect with different numbers of enhancers. Data were analyzed by Wilcoxon rank sum test with continuity correction. **H.** Functional annotation (RNA-Seq_Disease_Gene_and_Drug_Signatures_from_GEO) for the genes with gained H3K27ac HiChIP loops. **I.** Functional annotation (Mammalian phenotypes) for genes with gained H3K27ac HiChIP loops. **J.** Functional annotation (GO-Biological Process) for genes with gained H3K27ac HiChIP loops. **K-N.** Examples of H3K27ac gained interactions (identified by H3K27ac HiChIP and ChIP-seq), HAND1 binding (identified by HAND1 ChIP-seq) and transcription (identified by RNA-seq) in HAND1^{OE} versus Ctrl hiPSC-CM. Figure adapted with permission from Feng et al., *Circulation*, 2022.

anchors of HAND1^{OE} compared to Ctrl hiPSC-CM (**Figure 11D**). However, HAND1 binding intensity is minimal on permuted regions (**Figure 12E**) while equivalent on common loop anchors (**Figure 11F**), and subtly less on lost loops (**Figure 12G**) in HAND1^{OE} compared to Ctrl hiPSC-CM, which further supported that gained loops are

attributed to specific HAND1 binding on the loop anchors. Moreover, HAND1 binding intensity on the gained loop anchors is moderately correlated with loop intensity (**Figure 11E**). Taken together, these data indicated that the induced enhancer/promoter rewiring is a direct effect of the upregulation of HAND1.

By integrating RNA-seq data, we found that genes within the gained and lost loops were significantly up- and downregulated, respectively (*HAND1*^{OE} compared with Ctrl hiPSC-CM) (**Figure 11F**). In addition, in gained loops, the numbers of enhancers connected to promoters were significantly correlated with expression level (**Figure 11G**). Functional annotation analysis indicated that gained loops-associated genes overlapped with the transcriptional signatures in public human DCM RNA-seq data (**Figure 11H**). In line with this, we found that the genes (e.g., *NPPA*, *NPPB*) both upregulated in hiPSC-*HAND1*^{OE} (vs. Ctrl) and DCM hearts (vs. NF) from this study are enriched in the transcriptional signatures of dilated cardiomyopathy and left ventricle hypertrophy (**Figure 12H**). Moreover, mammalian phenotypes for gained loops-associated genes showed significant enrichment in decreased survival, abnormal myocardial fiber, and abnormal muscle physiology/contractility (**Figure 11I**).

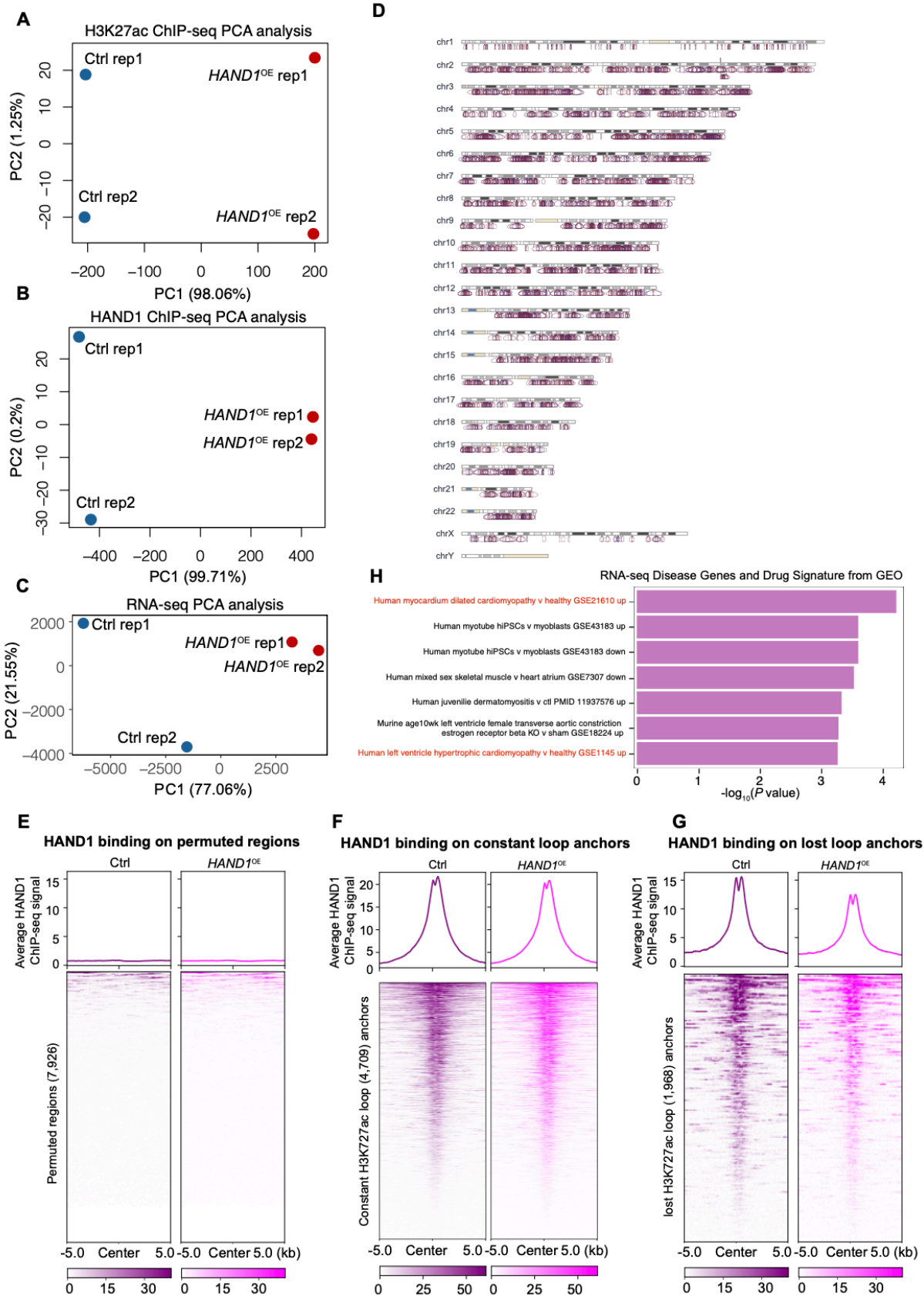


Figure 12. Overexpression of HAND1 Induces Genome-Wide Gain of Enhancer/Promoter Interactions.

A-C. PCA plot of H3K27ac ChIP-seq (A), HAND1 ChIP-seq (B) and RNA-seq (C) for Ctrl and HAND1^{OE} iPSC-CM. Red dots represent Ctrl samples. Blue dots represent HAND1^{OE} samples. **D.** Global view of gained H3K27ac HiChIP loops in hiPSC-CMs induced by HAND1. **E-G.** Aggregation plot and heatmap showing the intensity of HAND1 ChIP-seq peaks on the permuted regions (E), the constant (F) and lost (G) H3K27ac HiChIP loop anchors (HAND1^{OE} vs Ctrl hiPSC-CM). **H.** Functional annotation (RNA-Seq_Disease_Gene_and_Drug_Signatures_from_GEO) for the overlap between upregulated genes in hiPSC-HAND1^{OE} (vs. Ctrl) and upregulated genes in DCM hearts (vs. NF) from this study. Figure adapted with permission from Feng et al., *Circulation*, 2022.

Furthermore, Gene Ontology for gained loops-associated genes showed significant enrichment in biological processes associated with embryonic heart tube development, regulation of ventricular cardiac cell action potential, cardiac muscle hypertrophy, cardiac muscle contractility, and regulation of cardiac muscle contraction by regulating the release of sequestered calcium ion (**Figure 11J**), further validating that some HAND1 functions through activating chromatin looping and gene transcription involved in these processes during DCM pathogenesis. Examples of gained loops involving the well-established DCM driver gene *PDE1C*¹⁵¹ and unexplored genes *FREM2*, *SLC16A2*, *GAS1* are shown in **Figure 11K, L, M, N**. These results demonstrate that the upregulation of HAND1 in hiPSC-CM results in the genome-wide gain of enhancer/promoter contacts and concordant transcriptional activation of connected genes implicated in DCM development.

2.2.6 HAND1 Drives DCM Pathology in Human iPSC-CM and Mouse Models

To test the pathological effects of HAND1 in DCM, we performed systematic *in vitro* and *in vivo* functional assays (**Figure 13A**). For *in vitro* study, we investigated the effect of *HAND1* overexpression on human cardiac morphology and function using adenoviral vectors to overexpress *HAND1* in hiPSC-CMs. At 72 hrs after transduction, we found a significant higher percentage of *HAND1* overexpressing hiPSC-CMs (hiPSC-CM^{HAND1}) were larger than null adenovirus transfected hiPSC-CMs (hiPSC-CM^{Null}) (**Figure 14A**). Furthermore, overexpression of *HAND1* (hiPSC-CM^{HAND1}) resulted in significant increases in Ca²⁺ transient amplitude (**Figure 13B**), cell shortening, and relaxation velocity respectively, compared to hiPSC-CM^{Null} (**Figure 14B-D**). These results indicated acute overexpression of *HAND1* induced hiPSC-CM hypertrophy development with associated alteration in Ca²⁺-handling and contractility, which might be a compensation to maintain cardiomyocyte contraction¹⁵²⁻¹⁵⁴.

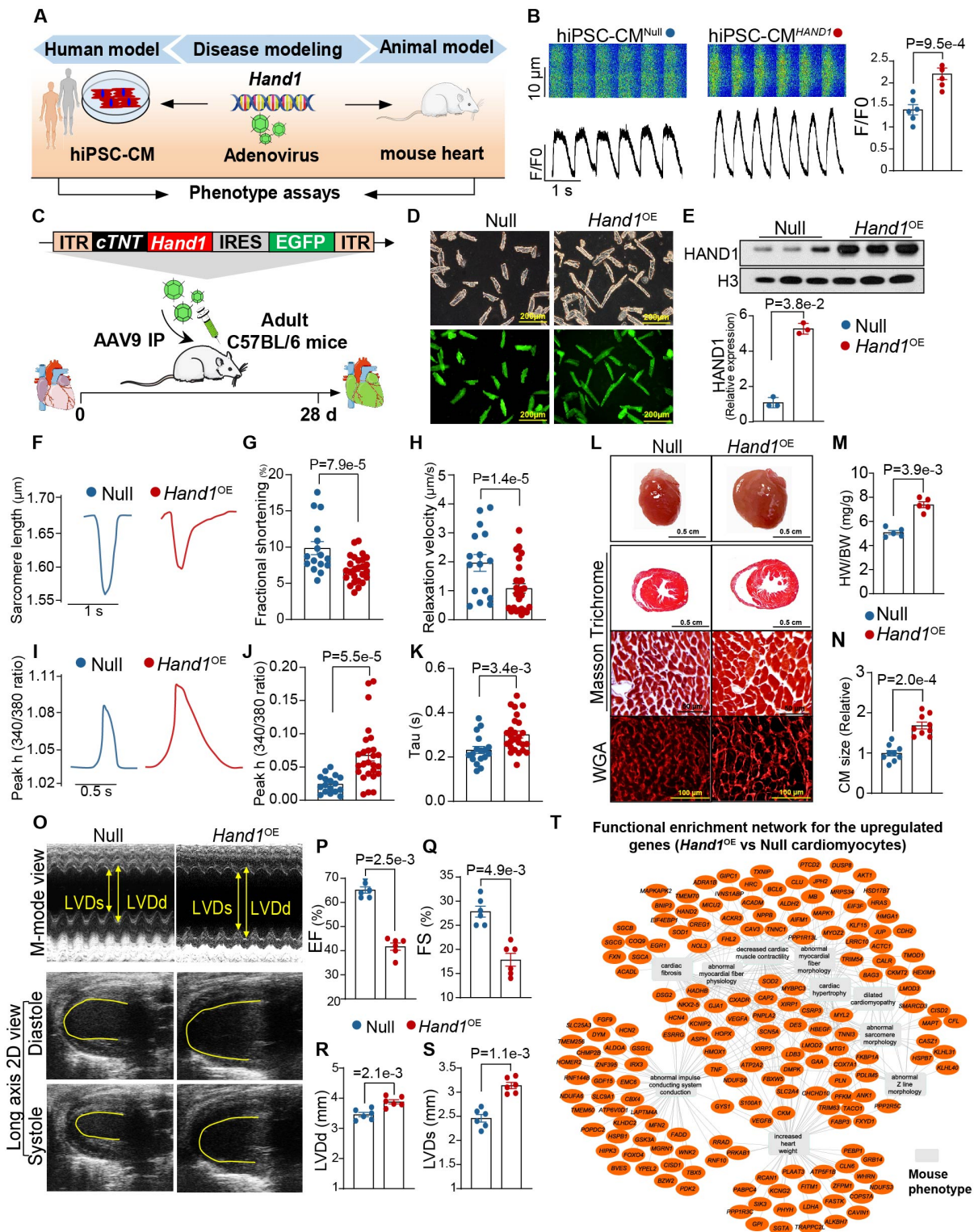


Figure 13. HAND1 Induces DCM Phenotypes in Human iPSC-Derived Cardiomyocytes and Mouse Hearts. A. Experimental scheme for assessing *Hand1* pathogenicity in both human and mouse models with

hiPSC-CMs and mouse hearts, respectively. **B.** Representative calcium imaging recording traces and calcium transients to electrical pacing at 1 Hz. HiPSC-CMs were transduced with adenoviruses carrying the CMV promoter driving HAND1 (hiPSC-CM^{HAND1}) or adenoviruses carrying the CMV promoter only (hiPSC-CM^{Null}). F/F₀, peak amplitude relative to baseline fluorescence. Data were analyzed by Student's *t* test. **C.** Experimental scheme for evaluating AAV9-driven Hand1 overexpression in cardiac morphology and function. Hand1 under the control of the cTNT promoter was cloned into an AAV9 backbone with an EGFP reporter (AAV9-cTNT-Hand1-EGFP). Animals were injected intraperitoneally and subsequently analyzed 28 days later. **D.** Representative images of CMs isolated from AAV treated hearts under light microscopy and fluorescence microscopy. The mice receiving injection of AAV9-cTNT-EGFP were designated as Null and the mice receiving AAV9-cTNT-Hand1-EGFP injection were designated as Hand1^{OE}. **E.** Western blot assay for HAND1 protein expression in Null and Hand1^{OE} hearts. Data were analyzed by Mann-Whitney U test. **F-H.** Contractility mechanic assay of mouse CMs isolated from Null and Hand1^{OE} hearts. (F) Representative images of sarcomere shortening tracing. (G) Cell shortening (%) at 0.5 Hz. (H) Relaxation velocity of sarcomere at 0.5 Hz. Data were analyzed by Student's *t* test. **I-K.** Ca²⁺ kinetics of mouse CMs isolated from Null and Hand1^{OE} hearts. (I) Representative images of Ca²⁺ transient tracing. (J) Ca²⁺ transient amplitude (Peak h) as indicated by Fura-2 ratio (340/380 nm) at 0.5 Hz. (K) Ca²⁺-decay time (Tau) at 0.5 Hz. Data were analyzed by Student's *t* test. **L.** Gross morphology (first row), Masson's trichrome staining of gross morphology (second row), Masson trichrome staining of CMs (third row) and wheat germ agglutinin (WGA) staining (bottom panel) of Null and Hand1^{OE} hearts. **M-N.** Measurement of heart weight (HW)-to-body weight (BW) ratio (M) and CM size (N) in Null and Hand1^{OE} mice. Data were analyzed by Mann-Whitney U test. **O-S.** Heart function was analyzed by echocardiography in Null and Hand1^{OE} mice. Representative images of M-mode and long axis 2D views (O). Quantifications of left ventricular (LV) ejection fraction (EF)(P), fraction shortening (FS)(Q), LV end-diastolic diameter (LVDd)(R), and LV end-systolic diameter (LVDs)(S). Data were analyzed by Mann-Whitney U test. **T.** Network representation of functional enrichment of the upregulated genes in Hand1^{OE} cardiomyocytes (vs Null). Functional enrichment was done using the ToppFun application. Orange nodes

represent DCM-enriched genes, while the grey colored rectangles represent enriched terms. Top10 enriched terms are shown here. Network was generated using the Cytoscape application. Figure adapted with permission from Feng et al., *Circulation*, 2022. The electrophysiology and in vivo experiments experiments for Figure 13B-S were designed by me and performed by Wei Huang (D,E,L,M,N,O,P,Q,R,S) and Mohit Kumar (B, F-K) at University of Cincinnati.

To validate these results *in vivo*, we engineered an AAV9 vector with the *cTNT* promoter to drive cardiac-specific overexpression of *Hand1* in mouse CMs (AAV9-*cTNT-Hand1*-EGFP) (**Figure 13C**). To assess the specificity and potency of the *cTNT* promoter, AAV9 virus lacking *Hand1* (AAV9-*cTNT*-EGFP) was delivered to adult wild-type (WT) mice (8 weeks old) (**Figure 14E**) at 3 different doses (5×10^{11} , 1×10^{12} , and 5×10^{12} viral genomes (vg) by intraperitoneal injection), and the mice receiving AAV9 were defined as Low, Medium, and High groups, respectively. The mice that received the same volume of saline were defined as the negative control group (designated as the Saline group). Mice injected with the AAV9-*cTNT*-EGFP viral vector exhibited robust expression of the GFP signal in the CMs (**Figure 14F, G**). As expected, no GFP signal was observed in saline-treated hearts. Although the AAV dosage was positively correlated with the number of GFP⁺ CMs, no difference in the GFP protein level was observed between the Medium and High groups (**Figure 14F**). These data indicate that one systemic dose of 1×10^{12} vg yields highly efficient transduction in the majority of CMs. We then used the selected dosage to intraperitoneally inject WT adult mice with AAV9-*cTNT-Hand1*-EGFP (designated as the *Hand1*^{OE} group) or AAV9-*cTNT*-EGFP (designated as the Null group). AAV9 transduction in *Hand1*^{OE} hearts (**Figure 13D**) resulted in significant upregulation of HAND1, as determined by Western blotting (**Figure 13E**). To interrogate the functional effect of *Hand1* on heart failure progression, contractile mechanics, amplitude and

kinetics of intracellular Ca^{2+} transients were measured in freshly isolated mouse CMs (mCMs) from *Hand1*^{OE} hearts and Null hearts after AAV9 treatment for 4 weeks *in vivo*. In contrast to acute overexpression of *HAND1* in hiPSC-CM, chronic overexpression of *Hand1* resulted in significant decreases in cell shortening and relaxation velocity respectively, compared with the Null CMs (**Figure 13F-H**). Although assessment of Ca^{2+} amplitude (peak h) revealed an increase in *Hand1*^{OE} CMs, the Tau was significantly prolonged as compared to Null CMs (**Figure 13I-K**), suggesting impaired Ca^{2+} handling. These results indicated the regulatory effect of *Hand1* on CM dysfunction progression, as evidenced by a decline in contractile function and abnormal Ca^{2+} handling.

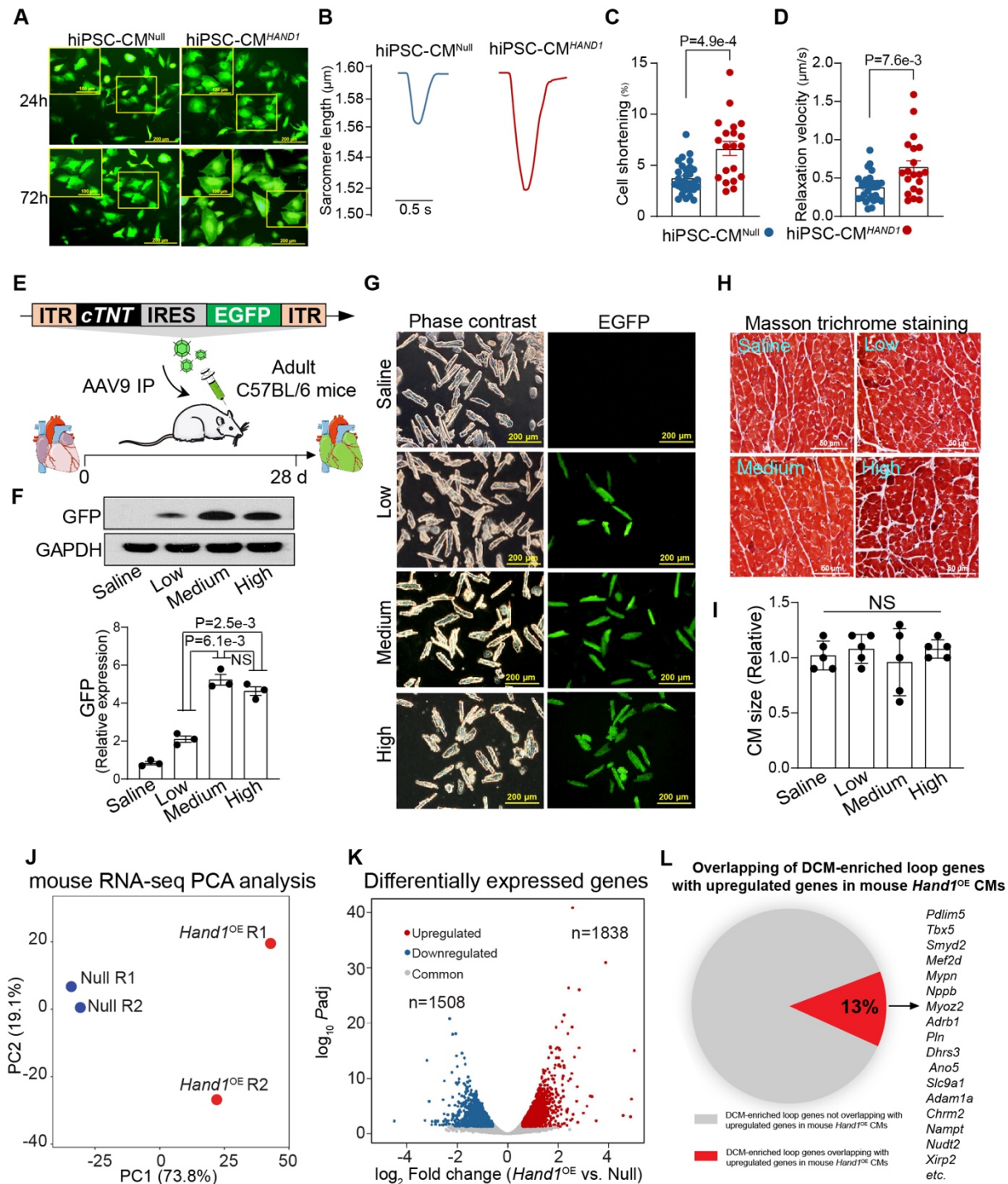


Figure 14. HAND1 Induces DCM Phenotypes in Human iPSC-Derived Cardiomyocytes and Mouse. *A.* Fluorescence microscopy showing GFP expression of hiPSC-CMs after adenovirus transduction for 24h or 72h. The cells transfected with Null adenovirus (Ad-GFP) were designated as hiPSC-CM^{Null} and the cells transfected with Ad-GFP-HAND1 were designated as hiPSC-CM^{HAND1}. *B-D.* Contractility mechanic

assay of hiPSC-CMs. (B) Representative images of sarcomere shortening tracing. (C) Cell shortening (%) at 0.5 Hz. (D) Relaxation velocity of sarcomere at 0.5 Hz. Data were analyzed by Student's *t* test. E. Schematic representation of the AAV9 transgene (AAV9-cTNT-EGFP) used for this study. F. Western blot analysis of EGFP expression in CMs of mice receiving different doses of AAV9-cTNT-EGFP or saline (upper panel). The mice receiving the different dose of AAV9 (5×10^{11} , 1×10^{12} , and 5×10^{12} viral genomes (vg)) by intraperitoneal injection were defined as Low, Medium, and High group, respectively. The mice receiving saline were defined as the Saline group. The quantification of protein expression in the western blot (lower panel). Data were analyzed by One-way ANOVA. G. Fluorescence microscopy showing EGFP expression in isolated CMs from mice receiving different dosages of AAV9-cTNT-EGFP. H. Masson trichrome staining of hearts of mice receiving AAV9-cTNT-EGFP or saline. I. Quantification of CM size. Data were analyzed by One-way ANOVA. N.S., no statistical significance. J. PCA plot of RNA-seq for the cardiomyocytes isolated from Null and *Hand1*^{OE} hearts (28 days). Red dots represent Null samples. Blue dots represent *Hand1*^{OE} samples. K. Volcano plot showing the differentially expressed genes in *Hand1*^{OE} cardiomyocytes (vs Null). Red dots represent upregulated genes samples. Blue dots represent downregulated genes. L. Pie chart showing the percentage of DCM-enriched loop genes overlapped with upregulated genes in mouse *Hand1*^{OE} CMs. Figure adapted with permission from Feng et al., *Circulation*, 2022. The electrophysiology and in vivo experiments experiments for Figure 14A-I were designed by me and performed by Wei Huang (A,F, G, H, I) and Mohit Kumar (C, D) at University of Cincinnati.

Accordingly, compared with the Null hearts, the explanted *Hand1*^{OE} hearts were markedly enlarged (**Figure 13L**). In addition, the heart-to-body weight ratios (HW/BW) were higher in *Hand1*^{OE} mice than in Null mice, indicating a significant increase in heart mass (**Figure 13M**). Using wheat germ agglutinin (WGA) staining, we also found that the *Hand1*^{OE} CMs were significantly larger than the Null CMs (**Figure 13N**), implying that cardiac hypertrophy was developed in the *Hand1*^{OE} CMs. These effects were not the result of

AAV9-induced toxicity since no changes in CM size were observed in mice transduced with AAV9 without *Hand1* (AAV9-*cTNT*-GFP) or in Saline control mice (**Figure 14H and I**). Next, cardiac function analysis was examined in *Hand1*^{OE} mice and Null mice by echocardiography (**Figure 13O**); compared to the Null mice, the *Hand1*^{OE} mice showed significant reductions in left ventricular (LV) ejection fraction (EF) and fractional shortening (FS) (**Figure 13P, Q**), which are key parameters of cardiac contractile functions. Cardiac remodeling was also significantly increased in *Hand1*^{OE} mice, with increased LV end-diastolic diameter (LVDd), and LV end-systolic diameter (LVDs) (**Figure 13R, S**). In order to interrogate underlying mechanisms for HAND1-induced cardiac dysfunction, we performed RNA-seq in the cardiomyocytes isolated from Null and *Hand1*^{OE} mice (**Figure 14J**). Differential analysis showed that HAND1 induced drastic transcriptome reprogramming in *Hand1*^{OE} cardiomyocytes (**Figure 14K**) and recapitulated the gene signatures in human DCM hearts (**Figure 14L**). Functional enrichment network analysis for the upregulated genes in *Hand1*^{OE} cardiomyocytes further showed that the most highly represented GO terms included dilated cardiomyopathy, decreased cardiac muscle contractility, abnormal impulse conducting system conduction, cardiac hypertrophy, increased heart weight etc. (**Figure 13T**), which are in line with our observations in functional studies.

These data indicated that overexpression of *Hand1* in the adult heart provokes impaired contractility, Ca²⁺ mishandling, CM hypertrophy, cardiac dilation and dysfunction. Taken together, our functional and mechanistic data uncovered HAND1 as a novel factor in the development of DCM that could also serve as a therapeutic target for this heart disease.

2.3 Discussion

Although one-dimensional epigenome mapping techniques (e.g., ATAC-seq and H3K27ac ChIP-seq) can identify CREs in the genome, they do not indicate the connectivity of CREs for transcription regulation. In contrast, high-resolution 3D genome techniques (e.g., H3K27ac HiChIP) can identify the connectome of active CREs, which is crucial since enhancers can skip their nearby promoters to regulate distal promoters.

In this study, we dissected, for the first time, the 3D chromatin architecture in human DCM at high-resolution and unraveled how this reorganization dictates transcriptional deregulation and its pathogenesis (**Figure 15**). We identified the DCM-enriched enhancer/promoter connectome, which is largely dependent on pre-existing chromatin accessibility and higher-order chromatin architecture, possibly priming their connected genes for rapid transcription in response to cardiac stress. We found that fetal cardiac and stress response genes (e.g., *PDLIM5* (**Figure 5E**), *TBX5* (**Figure 6E**), *SMYD2* (**Figure 6F**), *MEF2D* (**Figure 7F**), *MYPN* (**Figure 7G**)) were aberrantly activated via gain of enhancer-promoter loops, whose anchors are characterized by increased HAND1 binding. In line with this, overexpression of *Hand1* in adult mouse hearts also activated the expression of *PDLIM5*, *TBX5*, *SMYD2*, *MEF2D*, *MYPN* etc. (**Figure 14L**). This suggests that HAND1 may reactivate, at least in part, the early cardiac development and stress response genes in adult mice and humans to drive DCM pathogenesis.

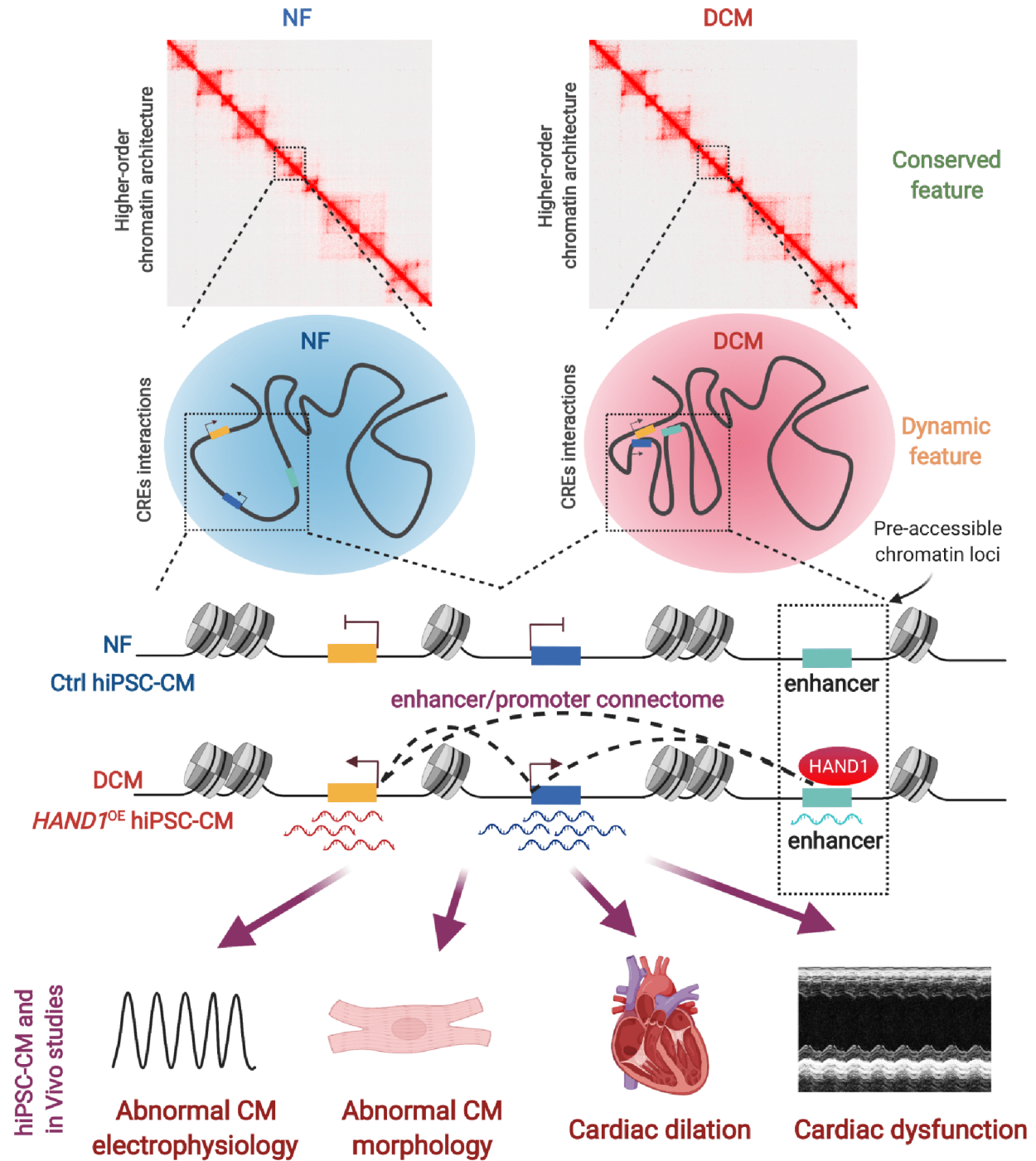


Figure 15. Model Illustrating 3D Epigenome Organization in Human DCM and How *HAND1* Rewires Enhancer/Promoter Connectome to Contribute to Transcriptome Reprogramming and DCM Pathogenesis. DCM-enriched enhancer/promoter connectome largely resides in pre-established, high-order chromatin architectures and pre-accessible chromatin sites bound by reactivated *HAND1*. Ectopic

expression of HAND1 in human iPSC-derived cardiomyocytes (hiPSC-CM) and DCM hearts induces a distinct gain of enhancer/promoter connectivity and increases the expression of their associated genes implicated for DCM etiology, subsequently leading to abnormal cardiomyocyte electrophysiology/morphology, cardiac dilation and dysfunction. This illustrates a model whereby a single transcription factor rewires genome-wide enhancer/promoter connectome to drive DCM pathogenesis.

HAND1 has previously not been investigated for a role in mediating long-range chromatin interactions. In order to test whether HAND1 could causally reprogram enhancer/promoter connectome and transcription, we performed extensive 3D epigenome mapping in hiPSC-CM. We found that HAND1 induces a genome-wide gain of distinct H3K27ac HiChIP loops. In addition, we showed that most of the gained loop anchors are characterized by increased HAND1 binding, indicating the direct effect of HAND1-induced chromatin looping. Moreover, the gained enhancer/promoter connections increase the expression of their associated genes implicated for DCM, suggesting a direct functional connection with the disease. Furthermore, we showed that reactivation of HAND1 in hiPSC-CM and mouse heart drives DCM phenotype and cardiac dysfunction, indicating that the precise temporal regulation of HAND1 is critical for cardiac homeostasis. Although the 3D epigenome mapping indicates differential chromatin architectures in DCM compared to NF, these datasets have some limitations. These are bulk sequencing data which were obtained from hearts at the end-stage of the disease at transplantation. As such, it is difficult to determine which of the many observed changes are disease casual, disease compensatory, or non-specific changes that accompany any form of heart failure. Such insight is particularly challenging to obtain in humans due to ethical limitations. Nevertheless, the identification of HAND1 as transcription factor

regulating the 3D chromatin connectomes, gene expression in DCM as well as in mouse cardiomyocytes highlights the importance of this factor and its effects in playing an important causal role in this disease.

Interestingly, while we identified the essential roles of HAND1 in chromatin looping, we also noticed that HAND1 binding is not always sufficient for loop formation and gene activation in iPSC-CM, as some enhancers/promoters have strong HAND1 binding, but are not involved in chromatin looping, suggesting that additional factors (e.g., architectural proteins /transcription co-factors or chromatin-associated RNA) are required to facilitate HAND1-mediated enhancer/promoter looping. Regarding this hypothesis, *in-situ* capture of chromatin interactions by biotinylated dCas9 in iPSC-CM, followed by proteomics and RNA-seq¹⁵⁵, could identify HAND1-interacting proteins/RNAs involved in chromatin looping. In addition, how HAND1 and its associated factors work together to regulate 3D chromatin contact (e.g., via phase-separated nuclear condensates⁴²) should be addressed in the future. Moreover, future work could be performed to compare the effect of HAND1 with other transcription factors (e.g., NFIB, NFIC and ZN563) which we explored from motif scan, and investigate whether they cooperate with HAND1 in regulating 3D genome organization in DCM.

An intriguing aspect is the possibility of using our discoveries for predicting the risk for developing DCM. A reactivation of HAND1 protein in DCM could indicate the early activation of signaling pathways involved in mesodermal-cardiac differentiation and HAND1 protein induction. Additionally, the unique HAND1-mediated H3K27ac looping of

genes is likely to play a central role in DCM. This could open novel therapeutic strategies for DCM. For example, signaling pathways blocking HAND1 protein levels (i.e., the E3 ligase Fbxo25 for HAND1¹⁵⁶) or chemical inhibitors that prevent HAND1 from reprogramming the enhancer-promoter connectomes leading to DCM could be attractive targets for therapeutic intervention. Our RNA-sequencing analysis of loop-specific genes uncovered several candidate pathways, such as EZH2 and TGF β , that could be studied further for their therapeutic potential in the hiPSC-CM model.

In summary, I generated a compendium of high-resolution 3D-epigenome maps (1.12×10^{10} sequencing reads) for human DCM, which are also the largest 3D genome datasets for human cardiovascular diseases. Moreover, I found that enhancer-promoter connectome is extensively rewired and contributes to the transcriptome reprogramming in human DCM. Furthermore, I uncovered that HAND1 acts on pre-accessible chromatin to drive the rewiring of enhancer-promoter connectome and induce DCM phenotypes. I believe that the 3D-chromatin datasets and research findings from this study will serve as a valuable resource for the biomedical community.

2.4 Materials and Methods

2.4.1 Data Availability

The data has been uploaded to GEO, the accession number for the data is GSE165303.

2.4.2 Culture of hiPSC-CMs and 293 cells

The human-induced pluripotent stem cell (hiPSC)-derived cardiomyocytes (hiPSC-CMs) obtained from Cellular Dynamics International were plated on fibronectin-coated dishes and maintained with maintenance medium according to the manufacturer's protocol. For adult mouse cardiomyocyte (CM) isolation, the hearts were digested with 15,000 U of type II collagenase (Worthington-Biochem) on a Langendorf apparatus. The 293 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Hyclone) supplemented with 10% FBS.

2.4.3 ChIP-seq on heart tissues and hiPSC-CMs

50 mg heart tissues were submersed in PBS with 1% (final concentration) formaldehyde (Thermo Fisher Scientific), cut into small pieces and incubated at room temperature for 15 min. Fixation was stopped by the addition of 0.125 M (final concentration) glycine. The tissue pieces were then treated with a TissueTearer and finally spun down and washed with PBS (twice). The chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. For hiPSC-CMs, 10 million cells were fixed by 1% formaldehyde at room temperature for 15 min, followed by cellular lysis and nuclear lysis. The lysates were sonicated and the DNA sheared to an average length of 300-500 bp. The pellets were resuspended and the resulting DNA was quantified on a Qubit spectrophotometer. An aliquot of chromatin (8 µg) was precleared with protein A/G agarose beads (Thermo Fisher Scientific). Genomic DNA regions of interest were precipitated using 4 µg of antibody against H3K27Ac (Active Motif) or 10 µg of antibody against HAND1 (Abclonal). For HAND1 ChIP-seq in hiPSC-CMs transfected with Ctrl or *HAND1*-overexpressing adenovirus, we employed Active Motif ChIP-seq spike in using *Drosophila* chromatin (Spike-in Chromatin, Active Motif) and an antibody against

Drosophila specific histone variant H2Av (Spike-in Antibody, Active Motif) to normal ChIP-seq signal, as per the manufacturer's instructions. The protein/DNA complexes were washed, eluted from the beads with 1% SDS buffer and subjected to RNase and proteinase K treatment overnight at 65°C. The ChIPed DNA was purified by phenol-chloroform extraction and ethanol precipitation. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K in a similar way as ChIPed DNA. Illumina sequencing libraries were prepared from the ChIP products by the standard consecutive enzymatic steps of end-repair, A tailing, and adaptor ligation using the KAPA Hyper Prep Kit (Roche). Following a final PCR amplification step, the PCR products were purified and the size was selected by 0.5X SPRI followed by 0.9X SPRI beads. The resulting DNA libraries were quantified and sequenced on an Illumina NextSeq 500 (75 bp, single end).

2.4.4 H3K27ac HiChIP (human heart tissues and hiPSC-CMs) and *in-situ* HiC (human heart tissues)

H3K27ac HiChIP/*in-situ* HiC Library preparation was performed by Arima Genomics (<https://arimagenomics.com/>). For heart tissues, each HiChIP and *in-situ* HiC library was generated from ~500 mg pulverized and crosslinked (2% formaldehyde, room temperature for 10 min) DCM and NF hearts. For hiPSC-CMs, each HiChIP library was generated from ~10 million crosslinked (2% formaldehyde, room temperature for 10 min) hiPSC-CMs transfected with Ctrl or *HAND1*-overexpressing adenovirus. For mouse-CMs, each HiChIP library was generated from cardiomyocytes isolated from adult mice transfected with the adenoviruses (Vector Biolabs) carrying cytomegalovirus (CMV)

promoter driving *Hand1* (Ad-GFP-*Hand1*, the cells are termed as mCM^{*Hand1*}) or with adenoviruses carrying CMV promoter (Ad-GFP, the cells are termed as mCM^{Null}). Initially, biotin-labeled proximally-ligated chromatin was prepared using the Arima-HiC kit (Arima Genomics). Briefly, the crosslinked chromatin was initially digested using a restriction enzyme (RE) cocktail. The 5'-overhangs were subsequently filled in and labeled with a biotinylated nucleotide. Subsequently, the spatially proximal digested ends of DNA were ligated. The resulting proximally-ligated chromatin was fragmented using the Diagenode Bioruptor Pico instrument, bound to an H3K27ac antibody (Abcam) overnight and immunoprecipitated on Protein A magnetic beads (Thermo Fisher Scientific). The immunoprecipitated chromatin was reverse crosslinked and purified using the Arima-HiC kit. Subsequently, the Illumina-compatible sequencing libraries were prepared by initial enrichment for biotin using C1 streptavidin beads (Thermo Fisher Scientific). The samples were tagged on-bead utilizing Tagment DNA Enzyme and Buffer (Illumina). Subsequently, DNA was PCR amplified utilizing NPM and the indexing primers from the Nextera XT DNA Library Prep kit (Illumina). The PCR products were purified and size selected by 0.5X SPRI beads followed by 0.7X SPRI beads. The purified libraries underwent standard QC (qPCR and Bioanalyzer) and were sequenced on the Illumina HiSeq Xten (150 bp, pair-end) following the manufacturer's protocols. *In situ* HiC was performed in a similar way as HiChIP with the exception of chromatin immunoprecipitation being performed using the H3K27ac antibody and protein A beads.

2.4.5 RNA-seq on human heart tissues, hiPSC-CMs and adult mouse CMs

Total RNA was extracted from the heart tissues, hiPSC-CMs, and mouse CMs using Trizol (Thermo Fisher Scientific) according to the manufacturer's instructions. Approximately 20 mg of tissue was converted into powder using liquid nitrogen and the powder samples were transferred into 2 ml Eppendorf tubes containing 1.5 ml Trizol reagent. The mix was centrifuged at 12,000 x g for 5 min at 4°C. The supernatant was transferred to a new 2.0 ml tube, which contained 0.3 ml Chloroform/isoamyl alcohol (24:1) per 1.5 ml of Trizol reagent. Following centrifugation at 12,000 x g for 10 min at 4°C, the aqueous phase was transferred to a new 1.5 ml tube which contained equal volume of supernatant of isopropyl alcohol. The mix was centrifuged at 12,000 x g for 20 min at 4°C and the supernatant was collected. Following washing with 1 ml 75% ethanol, the RNA pellet was air-dried in the biosafety cabinet and dissolved by addition of 25 µl nuclease-free water. Subsequently, total RNA was qualified and quantified using the Agilent 2100 bioanalyzer. Approximately 1 µg total RNA per sample was depleted from rRNA with the Ribo-Zero™ Magnetic Kit (Epicentre). The retrieved RNA was subjected to fragmentation by adding the First Strand Master Mix (Thermo Fisher Scientific). First-strand cDNA was generated using random primers and reverse transcription, followed by a second-strand cDNA synthesis. The synthesized cDNA was subjected to end-repair and was 3' adenylated. The adapters were ligated to the ends of these 3' adenylated cDNA fragments. A total of 10 rounds of PCR with PCR Primer Cocktail and PCR Master Mix were performed to amplify the cDNA fragments. The PCR products were purified and size selected by 0.5X SPRI beads followed by 0.9X SPRI beads. The identity of the final library was verified by the distribution of the fragments size using the Agilent 2100 bioanalyzer

and quantified by the Qubit High sensitivity assay. The Qualified libraries were sequenced pair end on the Illumina HiSeq Xten (150 bp, pair-end).

2.4.6 ATAC-seq on human heart tissues

A total of 20-50 mg human heart tissue sample was cut on dry ice for each ATAC-seq assay. Each sample was placed into the corresponding 5 cm petri dish (labeled on bottom of dish) with PBS on ice and minced as finely as possible with a razor (the use of a clean razor was ensured for each sample). The end was cut off with a 1 ml pipette tip at an angle and the samples were transferred to the corresponding conical tubes with PBS (a fresh tip was used for each sample). Subsequently, the tubes were centrifuged at 500 x g for 5 min at 4°C. The PBS was decanted as thoroughly as possible without losing the pellet and 1 ml ATAC lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630) was added to each sample. Each sample was added to a 1 ml dounce. The samples were dounced with a pestle 40-50 times slowly in order to avoid foaming. Following douncing, the samples were filtered through a 40 µM mesh strainer into a 1.5 ml tube on ice. Immediately following filtration of each sample, 10 µl aliquots were obtained for counting the nuclei. For each 10 µl of aliquot, 10 µl Trypan Blue was added, mixed, loaded onto the hemocytometer and counted. The tubes were inverted and flicked to mix. Subsequently, 100,000 cells were aliquoted from each sample into a PCR tube and stripped on ice. The ATAC lysis buffer was added to the PCR tube to ensure that the reaction volume was not less than 100 µl. The samples were spun down at 500 x g for 5 min at 4°C. The supernatant was removed very carefully, ensuring not to disturb the cell pellet. The tagmentation reaction was immediately continued. The transposition

reaction master mix was prepared and the samples were mixed by pipetting gently, spun quickly and left on ice. The tagmentation master mix included the following reagents: 25 μ l 2x TD Buffer (Illumina), 8 μ l Tn5 Transposase (Illumina), 17 μ l PBS, 5 μ l 1% Digitonin (dissolved in 50% DMSO in dH₂O) and 5 μ l 10% Tween 20. A total of 50 μ l tagmentation master mix was added to each sample, which was gently pipetted to resuspend the cells in the transposition reaction mix. The samples were incubated at 37°C for 30 min in a thermomixer at 800 rpm. The purification was conducted using a Qiagen MinElute Kit. The DNA was eluted with a 35 μ l EB buffer (10 mM Tris buffer, pH 8.0), allowed to settle at room temperature for 1 min and centrifuged at max speed for 30 sec. Each elution was transferred to a 250 μ l PCR tube and the PCR reactions were prepared by adding the components in the order shown. The PCR reaction contained the following components: 5 μ l Index mix (Illumina Nextera Kit), 10 μ l 5x Phusion HF reaction Buffer, 1.5 μ l dNTPs (10 mM), 0.5 μ l Phusion HF Polymerase (NEB) and 33 μ l tagmented DNA. The reagents were pipetted to mix. The PCR was performed using the following program on a thermal cycler (with heated lid): heating at 72°C for 5 min and at 98°C for 30 sec. A total of 10 cycles were performed at 98°C for 10 sec, at 63°C for 30 sec and at 72°C for 30 sec. A final extension was performed at 72°C for 1 min. The AMPure double size selection was performed and 80 μ l AMPure beads were added to the sample, vortexed briefly to mix and quickly spun. The solution was incubated at room temperature for 5 min. The samples were placed in a magnetic rack and left for 1 min. The liquid was removed with a pipette, taking care not to disrupt the pellet. The samples were washed by addition of 180 μ l fresh 70% ethanol to the tube, taking care not to disrupt the pellet. The ethanol was removed with a pipette and the procedure was repeated a second time. The ethanol was removed

slowly from the tube. The samples were allowed to dry for 30 sec and the pellet was resuspended with 22 µl Buffer EB (10 mM Tris, pH 8.0) and mixed by pipetting. The sample was allowed to stand for 1 min and placed in a magnetic rack on the magnetic side for 1 min. A total of 20 µl of the eluate were transferred to a new tube, taking care not to collect any magnetic beads. The aliquots of the libraries were run on an Agilent 2100 bioanalyzer and quantified with the Qubit High sensitivity assay. The libraries were sequenced by the Illumina HiSeq Xten (150 bp, pair-end).

2.4.7 Library QC metrics collection

All the QC metrics for ChIP-Seq, HiChIP, in-situ Hi-C, RNA-Seq and ATAC-Seq were collected using MultiQC¹⁵⁷. For HiChIP and Hi-C data, the raw read number, mappability, pairing statistics, filtering and contact statistics were documented. For the ATAC-seq and ChIP-seq data, the raw read number, mappability, duplicated reads and peak number were documented. For the RNA-seq data, the raw read number, average fragment length and mappability were collected.

2.4.8 ChIP-seq data analysis

Single-end reads were mapped to the hg19 reference genome using the bwa¹⁵⁸ mem. The duplicated reads were removed using the Picard Toolkit (<https://broadinstitute.github.io/picard/>), whereas the reads with a quality lower than 20 were filtered.

The ChIP-seq peaks were called using the default (narrow) setting in MACS2¹⁵⁹. The promoter regions were defined as +/- 2.5Kb around the TSSs regions. H3K27ac peaks corresponding to overlapping promoter regions were defined as “proximal promoter peaks” (P), whereas the others were defined as “distal enhancer peaks” (E). The peaks from 10 DCM H3K27ac ChIP-seq and 10 NF H3K27ac ChIP-seq datasets were loaded in R using the DiffBind package (<https://bioconductor.org/packages/release/bioc/html/DiffBind.html>). The overlapping peaks were merged and a single binding matrix was created. Differential peak analysis was performed using the edgeR¹⁶⁰ with a cutoff FDR < 0.05 and Fold change > 1.

For spike-in HAND1 ChIP-seq from hiPSC-CMs, raw sequencing reads were mapped to human hg19 and drosophila dm6 genome, respectively. Peaks were called using reads mapped to human hg19 genome. Normalization factors for each sample were calculated based on the reads mapped to dm6 genome. Reads mapped to human hg19 genome were then down-sampled based on the normalization factors. For each peak, number of reads were counted using down-sampled bam files and used as intensity for visualization.

2.4.9 Enhancer breadth analysis

H3K27ac binding sites (narrowPeak format) in 507 ENCODE cell lines / tissues were downloaded from ENCODE portal (<https://www.encodeproject.org>). H3K27ac enrichment breadth for each DCM-specific enhancer was calculated as the number of cell lines / tissues an enhancer peak overlapped. Common enhancer peaks were used as a

comparison. H3K27ac enrichment breadth were computed for all DCM-specific and common enhancers and visualized using density plot.

2.4.10 RNA-seq data analysis

The reads were mapped to hg19 using STAR¹⁶¹. Only properly mapped reads pairs were kept. Bam files were indexed and sorted by read names using samtools¹⁶². The reads mapped to gene exon regions were counted using featureCounts¹⁶³. Differential expressed genes were determined by DESeq2¹⁶⁴ using negative binomial generalized log-linear model followed by Wald Test. Age and gender were added to the model as blocking factors. Normalized read counts were calculated using variance stabilizing transformation (VST) method and then used for visualization in boxplot.

2.4.11 HiChIP and Hi-C data processing

HiChIP and Hi-C data were preprocessed using HiC-Pro¹⁶⁵ with default parameters with the exception of the LIGATION_SITE being set to GATCGATC, GANTGATC, GANTANTC, GATCANTC and the GENOME_FRAGMENT that was generated using digest_genome.py in HiC-Pro utilities. The following parameters were used: -r ^GATC G ^ANTC -o hg19_restriction_sites.bed (coordinates of restriction sites) hg19.fasta (reference sequence file). The raw reads were mapped to the hg19 reference genome and were assigned to restriction fragments. The raw reads were filtered for valid pairs (VI). The HiC files were generated using hicprotojuiebox.py utilities in HiC-Pro¹⁶⁵. All valid pairs served as input files and then imported into Juicebox¹⁶⁶ and Knight–Ruiz (KR) matrix–balanced interaction maps were generated for visualization. For HiChIP data, 1D

coverage files for the visualization in bigwig format were converted from bam format using deepTools¹⁶⁷ and RPKM (Reads Per Kilobase per Million reads to normalize the number of reads per bin) normalization was used.

2.4.12 Identification of chromatin interactions and differential loops

We first identified a set of significant loops for DCM and NF, respectively. Valid read pairs for each sample served as input to Mango¹⁶⁸ for stage 4 and 5 processing. The “peakinput” parameter was set to the corresponding ChIP-seq peak file. The read pairs were extended 500 bp on each side and subsequently clustered. The clusters in each sample were merged and clustered again to form a union set of clusters for DCM and NF, respectively. Significant loops were identified using the ChIA-PET Tool V3.0¹⁶⁹ with cutoff $FDR < 0.00001$. The loops with PET counts lower than 10 and spanning less than 5 Kb or larger than 1 Mb were further filtered.

Differential loops were identified following two steps: Initially, loops in DCM and NF were merged and subsequently the PET counts for each loop were counted in all the replicates in DCM and NF. A negative binomial model was fitted using glmFit function in edgeR R package and the loops with p value (< 0.1) were determined as differential loops³¹. We acknowledged that while using p value without correction for multiple testing will identify more differential loops, it could potentially generate more false positivity. We further employed 3C-qPCR to validate the fidelity of some differential loops we identified.

2.4.13 Validation of differential loops by 3C-qPCR

The 3C-qPCR assay was developed and modified from the previous publications^{170,171}. Briefly, human heart tissues were fixed with 2% formaldehyde for 10 min at room temperature and lysed, and then nuclei were subjected to in-situ digestion (Mbol, New England Biolabs) and ligation (T4 ligase, New England Biolabs). The 3C-qPCR primers were designed using unidirectional strategy. Briefly, the primers were designed on the “forward” strand, located near the Mbol restriction site (GATC) from 20~100 bp. Loading control (GAPDH) primers were designed on the Mbol fragment which does not across Mbol restriction site. The primers for 3C-qPCR are as follows: *TBX5* chr12:114819505-114820005: CCAATTCAGGCCAAAGAAGA, *TBX5* chr12:114842979-114843479: CTCCACCCCCTACTCATCAC; *MEF2D* chr1:156469903-156470403: GCAGGCGCATTATTTTTTCTC; *MEF2D* chr1:156478626-156479126; *SMYD2* chr1:214492375-214492875: TGATGGCATCTTCCTGACAA; *SMYD2* chr1:214449791-214450291: CATCTAGGCCCTTTGCTCAC; *MYPN* chr10:69847422-69847922: TAAAGCCATGTGGGAAAAGG; *MYPN* chr10:69869501-69870001: AATTACATATGACGGGACACGA; *GAPDH*-Forward: CCCACACACATGCACTTACC, *GAPDH*-Reverse: CCCACCCCTTCTCTAAGTCC.

2.4.14 Luciferase reporter gene assays

To explore the potential regulatory relationship between interacting chromatin loci, the dual luciferase assays for various combinations of promoter and enhancer regions were performed. The fragments of promoter and enhancer regions identified by HiChIP were PCR-amplified and cloned into pGL4.10 vector (Promega). For luciferase assay, cells were co-transfected with the reporter constructs and a constitutive *Renilla* luciferase

expressing vector pRL-SV40 (controlling for transfection efficiency) (Promega). Cells were harvested 48 hours after transfection and the luciferase activity was measured using the Dual-Luciferase reporter assay kit (Promega, E1910). For all measurements, firefly luciferase values were normalized to *Renilla* luciferase values.

2.4.15 Functional enrichment of specific active enhancers or differential loop anchors

The bed file of specific enhancers or differential loop anchor regions served as input of GREAT (version 4) (<http://great.stanford.edu/public/html/>). We used whole genome (hg19) as background. The Negative Log₁₀(score) for each GO (gene ontology) terms or Disease Ontology terms was presented as the bar. Functional enrichment of genes network was generated using the ToppFun application of the Toppgene Suite using a corrected (Benjamini and Hochberg) p-value cut-off of 0.05 (<https://toppgene.cchmc.org/>)¹³⁵. Cytoscape application (<https://cytoscape.org/>)¹⁷² was used to generate the functional enrichment network.

2.4.16 Identification of A/B compartments and topologically associated domains (TADs)

Interaction matrix of .hic format were converted to the cool format using HiCExplorer library hic2cool. Replicates of cool format files with 25 Kb resolution were merged using cooler¹⁷³. A and B compartments were identified using hicPCA command in HiCExplorer (<https://hicexplorer.readthedocs.io/en/latest/>). The topological associated domains (TADs) were identified using the hicFindTADs command in the HiCExplorer at 25 Kb resolution.

2.4.17 Differential TAD analysis

We used two methods to assess the conservation of genome organization at TAD level: 1) To define conserved TADs, we require that 70% fraction of reciprocal overlapping TADs in DCM and NF; 2) In addition, we used an alternative metrics, the correlation of insulation score to assess the overall TAD boundary conservation between DCM and NF, which has been used as cutoff to define TAD boundaries in multiple studies¹⁷⁴⁻¹⁷⁶.

2.4.18 ATAC-seq data analysis

The adapters were removed and the filtered reads were mapped to hg19 using bwa¹⁵⁸ mem. The duplicated reads were removed using Picard (<https://broadinstitute.github.io/picard/>) and the reads with quality lower than 20 were filtered.

2.4.19 Transcription factor motif analysis

The motif analysis was performed using Homer (<http://homer.ucsd.edu/homer/>, v4.11). We overlapped DCM-specific/NF-specific loop anchor regions with the ATAC-seq peak regions. Subsequently, we extracted the sequences from the overlapped regions using bedtools¹⁷⁷ getfasta and imported them into Homer findMotifs.pl with default setting for motif enrichment analysis.

2.4.20 Quantitative analysis of ATAC-seq and ChIP-seq

To calculate the differential chromatin accessibility and HAND1 binding on DCM-specific/NF-specific loop anchors, we generated aggregation plots. The reads were counted in a 6Kb window centered at an anchor with 10bp-bin resolution for the 10 DCM and 10 NF samples. The read counts for all anchors at each bin in each sample were summarized using the median value. Subsequently, the read counts in each bin across the samples were summarized using mean $\pm$ standard error. The Mann–Whitney U test was used to calculate the p-value.

2.4.21 Sample variation analysis for sequencing data

For all 1D sequencing data type (RNA-seq, ChIP-seq, ATAC-seq), we performed Principal Component Analysis (PCA) using genes or peaks with the highest variance across samples (top 2000). For 3D sequencing data type (HiChIP, Hi-C), we calculated stratum adjusted correlation coefficient (SCC) score across all chromosomes using HiCRep to quantify the similarity between interaction matrices in 25Kb resolution¹⁷⁸.

2.4.22 Multiple test correction

All p-values from 1D high-throughput sequencing data sets were adjusted using Benjamini-Hochberg Procedure. Total number of tests of each dataset are: Human heart tissue RNA-seq (N = 50,762), Human heart tissue H3K27ac ChIP-seq (N = 86,530) and Mouse cardiomyocytes RNA-seq (N = 39,905).

2.4.23 Visualization

The chromatin loops were prepared using bedpe format and then uploaded to the IGV¹⁷⁹ for arc-style visualization. Other 1D coverage signals were prepared using bigwig format and uploaded to IGV. The hic files were imported into Juicebox for heatmap visualization.

2.4.24 Statistical analysis for sequencing data

Statistical analysis was performed with R version 3.5.1 or Prism 7.0 (GraphPad Inc). The statistical tests performed for each statistical graph were listed in the individual figure legends. Hypergeometric test implemented in ChIA-PET Tool V3.0 was used to detect significant chromatin interactions in HiChIP data. Negative binomial generalized log-linear model was used to model read counts in sequencing data and quasi-likelihood (QL) F-test or Wald Test was used to identify differentially abundant features (chromatin interactions, expressed genes, transcription factor binding peaks, etc). HiCRep was used to calculate stratum adjusted correlation coefficient (SCC score) of HiChIP and Hi-C data.

2.4.25 Plasmid production

Mouse *Hand1* cDNA was cloned into adeno-associated virus serotype 9 (AAV9) expression vector plasmid (Addgene) flanked by inverted terminal repeats (ITRs), which expressed green fluorescent protein (GFP) driven by the cardiomyocyte-specific troponin T (*cTNT*) promoter (AAV9-*cTNT*-GFP). The AAV9 carrying *cTNT* promoter driving GFP was used as control.

2.4.26 AAV9 production and purification

AAV9 particles were packaged and produced by co-transfecting HEK293 cells (ATCC) with AAV9 expression plasmid, in the presence of other helper-plasmid DNAs. In brief, one day prior to transfection, HEK293 cells (80% confluence) were used. The cells were incubated for 2 days after the transfection and the cell pellets were harvested. The viruses were released by 3x cycles of freeze/thaw (-80°C for 10 min, 37°C for 20 min) in resuspension buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM MgCl_2)¹⁸⁰. The viruses were purified using CsCl-gradient ultra-centrifugation, followed by desalting. AAV9 was stored at -80°C . For titer quantification, purified virus was digested with DNase I (Sigma Aldrich) to remove unencapsidated DNA, followed by proteinase K digestion to release the DNA template from the AAV9 capsid. The viral titers (genome copies/ml) were determined by real-time PCR.

2.4.27 AAV9 delivery into the heart

Intraperitoneal delivery of AAV9 was performed accordingly as described¹⁸⁰. In brief, 8-12 weeks old wild type (WT) C57BL/6 mice (Jackson Laboratory) were anesthetized using light isoflurane and were prepared in a biosafety cabinet. To determine the effects of AAV9 dosage on cardiac phenotype, wild type mice were infected with control AAV9 virus (serial dilution) lacking *Hand1* (Null) and diluted in 50 μl of saline. We defined the mice receiving 5×10^{11} viral genomes (vg), 1×10^{12} vg, and 5×10^{12} vg as *Hand1_Low*, *Hand1_Mid*, and *Hand1_High*, respectively. The control mice received 50 μl of saline with AAV-Ctrl (1×10^{12} viral genomes). To determine the toxicity of AAV9, another group of WT mice was used that received 50 μl of saline without virus. The animals used after the procedure were kept warm using a heat lamp for recovery.

2.4.28 Adult mouse CM and fibroblast isolation

Adult mouse CMs were isolated from adult WT C57BL/6 mice with a modified protocol as described previously¹⁸¹. Briefly, the aorta of the excised heart tissue was clamped and mounted on a Langendorf apparatus. The tissue was perfused with calcium-free perfusion buffer (113 mM NaCl, 4.7 mM KCl, 0.6 mM KH₂PO₄, 0.6 mM Na₂HPO₄, 1.2 mM MgSO₄, 10 mM NaOH-HEPES, 12 mM NaHCO₃, 10 mM KHCO₃, 0.032 mM phenol red, 30 mM taurine, 10 mM 2,3-butanedione monoxime (BDM), and 5.5 mM glucose; pH-7.4.), followed by digestion with 50 ml perfusion buffer containing 15,000 U of type II collagenase (Worthington-Biochem) and 50 μ M CaCl₂. The digestion was terminated when the heart was slightly pale and flaccid. The heart was gently teased into small pieces with forceps and triturated with a Pasteur pipette to dissociate individual CM in the presence of 5 ml digestion buffer, followed by addition of 5 ml stop buffer (perfusion buffer, 10% FBS, and 12.5 μ M CaCl₂) to inhibit the enzyme activity. After passing through a 100- μ m filter, the cells were settled by five sequential rounds of gravity in the calcium solution. Then, the cell pellets in the bottom were enriched of highly purified CM fraction, and the supernatant fraction was enriched of non-CMs. To purify the cardiac fibroblasts, magnetic-activated cell sorting approach (MACS) using anti-Thy-1 antibody (Abcam, ab3105, 1: 100) [PMID: 27502479] was performed on the cell pellets collected from non-CM fraction after centrifuge (500 g, 5 mins). The pellets containing the CM fraction and non-CM fraction were subjected to downstream analysis.

2.4.29 Measurement of contractility and calcium transients in adult mouse CMs

Mouse CM (mCM) contractility and intracellular calcium transients in response to electrical field stimulation were recorded simultaneously using IonOptix MyoCam System (IonOptix Corp) at room temperature¹⁸². Freshly isolated mCMs were loaded 1 ml Tyrode's solution (1.8mM Ca^{2+}) with 2 μl Fura-2 Am (Stock solution, 1ug/ μl in DMSO, Thermo Fisher Scientific), the Ca^{2+} indicator dye in dark at 37°C with 5% CO_2 . After 10 mins, the medium was aspirated and the cells were washed twice with Tyrode's solution (10 mins/time). Then, these Fura-2 loaded cells were placed in a perfusion chamber mounted on the stage of an inverted microscope (Nikon eclipse TE2000-U fluorescence microscope) and superfused with Tyrode's solution. Sarcomere shortening was recorded using a video-based sarcomere length detection system. Ca^{2+} -bound-Fura-2 fluorescence was collected at an emission wavelength of 515 ± 10 nm that was excited at 340 and 380 nm using a spectrofluorometer (IonOptix). Data were analyzed by IonOptix LLC analyzing software.

2.4.30 hiPSC-CM culture and transduction

Functional hiPSC-CMs were obtained from Cellular Dynamics International and cultured according to the manufacturer's protocol. Briefly, hiPSC-CMs were thawed and plated on fibronectin-coated dishes. Following 4 h of incubation, the plating medium was aspirated and replaced with maintenance medium. The CMs were maintained for 7 days until the experiment, and the maintenance medium was replaced every 48 h. On day 7 post-plating, 4-6 hours prior to the infection, the old medium was aspirated and replaced with fresh maintenance medium. During the adenoviral transduction, the completed medium was removed and replaced with the medium with pre-mixed virus/basic medium (without growth factors and FBS, MOI=100) and incubated for 30 min at standard cell culture conditions. The cell culture was replaced with virus/basic medium and with fresh maintenance medium. HiPSC-CMs were transduced with adenoviruses (Vector Biolabs)

carrying cytomegalovirus (CMV) promoter driving *HAND1* (Ad-GFP-*HAND1*, the cells are termed as hiPSC-CM^{*HAND1*}) or with adenoviruses carrying CMV promoter (Ad-GFP, the cells are termed as hiPSC-CM^{Null}). The cells were harvested for subsequent experiments 72 h post transduction.

2.4.31 Measurement of contractility and calcium transients in hiPSC-CMs

HiPSC-CM contractile function was analyzed after seeding the cells on fibronectin-coated (Sigma Aldrich, 5 µg) cover-slips with a modified protocol using IonOptix MyoCam System (IonOptix Corp)^{183,184}. Ca²⁺ transient was measured by Fluo-4-AM (Thermo Fisher Scientific) with a modified protocol as described previously^{184, 150}. The dye was dissolved in 20 µl DMSO (Sigma Aldrich) and added to 1 ml of cell culture medium. The myocytes were incubated in 5 µM Fluo-4-AM for 20 min at 37°C and washed three times with Tyrode solution containing 1.8 mM CaCl₂ at room temperature prior to Ca²⁺ transient measurements. Ca²⁺ transient recording required myocytes that were paced with field stimulation (Grass S48 stimulator; Grass Instruments) with 2-ms (1.5×) threshold pulses at a rate of 1 Hz at room temperature. Fluorescence signals from Fluo-4-AM-loaded myocytes were imaged for Ca²⁺ transients using an inverted confocal microscope (Zeiss LSM 710). Image processing and data analysis were performed using IDL 6.3 software (Exelis Visual Information Solutions).

2.4.32 Western blot analysis

Human or mouse heart samples (50 mg) were lysed with ice-cold cell lysis buffer (Cell Signaling Technology, 9803S) in the presence of protease inhibitor (Sigma Aldrich).

Following homogenization with the Bullet Blender Storm homogenizer (MIDSCI), the samples were centrifuged at 12,000 rpm for 20 min at 4°C. The supernatant was carefully transferred into a new eppendorf tube and kept on ice. Following determination of the protein concentration by the Bradford assay (Bio-Rad), the protein samples (40 µg) were resolved in 10x NuPAGE reducing agent in the presence of 4x LDS sample buffer and denatured at 70°C for 10 min prior to loading on the NuPAGE™ 4-12% Bis-Tris Protein Gels (Thermo Fisher Scientific). Following electrophoresis, the proteins were transferred from the gel to nitrocellulose membranes using semi-dry transfer (Power Blotter Station, Thermo Fisher Scientific). Equal loading and transfer of proteins was confirmed by the quantitative Ponceau red staining. The membranes were incubated for 60 min with 5% dry milk and Tris-buffered saline was used to block the nonspecific binding sites. The membranes were immunoblotted overnight at 4 °C with antibodies against HAND1 (eHAND) (F-7) (Santa Cruz, SC-390376, 1:500 dilution), GFP (Abcam, ab32146, 1:2000), Histone 3 (H3) (Cell Signaling Technology, 9715S, 1: 2000 dilution) or GAPDH (Sigma Aldrich, G9545, 1: 5000 dilution) on a rocking platform. Following washing for three times for 5 min each with Tris-buffered saline, the membranes were incubated for 60 min with an HRP-conjugated secondary antibody, washed three times with Tris-buffered saline and developed with the SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific).

2.4.33 Quantitative real-time PCR (qPCR)

The RNA samples (1.0 µg) were reverse transcribed to cDNA using SuperScript™ the First-Strand Synthesis System for RT-PCR kit (Thermo Fisher Scientific). qPCR was

performed on a CFX96 Real-Time PCR system (Bio-Rad) using Power SYBR Green Master Mix (Thermo Fisher Scientific). In each analysis, a 0.5 µl cDNA pool was used for qPCR. The fold changes of each target gene expression were compared with those of GAPDH under experimental and control conditions and were calculated based on the threshold cycle (CT) as follows: $r = 2^{-\Delta(\Delta CT)}$, where $\Delta CT = CT(\text{target}) - CT(\text{GAPDH})$ and $\Delta(\Delta CT) = \Delta CT(\text{experimental}) - \Delta CT(\text{control})$. The primers for qPCR are as follows: *Nppa*-Forward: 5' TGCAGCTTCCTGTCAACACT 3', *Nppa*-Reverse: 5' GAGGCGAGGAAGTCACCATC 3'; *Nppb*-Forward: 5' TGGAAACGTCCGGGTACAG 3' *Nppb*-Reverse: 5' CTTCCAGACACCTGTGGGAC 3'; *Myh7*-Forward: 5' AGACACACTTGAGTAGCCCAG 3'; *Myh7*-Reverse: 5' CCACCCTCTCGAGACACGAT 3'; *Gapdh*-Forward: 5' GTCAAGGCTGAGAACGGGAA 3'; *Gapdh*-Reverse: 5' AAATGAGCCCCAGCCTTCTC 3'.

2.4.34 Analysis of cardiac fibrotic area

Mouse heart tissues were dissected, rinsed with PBS and fixed in 10 % formalin for 48 h. Following dehydration through a series of ethanol solutions, the tissues were embedded in paraffin wax according to standard laboratory procedures. Subsequently, they were sectioned (5 µm thick) using a microtome. Following deparaffinization and rehydration, Masson's trichrome staining (Sigma Aldrich) was performed on the heart sections to quantify the fibrotic area. In brief, the sections were fixed with prewarmed Bouin's solution at 56 to 60 °C for 1 h and stained in Weigert's iron hematoxylin solution for 10 min. Following rinsing in tape water, the sections were stained in Biebrich scarlet-acid fuchsin solution for 10 min and further rinsed in distilled water. The sections were differentiated

in phosphomolybdic-phosphotungstic acid solution for 10 to 15 min. Once the red color was removed from the collagen-rich areas, the sections were directly transferred to an aniline blue solution and stained for 5 to 10 min. The sections were rinsed in distilled water and differentiated in 1% acetic acid solution for 2 to 5 min. Following rinsing in distilled water, the sections were dehydrated and cleared with xylene. The images were examined with an Olympus BX41 microscope equipped with a CCD (Magna-Fire™) camera and the fibrotic area was quantified using Image J software.

2.4.35 Wheat Germ Agglutinin Staining and Quantify Cell Size

Following deparaffinization and rehydration, the slides were boiled in citrate buffer (Sigma Aldrich) for 20 min for antigen retrieval. Following cooling down for 30 min in distilled water, the slides were permeabilized with 0.2% Triton, washed with PBS 3 times and incubated with wheat germ agglutinin (WGA) conjugated to Texas Red™-X (Thermo Fisher Scientific) for 1 h at room temperature. Following 3 times washing with PBS, the slides were mounted in antifade mounting medium (Thermo Fisher Scientific). Four fields of each section were examined for quantification using the Image J software. Fluorescent imaging was performed with an Olympus BX41 microscope equipped with an epifluorescence attachment.

2.4.36 Echocardiography

Transthoracic echocardiography (Visual Sonics Vevo 2100) was performed with a 40-MHz probe. The heart rate and LV dimensions, including diastolic and systolic wall thicknesses and LV end-diastolic and end-systolic chamber dimensions were measured

under M-mode tracings at the level of the papillary muscle. LV end-diastolic (LVDd) and end-systolic diameters (LVDs) were measured from M-mode recordings. LV EF was calculated as follows: $EF\% = [(LVDd)^3 - (LVDs)^3] / (LVDd)^3 \times 100$. LV FS was determined as follows: $FS\% = [(LVDd - LVDs) / LVDd] \times 100$.

2.4.37 Animal studies

All research protocols conformed to the Guidelines for the Care and Use of Laboratory Animals published by the National Institutes of Health (National Academies Press, eighth edition, 2011). All animal use protocols and experiments were performed according to the guidelines approved by the Institutional Animal Care and Use Committee at Guangdong Provincial People's Hospital, South China University of Technology, and University of Cincinnati.

2.4.38 Human studies

Acquisition of 101 human heart samples was approved by the Institutional Review Board (IRB) for the protection of human subjects at Duke University in US. De-identified frozen human heart samples with pathological characterization were provided by the Duke Human Heart Repository (DHHR), Department of Surgery at Duke University School of Medicine. In brief, the left ventricular (LV) heart samples were collected from DCM patients during surgery and the LV tissues were immediately dissected and snap frozen in liquid nitrogen. The LV samples from non-failing (NF) donor hearts were served as controls.

Chapter 3

3D genome atlas in pancreatic cancer stem cells

3.1 Introduction

Pancreatic cancer is one of the deadliest human malignancies. With an overall median survival time of 6-9 months and a similarly dismal 5-year survival rate of 6%¹⁸⁵, Pancreatic cancer is currently the fourth leading cause of cancer-related deaths in the US and Europe¹⁸⁶. Furthermore, with the increasing incidence of risk factors including obesity and metabolic syndrome, pancreatic cancer is projected to overtake colorectal, breast and prostate cancer and to become the second leading cause of worldwide cancer-related deaths by 2030¹⁸⁵.

The development of many cancers including pancreatic cancer involves a dedifferentiation of cellular identity with the acquisition of a stem cell-like state in a subpopulation of cancer cells. The arising cancer stem cells (CSCs) are exceptionally important because their developmental plasticity allows them to resist conventional therapies, metastasize and give rise to new tumours. The changes in cell identity are caused by transcriptional dysregulation which is a universal feature of tumorigenesis and impacts all cancer hallmarks¹⁸⁷. It is increasingly evident that spatiotemporal changes in 3D chromatin architecture have a central function in governing gene transcription and thereby cancer development and cellular heterogeneity.

From Christine Iacobuzio-Donahue Lab (Memorial Sloan Kettering Cancer Center, USA), our lab obtained a pair of adherent human PDAC cell lines (A13A and A13D), which were derived from primary and lung metastatic tumors in a PDAC patient respectively. My lab

colleague Dr. Stefania Milioti isolated the cancer stem cells organoid from these cell lines through CD133⁺ magnetic beads selection and 3D culture.

Based on this model, I generated the first compendium of 3D epigenomic datasets in human PDAC to uncover the novel mechanisms that regulate this disease at multiple levels: transcriptome, epigenome and 3D chromatin connectome. For this I combined cutting-edge techniques such as ChIP-seq (H3K27ac, H3K4me3) for mapping active *cis*-regulatory elements (CREs), HiChIP (H3K27ac) for identifying active CRE connectomes, CTCT *in-situ* ChIA-PET for mapping chromatin topological structures, ATAC-seq for mapping chromatin accessibility, as well as RNA-seq for linking the topological regulation of chromatin to transcription output (**Figure 1**). The sequencing depth for each library is shown in **Table 1** (~1600 million for HiChIP (in 8 libraries), ~1600 million reads for *in-situ* ChIA-PET(in 8 libraries), ~160 million reads for H3K27ac ChIP-seq(in 8 libraries), ~160 million reads for H3K4me3 ChIP-seq (in 8 libraries), ~240 million reads for ATAC-seq (in 12 libraries), ~240 million reads for RNA-seq (in 12 libraries)).

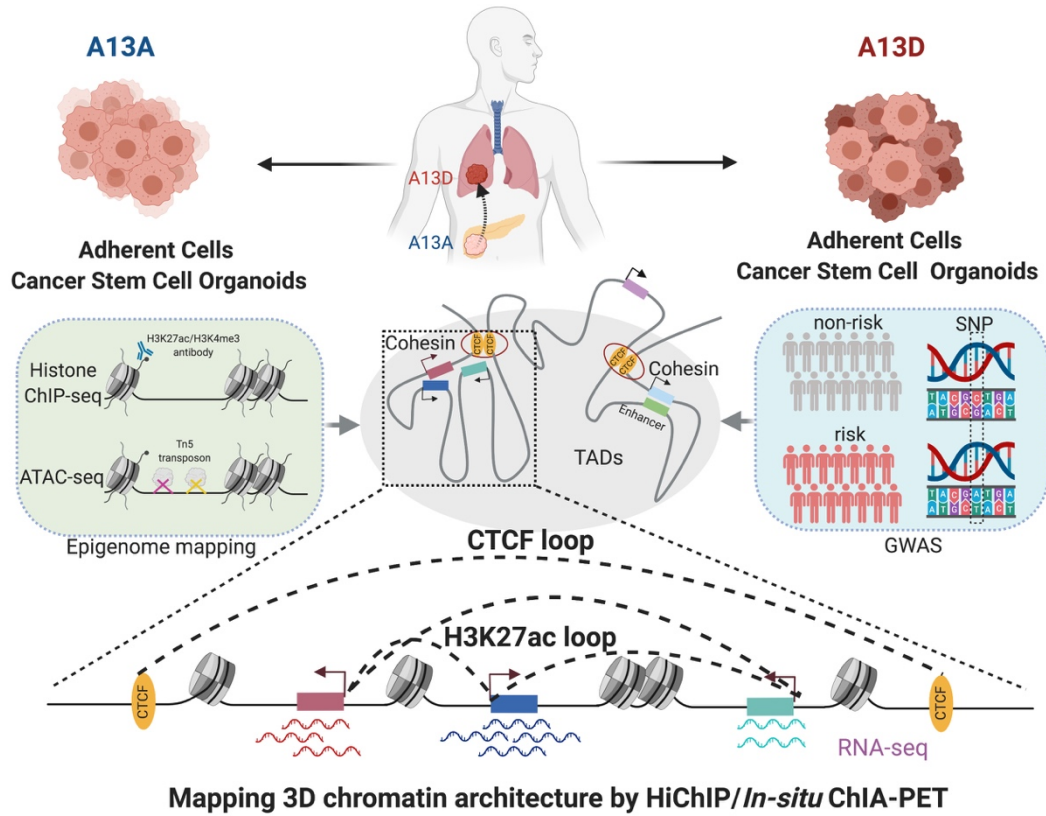


Figure 1. 3D Genome Mapping by H3K27ac HiChIP and CTCF in-situ Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) in A13A (primary pancreatic adenocarcinoma) and A13D (lung metastasis) cell lines derived from the same patient. ChIP-seq for active histone marks (H3K27ac, H3K4me3), ATAC-seq, RNA-seq and genome-wide association study (GWAS) SNPs are integrated with H3K27ac HiChIP and CTCF in-situ ChIA-PET data to reveal H3K27ac-associated chromatin interactions (enhancer–promoter interactions, promoter–promoter interactions) and their topological basis (e.g. topologically associating domains (TADs), and how genetic variations modulate promoter/enhancer activity, thus affecting enhancer–promoter interactions and leading to aberrant transcription in human PDAC metastasis. Figure adapted with permission from Feng et al. Trends Genet. 2020.

Table 1. Overview of 3D epigenome mapping in A13A/A13D cells

	H3K27ac HiChIP	CTCF ChIA-PET	H3K27ac ChIP-seq	H3K4me3 ChIP-seq	ATAC-seq	RNA-seq
A13A_Org	R1	R1	R1	R1	R1	R1
	R2	R2	R2	R2	R2	R2
					R3	R3
A13D_Org	R1	R1	R1	R1	R1	R1
	R2	R2	R2	R2	R2	R2
					R3	R3
A13A_Adh	R1	R1	R1	R1	R1	R1
	R2	R2	R2	R2	R2	R2
					R3	R3
A13D_Adh	R1	R1	R1	R1	R1	R1
	R2	R2	R2	R2	R2	R2
					R3	R3
Sequencing depth	200 million reads/Rep	200 million reads/Rep	20 million reads/Rep	20 million reads/Rep	20 million reads/Rep	20 million reads/Rep

3.2 Results

3.2.1 Identification of the interactions of cis-regulatory elements by H3K27ac HiChIP

The 3D connectome brings together regulatory chromatin regions and hence can potentially contribute to the regulation of gene expression. To generate high-resolution contact maps of active *cis*-regulatory elements (CREs), I performed H3K27ac HiChIP on the cells described above. The experimental steps for HiChIP encompassed *in-situ* chromatin digestion, *in-situ* chromatin ligation, sonication, chromatin immunoprecipitation by H3K27ac antibody, tagmentation (Tn5 transposase), streptavidin beads immunoprecipitation, PCR amplification, size selection and sequencing (**Figure 2A**). Data processing using HiC-Pro showed ~150 millions aligned reads (**Figure 2B**) and ~40 million *cis* (intrachromosomal) contacts and ~2-4 million *trans* (interchromosomal) on average from the HiChIP libraries (**Figure 2C**), suggesting the efficient capture of meaningful intra-chromosomal loops rather than interchromosomal loops, owing to the *in-*

situ digestion and ligation techniques. In addition, HiChIP data contain both 1D (binding sites of the targeted proteins) and 3D (chromatin interactions between binding sites) information. I found that the 1D signal of the HiChIP data was highly comparable to that of the H3K27ac ChIP-seq data sets (**Figure 2D**). Collectively, these QC metrics indicated the high quality of the HiChIP libraries.

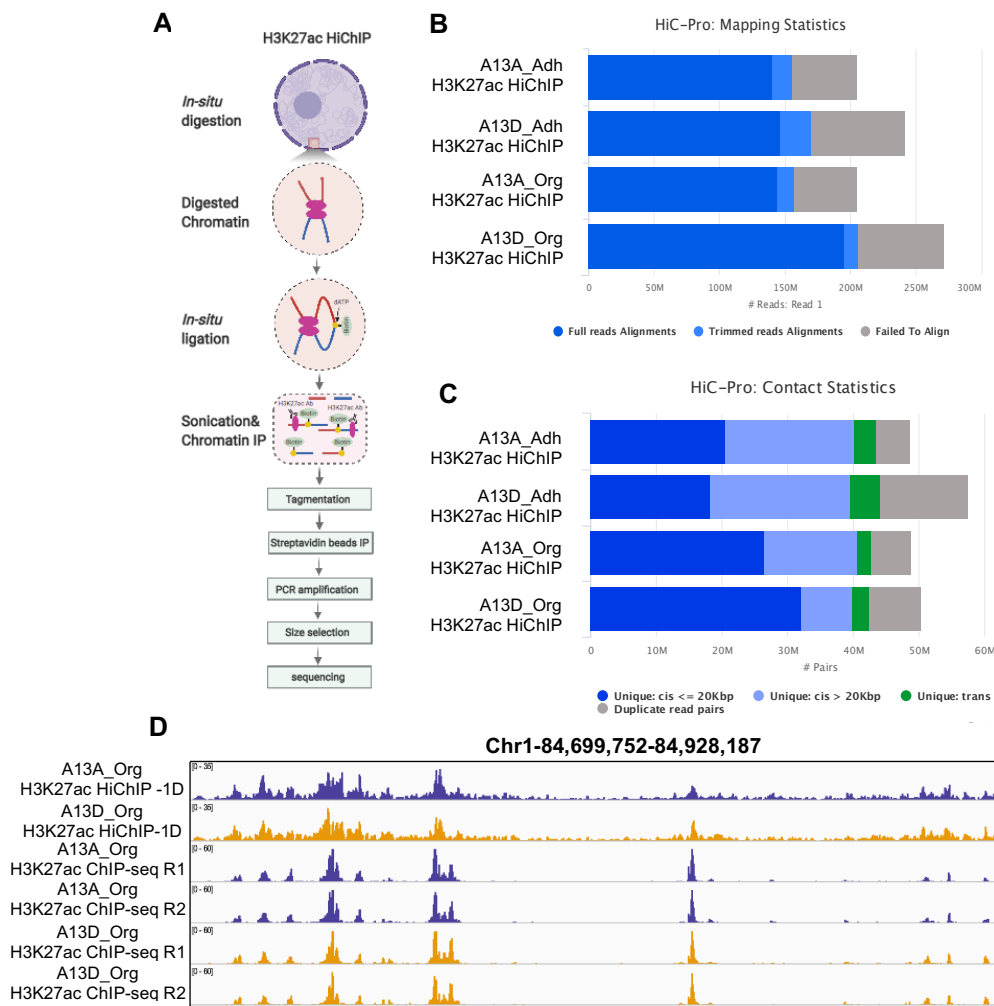


Figure 2. Interrogation of cis-regulatory element (CRE) by H3K27ac HiChIP. A. H3K27ac HiChIP mapping protocol: crosslinked cells are subjected to nuclei isolation. *In-situ* digestion of chromatin is performed using restricted enzymes, followed by *in-situ* ligation (biotin-labeled dATP) of two chromatin

fragments which are proximal in space. Thereafter, the nuclei are subjected to sonication. Chromatin immunoprecipitation is required to capture the H3K27ac-centric chromatin interactions via H3K27ac antibody. Subsequently, the chromatin is treated by decrosslinking and DNA purification. Then the purified DNA is tagged by transposon (Tn5) preloaded with sequencing adapters, followed by streptavidin bead immunoprecipitation (to capture the DNA fragments conjugated with biotin), PCR amplification, size selection and sequencing; Mapping statistics (B) and contact statistics (C) for H3K27ac HiChIP in A13A_Adh, A13A_Org, A13D_Adh, A13D_Org by HiC-Pro; D. Genome browser view of comparison H3K27ac ChIP-seq and H3K27ac HiChIP 1D signals in A13A_Org and A13D_Org (Chr1-84,699,752-84,928,187).

3.2.2 Specific CRE interactions regulate metastasis-specific transcription

To identify how CRE interactions contribute to A13A/A13D_Org-specific transcription, I performed differential analysis of HiChIP loops and identified 651 H3K27ac loops specific to A13A_Org (**Figure 3A**) and 363 H3K27ac loops specific to A13D_Org (**Figure 3B**).

By integration with RNA-seq data, I found that the genes with A13A_Org-specific H3K27ac loops were downregulated in A13D_Org while the genes with A13D_Org-specific H3K27ac loops were upregulated in A13D_Org (**Figure 3C**). In addition, the numbers of specific enhancers connected with promoters were positively correlated with increase in expression levels of their connected genes in A13D_Org, suggesting enhancer-promoter interactions contributed to gene transcription in a quantitatively manner (**Figure 3D**). To understand the functional implications of the genes regulated by A13D_Org-specific H3K27ac loops, I employed Enrichr and Cytoscape to represent their functional enrichment network. The most highly represented GO terms are displayed in **Figure 3E**, including neoplasm metastasis (e.g. *NPR1*, *BMI1*), cancer relapse (e.g.

NPR1), invasive carcinoma (e.g. *BMI1*), pathologic neovascularization (e.g. *NPR1*, *BMI1*) etc., suggesting that specific CRE interactions regulated specific transcription programs in A13D_Org implicated for human PDAC metastasis.

As exemplified in **Figure 4** and **Figure 5**, specific enhancer-promoter interactions were concomitantly observed for expression of *MAP1B* and *RP11-382D12.2* in A13A_Org, *NPR1* and *BMI1* in A13D_Org respectively. Interestingly, in addition to a specific enhancer connected to *BMI1* promoter, I also found that distinct promoters (*DNAJC1* and *MLLT10*) connected to *BMI1* promoter. *DNAJC1* and *MLLT10* were both upregulated in A13D_Org and their expression was correlated with *BMI1* expression as shown in the PDAC cohort from The Cancer Genome Atlas (TCGA)($R=0.39$ for *DNAJC1-BMI1* pair, $R=0.72$ for *MLLT10-BMI1* pair)(**Figure 6**). This suggested that *MLLT10-DNAJC1-BMI1* may form a regulatory hub for co-transcription in A13D_Org.

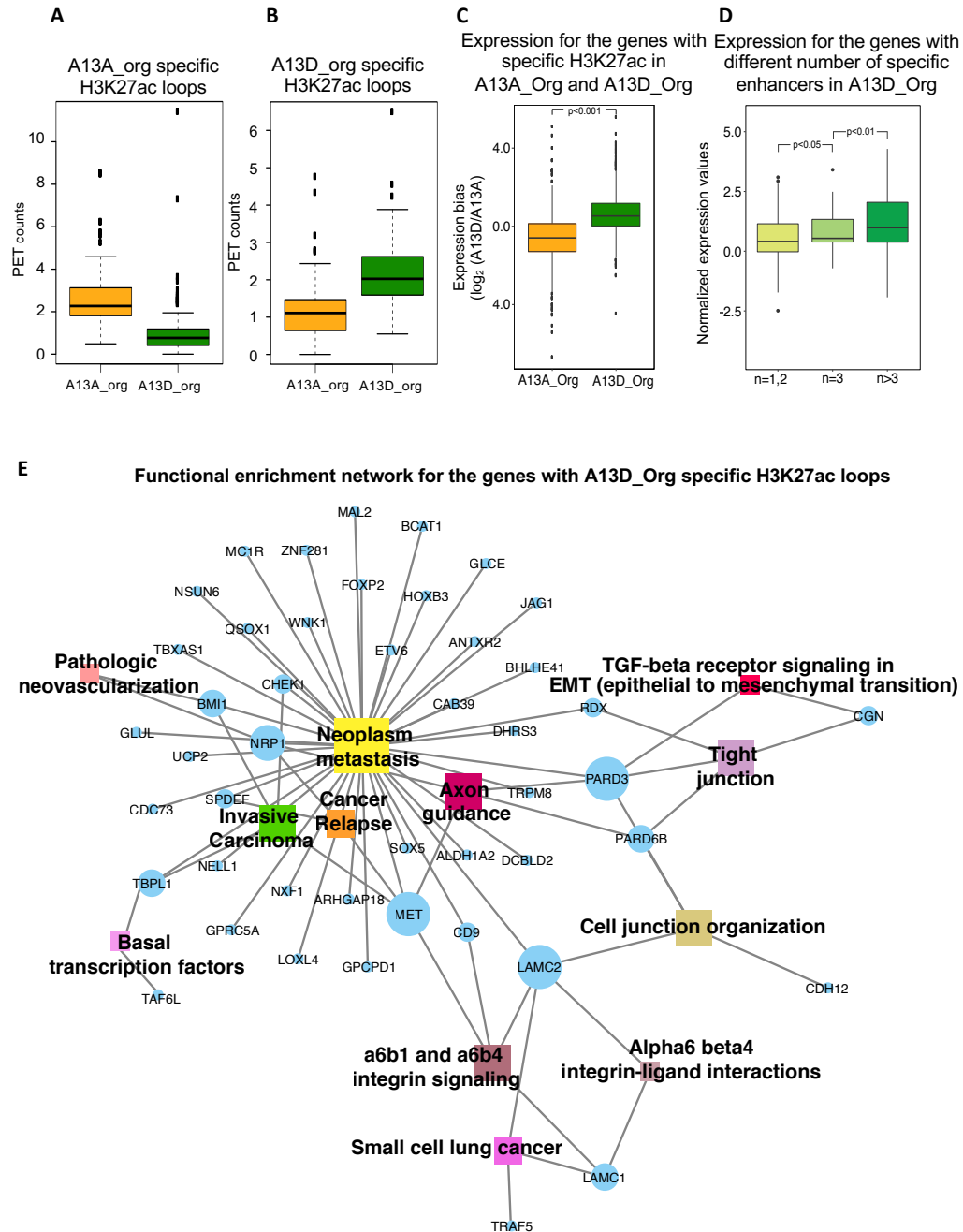


Figure 3. H3K27ac HiChIP identifies specific looping of the active CREs for A13A_Org and A13D_Org-specific transcription. **A.** PET counts for A13A_Org specific H3K27ac loops ($n=651$); **B.** PET counts for A13D_Org specific H3K27ac loops ($n=363$); **C.** Expression bias (Log_2 fold change (A13D/A13A)) of the genes with A13A_Org-specific and A13D_Org-specific H3K27ac loops. Data were analyzed by Mann-Whitney U test. $p<0.001$; **D.** Normalized expression value of the genes with different numbers of enhancers

connected to promoters in A13D_Org. Data were analyzed by Mann-Whitney U test. $n=1,2$ vs $n=3$ $**p<0.05$, $n=3$ vs $n>3$ $**p<0.05$. **E.** Network representation of functional enrichment of the genes with A13D_Org-specific H3K27ac loops. Functional enrichment was done using Enrichr (<https://maayanlab.cloud/Enrichr/>). Blue nodes represent the genes with A13D_Org-specific H3K27ac loops while the different colored rectangles represent enriched terms. Only select enriched terms are shown here. Network was generated using the Cytoscape application.

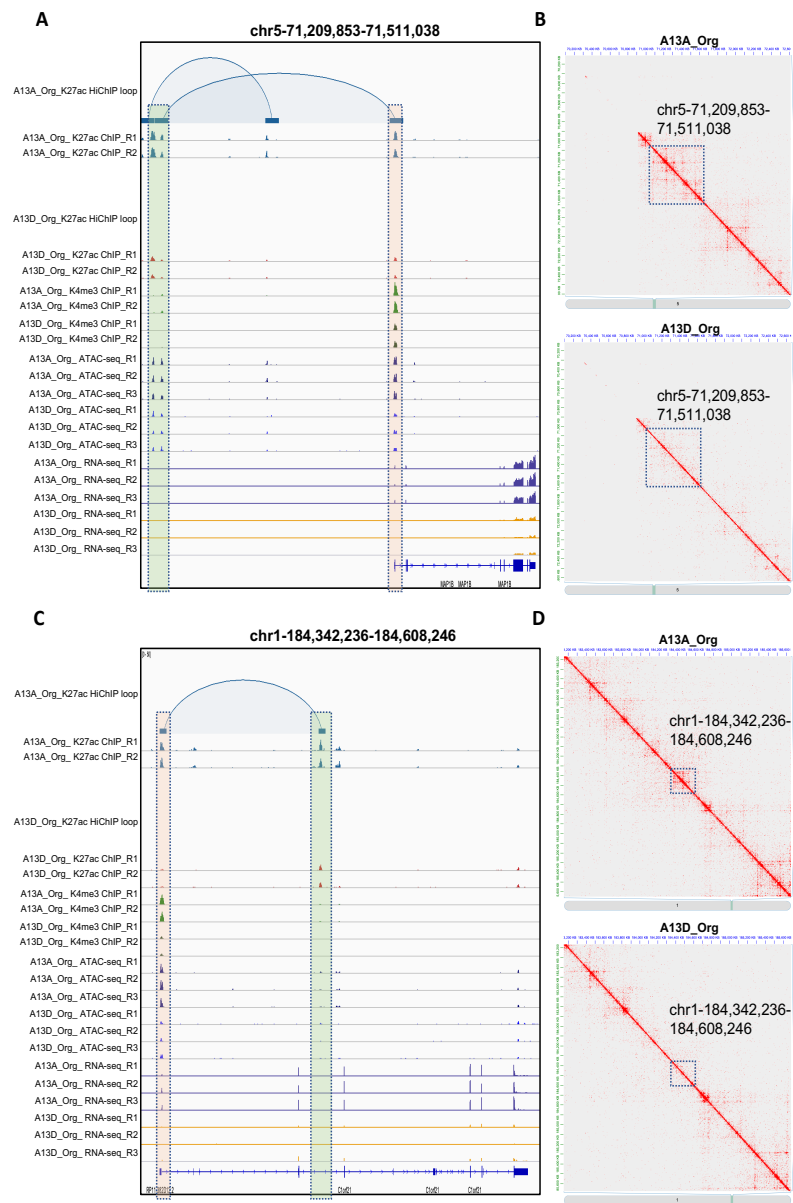


Figure 4. Examples of A13A_Org-specific H3K27ac loops. **A.** IGV browser screenshots showing *MAP1B* with A13A_Org-specific H3K27ac loops (chr5:71,209,853-71,511,038). Enhancer is highlighted in the green box, while promoter is highlighted in the orange box. **B.** 2D chromatin contact map (Juicebox) for genomic region (chr5:71,209,853-71,511,038). **C.** IGV browser screenshot showing *RP11-382D12.2* with A13A_Org-specific H3K27ac loops (chr1:184,342,236-184,608,246). **D.** 2D chromatin contact map (Juicebox) for genomic region (chr1:184,342,236-184,608,246).

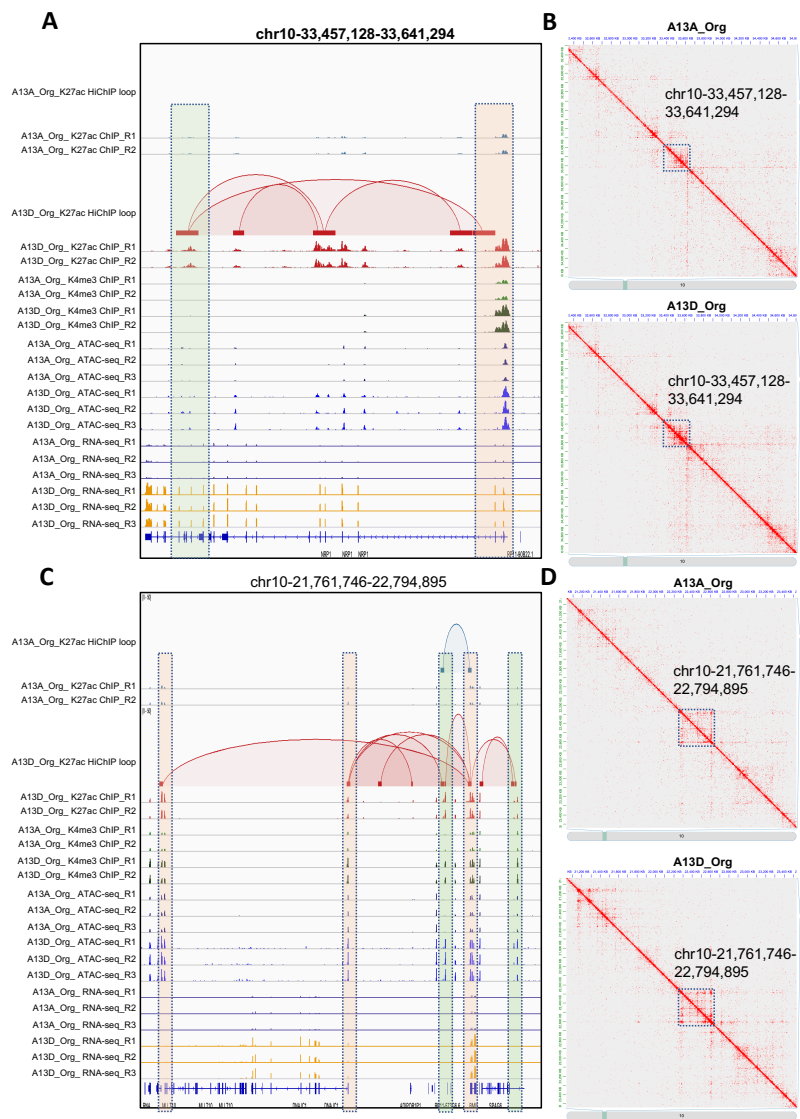


Figure 5. Examples of A13D_Org-specific H3K27ac loops. **A.** IGV browser screenshot showing *NRP1* with A13D_Org-specific H3K27ac loops (chr5:71,209,853-71,511,038). Enhancer is highlighted in the green box, while promoter is highlighted in the orange box. **B.** 2D chromatin contact map (Juicebox) for genomic region (chr5:71,209,853-71,511,038). **C.** IGV browser screenshot showing *MLLT10-DNAJC1-BMI1* with A13A_Org-specific H3K27ac loops (chr10:21,761,746-22,794,895). **D.** 2D chromatin contact map (Juicebox) for genomic region (chr10:21,761,746-22,794,895).

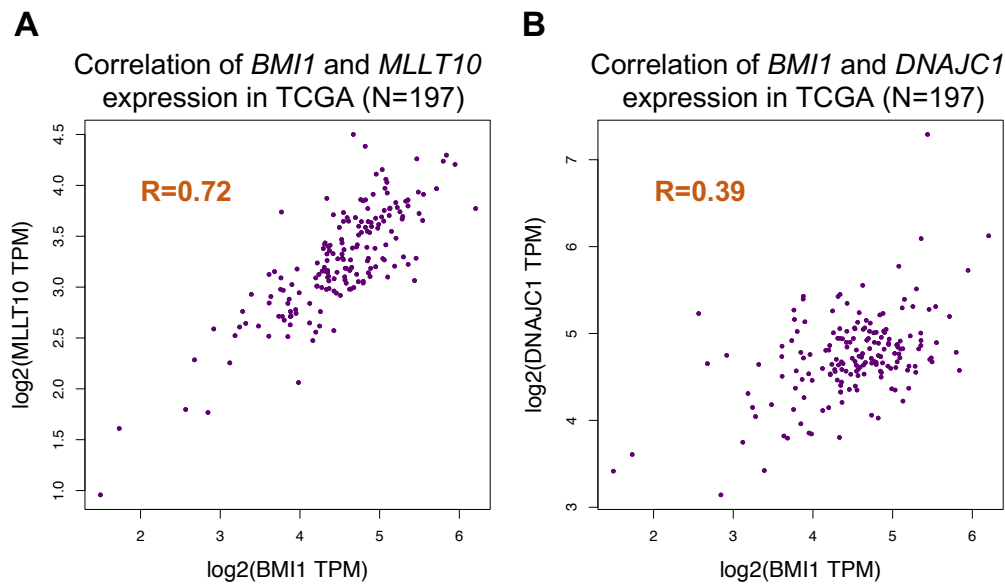


Figure 6. Co-transcription of *MLLT10-DNAJC1-BMI1* in PDAC patients. Dot-plots from TCGA data for pancreatic adenocarcinoma (N=197) to illustrate Pearson's correlation for genes pairs with promoter-promoter interactions (*MLLT10-BMI1* (A) and *DNAJC1-BMI1* (B)).

3.2.3 Chromatin accessibility is increased on specific H3K27ac loop anchors

Previous report from Howard Chang Lab showed that most cell-type-specific H3K27ac loop anchors had equivalent chromatin accessibility in each human T cell subtypes³¹. However, as observed for the examples in **Figure 4** and **Figure 5**, intensity of chromatin

accessibility is related to the specific H3K27ac loop anchors. This impelled me to perform global analysis to examine the chromatin accessibility on A13A_Org/A13D_Org-specific H3K27ac HiChIP loop anchors. In line with those examples, I found that on A13A_Org-specific H3K27ac loop anchors, the chromatin accessibility was higher in A13A_Org compared to A13D_Org (**Figure 7A**). Similarly, on A13D_Org-specific H3K27ac loop anchors, the chromatin accessibility was higher in A13D_Org compared to A13A_Org-specific H3K27ac loop anchors (**Figure 7B**). These results suggested that chromatin accessibility is associated with the formation of A13A_Org/A13D_Org-specific H3K27ac HiChIP loop anchors. This also suggested that the link of chromatin accessibility and specific H3K27ac loop anchors is cell-type dependent.

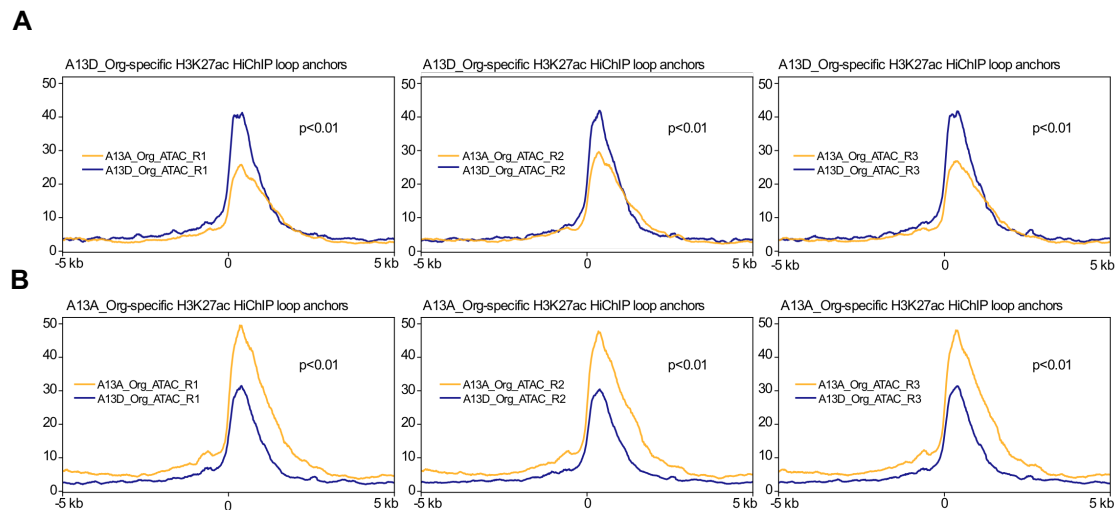


Figure 7. Chromatin accessibility on specific H3K27ac loop anchors in A13A_Org and A13D_Org. Aggregation plot of A13A_Org and A13D_Org normalized ATAC-seq signals ± 5 kb centered on ATAC-seq peaks overlapped with A13_Org-specific H3K27ac loop anchors (**A**) and A13_Org-specific H3K27ac loop anchors (**B**). Data were analyzed by Mann-Whitney U test. $p < 0.01$;

3.2.4 FOSL1 is enriched on specific H3K27ac loop anchors in A13D_Org

In order to investigate the potential transcription factors enriched on the specific H3K27ac loop anchors, I overlapped the ATAC-seq peaks with specific H3K27ac loop anchors and performed motif analysis by HOMER. The results showed that the motifs enriched on A13A_Org and A13D loop anchors are largely different (**Figure 8A**), highlighting the notion that these regions are functioning differently in A13A_Org and A13D_Org. For example, FOSL1 is the most significantly enriched motif in A13D_Org (**Figure 8B**). As FOSL1, FOSL2, ATF3, FOS and JUN belong to AP-1 family and share similar motifs, I further examined their expression and found that only FOSL1 was upregulated in A13D_Org as compared to A13A_Org (**Figure 8C**). In addition, FOSL1 but not FOSL2, ATF3, FOS and JUN was significantly correlated to poor survival in PDAC patients (N=187, $p=0.0058$) (**Figure 8D**). These data suggested that FOSL1 may be a potential transcription factor to regulate specific H3K27ac loops in A13D_Org.

A Transcription factor motif scan for A13A_Org-specific H3K27ac loop anchors

Consensus A13A_Org-specific H3K27ac loop anchors	Transcription factor	p value
	SMAD3	1e-602
	VEZF1	1e-594
	ETV6	1e-583
	WT1	1e-542
	LEF1	1e-497
	NFIB	1e-462
	ZIC2	1e-429
	FOXM1	1e-383
	HXD12	1e-265
	SMCA1	1e-264

B Transcription factor motif scan for A13D_Org-specific H3K27ac loop anchors

Consensus A13D_Org-specific H3K27ac loop anchors	Transcription factor	p value
	FOSL1	1e-202
	SMAD1	1e-200
	ZEB1	1e-171
	IKZF1	1e-165
	FOXJ2	1e-164
	SMAD3	1e-163
	ETV6	1e-159
	TWST1	1e-146
	MAFB	1e-137
	SALL4	1e-115

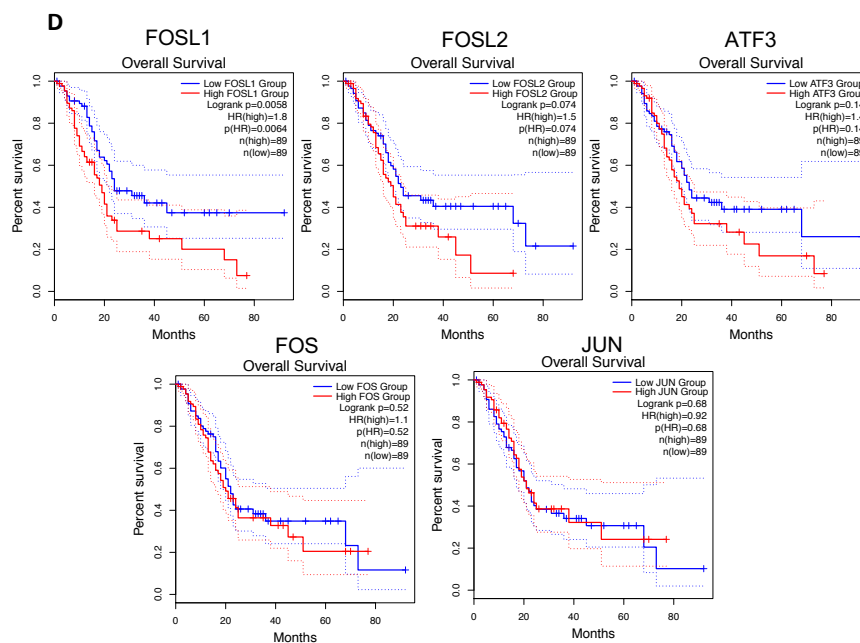
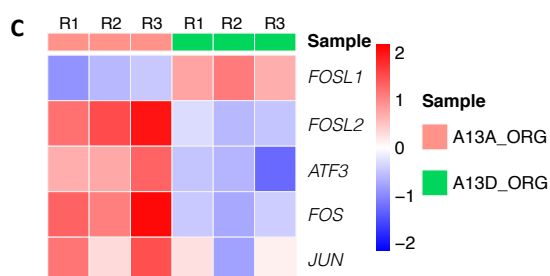
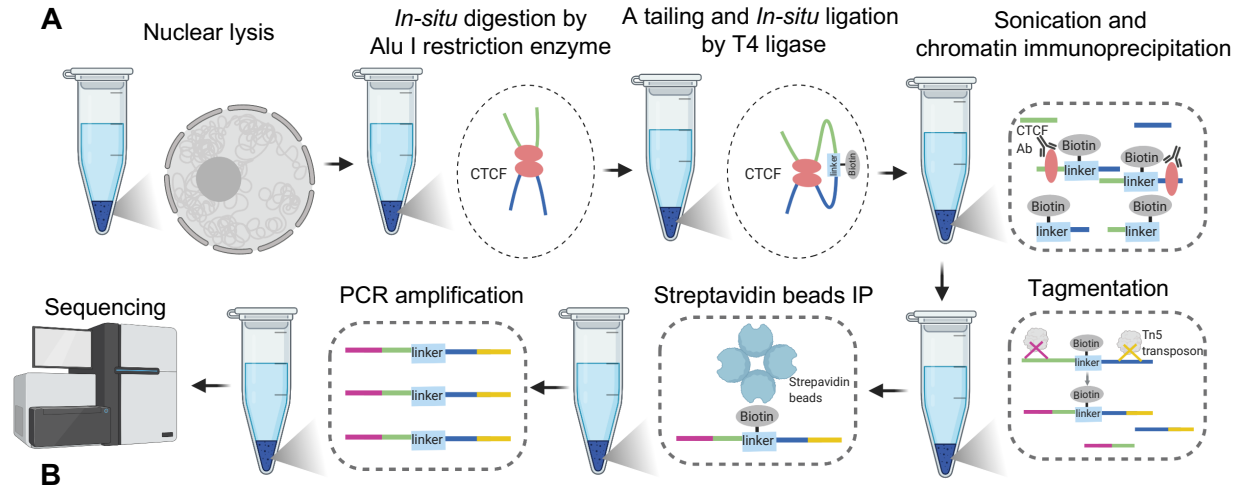


Figure 8. FOSL1 is enriched on A13D_Org-specific H3K27ac loop anchors. Transcription factor motif scan for A13A_Org-specific H3K27ac loop anchors (A) and A13D_Org-specific H3K27ac loop anchors (B); C. Gene expression for FOSL1, FOSL2, ATF3, FOS, JUN in A13A_Org and A13D_Org; D. Survival curves for FOSL1, FOSL2, ATF3, FOS, JUN in PDAC patients.

3.2.5 Topologically associating domains are conserved in A13A_Org and A13D_Org

Next, to investigate whether the altered H3K27ac looping is associated with changes in higher-order chromatin structures, I performed *in-situ* CTCF ChIA-PET (**Figure 9A**), which can capture the multilevel chromatin architecture, including TADs. All the QC metrics, in particular fraction of read pairs with linker (>0.80) and intrachromosomal/interchromosomal ratio (>2) (**Figure 9B**) indicated high quality of *in-situ* CTCF ChIA-PET libraries. Comparison of genome-wide insulation scores suggested that TADs were highly conserved between A13A_Org and A13D_Org (correlation coefficient = 0.93) (**Figure 10A**), and presented similar insulation profiles (**Figure 10B**). Hence, the overall higher-order chromatin structures are largely conserved between A13A_Org and A13D_Org.



Sample name:	A13A Org-R1	A13A Org-R2	A13D Org-R1	A13D Org-R2	A13A Adh-R1	A13A Adh-R1	A13D Adh-R1	A13D Adh-R2
Total read pairs	224,341,522	220,356,914	250,085,304	216,077,419	241,799,001	276,748,448	233,411,326	186,240,822
Read pairs with linker	183,722,905	178,385,418	211,134,598	187,450,340	216,401,354	245,213,210	205,212,232	162,265,895
Fraction read pairs with linker	0.82	0.81	0.84	0.87	0.89	0.89	0.88	0.87
One tag	53,637,753	44,671,709	61,250,751	58,241,351	61,900,423	69,522,478	49,287,085	39,847,504
PET	126,975,605	130,583,759	143,185,322	123,221,810	152,094,754	172,906,875	153,810,424	120,782,025
Uniquely mapped PET	102,270,798	107,017,617	115,270,412	97,862,143	118,905,556	135,414,136	123,949,556	97,028,826
Non-redundant PET	10,623,880	9,003,582	16,138,471	18,455,859	20,475,844	21,743,615	19,892,629	18,624,747
Redundancy	0.9	0.92	0.86	0.81	0.83	0.84	0.84	0.81
Non-redundant tag	519,150,226	362,122,631	737,988,262	841,917,622	853,928,572	916,932,942	761,644,005	715,040,491
Peak	200,157	170,246	225,936	222,145	274,675	271,517	259,377	224,909
Self-ligation PET	5,313,863	4,557,493	8,276,499	9,491,547	7,433,281	7,838,712	7,880,954	7,390,817
Inter-ligation PET	5,310,017	4,446,089	7,861,972	8,964,312	13,042,563	13,904,903	12,011,675	11,233,930
Intra-chr PET	3,459,020	2,901,038	5,359,211	6,114,091	8,247,648	8,764,827	7,313,683	6,844,171
Inter-chr PET	1,850,997	1,545,051	2,502,761	2,850,221	4,794,915	5,140,076	4,697,992	4,389,759
ratio of intra/inter PET	1.87	1.88	2.14	2.15	1.72	1.71	1.56	1.56
Singleton	3,658,229	2,884,322	5,308,850	6,371,832	10,076,374	10,638,964	9,303,072	8,835,607
Intra-chr singleton	2,266,305	1,806,548	3,425,718	4,095,204	6,067,871	6,368,881	5,360,209	5,106,824
Inter-chr singleton	1,391,924	1,077,774	1,883,132	2,276,628	4,008,503	4,270,083	3,942,863	3,728,783
PET cluster	548,734	522,933	872,642	899,005	1,064,620	1,159,104	935,428	852,289
ratio of intra/inter cluster (pets number 2)	2.27	2.1	2.68	2.43	2.45	2.43	2.27	2.29
Intra-chr PET cluster	381,106	354,090	636,437	668,690	755,610	820,673	649,312	592,958
pets number 2	251,711	234,162	423,991	453,945	519,485	560,134	435,356	405,537
pets number 3	58,217	53,559	103,898	105,329	115,175	125,150	99,326	90,081
pets number 4	24,349	22,070	40,643	42,326	48,205	53,104	43,380	38,535
pets number 5	13,657	12,881	21,374	22,392	25,959	28,708	24,211	20,817
pets number 6	8,753	8,325	12,933	13,544	15,384	17,059	14,730	12,349
pets number 7	6,085	5,794	8,811	8,453	9,614	10,919	9,265	7,609
pets number 8	4,355	4,130	5,868	5,497	6,096	7,179	6,121	4,922
pets number 9	3,151	3,067	4,261	3,647	3,985	4,632	4,128	3,214
pets number 10	2,305	2,299	3,010	2,639	2,669	3,194	2,834	2,104
pets number>10	8,523	7,803	11,648	10,918	9,038	10,594	9,961	7,790
Inter-chr PET cluster	167,628	168,843	236,205	230,315	309,010	338,431	286,116	259,331
pets number 2	120,429	120,253	172,135	175,636	231,268	252,135	207,398	193,496
pets number 3	22,636	23,178	34,378	30,094	39,128	43,276	37,830	33,398
pets number 4	8,800	8,865	12,119	10,832	16,808	17,996	16,520	13,962
pets number 5	5,038	5,203	6,087	5,625	9,049	9,933	9,340	7,796
pets number 6	3,165	3,350	3,691	3,266	5,329	5,820	5,702	4,435
pets number 7	2,215	2,299	2,487	1,992	3,079	3,606	3,566	2,542
pets number 8	1,537	1,582	1,696	1,189	1,863	2,176	2,217	1,565
pets number 9	1,110	1,208	1,143	687	1,054	1,373	1,381	872
pets number 10	791	860	749	413	609	845	872	544
pets number>10	1,907	2,045	1,720	581	823	1,271	1,290	721

Figure 9. Mapping of chromatin topology by CTCF in-situ ChIA-PET. *A. In-situ ChIA-PET protocol:* crosslinked cells are subjected to nuclei isolation. In-situ enzymatic digestion of chromatin is performed, followed by in-situ ligation (bio-labeled bridge linker) of two chromatin fragments which are proximal in space. Thereafter, the nuclei are subjected to sonication., chromatin immunoprecipitation is required to capture the specific protein-centric chromatin interactions (e.g. CTCF). Subsequently, the chromatin for all these assays is treated by decrosslinking and DNA purification. Then the purified DNA is tagged by

transposon (*Tn5*) preloaded with sequencing adapters, followed by streptavidin bead immunoprecipitation (to capture the DNA fragments conjugated with biotin), PCR amplification, size selection and sequencing.

B. QC metrics for CTCF in-situ ChIA-PET in *A13A_Adh*, *A13A_Org*, *A13D_Adh*, *A13D_Org*.

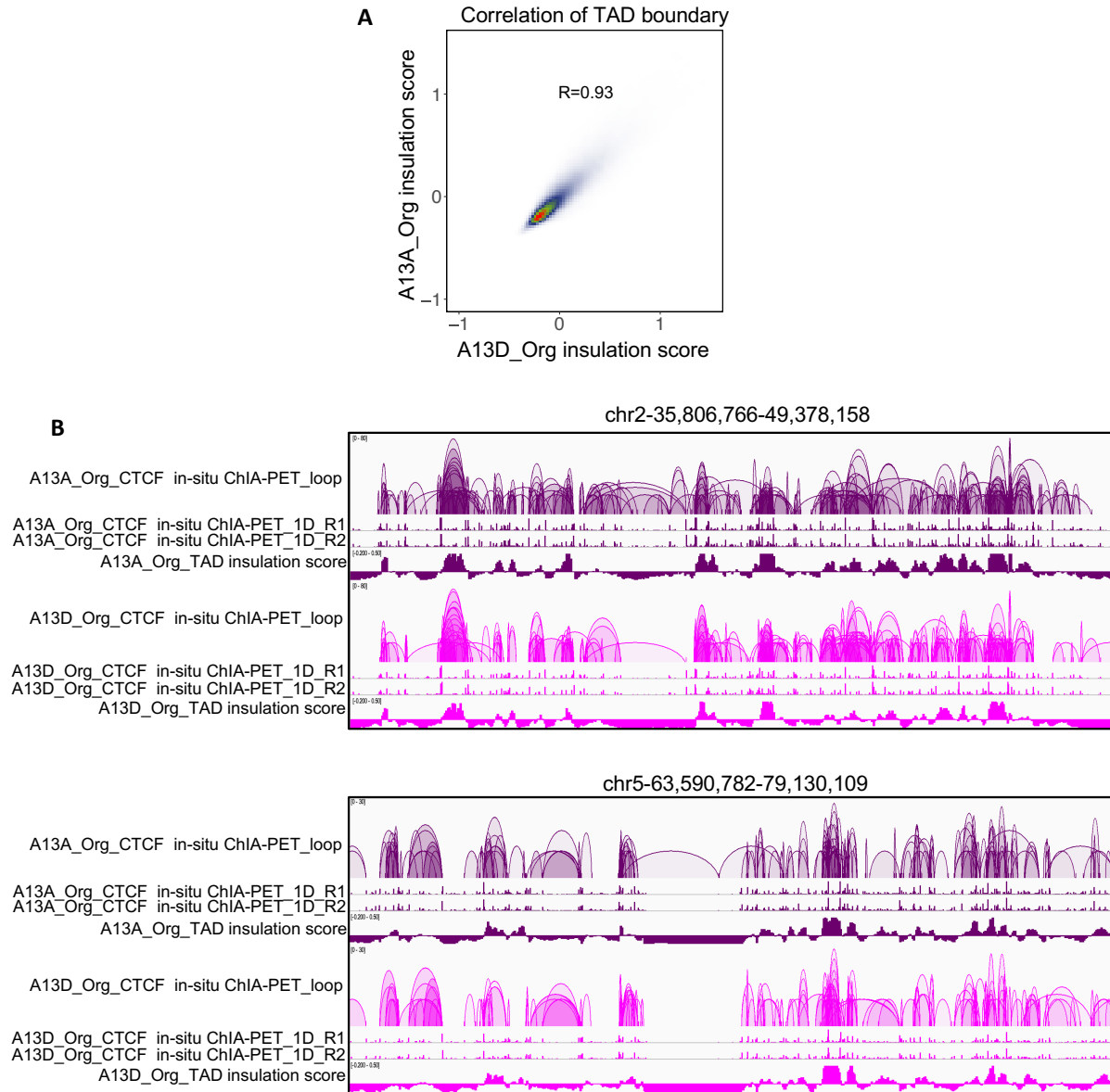


Figure 10. Conserved TAD boundaries in *A13A_Org* and *A13D_Org*. **A.** Correlation of insulation scores for *A13A_Org* and *A13D_Org*. The Pearson correlation coefficient (R) is indicated ($R=0.93$). **B.** CTCF

ChIA-PET loops, 1D (peaks) and corresponding insulation profiles for A13A_Org and A13D_Org in the genomic regions (chr2:35,806,766-49,378,158, chr5: 63,590,782-79,130,109).

3.2.6 Specific CTCF binding on the promoters links a portion of specific CRE interactions

As TADs are conserved in A13A_Org and A13D_org, I examined the chromatin architecture at high resolution. I found that 83 A13A_Org-specific vs. 10 A13D_Org specific CTCF peaks were enriched on the promoters with specific H3K27ac loops in A13A_Org (**Figure 11A**) while 72 A13D_Org-specific vs. 9 A13A_Org specific CTCF peaks were enriched on the promoters with specific H3K27ac loops in A13D_Org (**Figure 11B**). This suggested that specific CTCF binding on the promoters might link to specific CRE interactions, which is in line with a previous report showing that CTCF anchor genes were associated with active transcription¹⁸⁸.

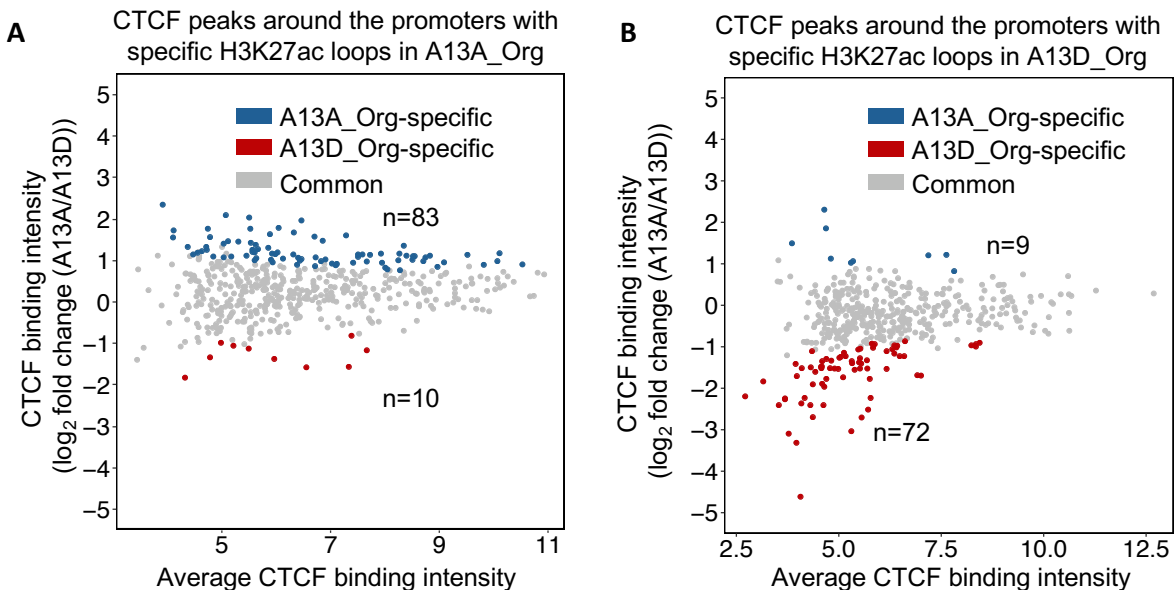


Figure 11. A portion of CRE contacts is associated with specific CTCF binding on the promoters. *A.* MA plot showing CTCF binding peak intensity on the promoters connected with specific H3K27ac loops in

A13A_Org (A) and A13D_Org (B). A13A_Org-specific, common and A13D_Org-specific CTCF bindings are highlighted in blue, grey and red respectively.

3.3 Discussion

PDAC is one of the most lethal cancers in the world and is featured by transcription dysregulation. However its mechanisms remain elusive. In this project, I aim to decode the 3D genome organization in PDAC and investigate how it links to transcription dysregulation. I used two unique cell lines, A13A and A13D, which were derived from primary and metastatic PDAC tumors respectively from the same patient. I performed comprehensive 3D epigenome mapping on the adherent cells and also the isolated cancer stem cells marked by CD133: histone ChIP-seq to map the active promoter and enhancer and RNA-seq data to measure the transcription output: in-situ ChIA-PET and HiChIP sequencing to map the 3D genome architecture. Since H3K27ac is the most important histone mark for active enhancers and promoters, capturing the H3K27ac-centric chromatin interactions will enable us to identify the interactions of active enhancers and promoters at high resolution. Similarly, capturing the CTCF-centric chromatin interactions will enable us to identify the topologic structures framing the enhancer-promoter interactions. And I uncovered the specific enhancers for each cell types. In particular, I found that the specific enhancers for A13D cancer stem cells are implicated for transcriptional misregulation of cancer, stem cell pluripotency, and TGF-beta signaling pathway, which is an emerging player in PDAC development and progression. By further integrating about 900 H3K27ac ChIP-seq data from different cell types, I found that these enhancers are more cell-specific compared to the shared enhancers with the other A13A/A13D cells, suggesting their unique function. Furthermore,

I found that FOSL1 and other AP-1 family transcription factor motifs were enriched on these enhancers. And high expression of FOSL1 is associated with poor survival of PDAC patients.

The HiChIP data identified the interactions of the *BMI1* promoter with the *DNAJC1* and *MLLT10* promoters and suggested that this might be a regulatory hub for co-transcription. However, like other proximity-ligation-based 3D genome mapping approaches (Hi-C, ChIA-PET), HiChIP can only detect binary interactions and cannot reflect whether the uncovered multigene clusters coexist in individual cells or whether they are just an ensemble of diverse chromatin interactions from millions of cells as demonstrated in **Figure 2C**. 3D DNA fluorescence in situ hybridization (FISH) followed by super resolution microscopy can the coexistence of this multigene cluster. In addition, future work can further address whether the functional roles of FOSL1 in regulating specific enhancer promoter connectome in A13D_Org.

3.4 Materials and Methods

3.4.1 HiChIP (H3K27ac)

HiChIP were performed on 10 million cells with pretreated by 1% formaldehyde. Initially, biotin-labeled proximally-ligated chromatin was prepared using the Arima-HiC kit (Arima Genomics Cat # A510008). Briefly, the crosslinked chromatin was initially digested using a restriction enzyme (RE) cocktail. The 5'-overhangs were subsequently filled in and labeled with a biotinylated nucleotide. Subsequently, the spatially proximal digested ends

of DNA were ligated. The resulting proximally-ligated chromatin was fragmented using the Diagenode Bioruptor Pico instrument, bound to an H3K27ac antibody (Abcam Cat # ab4729) overnight and immunoprecipitated on Protein A magnetic beads (Thermo Cat # 10002D). The immunoprecipitated chromatin was reverse crosslinked and purified using the Arima-HiC kit. Subsequently, the Illumina-compatible sequencing libraries were prepared by initial enrichment for biotin using C1 streptavidin beads (Thermo Cat # 65001). The samples were tagged on-bead utilizing Tagment DNA Enzyme and Buffer (Illumina Cat # 20034197). Subsequently, DNA was PCR amplified utilizing NPM and the indexing primers from the Nextera XT DNA Library Prep kit (Illumina Cat # FC-131-1024). The PCR products were purified and size selected by 0.5X SPRI beads followed by 0.7X SPRI beads. The purified libraries underwent standard QC (qPCR and Bioanalyzer) and were sequenced on the Illumina HiSeq Xten (150 bp, pair-end) following the manufacturer's protocols.

3.4.2 *In-situ* ChIA-PET (CTCF)

Cells were grown to 80% confluency, and treated by 1% formaldehyde for 20 min and ethylene glycol bis (EGS)(Thermo Fisher Scientific, cat.#21565) for 45 min. 10 million of dual crosslinked A13A/A13D cells were used to generate one in-situ ChIA-PET library. Briefly, after cell lysis and nuclear lysis, cells were subjected to in-situ digestion with Alu I restriction enzyme for overnight. On the next day, A-tailing was performed using Klenow Fragment (3'>5' exo-) (NEB, cat.#M0212L) and dATP (100 mM; NEB, cat.#N0440S) for 1 hr at room temperature. In-situ ligation was performed using in-house bridge linker (F: 5'- /5Phos/CGCGATATC/iBIOdT/TATCTGACT -3', R: 5'-

/5Phos/GTCAGATAAGATATCGCGT -3'. HPLC purified, from Integrated DNA Technologies) and T4 DNA ligase (Thermo Fisher Scientific, cat.#EL0013) for RT 1 hr and then 16°C for overnight. The nucleus was sonicated into ~1 kb size. Then chromatin was precleared using protein G beads for 1 hr and immunoprecipitation was performed using anti-CTCF antibody (Abclonal, cat.# A1133) coated protein G beads (Life Technologies, cat. no. 10009D) for overnight.

After that, proximity-ligated chromatin complexes were washed sequentially by low salt, high salt, LiCl buffer and TE buffer. Chromatin complexes were eluted from the beads by incubation with ChIP elution buffer (1% SDS+TE) at 65 °C for 1 hr. After that, proteinase K (Life Technologies, cat.#AM2548) was added to the elution reverse cross-linked chromatin complexes for overnight. After DNA purification using QIAquick PCR Purification Kit, tagmentation was performed on the proximity-ligated DNA by Tn5 transposase from Nextera DNA Sample Prep Kit (Illumina, cat.#FC-121-1031). Then tagmented DNA was immobilized on M280 streptavidin dynabeads (Invitrogen, cat.#11205D). PCR amplification was performed on beads using Nextera DNA Sample Prep Kit and the products were purified by AMPure XP beads (Beckman, cat.#A63881) and subjected to size selection (350-600 bp) on BluePippin instrument (Sage Science) using Blue Pippin Cassette Kit (Sage Science, cat.#BDF2010). DNA library was sequenced on Novaseq 6000 by paired-end 150 bp.

3.4.3 ChIP-seq

Cells were treated with 1% (final concentration) formaldehyde (Thermo Fisher, Cat#

28906) and incubated at room temperature for 10 min. Fixation was stopped by the addition of 0.125 M (final concentration) glycine to the medium. The chromatin was isolated by the addition of cell lysis buffer for 1 hr and nuclear lysis buffer for 30 min. The lysates were sonicated and the DNA sheared to an average length of 300-500 bp. The pellets were resuspended and the resulting DNA was quantified on a Qubit spectrophotometer. An aliquot of chromatin (8 µg) was precleared with protein A/G agarose beads (Invitrogen). Genomic DNA regions of interest were precipitated using 5 µg of antibody against H3K27Ac (Active Motif, Cat# 39133), H3K4me3 (Abcam, Cat#8580). The complexes were washed, eluted from the beads with 1% SDS buffer and subjected to RNase and proteinase K treatment overnight at 65°C. The ChIPed DNA was purified by QIAquick PCR Purification. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K in a similar way as ChIPed DNA. Quantitative PCR (qPCR) reactions were carried out in triplicate on specific genomic regions (ACTB-Forward: 5' TTGCCGACTTCAGAGCAACT 3'; ACTB-Reverse: CCCCAACACCACACTCTACC; GAPDH-Forward: 5' CAGGTTTCCAGGAGTGCCTT 3'; GAPDH-Reverse: 5' ATTAGGGCAGACAATCCCGG 3'. Neg-Forward: 5' GGCCCTGACTGATGGAAATTAC 3'; Neg-Reverse: 5' CAGCAGGTGACCATCCCTTTAG 3'), using SYBR Green Supermix (Life Technologies, Cat# 4368702). The resulting signals were normalized for primer efficiency by carrying out qPCR for each primer pair using input DNA. Illumina sequencing libraries were prepared from the ChIP products by the standard consecutive enzymatic steps of end-repair, A tailing, and adaptor ligation using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Cat# E7645L). The PCR products were purified and the

size was selected by 0.5X SPRI followed by 0.9X SPRI beads. The resulting DNA libraries were quantified and sequenced on an Nextseq 500 (single- end 75 bp).

3.4.4 ATAC-seq

Nuclear lysis was performed on 100,000 cells followed by tagmentation. Briefly, the transposition reaction master mix was prepared and the nuclei were mixed by pipetting gently, spun quickly and left on ice. The tagmentation master mix included the following reagents: 25 μ l 2x TD Buffer (Illumina Cat #FC-121-1030), 8 μ l Tn5 Transposase (Illumina Cat #FC-121-1030), 17 μ l PBS, 5 μ l 1% Digitonin (dissolved in 50% DMSO in dH₂O) and 5 μ l 10% Tween 20. A total of 50 μ l tagmentation master mix was added to each sample, which was gently pipetted to resuspend the cells in the transposition reaction mix. The samples were incubated at 37°C for 30 min in a thermomixer at 800 rpm. The purification was conducted using a Qiagen MinElute Kit. The DNA was eluted with a 35 μ l EB buffer (10 mM Tris buffer, pH 8.0), allowed to settle at room temperature for 1 min and centrifuged at max speed for 30 sec. Each elution was transferred to a 250 μ l PCR tube and the PCR reactions were prepared by adding the components in the order shown. The PCR reaction contained the following components: 5 μ l Index mix (Illumina Nextera Kit, 10 μ l 5x Phusion HF reaction Buffer, 1.5 μ l dNTPs (10 mM), 0.5 μ l Phusion HF Polymerase (NEB) and 33 μ l tagmented DNA. The reagents were pipetted to mix. The PCR was performed using the following program on a thermal cycler (with heated lid): heating at 72°C for 5 min and at 98°C for 30 sec. A total of 10 cycles were performed at 98°C for 10 sec, at 63°C for 30 sec and at 72°C for 30 sec. A final extension was performed at 72°C for 1 min. The AMPure double size selection was performed and 80 μ l

AMPure beads were added to the sample, vortexed briefly to mix and quickly spun. The solution was incubated at room temperature for 5 min. The samples were placed in a magnetic rack and left for 1 min. The liquid was removed with a pipette, taking care not to disrupt the pellet. The samples were washed by addition of 180 μ l fresh 70% ethanol to the tube, taking care not to disrupt the pellet. The ethanol was removed with a pipette and the procedure was repeated a second time. The ethanol was removed slowly from the tube. The samples were allowed to dry for 30 sec and the pellet was resuspended with 22 μ l Buffer EB (10 mM Tris, pH 8.0) and mixed by pipetting. The sample was allowed to stand for 1 min and placed in a magnetic rack on the magnetic side for 1 min. A total of 20 μ l of the eluate were transferred to a new tube, taking care not to collect any magnetic beads. The aliquots of the libraries were run on an Agilent 2100 bioanalyzer and quantified with the Qubit High sensitivity assay. The Qualified libraries were sequenced pair-end on the Novaseq 6000 (150 bp, pair-end).

3.4.5 RNA-seq

Total RNA was extracted from using Trizol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Approximately 1 million cells were transferred into 2 ml Eppendorf tubes containing 1.5 ml Trizol reagent. The mix was centrifuged at 12,000 x g for 5 min at 4°C. The supernatant was transferred to a new 2.0 ml tube, which contained 0.3 ml Chloroform/isoamyl alcohol (24:1) per 1.5 ml of Trizol reagent. Following centrifugation at 12,000 x g for 10 min at 4°C, the aqueous phase was transferred to a new 1.5 ml tube which contained equal volume of supernatant of isopropyl alcohol. The mix was centrifuged at 12,000 x g for 20 min at 4°C and the supernatant was collected.

Following washing with 1 ml 75% ethanol, the RNA pellet was air-dried in the biosafety cabinet and dissolved by addition of 25 µl nuclease-free water. Subsequently, total RNA was qualified and quantified using the Agilent 2100 bioanalyzer. Approximately 1 µg total RNA per sample was depleted from rRNA with the Ribo-Zero™ Magnetic Kit (Epicentre). The retrieved RNA was subjected to fragmentation by adding the First Strand Master Mix (Invitrogen). First-strand cDNA was generated using random primers and reverse transcription, followed by a second-strand cDNA synthesis. The synthesized cDNA was subjected to end-repair and was 3' adenylated. The adapters were ligated to the ends of these 3' adenylated cDNA fragments. A total of 10 rounds of PCR with PCR Primer Cocktail and PCR Master Mix were performed to amplify the cDNA fragments. The PCR products were purified and size selected by 0.5X SPRI beads followed by 0.9X SPRI beads. The identity of the final library was verified by the distribution of the fragments size using the Agilent 2100 bioanalyzer and quantified by the Qubit High sensitivity assay. The Qualified libraries were sequenced pair-end on the Novaseq 6000 (150 bp, pair-end).

3.4.6 ChIP-seq data analysis

Single-end reads were mapped to the hg19 reference genome using the bwa mem¹⁵⁸. The duplicated reads were removed using the Picard Toolkit (<https://broadinstitute.github.io/picard/>), whereas the reads with a quality lower than 20 were filtered. The ChIP-seq peaks were called using the default (narrow) setting in MACS2¹⁵⁹. The promoter regions were defined as +/- 2.5Kb around the TSSs regions. Differential peak analysis was performed using the edgeR¹⁶⁰ with a cutoff FDR < 0.05 and Fold change > 1.

3.4.7 RNA-seq data analysis

The reads were mapped to hg19 using STAR¹⁶¹. The reads mapped to gene exon regions were counted using featureCounts. Differential expression was determined using DESeq2¹⁶⁴.

3.4.8 HiChIP data processing

HiChIP data were processed using HiC-Pro¹⁶⁵ with default parameters with the exception of the LIGATION_SITE being set to GATCGATC, GANTGATC, GANTANTC, GATCANTC and the GENOME_FRAGMENT that was generated using digest_genome.py in HiC-Pro utilities. The following parameters were used: -r ^GATC G ^ANTC -o hg19_restriction_sites.bed hg19.fasta. The raw reads were mapped to the hg19 reference genome and were assigned to restriction fragments. The raw reads were filtered for valid pairs (VI). 1D coverage files for the visualization in bigwig format were converted from bam format using deepTools¹⁶⁷ and RPKM (Reads Per Kilobase per Million reads to normalize the number of reads per bin) normalization was used. The HiC files were generated using hicprotojuicebox.py utilities in HiC-Pro. All valid pairs served as input files and then imported into Juicebox¹⁶⁶ for visualization.

3.4.9 *In-situ* ChIA-PET data processing

We first identified a union set of loops for each A13A and A13D dataset. Valid read pairs for each sample served as input to Mango¹⁶⁸ for stage 4 and 5 processing. The “peakinput” parameter was set to the ChIP-seq peak file. The read pairs were extended 500 bp on

each side and subsequently clustered. Distance based correction was then performed to identify significant chromatin loops for each sample. Significant loops were identified using the cutoff $FDR < 0.1$. The clusters in each sample were also merged and clustered again to form a union set of clusters for A13A and A13D, respectively. Significant loops were identified using the ChIA-PET Tool V3.0¹⁶⁹ with cutoff $FDR < 0.00001$. The loops with PET counts lower than 10 and spanning less than 5 Kb or larger than 1 Mb were further filtered.

Specific loops were identified following two steps: Initially, loops in A13D and A13D were merged and subsequently the PET counts for each loop were counted in all the replicates in A13A and A13D. A negative binomial model was fitted using glmFit function in edgeR R package and the loops with a p value less than 0.1 were determined as specific loops.

3.4.10 Functional enrichment of genes with specific H3K27ac loops

Functional enrichment of genes network was generated using Enrichr (<https://maayanlab.cloud/Enrichr/>). Cytoscape application (<https://cytoscape.org/>) was used to generate the functional enrichment network.

3.4.11 Identification of topologically associated domains (TADs)

Replicates of cool format files were merged using cooler¹⁷³. A and B compartments were identified using hicPCA command in HiCExplorer (<https://hicexplorer.readthedocs.io/en/latest/>). The 25 Kb resolution was used to merge

the cool files. The topological associated domains (TADs) were identified using the hicFindTADs command in the HiCEXplorer at 25 Kb resolution.

3.4.12 ATAC-seq data analysis

The adapters were removed and the filtered reads were mapped to hg19 using bwa¹⁵⁸ mem. The duplicated reads were removed using Picard (<https://broadinstitute.github.io/picard/>) and the reads with quality lower than 20 were filtered.

3.4.13 Transcription factor motif analysis

The motif analysis was performed using Homer (<http://homer.ucsd.edu/homer/>, v4.11). We overlapped A13A_Org-specific/A13D_Org-specific loop anchor regions with the ATAC-seq peak regions. Subsequently, we extracted the sequences from the overlapped regions using bedtools¹⁷⁷ getfasta and imported them into Homer findMotifs.pl with default setting for motif enrichment analysis.

3.4.14 Quantitative analysis of ATAC-seq and ChIP-seq

To calculate the differential chromatin accessibility on A13A_Org-specific/A13D_Org-specific loop anchors, we generated aggregation plots. The reads were counted in a 5 Kb window centered at an anchor with 10 bp-bin resolution for A13A_Org and A13D_Org. The read counts for all anchors at each bin in each sample were summarized using the median value. Subsequently, the read counts in each bin across the samples were

summarized using mean $\pm$ standard error. The Mann–Whitney U test was used to calculate the p-value.

3.4.15 Visualization

The chromatin loops were prepared using bedpe format and then uploaded to the IGV¹⁷⁹ for arc-style visualization. Other 1D coverage signals were prepared using bigwig format and uploaded to IGV. The hic files were imported into Juicebox for heatmap visualization.

Chapter 4

BRD9 orchestrates stemness and
tumorigenesis in pancreatic ductal
adenocarcinoma via enhancer-promoter
looping

(Small molecule screening and in vitro functional assays are in collaboration with Andrei Stoica. In vivo tumorigenesis assay is in collaboration with Wei Huang)

4.1 Introduction

Pancreatic cancer with its most common type, pancreatic ductal adenocarcinoma (PDAC), is one of the deadliest human malignancies^{186,189,190}. It has an overall median survival time of only 6-9 months and a similarly dismal 5-year survival rate of 6%, making it currently the fourth leading cause of cancer-related deaths in the western countries^{191,192}. Due to the increasing incidence of risk factors including obesity and metabolic syndrome, pancreatic cancer is projected to overtake colorectal, breast and prostate cancer and become the second leading cause of worldwide cancer-related deaths by 2030¹⁹³. The disease owes its exceptional lethality to multiple factors. Pancreatic cancer often has no early symptoms and presents itself in an advanced stage at diagnosis: only 20% of newly diagnosed pancreatic cancers are amenable to surgery¹⁹⁴. In turn, the disease's poor response to chemotherapy and radiotherapy results in disease re-emergence for 90% of the surgically treated patients¹⁹⁵. Treatment options for PDAC are limited and inefficient. Gemcitabine, an antimetabolite drug of the nucleoside analogue class is the standard of care in PDAC therapy¹⁹⁵. It is currently the only approved single-agent drug for pancreatic cancer, with still modest improvements in survival rate¹⁹⁶, whether the drug is administered alone or in combination with adjuvant drugs such as the EGFR inhibitor erlotinib¹⁹⁷ or the tubulin-targeting drug Nab-paclitaxel¹⁹⁶. FOLFIRINOX is superior to gemcitabine-based regimens in restraining the progression of metastatic PDAC but has lower tolerability¹⁹⁸.

Precancerous lesions and dedifferentiation of the cells to a progenitor-like or stem cell-like state with increased cellular plasticity frequently occur during pancreatic tissue transformation^{191,199}. A cell population in various tumours seems to acquire a stem cell-

like state partially resembling naturally occurring stem cells²⁰⁰⁻²⁰², which allows them to give rise to the whole tumour with its entire cellular heterogeneity and thereby support metastases formation and resistance to current cancer therapeutics – these cells are known as cancer stem cells (CSCs). The existence of developmentally plastic cancer stem cells has been discovered in the brain, breast, colon, oesophagus, liver, lung, ovarian, prostate, stomach and thyroid cancers, among others. In the case of PDAC, the first reports of cancer stem cells date back to 2007^{200,201}. Since then, pancreatic CSCs have been conclusively shown to be involved in PDAC resistance to chemotherapy, displaying increased prevalence within the tumour after treatment with gemcitabine^{203,204}. Annihilating CSCs is thus emerging as an essential aim of PDAC therapeutics. CSCs are thought to have specific epigenetic mechanisms²⁰⁵ that regulate their self-renewal, and the formation of CSCs has been postulated to occur as a result of epigenetic events²⁰⁶. Therefore, cancer epigenetics has established itself as a promising area of oncology research²⁰⁷. After a decade in which epigenetic cancer drugs were approved only for haematological malignancies, the first FDA approval for an epigenetic drug targeting solid tumours was granted in 2020 for an EZH2 inhibitor. Despite their paradigm-shifting novel mechanism of action, early results of epigenetic modulators have identified the need for better selection of targets, improved intratumoral drug penetration and killing of the heterogeneous cancer cell populations, particularly CSCs.

Pancreatic cancers are complex tumours with significant heterogeneity in their molecular and cellular make-up that are controlled by various signalling pathways that crosstalk with epigenetic regulators. Among these pathways is TGF β /Activin/Nodal-SMAD2/3. The TGF β /Activin/Nodal-SMAD2/3 developmental signalling pathway plays a

central role in early development by regulating the self-renewal of human pluripotent stem cells (hPSCs), Epithelial-to-mesenchymal transition (EMT) and pancreatic tissue homeostasis²⁰⁸⁻²¹⁰. The TGF β /Activin/Nodal-SMAD2/3 pathway regulates epigenetic mechanisms, for instance by cooperating with core pluripotency protein NANOG and epigenetic modifiers DPY30-COMPASS to control pluripotency and differentiation of hPSCs²¹¹. Aside from the key role of TGF β /Activin/Nodal-SMAD2/3 in pluripotent stem cells and developmental processes, this signalling pathway is directly involved in the formation of PDAC²⁰⁹ and is frequently deregulated in PDAC^{212,213}. The function of the TGF β /Activin/Nodal pathway in PDAC is particularly interesting, because it confers dedifferentiated stem cell-like features to CSCs in PDAC²⁰², although the mechanisms are still largely unknown.

Here we performed a small molecule compound screening to identify epigenetic enzymes that regulate the maintenance of CSCs, and identified BRD9. Chemical and genetic inactivation of BRD9 revealed its importance in CSCs by regulating stem cell self-renewal and chemoresistance characteristics. BRD9 as an enzymatic subunit of the BAF complex induces stem cell factor expression including SOX4 and EMT regulators such as SNAI2 via Enhancer-Promoter (E-P) loop induction. Mechanistic characterization showed BRD9 orchestrates pancreatic CSC characteristics in a transcription regulatory complex with SMAD2/3 transcription factors that facilitate BRD9 binding to regulatory regions on chromatin. This governs CSC self-renewal and CSC invasiveness and gemcitabine/5-FU chemoresistance due to inducing stem cell-like characteristics. BRD9 also controls the entry of CSCs to the G0 phase which increases the chemoresistance to gemcitabine. Mouse in vivo experiments showed that BRD9 loss of function ameliorates

in vivo PDAC tumour growth. In patient tumours, BRD9 inhibition eliminates CSCs and synergizes with gemcitabine by resensitizing cells to its cytotoxicity. Collectively, our results indicate that BRD9 is an attractive therapeutic target for specifically eliminating CSCs in PDACs.

4.2 Results

4.2.1 Unbiased compound screening for epigenetic regulatory enzymes identifies BRD9 as an important factor for governing PDAC stemness and tumor growth

Due to the importance of epigenetic pathways in tumorigenesis, we hypothesized that the formation and maintenance of pancreatic CSCs are promoted by epigenetic mechanisms, and these could be used as therapeutic targets for eliminating CSCs. Since PDAC cells are heterogeneous with populations of CSCs and non-CSCs, we first decided to establish a suitable small molecule screening platform in PDAC cells (**Figure 1A**). To uncover novel regulators of CSCs, we used three CSC markers (OCT4, CD133, SSEA4) for identifying stem cell-like cells in the experiments. OCT4 is a transcription factor crucial to the self-renewal and pluripotency of embryonic stem cells²¹⁴ and is associated with worse outcomes in PDAC²¹⁵. Besides OCT4, we used two other CSC markers: surface protein CD133 and glycolipid carbohydrate epitope SSEA4. A high expression of CD133 in PDAC patients (over 5%) has been linked to a 0% 5-year patient survival rate, compared to 23.5% for CD133-negative tumors²¹⁶, and PDAC metastasis was found to depend upon a subset of CD133-expressing cells²⁰⁰ which trigger early metastasis to lymph nodes²¹⁶. Finally,

the third stem cell marker, stage-specific embryonic antigen-4 (SSEA4), is known to be expressed in a wide range of cancers, including PDAC²¹⁷.

Next, we performed the small molecule compound screening in PDAC cells. These experiments aimed to discover molecular targets of small molecule compounds which affect the presence of CSC marker expressing cells (**Figure 1A-C**). For the small molecule screening, we measured CSC marker expression, cell growth and cell death by flow cytometry (CD133+/OCT4+/SSEA4+) after incubating the cells with the compounds for 5 days, which allowed us to identify effective compounds that impact the percentage of cells that express each of the CSC marker while also detecting cells that do not express these CSC markers. Hence the screening allows for the detection of differential effects on the subpopulations of non-CSCs and CSCs. We used a library of 142 small molecule compounds that have been verified to be active and targeting specific epigenetic modifying enzymes (**Figure 1C**). The most extensive inhibitor collection contained in the library were bromodomain inhibitors that represented approximately 26% of the compounds, followed by HDAC inhibitors (14%), Histone methyltransferase inhibitors (11.3%), compounds targeting methyl lysine binders/tudor domains (9.9%) and arginine methyltransferases (9.2%) and lysine demethylases (8.5%). Overall, this large screening identified compounds that significantly and reliably reduced the relative percentage of triple+ marker (CD133+/OCT4+/SSEA4+) expressing CSCs and also reduced cancer cell survival (**Figure 1C**). Our study identified I-BET 672 and JQ1 as candidate compounds, both of which have shown antitumour activity²¹⁸ and synergy with gemcitabine²¹⁹. This confirms the functionality and reliability of our screening results.

Since CSCs have lower sensitivity to chemotherapeutics besides their self-renewing capacity, we investigated the impact of BRD9 inhibition on the chemoresistance of CSCs in combination with Gemcitabine treatment (**Figure 1D**). BRD9 inhibition together with Gemcitabine treatment led to a stronger reduction in CSC sphere formation compared to either I-BRD9 or Gemcitabine treatment alone, suggesting that BRD9 inhibition sensitizes CSCs for Gemcitabine-mediated cell killing. In addition, BRD9 inhibition as well as TGF β /Activin pathway inhibition strikingly increased the sensitivity of CSCs to Gemcitabine, while the combination of I-BRD9 and Gemcitabine treatment together with TGF β /Activin pathway inhibition very efficiently eliminated the formation of spheres (**Figure 1E**).

To study the effect of BRD9 on PDAC cell growth *in vivo*, the xenograft PDAC mouse models were developed by subcutaneous injection of A13A CSC with different transduction/treatment (**Figure 1F**). At 2 months after cell inoculation, only mice implanted with the shRNA scramble adenovirus transduced cells (A13A^{Null}) developed tumours. No tumours were observed in other mice implanted with Gemcitabine-treated cells (A13A^{Gemcitabine}), shRNA-BRD9adenovirus-transduced cells (A13A^{BRD9-KD}), or A13A^{BRD9-KD} plus with Gemcitabine treatment (A13A^{BRD9-KD + Gemcitabine}), indicating the slower growth rate of these cells in contrast to A13A^{Null}. After 3 months of cell inoculation, although no tumours were observed in mice with A13A^{BRD9-KD + Gemcitabine}, the other mice implanted with A13A^{Null}, or A13A^{Gemcitabine}, or A13A^{BRD9-KD} developed tumours respectively. To further characterize these tumours, calliper measurement and PET-¹⁸F-FDG were performed to analyze the size and metabolic activity of these tumours. As expected, the

A13A^{Null} tumours grow much faster than other tumours from month 3 to month 4, as evidenced by the significantly increased tumour volume (**Figure 1G**). Although tumours formed by A13A^{Gemcitabine} or A13A^{BRD9-KD} were detected at month 3, they were moderate and grew slowly from month 3 to month 4. Importantly, BRD9-KD further delayed tumour growth compared with A13A^{Null} and A13A^{Gemcitabine} groups. These results were consistent with the observation obtained from PET-¹⁸F-FDG. At month 4 after cell inoculation, the ¹⁸F-FDG maximum standardized uptake value (SUVmax) in tumours formed by A13A^{Null} is much higher than in other tumours (**Figure 1H-I**). Gemcitabine significantly decreased the SUVmax in A13A^{Gemcitabine} tumours compared with A13A^{Null} tumours, and BRD9-KD further reduced the SUVmax of A13A^{BRD9-KD} tumours compared with Gemcitabine treatment. Consistently, the tumours harvested from A13A^{BRD9-KD}-implanted mice at month 4 was markedly smaller compared with tumours from A13A^{Null} or A13A^{Gemcitabine} implanted mice (**Figure 1J**). The lower tumour-weight to body-weight ratios of A13A^{BRD9-KD}-transplanted mice further demonstrated a progressive decrease in tumour mass (**Figure 1K**). In contrast, no tumours were detected in all mice with A13A^{BRD9-KD + Gemcitabine} and these mice were alive without any signs of sickness at month 4. When we further extended the time window to 6 months, only one of 6 mice in the A13A^{BRD9-KD + Gemcitabine} groups developed a small tumour (data not shown). At necropsy, no metastases were observed among these tumour-bearing mice and non-tumour bearing mice. These data suggested that inhibition of BRD9 reduces the tumorigenesis in PDAC.

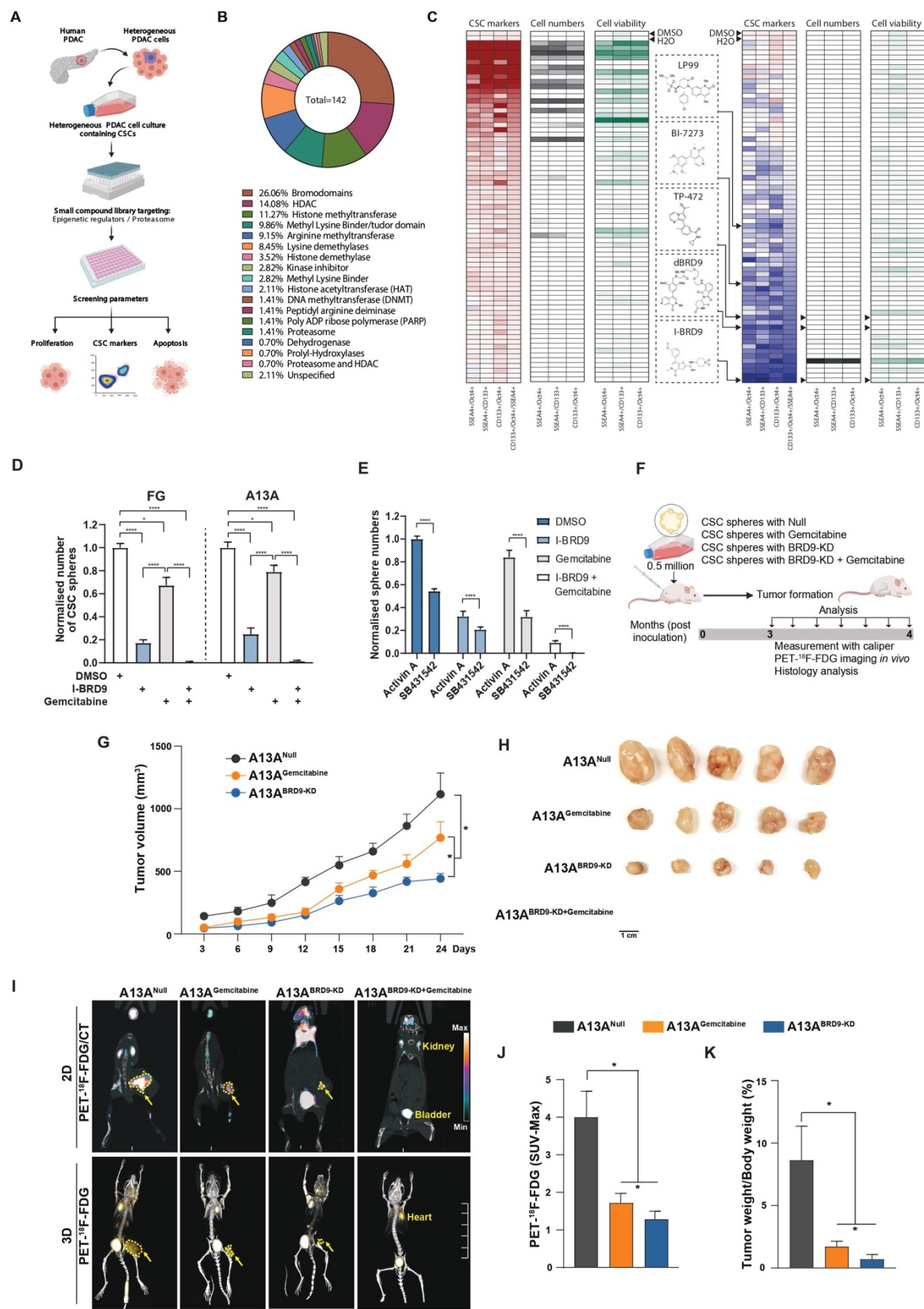


Figure 1. Inhibition of BRD9 reduces CSCs formation and tumorigenesis in PDAC. (A) Schematic depiction of the small molecule compound screening process on pancreatic cancer cells. (B) The classification of the compound library based on the percentage of the 142 compounds belonging to each class of enzymes they target. (C) BRD9 inhibitors decrease the relative numbers cells that express CSC markers OCT4-GFP, CD133 and SSEA4 as double positive cells or triple positive cells. Heat maps of chemical screening depicting the relative change in the expression of CSC markers, cell numbers and cell viability. Various BRD9 inhibitors, I-BRD9, dBRD9, TP-472, BI-7273, LP99, are marked in the heat map results. (D) TGF β /Activin signalling inhibition and BRD9 inhibition impair CSC self-renewal and sensitize CSCs for Gemcitabine-mediated destruction. (E) BRD9 inhibition sensitizes CSCs for Gemcitabine mediated elimination. Tumour sphere assays on cells treated with I-BRD9, Gemcitabine or by both simultaneously. (F) Schematic depiction of the mouse experiments. (G) Tumour volume measurements indicate reduced tumour sized upon BRD9 knockdown. (H) PET imaging of tumours in mice indicates reduced tumour sizes upon BRD9 knockdown and absence of visible tumours upon BRD9 knockdown with Gemcitabine treatment. The images show PET scan in 2D and in 3D with the tumour highlighted in yellow dotted areas. (I) PET- ^{18}F -FDG measurements of tumours indicate the abolishing of tumour growth in response to BRD9 knockdown. (J) Tumour size comparison in different treatments. Tumours from non-treated cells are larger compared to Gemcitabine treated and BRD9 knockdown PDAC cells, whereas BRD9 knockdown with Gemcitabine treatment resulted in undetectable tumours at 3 months of growth. (K) Relative tumour weight measurements in the different treatment conditions. The experiments for Figure 1A-E were designed by Siim Pauklin and performed by Andrei Stoica (Pauklin Lab) at University of Oxford. The experiments for Figure 1G-K was designed by me and performed by Wei Huang at University of Oxford.

4.2.2 BRD9 inhibition eliminates CSCs from patient tumour

CSCs are particularly challenging to eliminate due to their highly chemoresistant characteristics. Our experiments so far showed that treatment of PDAC cells with BRD9

inhibitors reduced the presence of CSCs while this effect was further enhanced by the combined treatment of BRD9 inhibitors and Gemcitabine. Although encouraging, these results were observed in PDAC cell lines while cells from primary pancreatic tumours could respond differently. To further examine the translational relevance of BRD9 as a candidate target gene, we utilized freshly isolated primary cancer patient samples and treated them with I-BRD9 or in combination with Gemcitabine for 72h followed by single-cell RNA-sequencing (**Figure 2A**) for ~12,000 cells. The analysis of this scRNA-sequencing data indicated different populations of cells based on their expression of genes characteristic for ductal cells (*KRT19*), fibroblasts (*LUM*) and stellate cells (*THY1*), and the further separation of pancreatic ductal cells to different sub-population of cancer cells (**Figure 2B, D**). One of these ductal cell subpopulations, designated as Ductal cells 3, expressed a range of cancer stem cell markers. This cancer cell subpopulation was reduced by I-BRD9 treatment (from 19.4% to 6.8%) and showed a ten-fold reduction in numbers (from 19.4% to 1.9%) upon the co-treatment with I-BRD9 and Gemcitabine (**Figure 2C**). The enriched Gene Ontology terms for downregulated genes in conditions I-BRD9/Gemcitabine and I-BRD9 versus DMSO and Gemcitabine indicated the central mechanisms that are relevant for PDAC development and CSCs including extracellular matrix organization, cell proliferation, cell migration, cell adhesion, regulation of apoptotic signalling pathways, and regulation of angiogenesis (**Figure 3A**), which are particularly relevant for tumour aggressiveness and metastatic processes. The enriched Pathways for downregulated genes in these same conditions uncovered TGF β /Activin A signalling pathway, fibrosis, VEGFA signalling, wound healing, autophagy pathways in cancer, proinflammatory/profibrotic mediators, NF κ B and YAP1/ECM axis (**Figure 3B**). Among

the genes that are specifically repressed by I-BRD9 treatment in the CSC population of cells were *SOX4* and *TWIST1* (**Figure 2E**), and *SNAIL2*, *JUN*, *IL6* and *IGF1* (**Figure 3C**). The combined treatment of I-BRD9 and Gemcitabine resulted in a synergistic reduction of these CSC factors (**Figure 2E and Figure 3C**). Altogether, the results using resected patient tumour samples indicated that BRD9 inhibition by a small molecule compound can efficiently target and eliminate the CSC subpopulation of pancreatic cancer cells, thus confirming our prior discoveries on established PDAC cell lines.

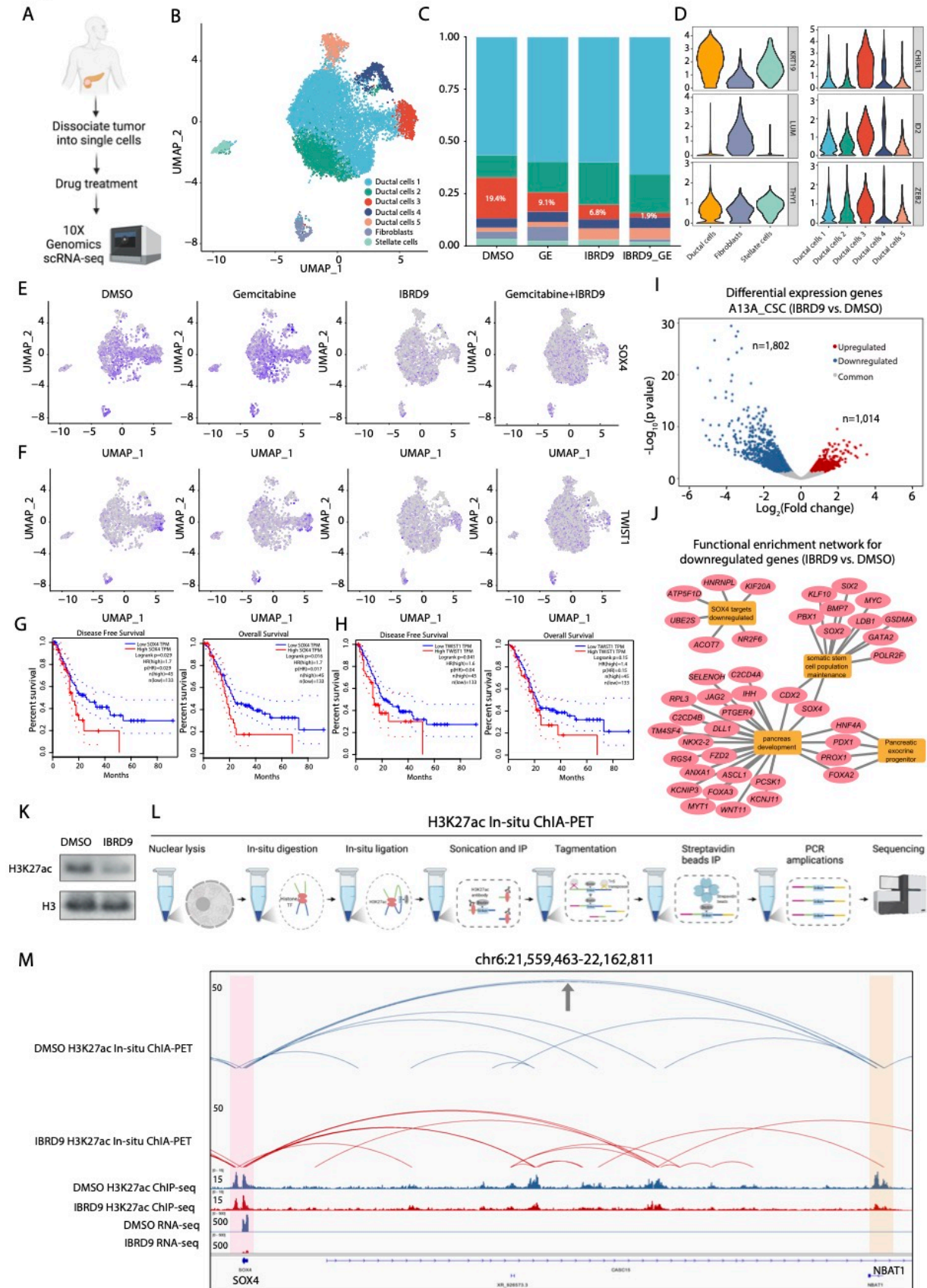


Figure 2. Inhibition of BRD9 reduces CSCs in patient tumors and the 3D enhancer-promoter connectome in CSCs. (A) Schematic depiction of single-cell RNA-sequencing strategy. The surgically removed PDAC tumour was dissociated to single cells and divided equally into four treatment conditions, control, Gemcitabine, I-BRD9 and Gemcitabine + I-BRD9. After 72 hours the samples were processed for single-cell RNA-sequencing. (B) Single-cell RNA-sequencing analyses of patient derived primary PDAC cells with clustering of cells. (C) BRD9 inhibition reduces the relative abundance of stem cell-like cells and enhances the effects of Gemcitabine. (D) Violin plots corresponding to marker gene expression in the different cell types in clusters. (E-F) BRD9 inhibition leads to the elimination of cells that express CSC genes (E) SOX4 and (F) TWIST1. Expression levels of SOX4 and TWIST1 in the CSC population is shown as dot plot graphs where blue colour indicates higher expression level compared to gray colour. (G-H) Higher expression of SOX4 and TWIST1 correlate with lower survival of pancreatic cancer patients, and lower disease-free survival. (I) Differential gene expression analysis upon BRD9 inhibition in CSCs. RNA-sequencing identified 1,802 downregulated genes and 1,014 upregulated genes in I-BRD9 treated samples compared to control cells. (J) Functional enrichment network of downregulated genes upon BRD9 inhibition uncover factors regulating stem cell maintenance, pancreatic development and pancreatic exocrine progenitors, as well as SOX4 target genes. (K) BRD9 inhibition reduces the abundance of H3K27ac in CSCs. Western blot analysis of the H3K27ac in control and I-BRD9 treated samples while total Histone H3 acts as a loading control. (L) Schematic depiction of analysing enhancer-promoter connectivity by H3K27ac in-situ ChIA-PET, H3K27ac ChIP-seq and RNA-seq. (M) BRD9 inhibition leads to the loss of enhancer-promoter looping at stem cell loci and loss of expression such as the SOX4 locus. The genomic locus of SOX4 shows H3K27ac abundance and SOX4 transcription together with 3D chromatin interactions.

Next, we performed bulk RNA-sequencing upon I-BRD9 inhibition in pancreatic CSCs which identified 1,802 downregulated genes and 1,014 upregulated genes (P-value

< 0.05) (**Figure 2I**). The functional enrichment network analysis of the downregulated genes by BRD9 inhibition revealed a number of central transcriptional regulators of somatic stem cell maintenance (e.g. *SOX4*, *SOX2*, *MYC*, *BMP7*), pancreatic exocrine progenitors/pancreatic development (e.g. *DLL1*, *JAG2*, *WNT11*, *NKX2-2*) and *SOX4* target genes (**Figure 2J**). GSEA pathway analysis of I-BRD9 downregulated genes indicated among the enriched pathways SMAD3 mediated transcription and Epithelial-to-Mesenchymal Transition (**Figure 4A**). Altogether these results emphasize the importance of key processes involved in regulating stem cell maintenance and metastasis that are downregulated upon BRD9 inhibition.

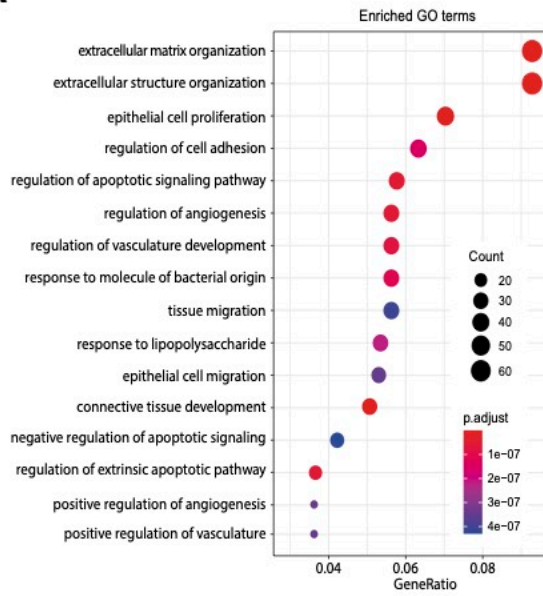
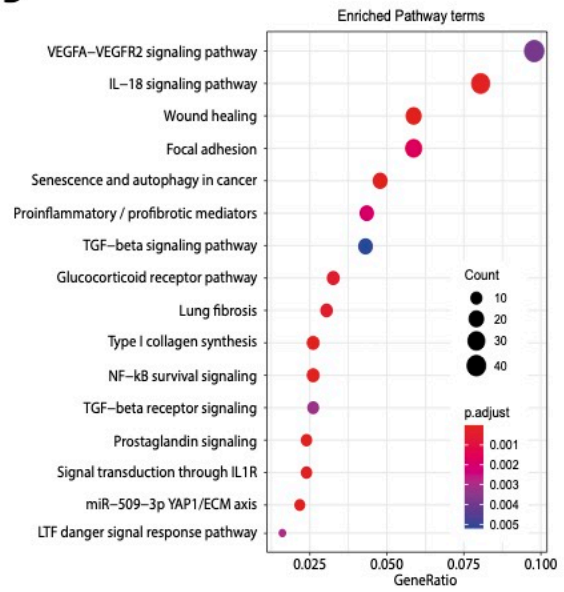
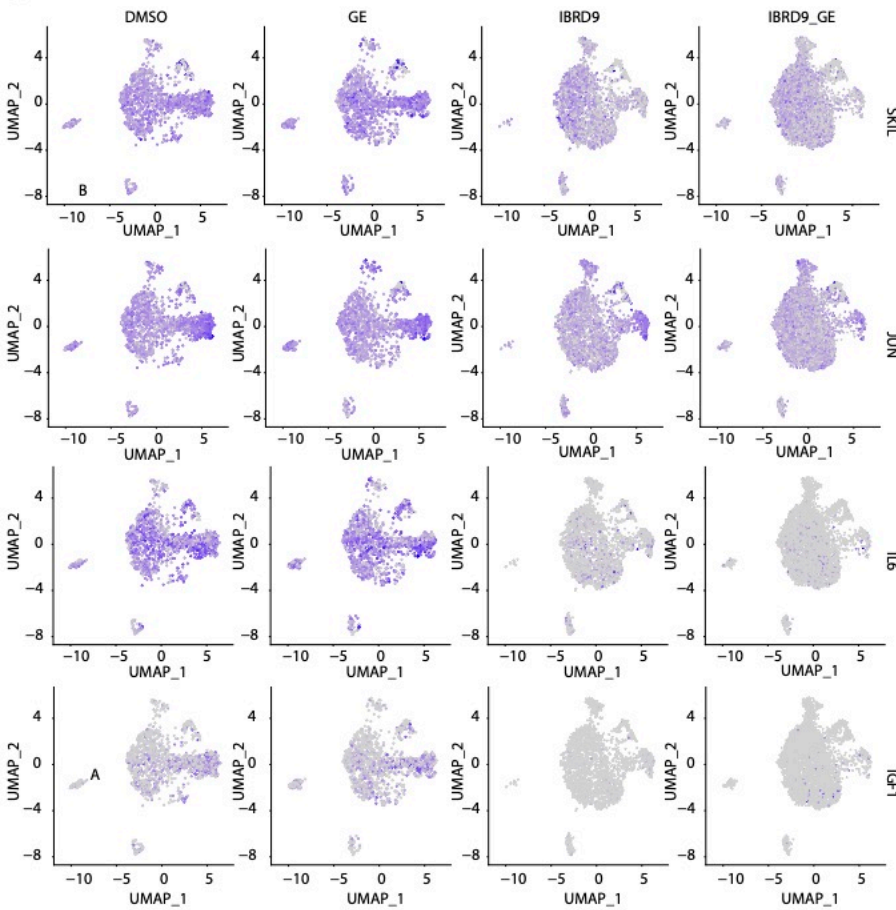
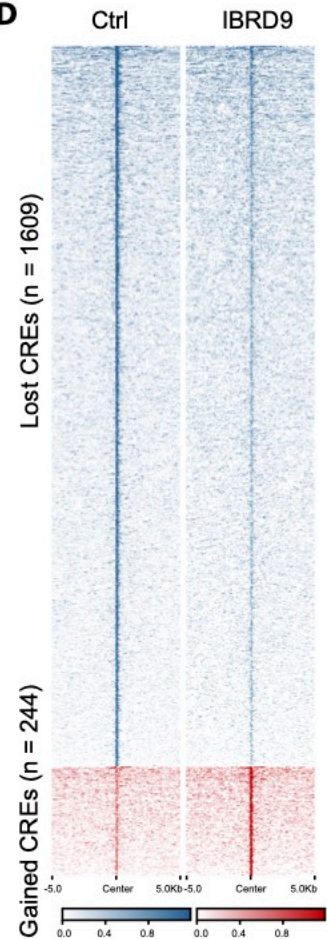
A**B****C****D**

Figure 3. BRD9 inhibition eliminates CSCs in patient-derived primary PDAC tumor cells. (A) Enriched GO terms for genes that are downregulated in I-BRD9 treated and I-BRD9/Gemcitabine treated sample compared to DMSO and Gemcitabine treated samples. (B) Enriched Pathway terms for genes that are downregulated in I-BRD9 treated and I-BRD9/Gemcitabine treated sample compared to DMSO and Gemcitabine treated samples. (C) BRD9 inhibition leads to the elimination of cells that are enriched in CSC population such as JUN, SOX4 and SKIL. Expression levels of JUN, SOX4 and SKIL in the CSC population is shown as dot plot graphs where blue colour indicates higher expression level compared to gray colour. (D) Heatmap of the lost and gained CREs upon BRD9 inhibition.

4.2.3 BRD9 inhibition disrupts enhancer-promoter connectome of stemness-related genes in pancreatic CSCs

In a cell, transcriptional output is largely orchestrated by cis-regulatory elements (CREs), in particular enhancers and promoters. By performing ATAC-seq assay, we identified that I-BRD9 treatment results in extensive loss of CREs (n=1,609) compared to control samples (**Figure 3D**). In line with this, western blot assay demonstrated that the enrichment of H3K27ac¹¹⁶ (the most important active enhancer and promoter histone mark) was significantly decreased in response to BRD9 inhibition (**Figure 2K**). In most cases, enhancers control gene expression through long-range interactions with promoters^{220,221}, but very little is known about the enhancer/promoter connectome in pancreatic CSCs. In order to study the effects of BRD9 inhibition on the enhancer/promoter connectome of CSCs, we performed H3K27ac In-situ ChIA-PET^{62,222,223} experiments, which capture H3K27ac-centric chromatin interactions (i.e. enhancer/promoter connectome) (**Figure 2L**). We uncovered that the transcriptional downregulation of stemness-related genes is possibly due to the loss of

enhancer/promoter connectome. For example, we found that in Control CSCs, *NBAT1* is characterized by H3K27ac enrichment on the promoter but it is transcriptionally silenced. In addition, the *NBAT1* promoter has strong chromatin interaction with the *SOX4* promoter, suggesting that this non-transcribing promoter may function as an enhancer¹³², to regulate the expression of *SOX4*. However, I-BRD9 treatment diminished the H3K27ac enrichment on the *NBAT1* promoter and its physical contact with the *SOX4* promoter, which led to transcriptional downregulation of *SOX4* (**Figure 2M**). Similarly, the transcriptional downregulation of other stemness-related genes (e.g. *CD133/PROM1*, *SMAD1* and *SNAI2*) are linked to the loss of enhancer-promoter connectome (**Figure 4B**). By analysing the cleavage profiles on ATAC-seq data we found SMAD2/3 footprints at the anchor regions (**Figure 4C-D**). SMAD3 has several motifs in JASPAR motif database and we found that MA0513 has 254 hits in Ctrl-specific ATAC peaks and 211 hits in I-BRD9-specific ATAC peaks, while MA1622 has 1370 hits in total on anchor regions with approximately similar number of hits in both conditions. Hence, SMAD2/3 footprints were not significantly lost upon BRD9 inhibition. On the other hand, for instance, the EMT inducer TWIST2 footprint was lost upon BRD9 inhibition (**Figure 4E**), suggesting that SMAD2/3 and BRD9 could co-bind to genomic regions and regulate the promoter-enhancer interactions in pancreatic CSCs through SMAD2/3 mediated complex recruitment. To further study the binding dependency of BRD9 and SMAD2/3 on TGF β /Activin pathway we performed BRD9 and SMAD2/3 ChIP-qPCR on representative SMAD2/3 binding regions in the proximity of *SOX4*, *CD133/PROM1*, *SNAI2* and *SMAD1* loci upon TGF β /Activin pathway inhibitor SB431542 treatment of pancreatic CSCs

(Figure 4F). BRD9 and SMAD2/3 bound to the anchor regions, while the binding of BRD9 on the anchor regions was lost following TGF β /Activin pathway inhibition.

These results indicate that BRD9 and SMAD2/3 are co-binding to these regions and SMAD2/3 transcription factors recruit BRD9 to the gene promoter and enhancer regions **(Figure 4G)**. Collectively, our results revealed the cooperation of BRD9 and SMAD2/3 in regulating the enhancer-promoter connectome and gene expression of stemness-related genes in pancreatic CSCs.

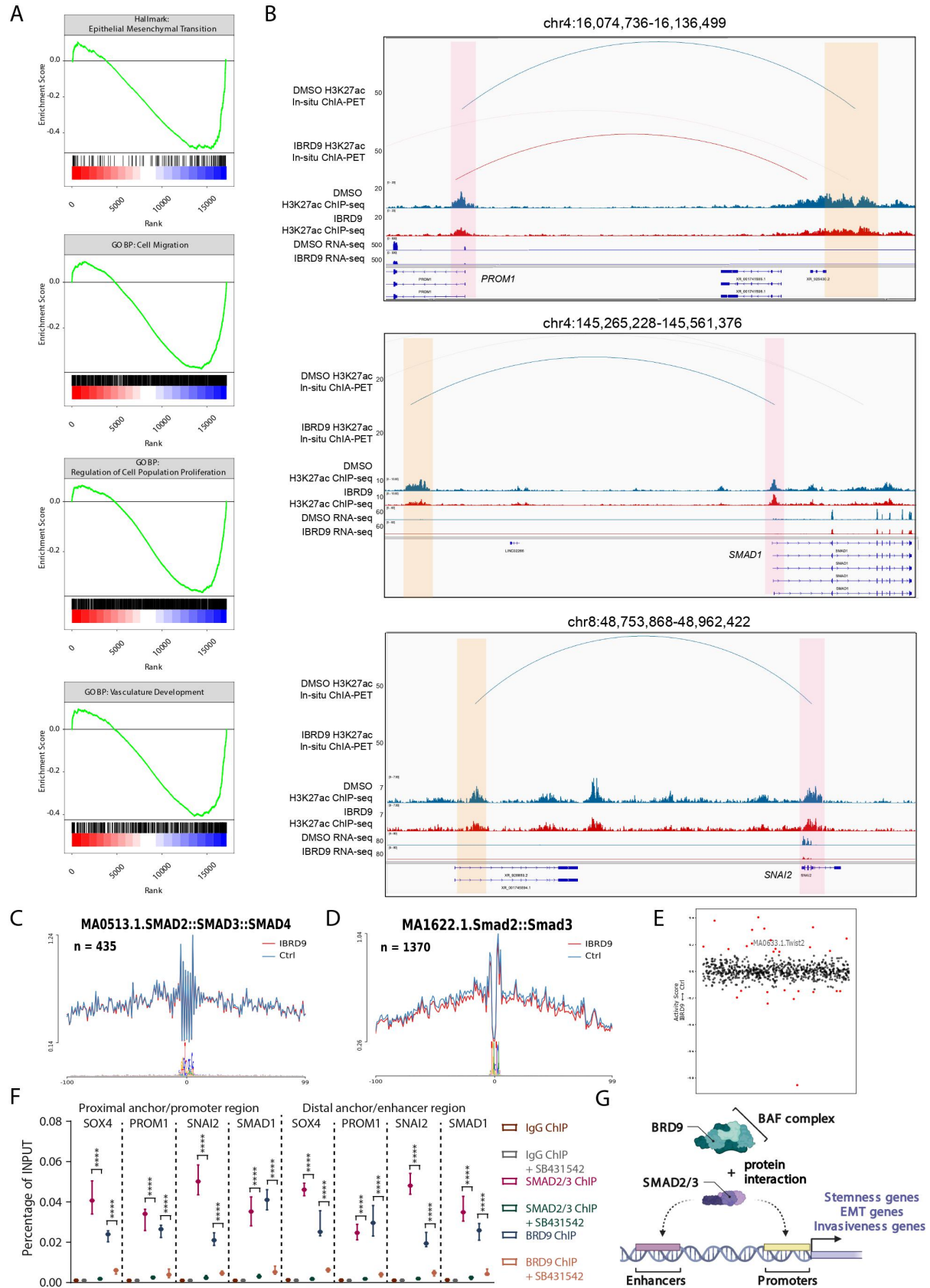


Figure 4. BRD9 inhibition changes the enhancer-promoter connectome of CSCs. (A) GSEA analysis of genes regulated by BRD9 inhibition indicate the enrichment of genes involved in epithelial to mesenchymal transition, cell migration, regulation of cell population proliferation and vasculature development. (B) BRD9 inhibition leads to the loss of enhancer-promoter looping and loss of expression at stem cell and EMT loci such as CD133/PROM1, SMAD1 and SNAI2. The genomic loci show H3K27ac abundance and gene transcription together with 3D chromatin interactions. (C-D) Transcription factor footprint analysis identified SMAD3 footprints based on JASPAR motif database at the ATAC-seq peaks of the anchor regions. (C) MA0513 has 254 hits in Ctrl-specific ATAC peaks and 211 hits in I-BRD9-specific ATAC peaks on anchors. (D) MA1622 has a total of 1370 hits in Ctrl-specific ATAC and I-BRD9-specific ATAC peaks on anchors. The depth of the footprinting is not statistically different in control and BRD9-I treated samples. (E) The presence of transcription factor cleavage sites at anchor regions. The red dots above zero (y axis) are the Ctrl specific TF footprinting profiles and the read dots below zero (y axis) are the I-BRD9 sample specific TF footprinting profiles. The footprint profile of TWIST2 is highlighted. (F) TGF β /Activin signalling leads to the loss of SMAD2/3 and BRD9 binding to the 3D chromatin looping anchors at promoters and enhancers of stem cell and EMT loci in CSCs. ChIP-qPCR of proximal and distal anchor regions at SOX4, CD133/PROM1, SNAI2, and SMAD1 loci in CSCs +/- SB431542. (G) Schematic depiction of the function of SMAD2/3 and BRD9 in regulating the 3D chromatin looping between enhancer and promoter regions near stemness genes, EMT and invasiveness genes. ChIP-qPCR was performed by Siim Pauklin.

4.3 Discussion

Epigenetic regulation of gene expression is essential for guiding developmental processes throughout embryogenesis and tissue homeostasis in the adult organism. The cellular identity of differentiated somatic cells can also be epigenetically reprogrammed through erasure and rewriting of epigenetic memory on chromatin to a pluripotent stem

cell state, or the conversion of cell identities from one differentiated cell type to another. Similar differentiation and dedifferentiation processes mediated by deregulated epigenetic regulatory enzymes contribute to tumour formation, invasiveness, metastatic processes, dormancy and reactivation of cancer cells, clonal evolution of tumour cells and the development of therapeutic resistance in cancers. Of particular importance to all these processes, epigenetic mechanisms regulate stem cell-like phenotypic plasticity and self-renewal capacity of cancer stem cells. Therefore, targeting the epigenetic mechanisms offers an attractive strategy to eliminate cancer stem cells through reducing their self-renewal characteristics or re-sensitizing them for chemotherapeutic drugs.

We used an unbiased approach for identifying candidate target enzymes that regulate the stem cell-like characteristics of pancreatic cancer stem cells by performing an extensive screening experiment in PDAC cells. The screening library that we used consisted of functionally verified small molecule compounds covering all main epigenetic regulatory enzyme classes with compounds that have bioactivity in the cellular context. Most of the epigenetic regulatory enzymes were targeted by multiple small molecules which made the screening process more robust due to expected similar effects for compounds targeting the same enzyme. Besides this, the library had negative control in the form of inactive compounds. Using this approach, we identified BRD9 as a novel epigenetic regulator of pancreatic CSC and the corresponding small molecule inhibitors for this enzyme.

Our research validated the translational relevancy of our preclinical models by using surgically removed tumor samples from PDAC patients and also by tumorigenesis experiments in mice. The BAF complex with the BRD9 epigenetic modulator serves as

novel, first-in-class anti-cancer compound that is expected to enter clinical testing in the future ²²⁴. Its paradigm-shifting novel mechanism of action is particularly attractive in pancreatic cancer that has been in urgent need for better selection of targets for the bromodomains and improved pharmacology to enhance exposure, in particular intratumoral drug penetration and killing of the heterogeneous cancer cell populations including the chemoresistant CSCs. The BRD9 could be used in combination with currently available treatments such as Gemcitabine or FOLFIRINOX with the objective of re-sensitizing pancreatic CSCs that are particularly resistant to currently used chemotherapies. Hence, the improved molecular target identification serves a dual purpose; in addition to improving drug selectivity, our discoveries will guide effective therapy combination with either standard of care or other investigational agents. The 3D chromatin architecture provides a higher level of epigenetic regulation to gene expression. Our results indicated upon BRD9 inhibition, enhancer-promoter looping at various cancer stemness genes such as SOX4 are lost, thus leading to stem cell gene repression and loss of CSC self-renewal. Targeting the regulators of 3D connectomes is thus emerging as an attractive therapeutic strategy for eliminating the stem cell-like populations in PDACs. The research findings from our project could be imminently used to benefit these PDAC patients that have a surgically unresectable cancer and patients that have already undergone standard chemotherapy treatments without benefit. Since this strategy is not limited to PDAC therapeutics, our project will provide valuable information for the broader field of cancer researcher and oncologist beyond PDAC.

Collectively, our research discoveries will be extremely important, because they will make it possible to use novel epigenetic machineries as therapeutic targets for specifically killing the harmful cancer-forming cells in patients.

4.4 Materials and Methods

4.4.1 Cell lines and cell culture

The A13A cells were provided by Christine Iacobuzio-Donahue at The Memorial Sloan Kettering Cancer Center. FG cells were provided by Isaiah J. Fidler at The University of Texas MD Anderson Cancer Center. For standard cell cultures, cells were grown at 37°C humidified incubator containing 5% CO₂ in Dulbeccos modified Eagles medium (DMEM), high glucose, GlutaMAX™ Supplement, pyruvate (Thermo Fisher Scientific, Inc.), 100 U/ml penicillin, 100 mg/ml streptomycin (Thermo Fisher Scientific, Inc.), MEM Non-Essential Amino Acids Solution 1X (Life Technologies; Thermo Fisher Scientific), MEM Vitamin Solution 1X (Life Technologies; Thermo Fisher Scientific) and 10% inactivated fetal calf serum (FCS; Life Technologies). For 3D cultures, cells were grown in ultra-low attachment plates (Thermo Fisher Scientific, Inc.) in Dulbecco's Modified Eagle Medium/F12 (Sigma-Aldrich) supplemented with 2.5mM L-Glutamine (Thermo Fisher Scientific, Inc.), 1 % B27 supplement (Life Technologies), 100 U/ml penicillin, 100 mg/ml streptomycin (Thermo Fisher Scientific, Inc.), 20ng/ml basic thermostable fibroblast growth factor (FGF-Basic TS, Proteintech) (5,000 cells/ml).

4.4.2 Tumour sphere assays

Cells are seeded into ultra-low attachment plates in Dulbecco's Modified Eagle Medium/F12 (Sigma-Aldrich) supplemented with 2.5mM L-Glutamine (Thermo Fisher Scientific, Inc.), 1 % B27 supplement (Life Technologies), 100 U/ml penicillin, 100 mg/ml streptomycin (Thermo Fisher Scientific, Inc.), 20ng/ml basic thermostable fibroblast growth factor (FGF-Basic TS, Proteintech)(5,000 cells/ml). For tumour sphere formation assay, cells were passaged after one-week incubation, and grown for another week after which the tumour sphere numbers were counted under a phase-contrast microscope using the 40x magnification lens or by Celigo Image Cytometer (Nexcelom). For RNA-seq and ATAC-seq experiments, serial passaging of the first-generation tumour spheres is required. After 7 days of incubation the tumour spheres were harvested by using a 40 µm cell strainer and centrifuged for 5 min at 200 x g at RT. Dissociate the pellet of tumour spheres to single cells using trypsin, and then expanded for another 4 days before performing the treatments of samples.

4.4.3 The small molecule screening library

The screening library contained concentrated small molecule compounds with verified biochemical activity against their targets. Most of the compounds target epigenetic regulators with high specificity.

Table 1. List of small molecule compounds

Vial	Source Name	Alternative Name/Notes	[Working] uM	Class/Target
1	(+)-JQ1	SGC PROBE, (BET Probe), (+)-SGCBD01, JQ1S	1	Bromodomains - BRD2, BRD3, BRD4, BRDT (BET)
2	(-)-JQ1 (inactive)	(-)-SGCBD01, JQ1R	1	Bromodomains - Negative control
3	PFI-1	SGC PROBE, (BET probe), Pfizer-1	5	Bromodomains - BRD2, BRD3, BRD4, BRDT (BET)
4	I-BET	GSK525762A	1	Bromodomains - BRD2/3/4
5	Bromosporine	"E12348"	1	Bromodomains - pan-Bromodomain

6	CBP/BRD4 (0383)	DH1-65-3 isoxazole series "CBP/EP300" [10x selectivity for BRD4 only], Global ID from vial	5	Bromodomains - CBP, BRD4(1)
9	SGC-CBP30	SGC PROBE, (CREBBP/EP300 Probe)	1	Bromodomains - CREBBP, EP300
10	I-CBP112	SGC PROBE, (CREBBP/EP300 Probe), (Racemate) 113071004, benzoxazepine series "E12179a"	1	Bromodomains - CREBBP, EP300
15	RVX-208	RVX 208, RVX000222, Clinic	5	Bromodomains - BRD2, BRD3, BRD4, BRDT (BET, BD2)
16	SMARCA	SMARCA2/4 & PB1 probe candidate, "E13202a"	2.5	Bromodomains - SMARCA, PB1
17	PB1/SMARCA	PB1/SMARCA, "E39191a"	1	Bromodomains - SMARCA, PB1
18	PFI-3	SGC PROBE, SMARCA2/4 and PB1(5)	1	Bromodomains - SMARCA2/4, PB1(5)
19	GSK2801	SGC PROBE, (BAZ2B/A Probe), Formally "BAZ2B" "E12871"	1	Bromodomains - BAZ2A, BAZ2B
20	PFI-4	E47658a, Pfizer BRPF1	1	Bromodomains - BRPF1B
21	TRIM24/BRPF	TIF1a, E23057a	10	Bromodomains - TRIM24/BRPF
22	OF-1	E21612a, pan-BRPF	5	Bromodomains - pan-BRPF
23	Belinostat	PXD101	5	HDAC - hydroxamic acids
24	CXD101	previously AZD 9468	1	HDAC -
25	Valproic acid		1000	HDAC - aliphatic acid compounds
26	Entinostat	SNDX-275, MS-275	0.5	HDAC - ortho-amino anilides
27	SAHA	Zolinza, Vorinostat	2.5	HDAC - hydroxamic acids
28	Trichostatin A		0.5	HDAC - hydroxamic acids - Class I & II
29	SRT1720		1	HDAC - SIRT1 (indirect?) activator
30	EX 527		1	HDAC - SIRT1
32	CI-994	PD 123654, Goe 5549, Tacedinaline, N-Acetyldinaline	1	HDAC - 1,2,3,(8)
33	CPI-360		10	Histone methyltransferase - EZH2 and EZH1
34	CPI-413		10	Histone methyltransferase - EZH2 and EZH1
35	UNC0638	SGC PROBE, (G9a Probe) [BIX01294 Derivative]	1	Histone methyltransferase - G9a, GLP
37	UNC0642	SGC PROBE, (G9a/GLP Probe) (G9a-brain penetrant)	1	Histone methyltransferase - G9a, GLP
38	A-366	SGC PROBE, (G9a/GLP Probe)	2	Histone methyltransferase - G9a, GLP
39	Chaetocin		0.05	Histone methyltransferase - SUV39H1
43	PFI-2	SGC PROBE, (SETD7 Probe)	2	Histone methyltransferase - SETD7
45	SGC0946	SGC PROBE, (DOT1L Probe), APS-V-33	7.5	Histone methyltransferase - DOT1L
46	GSK343	SGC PROBE, (EZH2 Probe)	3	Histone methyltransferase - EZH2
47	UNC1999	SGC PROBE, (EZH2/1 Probe)	1	Histone methyltransferase - EZH2
49	LLY-507	SGC PROBE, (SMYD2 Probe)	1	Histone methyltransferase - SMYD2
50	Tranylcypromine		20	Lysine demethylases - LSD1
51	GSK-LSD1 (irreversible)	SGC PROBE, (LSD1 Probe) (trans-racemic) Di-HCl	0.5	Lysine demethylases - LSD1
52	GSK690	GSK-LSD1 (reversible)	5	Lysine demethylases - LSD1
53	GSK J4	Var GSK-J1 SGC PROBE, (KDM probe)	10	Lysine demethylases - JMJD3, UTX, JARID1B
54	GSK J5 (inactive)		10	Lysine demethylases - Negative control
55	IOX1 (5-carboxy-8HQ)	SGC TOOL, 8-Hydroxy-5-quinolinecarboxylic acid, 5-Carboxy-8-hydroxyquinoline	40	Lysine demethylases - pan-2-OG

57	Methylstat (Ester)		2.5	Histone demethylase
58	(E)-JIB-04	(also synthesis KSE270E)	0.05	Histone demethylase - Pan JmjC
60	ML324	NCGC00183808-03, "S05185a"	5	Histone demethylase - JMJD2E
61	IOX2	SGC PROBE, (PHD2 probe)	10	Prolyl-Hydroxylases - PHD2 (EGLN1)
62	OICR-9429	WDR5	1	Methyl Lysine Binder? - WDR5
63	UNC1215	SGC PROBE, (L3MBTL3 Probe)	5	Methyl Lysine Binder - L3MBTL3
64	5-Azaciditidine	Vidaza, azacytidine	10	DNA methyltransferase (DNMT) -
65	5-Azadeoxycytidine	Decitabine, 2'-deoxy-5-azacytidine, DAC	5	DNA methyltransferase (DNMT) - DNMT1/3
66	Olaparib	KU-0059436	1	Poly ADP ribose polymerase (PARP)
67	Rucaparib	AG-014699, PF-01367338	10	Poly ADP ribose polymerase (PARP)
68	K00135		1	Kinase inhibitor - ATP competitive - PIM
69	5-Iodotubercidin	"Haspin"	1	Kinase inhibitor - ATP mimetic - Haspin
70	C646	4-[4-[[5-(4,5-Dimethyl-2-nitrophenyl)-2-furanyl]methylene]-4,5-dihydro-3-methyl-5-oxo-1H-pyrazol-1-yl]benzoic acid	1	Histone acetyltransferase (HAT) p300/CBP
74	DUAL1946		1	
75	GSK484		1	Peptidyl arginine deiminase (PAD4)
76	KDOBA67	GSK-J1 analog (no esterase), J498-140, J552-62	10	Histone demethylase
77	BAZ2-ICR	SGC PROBE, (BAZ2B/A Probe)	1	Bromodomains - BAZ2A, BAZ2B
78	NI-57	SGC PROBE, (pan-BRPF Probe)	1	Bromodomains - pan-BRPF
79	LP99	SGC PROBE, (BRD9/7 Probe)	1	Bromodomains - BRD9, BRD7
80	SGC707	SGC PROBE, (PRMT3 Probe)	1	Arginine methyltransferase - PRMT3
81	RGFP966		10	HDAC - HDAC3
82	PCI-34051		5	HDAC - HDAC8
83	Rocilinostat		10	HDAC - HDAC6
84	Tubastatin A HCl		10	HDAC - HDAC6
85	KDOAM-25a		1	Lysine demethylases - JARID
86	KDM5-C70		10	Histone demethylase - JARID1
88	MAZ1805	Halofuginol (inactive)	1	t-RNA sythetase
89	MAZ1392	Halofuginone (Active)	1	t-RNA sythetase
90	BI-9564	BDH35145960	1	Bromodomains - BRD9, BRD7
91	NVS-CECR2-1	BDV00052666	1	Bromodomains - CECR2
92	GSK106	inactive control for GSK484	1	Peptidyl arginine deiminase (PAD4)
93	J556-42R		1	Arginine methyltransferase - PRMT5
94	J556-63R		1	Arginine methyltransferase - PRMT5
95	J556-70R		1	Arginine methyltransferase - PRMT5
96	A-196		1	Histone methyltransferase - SUV420H1/H2
97	BAY-598		1	Histone methyltransferase - SMYD2
98	J556-143		1	Arginine methyltransferase - PRMT5
99	MS049		1	Arginine methyltransferase
100	MS023		1	Arginine methyltransferase - Type I PRMTs
101	MS003	inactive control for MS023	1	Arginine methyltransferase - negative control

104	SGI-1776		10	Kinase inhibitor - Haspin
105	CHR-6494		1	Kinase inhibitor - Haspin
106	CPI-169		10	Histone methyltransferase - EZH2, EZH1
107	UNC2400	inactive control for UNC1999	1	Histone methyltransferase - EZH2
108	GSK864		5	Dehydrogenase
109	GSK8814	GSK814	10	Bromodomains - ATAD2
110	GSK8815	inactive control for GSK814, GSK815	10	Bromodomains - ATAD2
111	GSK959		1	Bromodomains - BRPF1
112	NVS-CECR2-C	inactive control for NVS-CECR2-1	1	Bromodomains - CECR2
113	BAY-299		1	Bromodomains - BRD1, TAF1
114	PCI-24781	Abexinostat	10	HDAC -
115	Romidepsin		1	HDAC -
116	Mocetinostat	MGCD0103	10	HDAC -
117	Santacruzamate		50	HDAC 2?
118	KDOAM32	inactive control for KDOAM25	1	Lysine demethylases - JARID
120	MS409N	inactive control for MS049	1	Arginine methyltransferase - PRMT4, PRMT6 inactive control
121	TP-064		1	Arginine methyltransferase - PRMT4
122	TP-064N	inactive control for TP-064	1	Arginine methyltransferase - PRMT4
123	A-395		1	Methyl Lysine Binder - EED
124	A-395N		1	Methyl Lysine Binder - EED
125	I-BRD9		10	Bromodomains - BRD9
126	TP-472		1	Bromodomains - BRD9
127	TP-472N	inactive control for TP-472	1	Bromodomains - BRD9
128	KDOPZ-32a		1	Lysine demethylases - KDM5
129	KDOOA012000		1	Lysine demethylases KDM2
130	AMI-1		50	Arginine methyltransferase - PRMT
131	TMP269		10	HDAC -4, 5, 7 & 9
132	AGK2		10	HDAC - SIRT2
133	GSK6853		1	Bromodomains - BRPF1/2/3
134	GSK9311	inactive control for GSK6853	1	Bromodomains - BRPF1/2/3
135	LLY-283		1	Arginine methyltransferase - PRMT5
136	TD001851a	Spin1, TDOSI000054a	1	Methyl Lysine Binder/tudor domain - Spin1
137	TDOSI000058a	Spin1	1	Methyl Lysine Binder/tudor domain - Spin1
138	TD001863a	Spin1, TDOSI000060a	1	Methyl Lysine Binder/tudor domain - Spin1
139	TDOSI000062a	Spin1	1	Methyl Lysine Binder/tudor domain - Spin1
140	TD001857a	less active 1851 ctrl spin1, TDOSI000063a	1	Methyl Lysine Binder/tudor domain - Spin1
141	TD001856a	less active 1851 ctrl spin1, TDOSI000064a	1	Methyl Lysine Binder/tudor domain - Spin1
142	TD001858a	less active 1863 ctrl spin1, TDOSI000065a	1	Methyl Lysine Binder/tudor domain - Spin1
143	TMP195		1	HDAC -4,5,7,9
144	GSK2879552	irreversible	10	Lysine demethylases - LSD1

145	TDO20821a	Spin1	1	Methyl Lysine Binder/tudor domain - Spin1
146	TDO20824a	Spin1 VinSpinIn SGC probe	1	Methyl Lysine Binder/tudor domain - Spin1
147	TDO20823a	Spin1	1	Methyl Lysine Binder/tudor domain - Spin1
148	A-485		1	Histone acetyltransferase (HAT) p300/CBP
149	A-486	inactive control	1	Histone acetyltransferase (HAT) p300/CBP
150	GSK4027		1	Bromodomains - PCAF, GCN5
151	GSK4028	inactive control	1	Bromodomains - PCAF, GCN5
152	L-Moses		1	Bromodomains - PCAF, GCN5
153	D-Moses	inactive control	1	Bromodomains - PCAF, GCN5
154	PFI-5		1	Histone methyltransferase - SMYD2
155	YX39-31b		1	Methyl Lysine Binder/tudor domain - Spin1
156	TDO208229	Spin1	1	Methyl Lysine Binder/tudor domain - Spin1
157	TDO01856a	Neg control for TDO01000064a	1	Methyl Lysine Binder/tudor domain - Spin1
158	TDO20826a	Spin1 (negative control)	1	Methyl Lysine Binder/tudor domain - Spin1
159	Bortezomib		0.1	Proteasome
160	Carfilzomib		0.1	Proteasome
161	RTS-V5	Bort/Hdac	1	Proteasome and HDAC
162	dBRD9	from Qi Lab	1	Bromodomains - BRD9
163	BI-7273		0.1	Bromodomains - BRD9/7 (IC50 19/117nM)
164	CPI-621		0.1	Lysine demethylases - KDM5

4.4.4 Screening of the chemical compounds

The cells were grown in 96-well plates in standard growth medium with puromycin (1 µg/ml stock). Three technical replicates and three biological replicates were used for the screening. Cells were plated at a concentration of 10,000 cells in 100 µl of media per well in a 96-well plate. One day after plating the cells, the medium was changed to standard growth medium supplemented with puromycin (1 µg/ml) and Activin A (5 ng/ml). On the same day, the compounds were added: first, 100x compound library dilutions were made, and 10uL of 100x diluted chemical was added to each well. Cells were then cultured with chemical compounds for five days with one media exchange supplemented by fresh compounds on day 3. Each replicate was analyzed using Celigo Image Cytometer

(Nexcelom) and flow cytometry. Cells were lifted and dissociated to single cells with TrypLE. Details on the antibodies that were used for flow cytometry are listed in Table 2. The cells were incubated with antibodies in 1% BSA-PBS for 30 minutes on ice, and washing was performed using 1% BSA-PBS. Cells were then incubated with secondary antibodies for 30 min on ice and washing was repeated as before. The cells were then suspended in 300 uL 1% BSA-PBS with DAPI (1:2000) for live/dead separation and kept on ice to be used for the flow cytometry analysis. Antibody for flow cytometry in the compound screening experiment is listed below:

Table 2. List of antibodies for flow cytometry

Antibody	Target/function	Ratio	Number of cells
CD133-BV786, Mouse Anti-Human, clone W6B3C1, BD 747640	CD133	1.2 uL in 100 uL	300,000
Mouse IgG1-BV786l, BD 563330	isotype control for CD133	1.2 uL in 100 uL	300,000
Alexa647 Mouse anti- SSEA-4 clone MC813-70	SSEA4	1.5 uL to 100 uL	300,000
Alexa647 Mouse IgG3	isotype control for SSEA4	1.5 uL to 100 uL	300,000

4.4.5 Establishment of xenograft models in NSG mice

The A13A cells for subcutaneous injection were divided into the following groups, 1) Null group, cells were transduced with scrambled shRNA adenovirus (Ad-GFP-U6-shScram, Vector Biolabs) containing a scramble sequence under the control of U6 promoter, with the GFP co-expression under a cytomegalovirus (CMV); 2) BRD9 KD group, cells were

transduced with BRD9-targeted shRNA adenovirus (Ad-GFP-U6-shBRD9, Vector Biolabs); 3) Gemcitabine group, cells were treated with gemcitabine (7.5 μ M, Selleckchem, LY-188011); 4) BRD9 KD + Gemcitabine group, cells transduced with Ad-GFP-U6-shBRD9 were treated with gemcitabine additionally. These cells were maintained at a low-density culture (1250 cells/ml, 312.5 cells/cm²) and harvested for subsequent experiments 72 h post transduction/treatment.

Tumor studies were performed in female 8-10 weeks old NSG mice (Jackson lab, 005557). For tumor formation, 0.5×10^6 cells in 200 μ l of Matrigel/DMEM media mixture (1:1 ratio) were subcutaneously injected into the right flank of NSG mice.

4.4.6 Tumor growth and mouse survival

For tumor growth, mice were monitored twice a week by observation and palpation. Tumor size was measured with a digital caliper and tumor volume was calculated using the formula $L \times W^2 \times 0.5$. Mouse body weight was measured twice weekly. Death, weight loss of 15% or more of body weight, tumor volume with 2,000 mm³ or more, or severe labored breathing or dyspnea were considered endpoints and all mice were euthanized.

4.4.7 In vivo PET imaging

Mice were fasted for 12 h before ¹⁸F- fluorodeoxyglucose (FDG) injection but allowed free access to water. After anaesthesia with 2% isoflurane, mice were injected with ¹⁸F-FDG (11 to 16 MBq, German Cancer Research Center, Heidelberg, Germany) by tail vein and kept at 37°C until imaging. Imaging was started at 60 minutes after ¹⁸F-FDG injection with non-invasive micro-positron emission tomography (μ PET, Siemens Inveon). A CT

scan (80 kVp, 500 μ A, at 120 projections; approximately 4-minutes) was acquired for anatomical reference and enabled PET attenuation correction during reconstruction. The ^{18}F -FDG maximum standardized uptake value (SUV_{max}) in tumors was calculated as being the ratio of maximal FDG uptake to the administered FDG dose normalized to body weight. Postprocessing was carried out with Inveon Research Workplace.

4.4.8 Preparation and Sequencing of Illumina RNA libraries

Total RNA was isolated by RNeasy RNA Extraction Kit (Qiagen) according to manufacturer's guidelines. RNA-Seq libraries were created using the NEBNext Ultra RNA library prep kit using TruSeq indexes, following the manufacturer's recommendations. In summary, 500 ng of total RNA was used to isolate mRNA poly(A) by two rounds of purification using oligo dT magnetic beads followed by fragmentation and cDNA synthesis by random primers and reverse transcriptase. Bar-coded adapters were ligated to the cDNA fragments and a PCR reaction was performed to produce the sequencing libraries. To verify the library concentration and the library fragments length was used Agilent 2200 Tape-Station System. Adapter-ligated cDNA fragment libraries were sequenced on a NextSeq 500 platform (Illumina) using a paired-end run (2 $\times$ 41 bp).

4.4.9 RNA-sequencing analysis

Sequencing reads from the RNA-seq experiment were aligned to the human genome (hg38) using HISAT2 with the default parameters. FeatureCounts was used to assign mapped reads to genes with annotation gtf file Ensembl94. Differential gene expression analysis was performed using DESeq2 with IHW method for p value

adjustment, apegln method for effect size shrinkage, FDR-adjusted p value < 0.05 and log2 fold change > 0. Functional analysis of differential expressed genes was performed using Metacore. All plots were generated using R package 3.6.

4.4.10 Chromatin Immunoprecipitation (ChIP)

All steps were performed on ice or at 4°C and ice-cold buffers and PBS were supplemented with 1mg/ml Leupeptin, 0.2mM PMSF, and 10mM NaButyrate were used unless otherwise stated. Approximately 5×10^6 cells were used per sample and cross-linked with 1% formaldehyde for 15 minutes. Cross-linking was stopped by incubating samples with glycine at a final concentration of 0.125M for 5 minutes at room temperature, and the cells were washed with PBS followed by pelleting at 250g for 5 minutes. The pellet was re-suspended in 2ml ChIP Cell Lysis Buffer (CLB: 10 mM Tris pH8, 10 mM NaCl, 0.2% NP-40) and incubated for 10 minutes to lyse the plasma membranes. Nuclei were pelleted at 600g for 5 min, lysed in 1.25ml of ChIP Nuclear Lysis Buffer (NLB: 50 mM Tris pH8, 10mM EDTA, 1% SDS) for 10 minutes, and then 0.75ml of ChIP Dilution Buffer (DB: 20 mM Tris pH8, 2mM EDTA, 150mM NaCl, 0.01% SDS, 1% Triton X-100) was added to the samples. Chromatin was sonicated in 15ml Diagenode Bioruptor Pico water bath sonicator with an automated water cooling system, by performing 30 cycles of 30 seconds ON, 45 seconds OFF. This protocol resulted in the homogeneous generation of fragments of 100-400bp. Samples were clarified by centrifugation at 16000g for 10 minutes, and diluted with 3.5ml of DB. After pre-clearing with 10µg of non-immune IgG for 1h and 50µl of Protein G-Agarose for 2h, ChIP was performed overnight in rotation using specific antibodies or non-immune IgG as a control.

After incubation for 1 hour with 30µl of Protein G-Agarose, beads were washed twice with ChIP Washing Buffer 1 (WB1: 20mM Tris pH8, 2mM EDTA, 50mM NaCl, 0.1% SDS, 1% Triton X-100), once with ChIP Washing Buffer 2 (WB2: 10mM Tris pH8, 1mM EDTA, 0.25M LiCl, 1% NP-40, 1% Deoxycholic acid), and twice with Tris-EDTA (TE: 10mM Tris pH8, 1mM EDTA). Precipitated DNA was eluted with 150µl of ChIP Elution Buffer (EB: 100mM NaHCO₃) twice for 15 minutes at room temperature in rotation, and processed as follows in parallel with 300µl of sonicated chromatin non-used for ChIP (Input). Cross-linking was reverted by adding NaCl to a final concentration of 300mM for protein-DNA de-crosslinking and incubated at 65°C for 5 hours and 1µg RNase A (Sigma) to digest contaminating RNA. Finally, 60µg of Proteinase K (Sigma) were added overnight at 45°C. DNA was extracted by sequential phenol-chloroform and chloroform extractions, and precipitated overnight at -80°C in 100mM NaAcetate, 66% ethanol and 50µg of glycogen (Ambion) as a carrier. After centrifugation at 16,000g for 1 hour at 4°C, DNA pellets were washed once with ice-cold 70% ethanol, and finally air dried. ChIP samples were resuspended in 30µl and 1:10 of the samples were used in qPCR for verifying the ChIP samples.

4.4.11 Chromatin Immunoprecipitation (ChIP) sequencing

Tagmentation was performed on ChIPed DNA by Tn5 transposase from Nextera DNA Sample Prep Kit (Illumina, cat.#FC-121-1031). Tagmented DNA purified by DNA Clean & Concentrator-5 kit (Zymo research, Cat# 4013). Tagmented DNA was subjected to PCR amplification and double-size selection by AMPure XP beads (Beckman, cat.#A63881) (0.55x/0.9x). The library were multiplexed, quantified using a High-sensitivity d1000

TapeStation (Agilent) and then sequenced using a NextSeq 500 (Illumina) (paired-end, 2 × 41 bp). Sequencing depth was >20 million reads per sample.

4.4.12 ATAC-sequencing

Cells were washed once with PBS, collected in Cell Dissociation Buffer (Gibco 13150-016) or TrypLE and centrifuged at 300g for 3 min. The cell pellets were then resuspended in 2 ml of 4°C PBS and counted by haemocytometer for using 100,000 cells in the subsequent step. Cells were centrifuged at 300g for 3 min, the supernatant aspirated, the cell pellet resuspended in 150 µl of Isotonic Lysis Buffer (10 mM Tris-HCl pH 7.5, 3 mM CaCl₂, 2 mM MgCl₂, 0.32 M Sucrose and Protease Inhibitors, Roche), and incubated for 12 min on ice. Triton X-100 from a 10% stock was then added at a final concentration of 0.5%, the samples were vortexed briefly and incubated on ice for 6 min. The samples were centrifuged for 5 min at 400g at 4°C, and the cytoplasmic fraction removed from the nuclear pellet. The samples were resuspended gently in 625 µL of PBS and transferred to a fresh 1.5 ml eppendorf tube. The nuclei were centrifuged at 1500g for 3 min at 4°C and the supernatant aspirated thoroughly from the nuclear pellet. This step was immediately followed by tagmentation (Nextera DNA Sample Preparation Kit for 24 Samples, FC-121-1030) by resuspending each sample in 100 µL Nextera mastermix (52.5 µl TD buffer, 42.5 µl of water and 5 µl of TDE1 per reaction). The nuclear pellet was resuspended thoroughly by pipetting and incubated at 37 °C for 1 hour shaking at 300rpm. The reaction was stopped with 300 µL of buffer PB from the Qiagen PCR purification kit, followed by Qiagen PCR clean up protocol using MinElute columns and eluting each sample in 18 µl buffer EB. For the control sample, the nuclear pellet was subjected to genomic DNA

isolation with GenElute Mammalian Genomic DNA Miniprep Kit (Sigma, G1N70) according to manufacturer's protocol, and the purified genomic DNA was thereafter immediately used for tagmentation as for other ATAC-seq samples.

Next a PCR reaction (for all samples including control sample) was performed with the following constituents: 10 µl template from tagmentation, 2.5 µl I7 primer (Nextera® Index Kit with 24 Indices for 96 Samples, FC-121-1011), 2.5 µl I5 primer, 10 uL Nudex Free H₂O 25µl NEBNext High-Fidelity 2x PCR Master Mix (New England Labs Cat #M054 and 10 uL Nuclease Free H₂O. The PCR settings were as follows: at 72 °C for 5 minutes, initial denaturation at 98 °C for 30 seconds, then 12 cycles of 98 °C for 10 seconds, primer annealing at 63 °C for 30 seconds and elongation at 72 °C for 1 minute, and holding at 4 °C. After completing the PCR, the sample were stored at -20 °C. The PCR primers were removed with 1 x 0.9:1 SPRI beads (Beckman Coulter, Cat no. A63880) according to manufacturer's protocol and samples eluted in 20 µl. 2 µl of the samples were run on Agilent HS Bioanalyzer HS for confirming the size selection of the ATAC libraries. ATAC-sequencing was performed by Illumina HiSeq 2000 sequencing with 75 bp PE for obtaining more than 40 million mapped reads per library.

4.4.13 ATAC-sequencing analysis

Sequencing reads from the ChIP-seq and ATAC seq experiment were aligned to the human genome (hg38) using bowtie with reporting mode," --best --strata --v2". DeepTools was used to generate coverage track(bigwig). Coverage track was visualized by using UCSC genome browser. Peak calling was performed by using macs2 peak caller with default parameters for ChIP seq, and with parameter "--nomodel --shift -100 --extsize 200"

for ATAC seq. Peaks annotated with nearest gene information by using BEDTools. Peak distribution over different genomic features were summarized by using Bioconductor package ChiPpeakAnno. Motif enrichment analysis within peak regions was performed using HOMER. All plots were generated using R package 3.6.

For the TF footprint analysis, motifs in JASPAR database were scanned using ATAC bam file and peak coordinates for Ctrl and IBRD9 separately. We then compared the cleavage profiles from Ctrl and IBRD9 for each TF: 2027 IBRD9 ATAC peaks within IBRD9 specific loop anchors and 2372 Ctrl ATC peaks within Ctrl specific loop anchors.

4.4.14 Single-cell gene expression analysis

Sequencing data was aligned and quantitated using CellRanger (10x Genomics) and hg38 reference genome sequence. Raw gene expression matrices were then analysed using Seurat R package (v4.0.5). Quality control was performed individually for each sample. All cells with <500 detected genes were removed, as well as cells contained <600 unique molecular identifiers (UMIs) and >10% mitochondrial counts. Raw read counts were normalized using *NormalizeData* function, based on which 2000 variable genes were identified. 2000 genes across samples for integration were selected using *SelectIntegrationFeatures* function. To perform data integration using reciprocal PCA method in *FindIntegrationAnchors* function, data scaling and PCA were performed on each sample first. 50 dimensions were used in integration process.

In order to identify cell types, we performed principal component analysis (PCA) on integrated data with default parameters in Seurat to reduce dimensionality. The first 50 principal components (PCs) were used for Uniform Manifold Approximation and

Projection (UMAP) dimensionality reduction. The *FindNeighbours* (30 nearest neighbours) and *FindCluster* function (resolution = 2.5) were used to cluster cells. Clusters in 2D UMAP were used to identify cell types based on marker genes. Unless specified, default parameters were used for each function.

4.4.15 ChIP-sequencing data analysis

The adapters were removed and the filtered reads were mapped to hg38 reference genome using *bwa*²²⁵ *mem*. The duplicated reads were removed using Picard (<https://broadinstitute.github.io/picard/>) and the reads with quality lower than 20 were filtered. Bam files were converted to bigwig format file using *deepTools*²²⁶ *bamCoverage*. The parameter 'scaleFactor' for each sample were determined by *edgeR*²²⁷ function *calcNormFactors* using "TMM" normalization method. The ChIP-seq peaks were called using the default (narrow) setting in MACS2. The promoter regions were defined as +/- 2.5Kb around the TSSs regions.

4.4.16 In-situ ChIA-PET (H3K27ac)

Cells were grown to 80% confluency, and treated by 1% formaldehyde for 20 min and ethylene glycol bis (EGS)(Thermo Fisher Scientific, cat.#21565) for 45 min. 10 million of dual crosslinked A13A CSC cells were used to generate one in-situ ChIA-PET library. Briefly, after cell lysis and nuclear lysis, cells were subjected to in-situ digestion with *AluI* restriction enzyme for overnight. On the next day, A-tailing was performed using *Klenow* Fragment (3'>5' exo-) (NEB, cat.#M0212L) and dATP (100 mM; NEB, cat.#N0440S) for 1 hr at room temperature. In-situ ligation was performed using in-house bridge linker (F:

5'- /5Phos/CGCGATATC/iBIOdT/TATCTGACT -3', R: 5'-
/5Phos/GTCAGATAAGATATCGCGT -3'. HPLC purified, from Integrated DNA Technologies) and T4 DNA ligase (Thermo Fisher Scientific, cat.#EL0013) for RT 1 hr and then 16°C for overnight. The nucleus was sonicated into ~1 kb size. Then chromatin was precleared using protein G beads for 1 hr and immunoprecipitation was performed using anti-H3K27ac antibody (Active Motif, cat.# 39135) coated protein G beads (Life Technologies, cat. no. 10009D) for overnight.

After that, proximity-ligated chromatin complexes were washed sequentially by low salt, high salt, LiCl buffer and TE buffer. Chromatin complexes were eluted from the beads by incubation with ChIP elution buffer (1% SDS+TE) at 65 °C for 1 hr. After that, proteinase K (Life Technologies, cat.#AM2548) was added to the elution reverse cross-linked chromatin complexes for overnight. After DNA purification using QIAquick PCR Purification Kit, tagmentation was performed on the proximity-ligated DNA by Tn5 transposase from Nextera DNA Sample Prep Kit (Illumina, cat.#FC-121-1031). Then tagmented DNA was immobilized on M280 streptavidin dynabeads (Invitrogen, cat.#11205D). PCR amplification was performed on beads using Nextera DNA Sample Prep Kit and the products were purified by AMPure XP beads (Beckman, cat.#A63881) and subjected to size selection (350-600 bp) on BluePippin instrument (Sage Science) using Blue Pippin Cassette Kit (Sage Science, cat.#BDF2010). DNA library was sequenced on Illumina Novaseq 6000 by paired-end 150 bp.

4.4.17 *In-situ* ChIA-PET data processing

We used ChIA-PIPE²²⁸ to process ChIA-PET data. Pair-end reads containing bridge linker were kept and trimmed, the flanking sequences were mapped to human hg38 reference genome and uniquely aligned pair-end tags (PETs) with MAPQ ≥ 30 were retained. Redundant reads were then removed and bam format files were used for downstream analysis. Each PET was classified as self-ligation PET or inter-ligation PET using a genomic span cutoff of 8Kb. Inter-ligation PETs were extended by 500 bp on both ends and clustered. The PET clusters with counts lower than 10 and spanning less than 5 Kb or larger than 1Mb were filtered. We referred the remaining PET clusters as chromatin interaction loops. Un-clustered individual inter-ligation PETs were referred as PET singletons. All PETs (including self-ligation and inter-ligation PETs) were used to call peaks using MACS2²²⁹.

Chromatin interaction loops from all samples were merged to form a union set. Read counts for each merged loop were counted across all samples to form a loop – sample count matrix. To identify control-specific and IBRD9-specific loops, the read matrix was imported into R and a negative binomial model was fitted using glmFit function in edgeR²²⁷ R package. The loops with p value < 0.1 were determined as differential loops³¹. Loop anchors proximal to (less than 3Kb) known transcription start sites (TSSs) were defined as promoters and the remaining loop anchors were defined as distal regulatory elements.

4.4.18 Functional enrichment of downregulated genes with IBRD9 treatment

Functional enrichment of genes network was generated using the ToppFun application of the Toppgene Suite using a corrected (Benjamini and Hochberg) p-value cut-off of 0.05

(<https://toppgene.cchmc.org/>)¹³⁵. Cytoscape application (<https://cytoscape.org/>)¹⁷² was used to generate the functional enrichment network.

4.4.19 Visualization

Statistical plots were created using R. Screenshots of chromatin interactions was visualized using WashU Genome Browser.

4.4.20 ChIP-qPCR of anchor regions

The cells were treated as for RNA-sequencing experiment and the ChIP was performed as described above. The primers used for ChIP-qPCR of anchor regions are listed below:

Table 3. List of ChIP-qPCR primers

Primer name	Primer sequence
SOX4 promoter anchor F	AAAAAGTTTGCTCTGTAAATTGGAA
SOX4 promoter anchor R	CAGCCAATCAGCTGCTGTAA
SOX4 enhancer anchor F	TATGTTTGGCCAGCTGTTGA
SOX4 enhancer anchor R	GGAGAAAGAGGTAAGAATGATCAGA
PROM1 promoter anchor F	CGGAGTAACTTGGCTGCTCT
PROM1 promoter anchor R	GATCGGAGGGAGGAACTAGC
PROM1 enhancer anchor F	TCTTATTGCAGGGTCTGAGC
PROM1 enhancer anchor R	TTGTTTCAGTAAATGAATACATGAAAGA
SMAD1 enhancer anchor F	CTTCCACATCAAGGATGAGC
SMAD1 enhancer anchor R	GCAATATAACTGGGCTGCAC
SMAD1 promoter anchor F	CCGCACGGGTTTTCTTCT
SMAD1 promoter anchor R	CCGCCAAACCTTTGAATAAAT

SNAI2 enhancer anchor F	CACTTGACTGGCTTCCAAGG
SNAI2 enhancer anchor R	AAGGAGGGTGGAGCATGATT
SNAI2 promoter anchor F	GGAAGAATCTGTTGTCAGACTAAAT
SNAI2 promoter anchor R	CCCTCCTAGCTCCCAGAGA

Chapter 5

Conclusions and future direction

In this thesis, we performed systemic 3D genome analysis to interrogate 3D chromatin architectures in transcription regulation and human disease pathogenesis.

In Chapter 2, I generated a compendium of 3D-epigenome and transcriptome maps from 101 biobanked human DCM and non-failing heart tissues through HiChIP (H3K27ac), *in situ* Hi-C, ChIP-seq, ATAC-seq and RNA-seq profiling. I employed human iPSC-derived cardiomyocytes (hiPSC-CMs) and mouse models to further interrogate the key transcription factor implicated in 3D chromatin organization and transcriptional regulation in DCM pathogenesis. I discovered that the active regulatory elements (H3K27ac peaks) and their connectome (H3K27ac loops) were extensively reprogrammed in DCM hearts and contributed to transcriptional dysregulation implicated for DCM development. For example, I identified that non-transcribing *NPPA-AS1* promoter functions as an enhancer and physically interacts with the *NPPA* and *NPPB* promoters, leading to the co-transcription of *NPPA* and *NPPB* in DCM hearts. We uncovered that DCM-enriched H3K27ac loops largely resided in conserved high-order chromatin architectures (Compartments, Topologically Associating Domains) and unexpectedly their anchors had equivalent chromatin accessibility. Intriguingly, we discovered that the DCM-enriched H3K27ac loop anchors exhibited a strong enrichment for *Heart and Neural Crest Derivatives Expressed 1* (HAND1), a key transcription factor involved in early cardiogenesis. In line with this, its protein expression was upregulated in human DCM and mouse failing hearts. To further validate whether HAND1 is a causal driver for the reprogramming of enhancer/promoter connectome in DCM hearts, we performed comprehensive 3D epigenome mappings in hiPSC-CMs. I found that forced

overexpression of *HAND1* in hiPSC-CM induced a distinct gain of enhancer/promoter connectivity and, correspondingly, increased the expression of their connected genes implicated in DCM etiology, thus recapitulating the transcriptional signature in human DCM hearts. Moreover, electrophysiology analysis demonstrated that forced overexpression of *HAND1* in hiPSC-CM induced abnormal calcium handling. Furthermore, cardiomyocyte-specific overexpression of *Hand1* in the mouse hearts resulted in a dilated cardiac remodeling with impaired contractility/ Ca^{2+} handling in cardiomyocytes, increased ratio of heart weight/body weight and compromised cardiac function, which were ascribed to recapitulation of transcriptional reprogramming in DCM. In summary, this study provided novel chromatin topology insights into DCM pathogenesis and illustrated a model whereby a single transcription factor (HAND1) reprograms the genome-wide enhancer/promoter connectome to drive DCM pathogenesis.

In Chapter 3, I performed integrative 3D-epigenomic analyses in A13A_Org (primary) and A13D_Org (metastatic) PDAC tumor spheres from the same patient: H3K27ac HiChIP (enhancer connectome), CTCF in-situ ChIA-PET (chromatin topology), H3K27ac ChIP-seq (active regulatory elements), ATAC-seq (chromatin accessibility) and RNA-seq (transcriptional output) to reveal potential 3D-chromatin mechanisms that regulate metastasis-specific transcription in PDAC. By comparing A13D_Org to A13A_Org, I found that H3K27ac-associated enhancer connectome significantly contributes to A13D_Org-specific gene transcription, especially to those prognostic molecular markers in PDAC (e.g. *NRP1*) and cancer stem cell markers (e.g. *BMI1*). In addition, I demonstrated that

chromatin accessibility is linked to the formation of A13D_Org-specific H3K27ac HiChIP loop anchors. Moreover, by transcription factor motif scan analysis, I showed that FOSL1 is enriched on A13D_Org-specific H3K27ac HiChIP loop anchors, the expression of which is upregulated in A13D_Org and associated with poor prognosis in PDAC patients. To further dissect the topological mechanism of enhancer-promoter interactions, we performed CTCF ChIA-PET analysis and uncovered that specific CTCF binding on the promoters is linked to the specific enhancer-promoter connectome. In sum, my comprehensive 3D-epigenomic analyses suggest integrated epigenomic and topological mechanisms that regulate PDAC-specific transcription, and provide novel insights into the interplay of epigenetics, and 3D chromatin architecture in PDAC progression.

In Chapter 4, by using a small molecule compound library targeting epigenetic enzymes I and my colleague identified BRD9 enzyme for eliminating CSCs. Chemical inhibition and genetic loss of function of BRD9 blocks the self-renewal of CSCs, reduces CSC invasiveness and resensitizes PDAC CSCs to conventional therapies. BRD9 inhibition abrogated tumour formation in mice and eliminated CSCs in tumours from pancreatic cancer patients. By employing H3K27ac in-situ ChIA-PET assay, I showed that inhibition of BRD9 reduced the enhancer-promoter interactions of stemness and EMT-associated genes in CSCs. Collectively, I uncovered BRD9 enzyme as an attractive therapeutic target for re-sensitizing and eliminating CSCs in pancreatic cancer patients.

In this thesis, although we have demonstrated the roles of active regulatory elements and their interactions orchestrates in transcriptional regulation of human diseases, we have not interrogate the functionality of repressive cis-elements, such as transcriptional

silencers marked by polycomb proteins, and their 3D chromatin structures in human disease pathogenesis. In addition, it is unclear the roles of topological structures (e.g. CTCF/cohesin mediated loopings) in orchestrating the interactions of active and repressive cis elements in disease pathogenesis. Collectively, I believe that by investigating disease development and evolution through the lens of 3D chromatin architecture, it will be possible to uncover many new etiological mechanisms that have so far been elusive. Such studies will not only help understand disease pathogenesis but also open new avenues for diagnostics and therapeutic interventions for human disease.

Appendix A Publications related to the thesis

Feng Y[#], Cai L[#], Hong W[#], Zhang C[#], Tan N, Wang M, Wang C, Liu F, Wang X, Ma J, Gao C, Kumar N, Mo Y, Geng Q, Luo C, Lin Y, Chen H, Wang S, Watson M, Jegga A, Pedersen R, Fu J, Wang Z, Fan G, Sadayappan S, Wang Y, Pauklin S^{*}, Huang F^{*}, Huang W^{*}, Jiang L^{*}. Rewiring of 3D Chromatin Topology Orchestrates Transcriptional Reprogramming and the Development of Human Dilated Cardiomyopathy. **Circulation**. 2022 Apr 11. doi: 10.1161/CIRCULATIONAHA.121.055781. (IF=39.9)([#]Co-first authors, ^{*}co-corresponding authors)(Research Article)

Deng S, **Feng Y**, Pauklin S. 3D chromatin architecture and transcription regulation in cancer. **J Hematol Oncol**. 2022 May 4;15(1):49. (IF=17.3)(Review)

Feng Y[#], Huang W[#], Paul C, Liu X, Sadayappan S, Wang Y, Pauklin S. Mitochondrial nucleoid in cardiac homeostasis: bidirectional signaling of mitochondria and nucleus in cardiac diseases. **Basic Res Cardiol**. 2021 Aug 14;116(1):49. (Co-first author)(IF=17.1)(Review)

Feng Y, Mouti M and Pauklin S^{*}. PAF1 regulates the stemness of pancreatic cancer stem cells. **Gastroenterology**. 2021 May;160(6):2224-2225. (IF=22.6)(Comment)

Feng Y, Liu X and Pauklin S^{*}. 3D chromatin architecture and epigenetic regulation in cancer stem cells. **Protein Cell**. 2021 Jun;12(6):440-454. (IF=14.8)(Review)

Lee B, Wang J, Cai L, Kim M, Namburi S, Tjong H, **Feng Y**, Wang P, Tang Z, Abbas A, Wei CL*, Ruan Y* and Li S*. ChIA-PIPE: A fully automated pipeline for comprehensive ChIA-PET data analysis and visualization. **Sci Adv.** 2020;6:eaay2078. (IF=16.8)(Research Article)

French R#, **Feng Y**# and Pauklin S*. Targeting TGFbeta Signalling in Cancer: Toward Context-Specific Strategies. **Trends Cancer.** 2020;6:538-540. (Co-first author)(IF=14.2)(Spotlight)

Feng Y and Pauklin S*. Two Sides of the Same Coin: The Roles of Transforming Growth Factor-beta in Colorectal Carcinogenesis. **Gastroenterology.** 2020;159:397-398. (IF=22.6)(Comment)

Feng Y and Pauklin S*. Demystifying Cancer Etiology via 3D Genome Mapping. **Trends Genet.** 2020;36:634-637.(IF=11.6)(Spotlight)

Feng Y and Pauklin S*. Revisiting 3D chromatin architecture in cancer development and progression. **Nucleic Acids Res.** 2020;48:10632-10647.(IF=16.9)(Review)

Appendix B Awards and Honors

Clarendon Fund (University of Oxford), 2019-2022

Winkler Oxford Award in Medicine or Psychology, 2019

Best Poster Award, Botnar Research Centre Student Symposium, University of Oxford, 2021

Richard J Bing Award for Young Investigator (Finalist), XXIV World Congress International Society for Heart Research (ISHR), Germany, 2022 (One of four finalists for Richard J Bing Award for Young Investigator in triennial ISHR world congress, which is the most prestigious award for global young cardiovascular investigator).



References

- 1 Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat Rev Cancer* **16**, 483-493, doi:10.1038/nrc.2016.62 (2016).
- 2 Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nature reviews. Molecular cell biology* **20**, 535-550, doi:10.1038/s41580-019-0132-4 (2019).
- 3 Hnisz, D., Schuijers, J., Li, C. H. & Young, R. A. Regulation and dysregulation of chromosome structure in cancer. *Annual Review of Cancer Biology* **2**, 21-40 (2018).
- 4 Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* **58**, 362-370, doi:10.1016/j.molcel.2015.02.014 (2015).
- 5 Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334, doi:10.1016/j.cell.2013.03.036 (2013).
- 6 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
- 7 Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613-626, doi:10.1038/nrg3207 (2012).
- 8 Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-381, doi:10.1016/j.cell.2014.02.019 (2014).
- 9 Northcott, P. A. *et al.* Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* **511**, 428-434, doi:10.1038/nature13379 (2014).

- 10 Beroukhim, R., Zhang, X. & Meyerson, M. Copy number alterations unmasked as enhancer hijackers. *Nat Genet* **49**, 5-6, doi:10.1038/ng.3754 (2016).
- 11 Takeda, D. Y. *et al.* A Somatic Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. *Cell* **174**, 422-432 e413, doi:10.1016/j.cell.2018.05.037 (2018).
- 12 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318, doi:10.1038/ng1966 (2007).
- 13 Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98, doi:10.1016/j.cell.2011.12.014 (2012).
- 14 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108, doi:10.1038/nature11233 (2012).
- 15 Nguyen, T. A. *et al.* High-throughput functional comparison of promoter and enhancer activities. *Genome Res* **26**, 1023-1033, doi:10.1101/gr.204834.116 (2016).
- 16 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).
- 17 Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**, 629-635, doi:10.1038/nmeth.4264 (2017).

- 18 Jung, I. *et al.* A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat Genet* **51**, 1442-1449, doi:10.1038/s41588-019-0494-8 (2019).
- 19 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. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
- 20 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).
- 21 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).
- 22 Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490, doi:10.1038/nature14590 (2015).
- 23 Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914, doi:10.1126/science.aab1601 (2015).
- 24 Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538-542, doi:10.1038/nature25981 (2018).
- 25 Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548 e1516, doi:10.1016/j.cell.2018.03.074 (2018).

- 26 Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-1203, doi:10.1038/ng.3646 (2016).
- 27 Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936, doi:10.1038/s41587-019-0206-z (2019).
- 28 Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* **37**, 1458-1465, doi:10.1038/s41587-019-0332-7 (2019).
- 29 Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380-1385, doi:10.1126/science.aau0730 (2018).
- 30 Roe, J. S. *et al.* Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell* **170**, 875-888 e820, doi:10.1016/j.cell.2017.07.007 (2017).
- 31 Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature genetics* **49**, 1602-1612, doi:10.1038/ng.3963 (2017).
- 32 Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930, doi:10.1038/s41467-019-09982-5 (2019).
- 33 Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol Cell* **76**, 206-216 e207, doi:10.1016/j.molcel.2019.07.015 (2019).

- 34 Zhang, T., Zhang, Z., Dong, Q., Xiong, J. & Zhu, B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol* **21**, 45, doi:10.1186/s13059-020-01957-w (2020).
- 35 Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep* **3**, 224-229, doi:10.1093/embo-reports/kvf053 (2002).
- 36 Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239-243, doi:10.1038/nature25461 (2018).
- 37 Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490-493, doi:10.1038/nature09158 (2010).
- 38 Maass, P. G., Barutcu, A. R. & Rinn, J. L. Interchromosomal interactions: A genomic love story of kissing chromosomes. *J Cell Biol* **218**, 27-38, doi:10.1083/jcb.201806052 (2019).
- 39 Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448-453, doi:10.1038/s41586-018-0845-0 (2019).
- 40 Bartman, C. R., Hsu, S. C., Hsiung, C. C., Raj, A. & Blobel, G. A. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell* **62**, 237-247, doi:10.1016/j.molcel.2016.03.007 (2016).
- 41 Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell* **76**, 473-484 e477, doi:10.1016/j.molcel.2019.07.038 (2019).

- 42 Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28-45, doi:10.1016/j.cell.2020.09.014 (2020).
- 43 Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345-1349, doi:10.1126/science.aaz4475 (2019).
- 44 Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345, doi:10.1126/science.aaz3418 (2019).
- 45 Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**, 1165-1178 e1120, doi:10.1016/j.cell.2018.03.072 (2018).
- 46 Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737-743, doi:10.1038/s41586-020-2151-x (2020).
- 47 Downen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387, doi:10.1016/j.cell.2014.09.030 (2014).
- 48 Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167**, 1188-1200, doi:10.1016/j.cell.2016.10.024 (2016).
- 49 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).
- 50 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922, doi:10.1016/j.cell.2017.05.004 (2017).

- 51 Johnstone, S. E. *et al.* Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* **182**, 1474-1489 e1423, doi:10.1016/j.cell.2020.07.030 (2020).
- 52 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 53 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
- 54 Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59-64, doi:10.1038/nature21429 (2017).
- 55 Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* **52**, 1151-1157, doi:10.1038/s41588-020-00716-8 (2020).
- 56 Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Biol* **33**, 265-289, doi:10.1146/annurev-cellbio-100616-060531 (2017).
- 57 Schwartzman, O. *et al.* UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat Methods* **13**, 685-691, doi:10.1038/nmeth.3922 (2016).
- 58 Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-212, doi:10.1038/ng.2871 (2014).
- 59 Davies, J. O. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* **13**, 74-80, doi:10.1038/nmeth.3664 (2016).

- 60 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods* **13**, 919-922, doi:10.1038/nmeth.3999 (2016).
- 61 Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* **26**, 1345-1348, doi:10.1038/cr.2016.137 (2016).
- 62 Bertolini, J. A. *et al.* Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. *Cell stem cell* **24**, 462-476 e466, doi:10.1016/j.stem.2019.02.004 (2019).
- 63 Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524, doi:10.1038/nature21411 (2017).
- 64 Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 e724, doi:10.1016/j.cell.2018.05.024 (2018).
- 65 Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558-562, doi:10.1038/s41586-019-0949-1 (2019).
- 66 Tian, S. Z. *et al.* ChIA-DropBox: a novel analysis and visualization pipeline for multiplex chromatin interactions. *bioRxiv*, 613034, doi:10.1101/613034 (2019).
- 67 Cook, P. R. A model for all genomes: the role of transcription factories. *J Mol Biol* **395**, 1-10, doi:10.1016/j.jmb.2009.10.031 (2010).
- 68 Fasolino, M. *et al.* Genetic Variation in Type 1 Diabetes Reconfigures the 3D Chromatin Organization of T Cells and Alters Gene Expression. *Immunity* **52**, 257-274 e211, doi:10.1016/j.immuni.2020.01.003 (2020).

- 69 Cardozo Gizzi, A. M., Cattoni, D. I. & Nollmann, M. TADs or no TADs: Lessons From Single-cell Imaging of Chromosome Architecture. *J Mol Biol* **432**, 682-693, doi:10.1016/j.jmb.2019.12.034 (2020).
- 70 Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, doi:10.1126/science.aau1783 (2018).
- 71 Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* **112**, 3002-3007, doi:10.1073/pnas.1420024112 (2015).
- 72 Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050-1055, doi:10.1126/science.aao3136 (2018).
- 73 Papp, B. & Plath, K. Epigenetics of reprogramming to induced pluripotency. *Cell* **152**, 1324-1343, doi:10.1016/j.cell.2013.02.043 (2013).
- 74 Ohnishi, K. *et al.* Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. *Cell* **156**, 663-677, doi:10.1016/j.cell.2014.01.005 (2014).
- 75 Shibata, H. *et al.* In vivo reprogramming drives Kras-induced cancer development. *Nat Commun* **9**, 2081, doi:10.1038/s41467-018-04449-5 (2018).
- 76 Pajerowski, J. D., Dahl, K. N., Zhong, F. L., Sammak, P. J. & Discher, D. E. Physical plasticity of the nucleus in stem cell differentiation. *Proc Natl Acad Sci U S A* **104**, 15619-15624, doi:10.1073/pnas.0702576104 (2007).
- 77 Bartova, E., Krejci, J., Harnicarova, A. & Kozubek, S. Differentiation of human embryonic stem cells induces condensation of chromosome territories and

- formation of heterochromatin protein 1 foci. *Differentiation* **76**, 24-32, doi:10.1111/j.1432-0436.2007.00192.x (2008).
- 78 Lorzadeh, A. *et al.* Nucleosome Density ChIP-Seq Identifies Distinct Chromatin Modification Signatures Associated with MNase Accessibility. *Cell Rep* **17**, 2112-2124, doi:10.1016/j.celrep.2016.10.055 (2016).
- 79 Bartova, E. *et al.* Epigenome and chromatin structure in human embryonic stem cells undergoing differentiation. *Dev Dyn* **237**, 3690-3702, doi:10.1002/dvdy.21773 (2008).
- 80 Underwood, J. M., Becker, K. A., Stein, G. S. & Nickerson, J. A. The Ultrastructural Signature of Human Embryonic Stem Cells. *J Cell Biochem* **118**, 764-774, doi:10.1002/jcb.25736 (2017).
- 81 Pagliara, S. *et al.* Auxetic nuclei in embryonic stem cells exiting pluripotency. *Nat Mater* **13**, 638-644, doi:10.1038/nmat3943 (2014).
- 82 Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642-654, doi:10.1016/j.cell.2012.12.033 (2013).
- 83 Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437-447, doi:10.1016/j.stem.2008.03.021 (2008).
- 84 Schlesinger, S. *et al.* A hyperdynamic H3.3 nucleosome marks promoter regions in pluripotent embryonic stem cells. *Nucleic Acids Res* **45**, 12181-12194, doi:10.1093/nar/gkx817 (2017).

- 85 Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-116, doi:10.1016/j.devcel.2005.10.017 (2006).
- 86 Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532-538, doi:10.1038/ncb1403 (2006).
- 87 Manukyan, M. & Singh, P. B. Epigenome rejuvenation: HP1beta mobility as a measure of pluripotent and senescent chromatin ground states. *Sci Rep* **4**, 4789, doi:10.1038/srep04789 (2014).
- 88 Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14**, 735-751, doi:10.1016/j.stem.2014.05.006 (2014).
- 89 You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22**, 9-20, doi:10.1016/j.ccr.2012.06.008 (2012).
- 90 Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39**, 232-236, doi:10.1038/ng1950 (2007).
- 91 Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39**, 237-242, doi:10.1038/ng1972 (2007).
- 92 Dunican, D. S. *et al.* Bivalent promoter hypermethylation in cancer is linked to the H327me3/H3K4me3 ratio in embryonic stem cells. *BMC Biol* **18**, 25, doi:10.1186/s12915-020-0752-3 (2020).

- 93 Adam, R. C. *et al.* Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366-370, doi:10.1038/nature14289 (2015).
- 94 Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21-33, doi:10.1038/nrg1748 (2006).
- 95 Fishedick, G. *et al.* Nanog induces hyperplasia without initiating tumors. *Stem Cell Res* **13**, 300-315, doi:10.1016/j.scr.2014.08.001 (2014).
- 96 Lengner, C. J. *et al.* Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell* **1**, 403-415, doi:10.1016/j.stem.2007.07.020 (2007).
- 97 Marucci, L. *et al.* beta-catenin fluctuates in mouse ESCs and is essential for Nanog-mediated reprogramming of somatic cells to pluripotency. *Cell Rep* **8**, 1686-1696, doi:10.1016/j.celrep.2014.08.011 (2014).
- 98 Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**, 1350-1353, doi:10.1038/ng.471 (2009).
- 99 Lu, X., Mazur, S. J., Lin, T., Appella, E. & Xu, Y. The pluripotency factor nanog promotes breast cancer tumorigenesis and metastasis. *Oncogene* **33**, 2655-2664, doi:10.1038/onc.2013.209 (2014).
- 100 Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121**, 465-477, doi:10.1016/j.cell.2005.02.018 (2005).

- 101 Xiong, X. *et al.* KLF4, A Gene Regulating Prostate Stem Cell Homeostasis, Is a Barrier to Malignant Progression and Predictor of Good Prognosis in Prostate Cancer. *Cell Rep* **25**, 3006-3020 e3007, doi:10.1016/j.celrep.2018.11.065 (2018).
- 102 de Wit, E. *et al.* The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**, 227-231, doi:10.1038/nature12420 (2013).
- 103 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
- 104 Novo, C. L. *et al.* Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell Rep* **22**, 2615-2627, doi:10.1016/j.celrep.2018.02.040 (2018).
- 105 Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602-616, doi:10.1016/j.stem.2013.08.013 (2013).
- 106 Di Giammartino, D. C. *et al.* KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat Cell Biol* **21**, 1179-1190, doi:10.1038/s41556-019-0390-6 (2019).
- 107 Cheng, J. *et al.* Tryptophan derivatives regulate the transcription of Oct4 in stem-like cancer cells. *Nat Commun* **6**, 7209, doi:10.1038/ncomms8209 (2015).
- 108 Ziaeeian, B. & Fonarow, G. C. Epidemiology and aetiology of heart failure. *Nature reviews. Cardiology* **13**, 368-378, doi:10.1038/nrcardio.2016.25 (2016).
- 109 Merlo, M. *et al.* Evolving concepts in dilated cardiomyopathy. *European journal of heart failure* **20**, 228-239, doi:10.1002/ejhf.1103 (2018).

- 110 Rosenbaum, A. N., Agre, K. E. & Pereira, N. L. Genetics of dilated cardiomyopathy: practical implications for heart failure management. *Nature reviews. Cardiology*, doi:10.1038/s41569-019-0284-0 (2019).
- 111 Yu, J., Zeng, C. & Wang, Y. Epigenetics in dilated cardiomyopathy. *Current opinion in cardiology* **34**, 260-269, doi:10.1097/HCO.0000000000000616 (2019).
- 112 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature reviews. Genetics* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).
- 113 Chatterjee, S. & Ahituv, N. Gene Regulatory Elements, Major Drivers of Human Disease. *Annual review of genomics and human genetics* **18**, 45-63, doi:10.1146/annurev-genom-091416-035537 (2017).
- 114 Gacita, A. M. *et al.* Genetic Variation in Enhancers Modifies Cardiomyopathy Gene Expression and Progression. *Circulation* **143**, 1302-1316, doi:10.1161/CIRCULATIONAHA.120.050432 (2021).
- 115 Gacita, A. M. *et al.* Altered Enhancer and Promoter Usage Leads to Differential Gene Expression in the Normal and Failed Human Heart. *Circ Heart Fail* **13**, e006926, doi:10.1161/CIRCHEARTFAILURE.120.006926 (2020).
- 116 Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

- 117 Tan, W. L. W. *et al.* Epigenomes of Human Hearts Reveal New Genetic Variants Relevant for Cardiac Disease and Phenotype. *Circ Res* **127**, 761-777, doi:10.1161/CIRCRESAHA.120.317254 (2020).
- 118 Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nature reviews. Genetics* **20**, 437-455, doi:10.1038/s41576-019-0128-0 (2019).
- 119 Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412 e1322, doi:10.1016/j.cell.2018.03.068 (2018).
- 120 Wu, S. *et al.* Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699-703, doi:10.1038/s41586-019-1763-5 (2019).
- 121 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).
- 122 Flavahan, W. A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229-233, doi:10.1038/s41586-019-1668-3 (2019).
- 123 Poleshko, A. *et al.* Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* **171**, 573-587 e514, doi:10.1016/j.cell.2017.09.018 (2017).
- 124 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).

- 125 Dirkx, E., da Costa Martins, P. A. & De Windt, L. J. Regulation of fetal gene expression in heart failure. *Biochimica et biophysica acta* **1832**, 2414-2424, doi:10.1016/j.bbadis.2013.07.023 (2013).
- 126 Chien, K. R., Knowlton, K. U., Zhu, H. & Chien, S. Regulation of cardiac gene expression during myocardial growth and hypertrophy: molecular studies of an adaptive physiologic response. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **5**, 3037-3046, doi:10.1096/fasebj.5.15.1835945 (1991).
- 127 Heinig, M. *et al.* Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome biology* **18**, 170, doi:10.1186/s13059-017-1286-z (2017).
- 128 Celik, S. *et al.* Antisense regulation of atrial natriuretic peptide expression. *JCI Insight* **4**, doi:10.1172/jci.insight.130978 (2019).
- 129 Sandhu, K. S. *et al.* Large-scale functional organization of long-range chromatin interaction networks. *Cell reports* **2**, 1207-1219, doi:10.1016/j.celrep.2012.09.022 (2012).
- 130 Huang, X., Qu, R., Ouyang, J., Zhong, S. & Dai, J. An Overview of the Cytoskeleton-Associated Role of PDLIM5. *Front Physiol* **11**, 975, doi:10.3389/fphys.2020.00975 (2020).
- 131 Wang, D. *et al.* Novel polymorphisms in PDLIM3 and PDLIM5 gene encoding Z-line proteins increase risk of idiopathic dilated cardiomyopathy. *Journal of cellular and molecular medicine* **23**, 7054-7062, doi:10.1111/jcmm.14607 (2019).

- 132 Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet* **53**, 110-119, doi:10.1038/s41588-020-00745-3 (2021).
- 133 Dao, L. T. M. *et al.* Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature genetics* **49**, 1073-1081, doi:10.1038/ng.3884 (2017).
- 134 Pan, Y. F. *et al.* Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. *The Journal of biological chemistry* **283**, 32977-32988, doi:10.1074/jbc.M802024200 (2008).
- 135 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research* **37**, W305-311, doi:10.1093/nar/gkp427 (2009).
- 136 Steimle, J. D. & Moskowitz, I. P. TBX5: A Key Regulator of Heart Development. *Current topics in developmental biology* **122**, 195-221, doi:10.1016/bs.ctdb.2016.08.008 (2017).
- 137 Yi, X., Jiang, X. J. & Fang, Z. M. Histone methyltransferase SMYD2: ubiquitous regulator of disease. *Clin Epigenetics* **11**, 112, doi:10.1186/s13148-019-0711-4 (2019).
- 138 She, P. *et al.* The Gridlock transcriptional repressor impedes vertebrate heart regeneration by restricting expression of lysine methyltransferase. *Development* **147**, doi:10.1242/dev.190678 (2020).

- 139 Lu, J., McKinsey, T. A., Zhang, C. L. & Olson, E. N. Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases. *Mol Cell* **6**, 233-244, doi:10.1016/s1097-2765(00)00025-3 (2000).
- 140 Kim, Y. *et al.* The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. *J Clin Invest* **118**, 124-132, doi:10.1172/JCI33255 (2008).
- 141 Bang, M. L. *et al.* Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies. *J Cell Biol* **153**, 413-427, doi:10.1083/jcb.153.2.413 (2001).
- 142 Ramaccini, D. *et al.* Mitochondrial Function and Dysfunction in Dilated Cardiomyopathy. *Front Cell Dev Biol* **8**, 624216, doi:10.3389/fcell.2020.624216 (2020).
- 143 Andersson, C. *et al.* Integrated Multiomics Approach to Identify Genetic Underpinnings of Heart Failure and Its Echocardiographic Precursors: Framingham Heart Study. *Circulation. Genomic and precision medicine* **12**, e002489, doi:10.1161/CIRCGEN.118.002489 (2019).
- 144 Yasuda-Yamahara, M. *et al.* AIF1L regulates actomyosin contractility and filopodial extensions in human podocytes. *PLoS One* **13**, e0200487, doi:10.1371/journal.pone.0200487 (2018).
- 145 Brody, M. J. *et al.* Ablation of the cardiac-specific gene leucine-rich repeat containing 10 (*Lrrc10*) results in dilated cardiomyopathy. *PLoS One* **7**, e51621, doi:10.1371/journal.pone.0051621 (2012).
- 146 Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Molecular cell* **62**, 668-680, doi:10.1016/j.molcel.2016.05.018 (2016).

- 147 Xu, J. *et al.* Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. *The Journal of biological chemistry* **281**, 9152-9162, doi:10.1074/jbc.M510217200 (2006).
- 148 Riley, P., Anson-Cartwright, L. & Cross, J. C. The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nature genetics* **18**, 271-275, doi:10.1038/ng0398-271 (1998).
- 149 Zhou, Y. M. *et al.* HAND1 loss-of-function mutation associated with familial dilated cardiomyopathy. *Clin Chem Lab Med* **54**, 1161-1167, doi:10.1515/cclm-2015-0766 (2016).
- 150 Sun, N. *et al.* Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Science translational medicine* **4**, 130ra147, doi:10.1126/scitranslmed.3003552 (2012).
- 151 Knight, W. E. *et al.* PDE1C deficiency antagonizes pathological cardiac remodeling and dysfunction. *Proc Natl Acad Sci U S A* **113**, E7116-E7125, doi:10.1073/pnas.1607728113 (2016).
- 152 Mork, H. K. *et al.* Increased cardiomyocyte function and Ca²⁺ transients in mice during early congestive heart failure. *J Mol Cell Cardiol* **43**, 177-186, doi:10.1016/j.yjmcc.2007.05.004 (2007).
- 153 Mork, H. K., Sjaastad, I., Sejersted, O. M. & Louch, W. E. Slowing of cardiomyocyte Ca²⁺ release and contraction during heart failure progression in postinfarction mice. *Am J Physiol Heart Circ Physiol* **296**, H1069-1079, doi:10.1152/ajpheart.01009.2008 (2009).

- 154 Zhao, S. M., Wang, Y. L., Guo, C. Y., Chen, J. L. & Wu, Y. Q. Progressive decay of Ca²⁺ homeostasis in the development of diabetic cardiomyopathy. *Cardiovasc Diabetol* **13**, 75, doi:10.1186/1475-2840-13-75 (2014).
- 155 Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170**, 1028-1043 e1019, doi:10.1016/j.cell.2017.08.003 (2017).
- 156 Jang, J. W. *et al.* A novel Fbxo25 acts as an E3 ligase for destructing cardiac specific transcription factors. *Biochem Biophys Res Commun* **410**, 183-188, doi:10.1016/j.bbrc.2011.05.011 (2011).
- 157 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
- 158 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 159 Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature protocols* **7**, 1728-1740, doi:10.1038/nprot.2012.101 (2012).
- 160 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 161 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 162 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

- 163 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 164 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 165 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).
- 166 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).
- 167 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187-191, doi:10.1093/nar/gku365 (2014).
- 168 Phanstiel, D. H., Boyle, A. P., Heidari, N. & Snyder, M. P. Mango: a bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* **31**, 3092-3098, doi:10.1093/bioinformatics/btv336 (2015).
- 169 Li, G. *et al.* Chromatin Interaction Analysis with Updated ChIA-PET Tool (V3). *Genes (Basel)* **10**, doi:10.3390/genes10070554 (2019).
- 170 Naumova, N., Smith, E. M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58**, 192-203, doi:10.1016/j.ymeth.2012.07.022 (2012).

- 171 Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature protocols* **2**, 1722-1733, doi:10.1038/nprot.2007.243 (2007).
- 172 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 173 Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311-316, doi:10.1093/bioinformatics/btz540 (2020).
- 174 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).
- 175 Wolff, J. *et al.* Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic acids research* **48**, W177-W184, doi:10.1093/nar/gkaa220 (2020).
- 176 Gong, Y. *et al.* Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nature communications* **9**, 542, doi:10.1038/s41467-018-03017-1 (2018).
- 177 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 178 Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome research* **27**, 1939-1949, doi:10.1101/gr.220640.117 (2017).

- 179 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
- 180 Carroll, K. J. *et al.* A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 338-343, doi:10.1073/pnas.1523918113 (2016).
- 181 Huang, W. *et al.* Loss of microRNA-128 promotes cardiomyocyte proliferation and heart regeneration. *Nature communications* **9**, 700, doi:10.1038/s41467-018-03019-z (2018).
- 182 Kumar, M., Haghighi, K., Kranias, E. G. & Sadayappan, S. Phosphorylation of cardiac myosin-binding protein-C contributes to calcium homeostasis. *The Journal of biological chemistry* **295**, 11275-11291, doi:10.1074/jbc.RA120.013296 (2020).
- 183 Dai, B. *et al.* Reduced collagen deposition in infarcted myocardium facilitates induced pluripotent stem cell engraftment and angiomyogenesis for improvement of left ventricular function. *Journal of the American College of Cardiology* **58**, 2118-2127, doi:10.1016/j.jacc.2011.06.062 (2011).
- 184 Ma, J., Hong, K. & Wang, H. S. Progesterone Protects Against Bisphenol A-Induced Arrhythmias in Female Rat Cardiac Myocytes via Rapid Signaling. *Endocrinology* **158**, 778-790, doi:10.1210/en.2016-1702 (2017).
- 185 Ying, H. *et al.* Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev* **30**, 355-385, doi:10.1101/gad.275776.115 (2016).

- 186 Von Hoff, D. D. *et al.* Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N Engl J Med* **369**, 1691-1703, doi:10.1056/NEJMoa1304369 (2013).
- 187 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
- 188 Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627, doi:10.1016/j.cell.2015.11.024 (2015).
- 189 Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA Cancer J Clin* **63**, 11-30, doi:10.3322/caac.21166 (2013).
- 190 Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324**, 1457-1461, doi:10.1126/science.1171362 (2009).
- 191 Stoica, A. F., Chang, C. H. & Pauklin, S. Molecular Therapeutics of Pancreatic Ductal Adenocarcinoma: Targeted Pathways and the Role of Cancer Stem Cells. *Trends Pharmacol Sci*, doi:10.1016/j.tips.2020.09.008 (2020).
- 192 Kleeff, J. *et al.* Pancreatic cancer. *Nat Rev Dis Primers* **2**, 16022, doi:10.1038/nrdp.2016.22 (2016).
- 193 Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* **74**, 2913-2921, doi:10.1158/0008-5472.CAN-14-0155 (2014).

- 194 Li, J., Wientjes, M. G. & Au, J. L. Pancreatic cancer: pathobiology, treatment options, and drug delivery. *AAPS J* **12**, 223-232, doi:10.1208/s12248-010-9181-5 (2010).
- 195 Sinn, M. *et al.* CONKO-005: Adjuvant Chemotherapy With Gemcitabine Plus Erlotinib Versus Gemcitabine Alone in Patients After R0 Resection of Pancreatic Cancer: A Multicenter Randomized Phase III Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **35**, 3330-3337, doi:10.1200/JCO.2017.72.6463 (2017).
- 196 Hoffman, R. M. & Bouvet, M. Nanoparticle albumin-bound-paclitaxel: a limited improvement under the current therapeutic paradigm of pancreatic cancer. *Expert Opin Pharmacother* **16**, 943-947, doi:10.1517/14656566.2015.1016912 (2015).
- 197 Halfdanarson, T. R. *et al.* A Phase II Randomized Trial of Panitumumab, Erlotinib, and Gemcitabine Versus Erlotinib and Gemcitabine in Patients with Untreated, Metastatic Pancreatic Adenocarcinoma: North Central Cancer Treatment Group Trial N064B (Alliance). *Oncologist* **24**, 589-e160, doi:10.1634/theoncologist.2018-0878 (2019).
- 198 Khushman, M. *et al.* Full dose neoadjuvant FOLFIRINOX is associated with prolonged survival in patients with locally advanced pancreatic adenocarcinoma. *Pancreatology* **15**, 667-673, doi:10.1016/j.pan.2015.08.010 (2015).
- 199 French, R. & Pauklin, S. Epigenetic regulation of cancer stem cell formation and maintenance. *Int J Cancer* **148**, 2884-2897, doi:10.1002/ijc.33398 (2021).

- 200 Hermann, P. C. *et al.* Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell stem cell* **1**, 313-323, doi:10.1016/j.stem.2007.06.002 (2007).
- 201 Li, C. *et al.* Identification of pancreatic cancer stem cells. *Cancer Res* **67**, 1030-1037, doi:10.1158/0008-5472.CAN-06-2030 (2007).
- 202 Lonardo, E. *et al.* Nodal/Activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell stem cell* **9**, 433-446, doi:10.1016/j.stem.2011.10.001 (2011).
- 203 Mueller, M. T. *et al.* Combined targeted treatment to eliminate tumorigenic cancer stem cells in human pancreatic cancer. *Gastroenterology* **137**, 1102-1113, doi:10.1053/j.gastro.2009.05.053 (2009).
- 204 Hermann, P. C. & Sainz, B., Jr. Pancreatic cancer stem cells: A state or an entity? *Semin Cancer Biol* **53**, 223-231, doi:10.1016/j.semcancer.2018.08.007 (2018).
- 205 Suva, M. L., Riggi, N. & Bernstein, B. E. Epigenetic reprogramming in cancer. *Science* **339**, 1567-1570, doi:10.1126/science.1230184 (2013).
- 206 Vincent, A. & Van Seuning, I. On the epigenetic origin of cancer stem cells. *Biochim Biophys Acta* **1826**, 83-88, doi:10.1016/j.bbcan.2012.03.009 (2012).
- 207 Rodriguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nat Med* **17**, 330-339, doi:10.1038/nm.2305 (2011).
- 208 Pauklin, S. & Vallier, L. Activin/Nodal signalling in stem cells. *Development* **142**, 607-619, doi:10.1242/dev.091769 (2015).

- 209 Abdel Mouti, M. & Pauklin, S. TGFB1/INHBA Homodimer/Nodal-SMAD2/3 Signaling Network: A Pivotal Molecular Target in PDAC Treatment. *Mol Ther* **29**, 920-936, doi:10.1016/j.ymthe.2021.01.002 (2021).
- 210 Chang, C. H. & Pauklin, S. ROS and TGFbeta: from pancreatic tumour growth to metastasis. *J Exp Clin Cancer Res* **40**, 152, doi:10.1186/s13046-021-01960-4 (2021).
- 211 Bertero, A. *et al.* Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev* **29**, 702-717, doi:10.1101/gad.255984.114 (2015).
- 212 Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806, doi:10.1126/science.1164368 (2008).
- 213 Zhang, H. *et al.* TGFbeta signaling in pancreatic ductal adenocarcinoma. *Tumour Biol* **36**, 1613-1618, doi:10.1007/s13277-014-2757-4 (2015).
- 214 Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics* **24**, 372-376 (2000).
- 215 Villodre, E. S., Kipper, F. C., Pereira, M. B. & Lenz, G. Roles of OCT4 in tumorigenesis, cancer therapy resistance and prognosis. *Cancer Treat Rev* **51**, 1-9, doi:10.1016/j.ctrv.2016.10.003 (2016).
- 216 Maeda, S. *et al.* CD133 expression is correlated with lymph node metastasis and vascular endothelial growth factor-C expression in pancreatic cancer. *Br J Cancer* **98**, 1389-1397, doi:10.1038/sj.bjc.6604307 (2008).

- 217 Kim, J. *et al.* An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. *Cell reports* **3**, 2088-2099, doi:10.1016/j.celrep.2013.05.036 (2013).
- 218 Leal, A. S. *et al.* Bromodomain inhibitors, JQ1 and I-BET 762, as potential therapies for pancreatic cancer. *Cancer letters* **394**, 76-87 (2017).
- 219 Xie, F. *et al.* The BET inhibitor I-BET762 inhibits pancreatic ductal adenocarcinoma cell proliferation and enhances the therapeutic effect of gemcitabine. *Scientific reports* **8**, 1-10 (2018).
- 220 Feng, Y., Liu, X. & Pauklin, S. 3D chromatin architecture and epigenetic regulation in cancer stem cells. *Protein Cell* **12**, 440-454, doi:10.1007/s13238-020-00819-2 (2021).
- 221 Feng, Y. & Pauklin, S. Revisiting 3D chromatin architecture in cancer development and progression. *Nucleic Acids Res* **48**, 10632-10647, doi:10.1093/nar/gkaa747 (2020).
- 222 Wang, P. *et al.* In situ Chromatin Interaction Analysis Using Paired-End Tag Sequencing. *Curr Protoc* **1**, e174, doi:10.1002/cpz1.174 (2021).
- 223 Zhu, Y. *et al.* Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer cell* **39**, 694-707 e697, doi:10.1016/j.ccell.2021.03.006 (2021).
- 224 Wanior, M., Kramer, A., Knapp, S. & Joerger, A. C. Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy. *Oncogene* **40**, 3637-3654, doi:10.1038/s41388-021-01781-x (2021).

- 225 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 226 Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* **42**, W187-W191, doi:10.1093/nar/gku365 (2014).
- 227 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 228 Lee, B. *et al.* ChIA-PIPE: A fully automated pipeline for comprehensive ChIA-PET data analysis and visualization. *Science Advances* **6**, eaay2078, doi:10.1126/sciadv.aay2078 (2020).
- 229 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).