KSBi-BIML 2026

Bioinformatics & Machine Learning(BIML) Workshop for Life Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

(Single-cell) 3D Epigenome Data Analysis

정인경 _ KAIST

KSBI KOREAN SOCIETY FOR BIOINFORMATICS | 한국생명정보학회

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2026 워크샵을 목적으로
제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우
발생하는 **모든 법적 책임은 행위자 본인에게 있음**을 알립니다.

KSBi-BIML 2026

Bioinformatics & Machine Learning (BIML) Workshop for Life Scientists

한국생명정보학회가 주최하는 BIML-2026 동계 Bioinformatics &Machine Learning 교육 워크숍에 여러분을 초대합니다.

BIML 워크숍은 생명정보학 연구자들이 최신 AI바이오 분야의 인공지능 기반 분석 기술과 바이오 데이터 분석 기법을 이론과 실습을 통해 체계적으로 배울 수 있는 전문 교육 프로그램입니다. 2015년에 시작된 BIML 워크숍은 올해로 12년 차를 맞이하며, 국내 생명정보학 분야의 최초이자 최고 수준의 교육 프로그램으로 자리 잡았습니다. 이번 워크숍은 크게 인공지능바이오(AI바이오) 분야와 디지털바이오 분야, 두 분야로 구성됩니다.

AI바이오 분야에서는 생명정보 분석에 폭넓게 응용되고 있는 다양한 인공지능 기반 자료 모델링 기법을 다룰 예정입니다. 특히, 인공지능 심층학습을 활용한 단백질 구조 예측, 유전체 분석, 신약 개발에 대한 이론 및 실습 강의가 진행됩니다.

또한 디지털바이오 분야에서는 단일세포오믹스, 공간오믹스, 멀티오믹스, 메타오믹스에 대한 강의도 마련되어 있어, 연구자들의 분석 역량 강화에 실질적인 도움을 줄 것으로 기대됩니다.

또한 2024년부터 추가된 의료정보 자료 분석을 다루는 강의를 올해도 지속해서 운영하고자 합니다. 이는 최근 의료정보 자료 분석에 관한 연구 수요 증가를 반영한 것으로, 관련 연구를 수행하는 의과학자 및 의료정보 연구자들에게 유용한 지침을 제공할 것입니다.

또한, 올해도 생명정보학 기술의 다양화에 발맞춰 온라인 강좌를 대폭 확대했습니다. 올해는 무료 강좌 10개를 포함한 총 40개 이상의 강좌가 개설되며, 연구 주제에 맞는 강좌 추천과 강연료 할인 혜택도 제공합니다.

BIML-2026는 국내 주요 연구 중심 대학의 전임 교수 및 각 분야 최고 전문가들의 강의로 구성되어 있으며, 기초 이론부터 최신 연구 동향까지 아우르는 심도 있는 교육의 장이 될 것으로 확신합니다.

여러분의 많은 관심과 참여를 기대합니다!

2026년 2월

한국생명정보학회장 류 성 호

강의개요

(Single-cell) 3D Epigenome Data Analysis

염색질 3차구조란 핵 내부에서 게놈이 3차원적으로 조직화되어 배열된 구조를 의미한다. 최근 연구에 따르면 염색질 3차구조는 무작위적으로 형성되는 것이 아니라, TAD(Topologically Associating Domain)나 loop domain과 같은 구조적 단위를 기본으로 하여 다층적인 계층 구조를 이루고 있다. 이러한 구조적 제약으로 인해 DNA 서열상 멀리 떨어져 있는 인핸서, 프로모터 등 다양한 전사 조절 요소들이 3차원 공간상에서는 서로 인접할 수 있으며, 이는 유전자 발현 조절의 핵심 원리로 제시되고 있다.

염색질 3차구조는 히스톤 변형, DNA 메틸화와 같은 후성유전적 변화와 밀접하게 연관되어 있으며, 이에 따라 염색질 3차구조와 후성유전체 정보를 통합적으로 이해하고자 하는 '3D epigenome' 연구가 최근 급격히 발전하고 있다. 특히나 최신의 연구들은 단일세포 멀티 오믹스 관점에서 single-cell 단위에서 게놈 3차구조와 유전자 발현 그리고 후성유전을 동시에 동정하려는 시도가 많이 이루어지고 있다.

본 강의에서는 염색질 3차구조를 중심으로 관련 이론, 주요 실험 기법, 그리고 기본적인 데이터 분석 방법을 실습과 함께 학습하고자 한다. 먼저 후성유전학의 기초 개념을 간략히 소개한 뒤, 염색질 3차구조의 전반적인 개념과 분석 방법을 다룬다. 이어서 최근 빠르게 발전하고 있는 염색질 3차구조 기반 단일세포 multi-omics 연구 동향을 소개하고, 이러한 데이터의 분석 workflow를 학습한다. 또한 본 연구팀이 개발한 3DIV 웹 기반 염색질 3차구조 분석 도구를 활용하여 Hi-C 데이터 분석 실습을 진행한다.

본 강의는 다음의 내용을 포함한다:

- 후성유전학 및 염색질 3차구조 개요
- 염색질 3차구조 기반 단일세포 multi-omics 개요
- 염색질 3차구조 데이터 분석 방법
- 3DIV 기반 Hi-C 데이터 분석 실습

* 교육생준비물: 노트북 (메모리 16GB 이상, 디스크 여유공간 30GB 이상)

* 강의 난이도: 중급

* 강의: 정인경 교수 (한국과학기술원 생명과학과)

Curriculum Vitae

Speaker Name: Inkyung Jung, Ph.D.

▶ **Personal Info**

Name	Inkyung Jung
Title	Associate Professor
Affiliation	KAIST

▶ **Contact Information**

Address	Department of biological sciences, KAIST
Email	ijung@kaist.ac.kr

Research Interest

Epigenetic gene regulation, 3D chromatin structure, single-cell multi-omics

Educational Experience

2006-2011	Ph.D. KAIST / Bio and Brain Engineering
2002-2006	B.S. KAIST / Biosystems

Professional Experience

2016-present	Assistant Professor, Associate Professor, Department of Biological Sciences, KAIST
2012-2016	Postdoctoral fellow, Ludwig Institute for Cancer Research
2011-2012	Postdoctoral fellow, KAIST

Selected Publications (5 maximum)

1. Wei X*, Xu Y*, Yang D*, Kim K, Lin X, Williams AB, Wang X, Srivas S, Li W, Li YE, Yue F, Huang ZH, Jung I#, Diao Y# (**2025**) scHiCAR: a tri-modal single-cell technology for integrated transcriptome, epigenome, and 3D genome analysis in complex tissues. **Nat Biotech.** (in press)
2. Park S*, Park H*, Byun G, Wei X, Eom J, Joo J, Lee AJ, Diao Y, Chung WS#, Jung I# (**2025**) NR3C1-mediated epigenetic regulation suppresses astrocytic immune responses. **Nat Commun.** Sep 22;16(1):8330
3. Song W*, Lee EE*, Park S*, Choi B, Kim MG, Choi SR, Kim JY, Kim SU, Kim JI, Shin EC, Jung I#, Lee JS#, Lee EY# (**2025**) Type 1 interferon signature and allograft inflammatory factor-1 contribute to refractoriness to TNF inhibition in ankylosing spondylitis. **Nat Commun.** July 1;16(1):5531
4. Lee AJ*, Kim C*, Park S, Jun K, Eom J, Lee S-J, Chung SJ, Rissman RA, Chung J, Masliah E#, Jung I# (**2023**) Characterization of altered molecular mechanisms in Parkinson's disease through cell type-resolved multi-omics analyses. **Sci Adv.** Apr 14;9(15):eabo2467
5. Joo J*, Cho S*, Hong S, Min S, Kim K, Kumar R, Choi J, Shin Y#, Jung I# (**2023**) Probabilistic establishment of speckle-associated inter-chromosomal interactions, **Nucleic Acids Res.** Apr 4; gkad211

KSBI | 한국생명정보학회

KSBi-BIML 2026

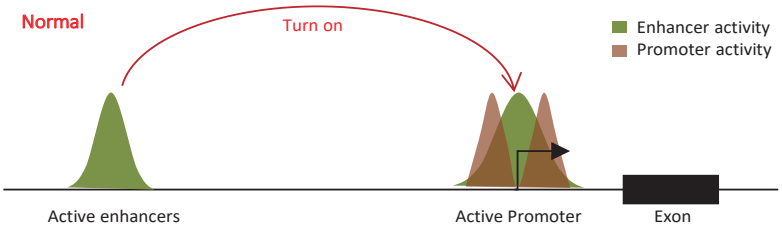
(Single-cell) 3D Epigenome Data Analysis

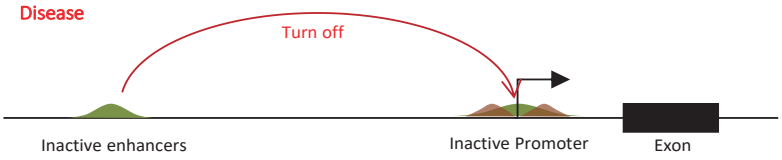
정인경(KAIST)

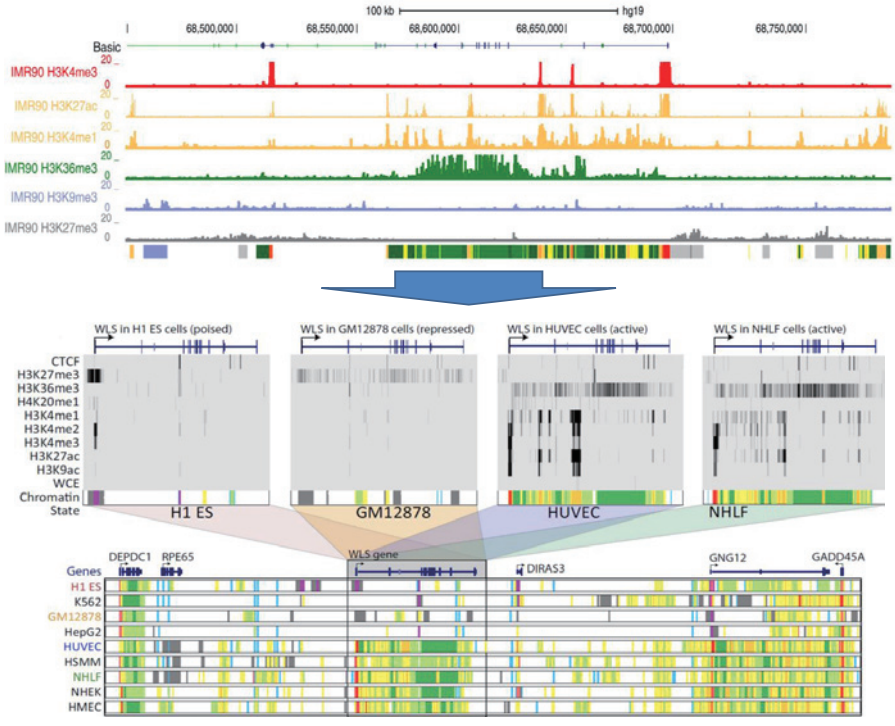
Contents

1. 후성유전학/염색질 3차구조 개요
2. 염색질 3차구조 중심의 단일세포 multi-omics 개요
3. 염색질 3차구조 데이터 분석 방법
4. 3DIV 기반 Hi-C 데이터 분석 실습

Gene Regulation

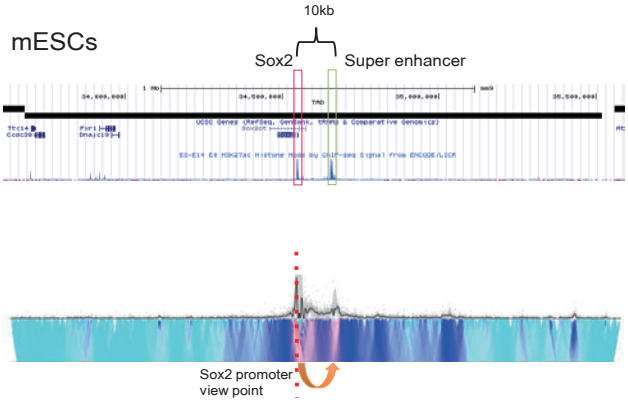


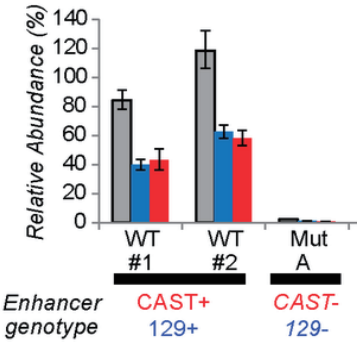




Nature (2011)

Long-range enhancer control of gene expression

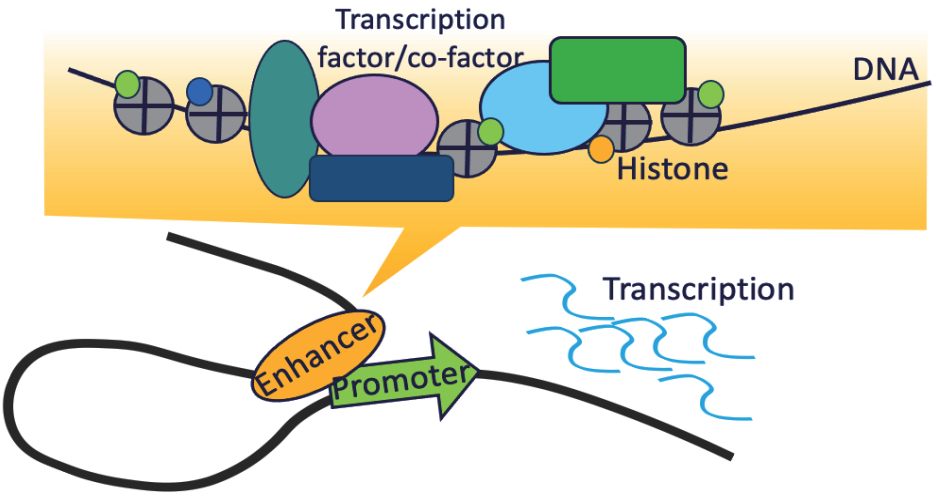




Li et al., 2014

Chromatin Looping

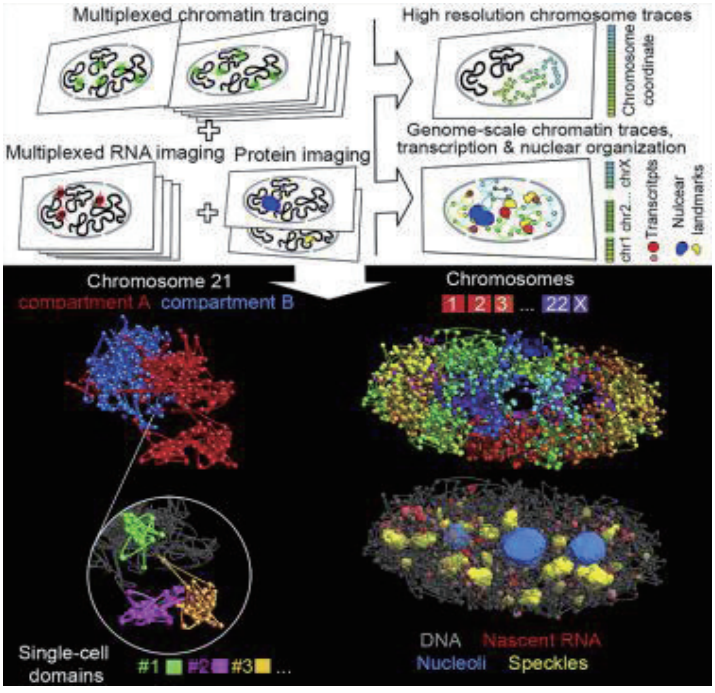
- a key principle of gene regulation -



How does enhancer control distal gene expression?

40km (Genome)

How can we investigate 3D genome organization?

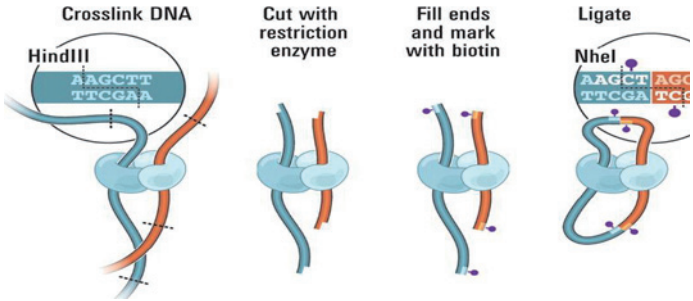


A multiplexed error-robust fluorescence *in situ* hybridization (MERFISH)

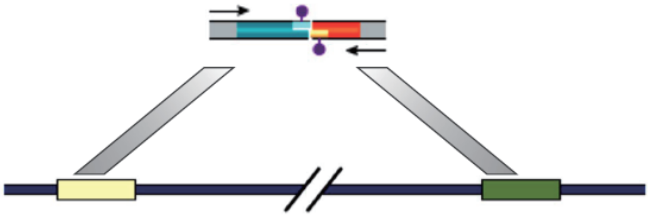
https://www.sciencedirect.com/science/article/pii/S0092867420309405

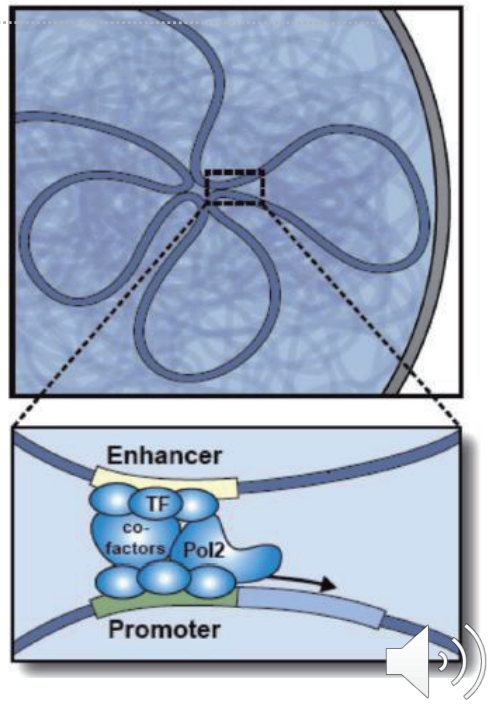
How can we investigate 3D genome organization?

Hi-C: **Hi**gh-throughput **c**hromosome conformation capture (3C)



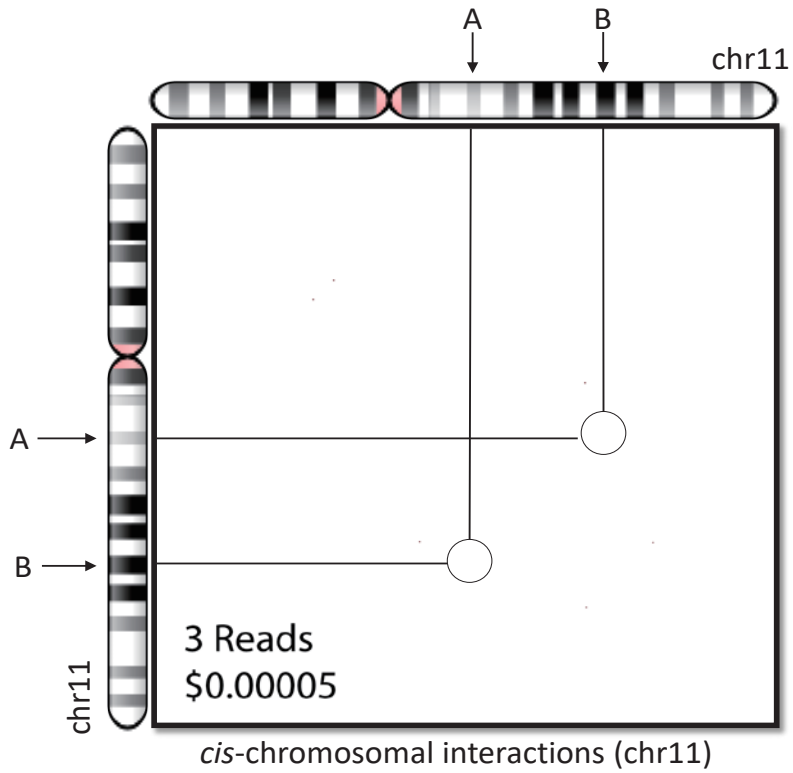
Cross-linking → Digestion → Re-lig





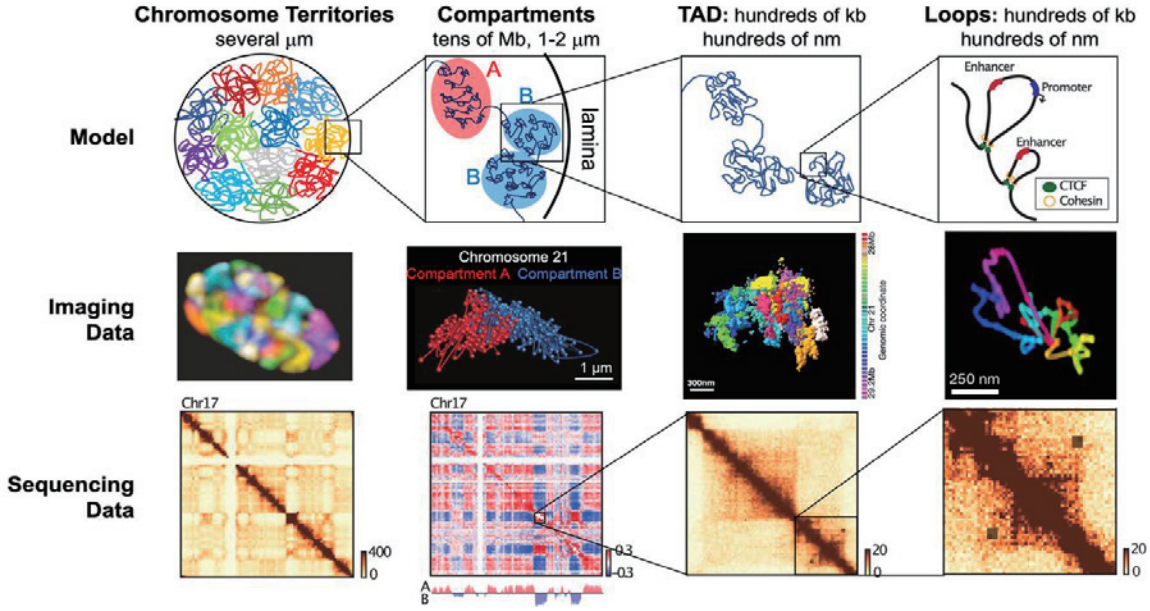
Lieberman et al., Science (2009)

Hi-C contact map to visualize 3D genome



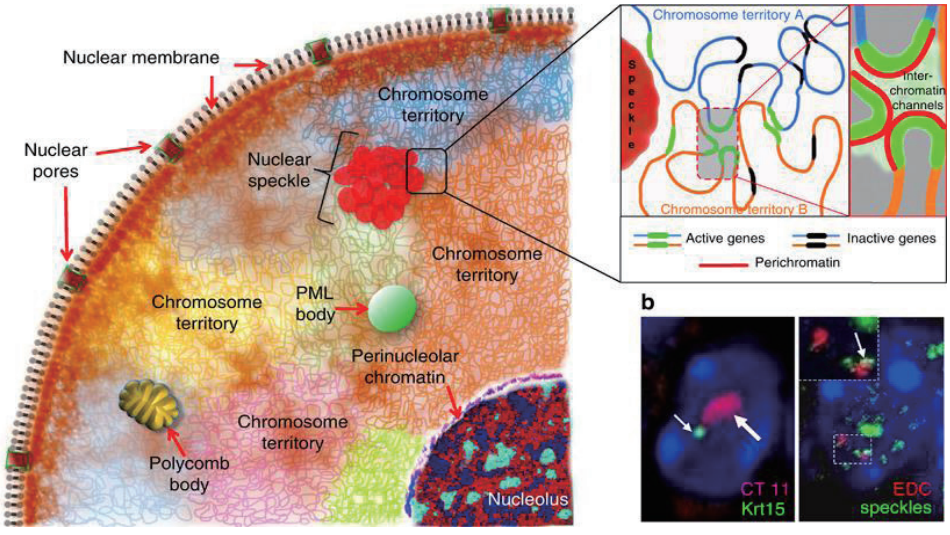
cis-chromosomal interactions (chr11)

Multi-layered 3D genome organization



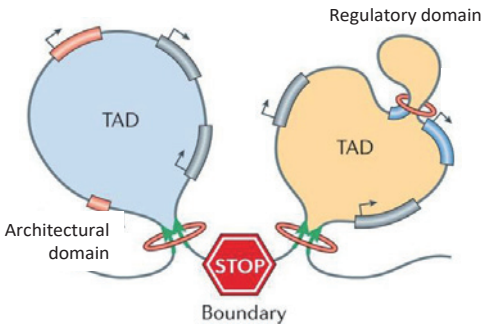
Shim SH (2021); Kim et al., (2019)

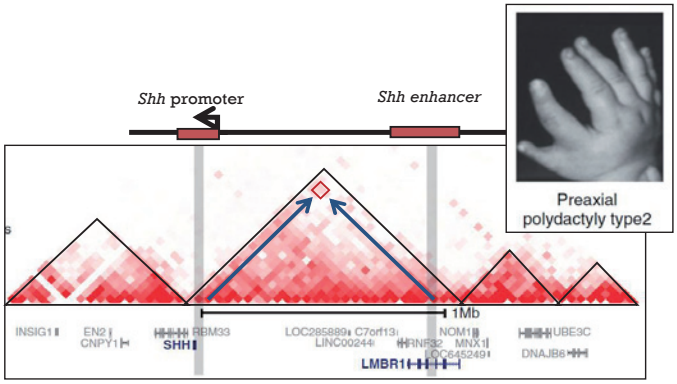
Genome organization in 3D nuclear space

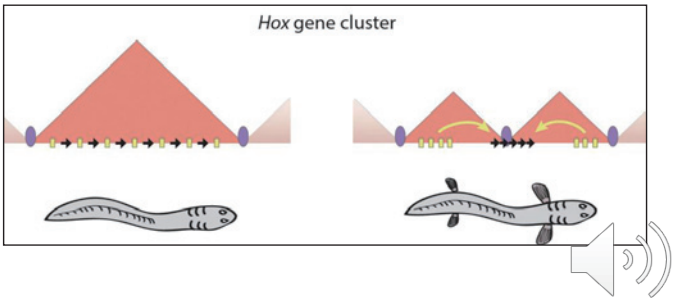


Botchkarev et al., 2012

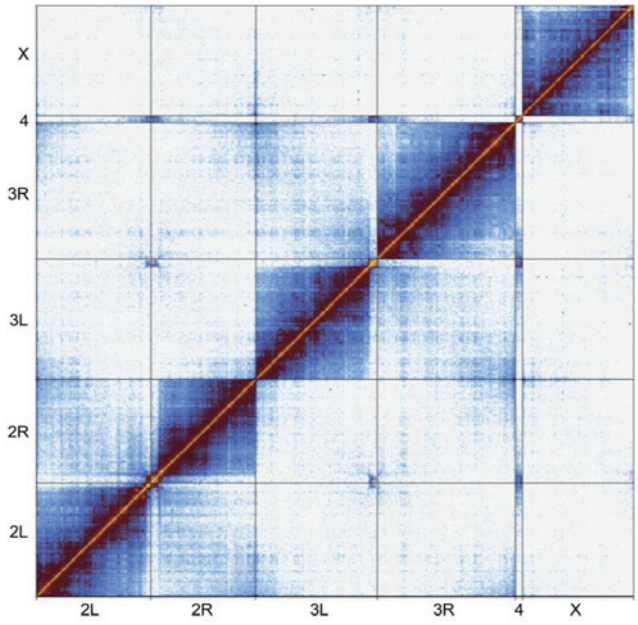
Topological Associating Domain (TAD)
a basic unit of gene regulation



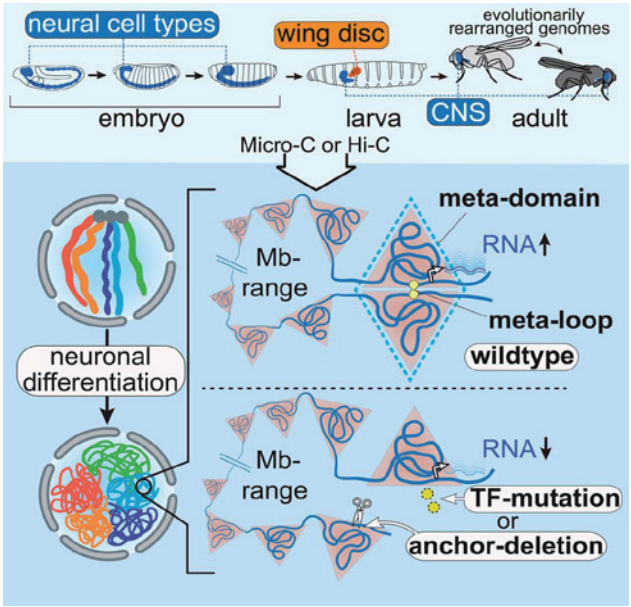




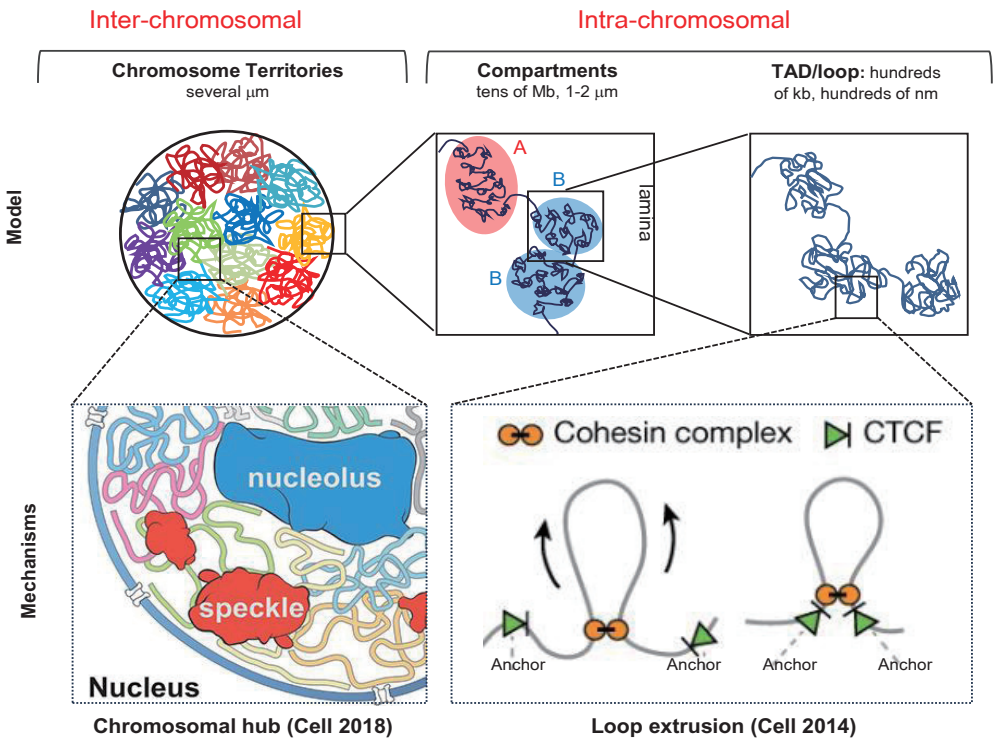
- **Architectural role of TAD**
  - A building block of 3D genome
  - A unit of DNA replication, DNA repair process, heterochromatin homeostasis
- **Gene regulatory role of TAD**
  - Gene regulatory domain
  - Disease-specific gene expression



Sexton, Cell 2012 from Giacomo Cavalli lab

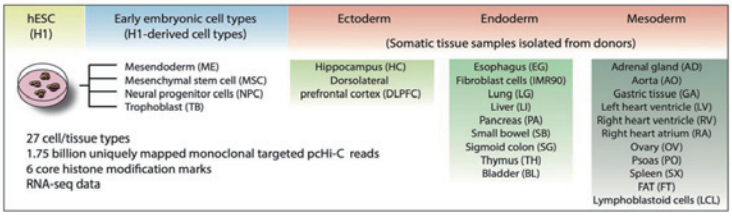


Mohana, Cell 2023 from Michael Levine & Maria Gambetta

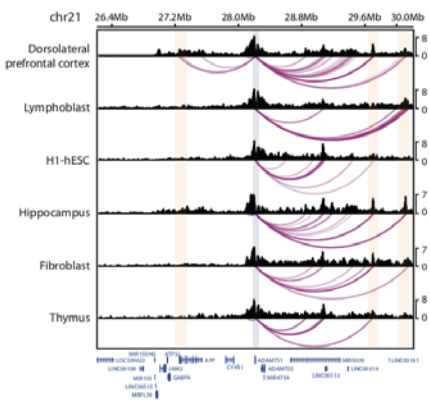


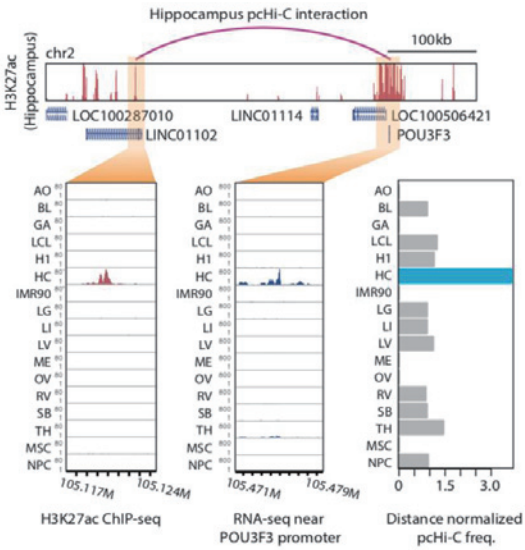
Dixon & Jung et al., Nature 2015; Leung & Jung Nature 2015; Dahl & Jung et al., Nature 2016
Kim, Nat Commun 2024; Kim, Cell Rep 2023; Lee, Sci Adv 2023; Kim, Nucleic Acid Res 2021; Jung et al., Nat Genet. 2019; Yang, Nucleic Acid Res 2018

Cell-type specific long-range gene regulation



1. **Interactions are tissue-specific**
2. **Tissue-specific interactions correlate with tissue-specific gene expression**



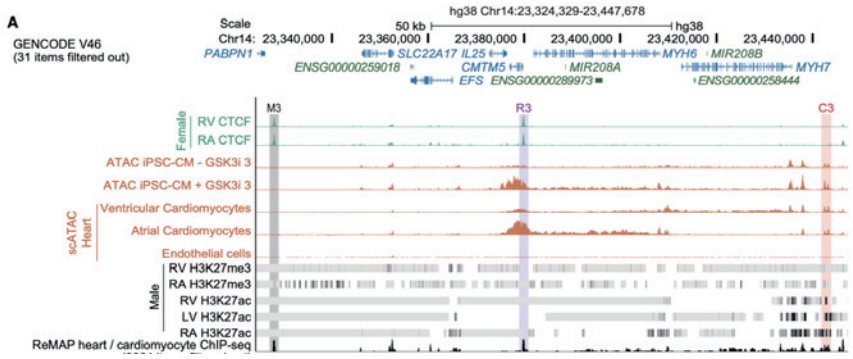


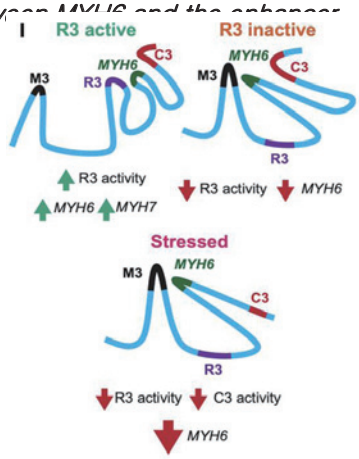
Yang et al., Nucleic Acids Res. (2018), Jung et al., Nat Genet. (2019)

Research

A gene regulatory element modulates myosin expression and controls cardiomyocyte response to stress

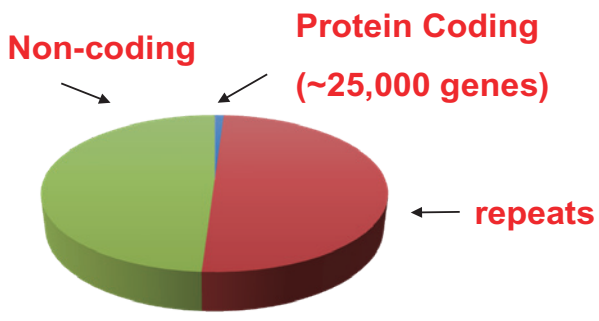
We identify and *characterize a gene regulatory element* essential for regulating MYH6 expression, which encodes human fetal myosin. Using chromatin conformation assays in combination with epigenome editing, we find that *gene regulation is mediated by a direct interaction betwe*



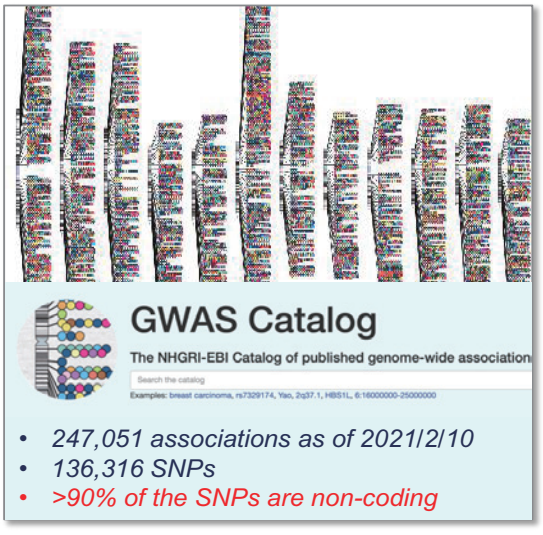


In collaboration with Charles Gersbach at Duke University (Genome Res. 2025)

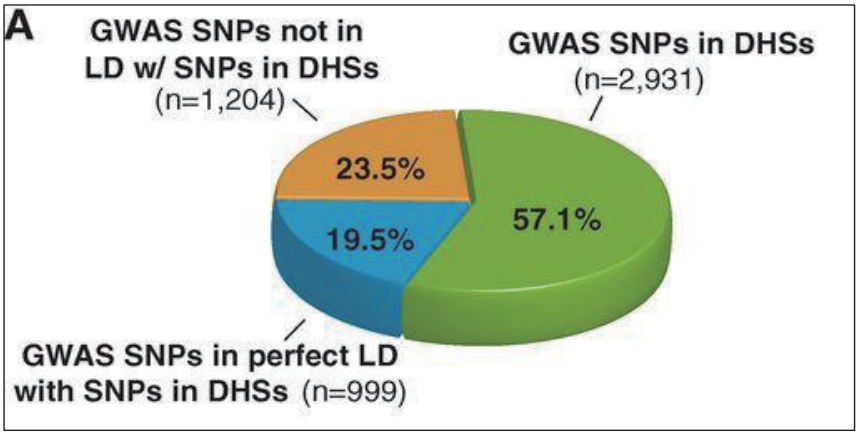
3D genome: a new way to interpret the human genome



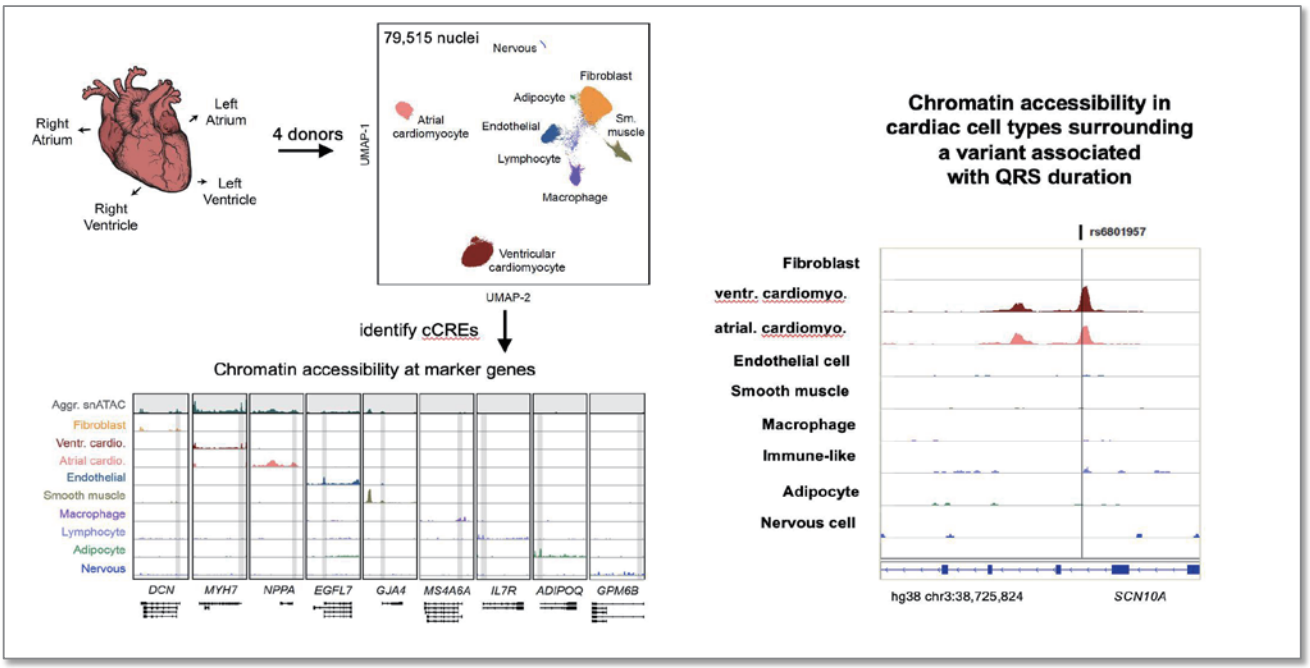
3x10^9 basepairs



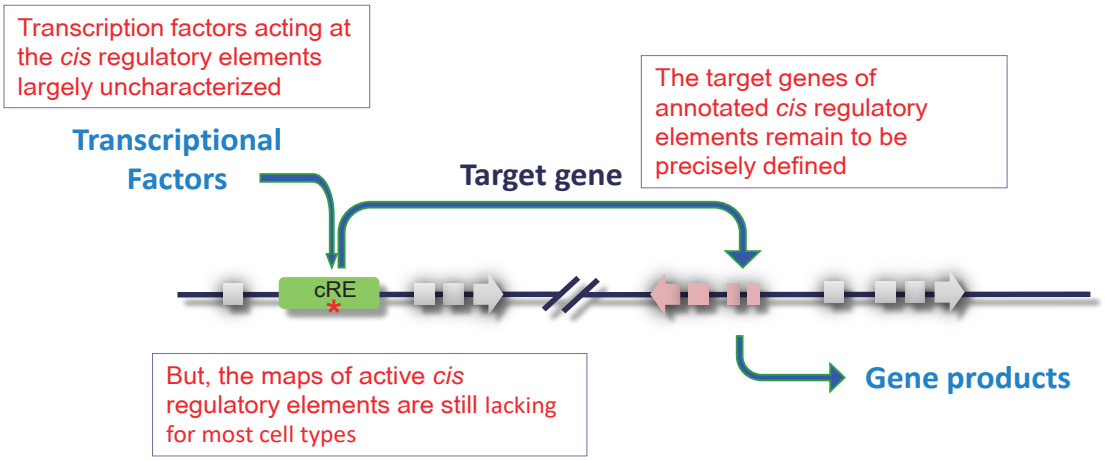
https://www.ebi.ac.uk/gwas/



https://www.ebi.ac.uk/gwas/, Maurano et al., Nature (2012)



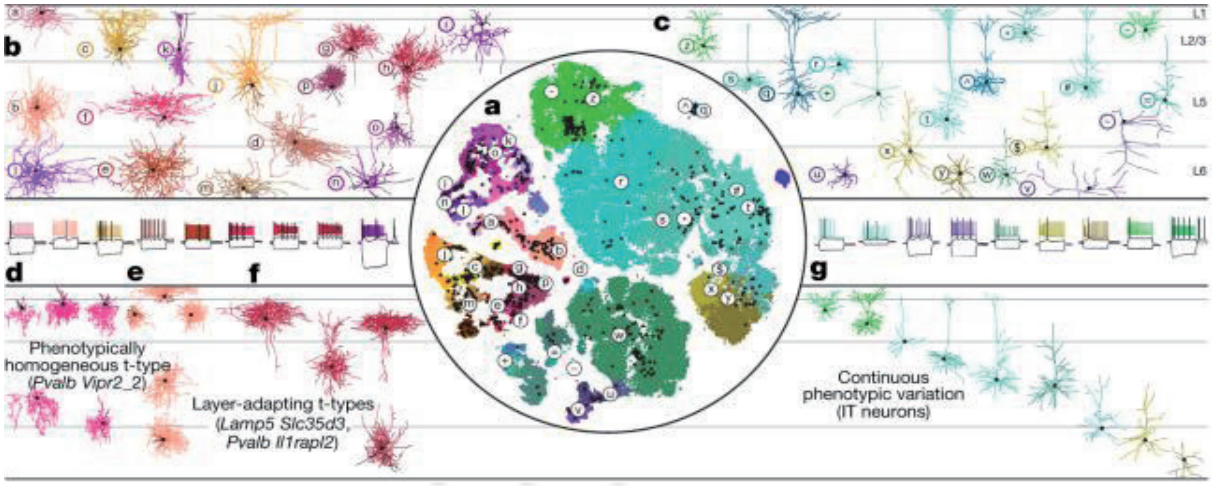
Hypothesis: noncoding variants disrupt *cis*-regulatory elements active in disease-relevant cell types

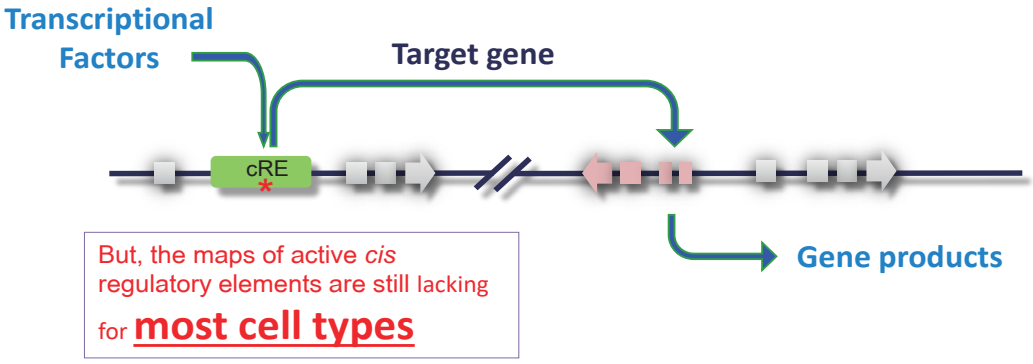


cRE: cis regulatory elements

Contents

1. 후성유전학/염색질 3차구조 개요
2. **염색질 3차구조 중심의 단일세포 multi-omics 개요**
3. 염색질 3차구조 데이터 분석 방법
4. 3DIV 기반 Hi-C 데이터 분석 실습

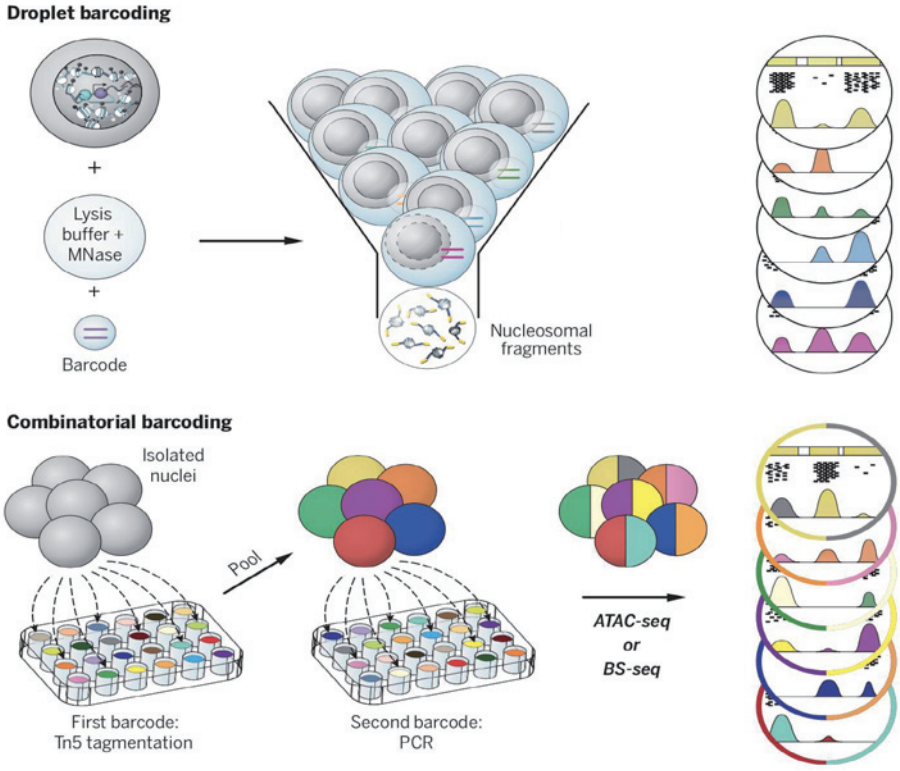




But, the maps of active *cis* regulatory elements are still lacking for **most cell types**

How can we map cell-type specific cREs?

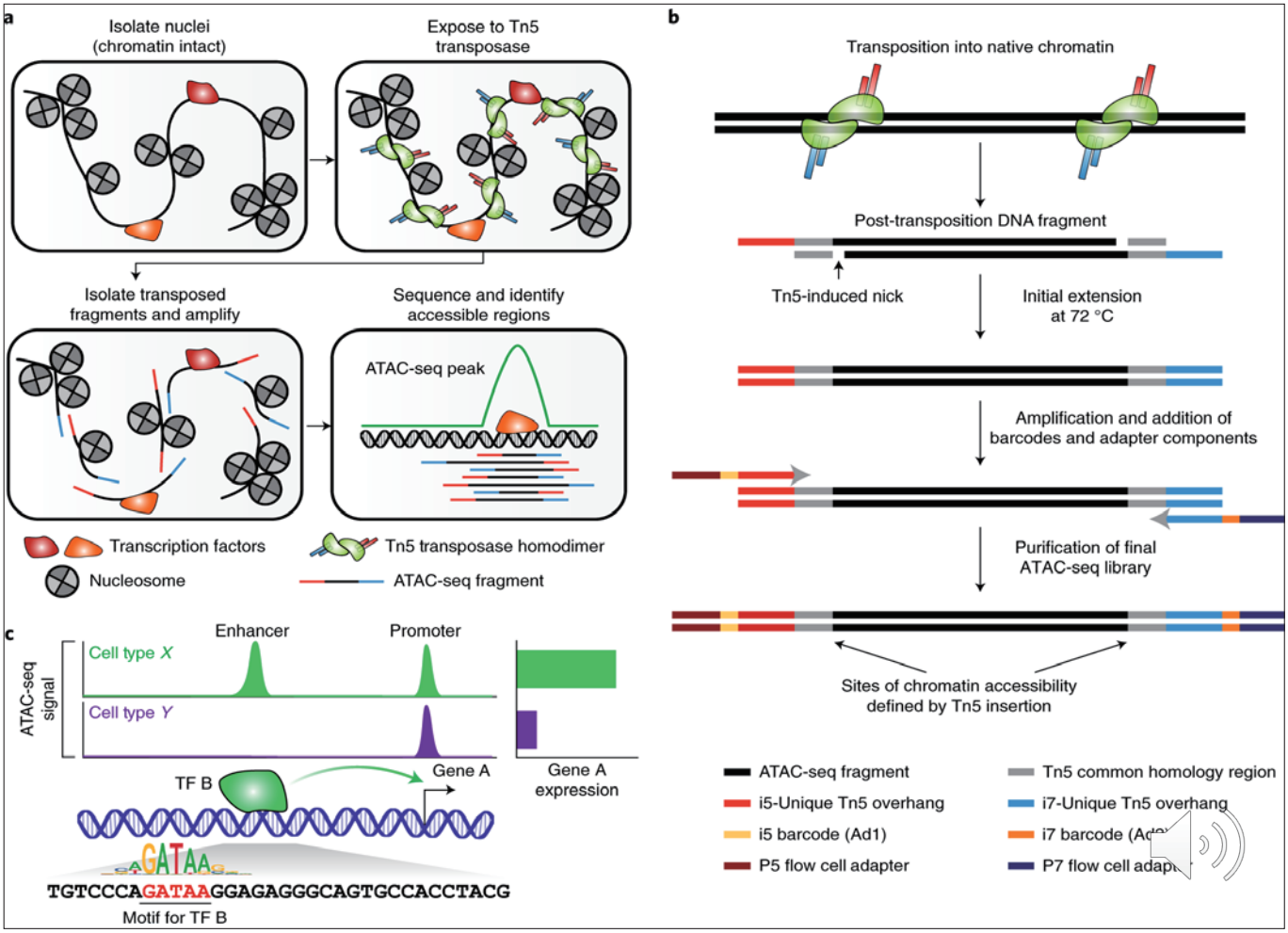
Single-cell Epigenomics



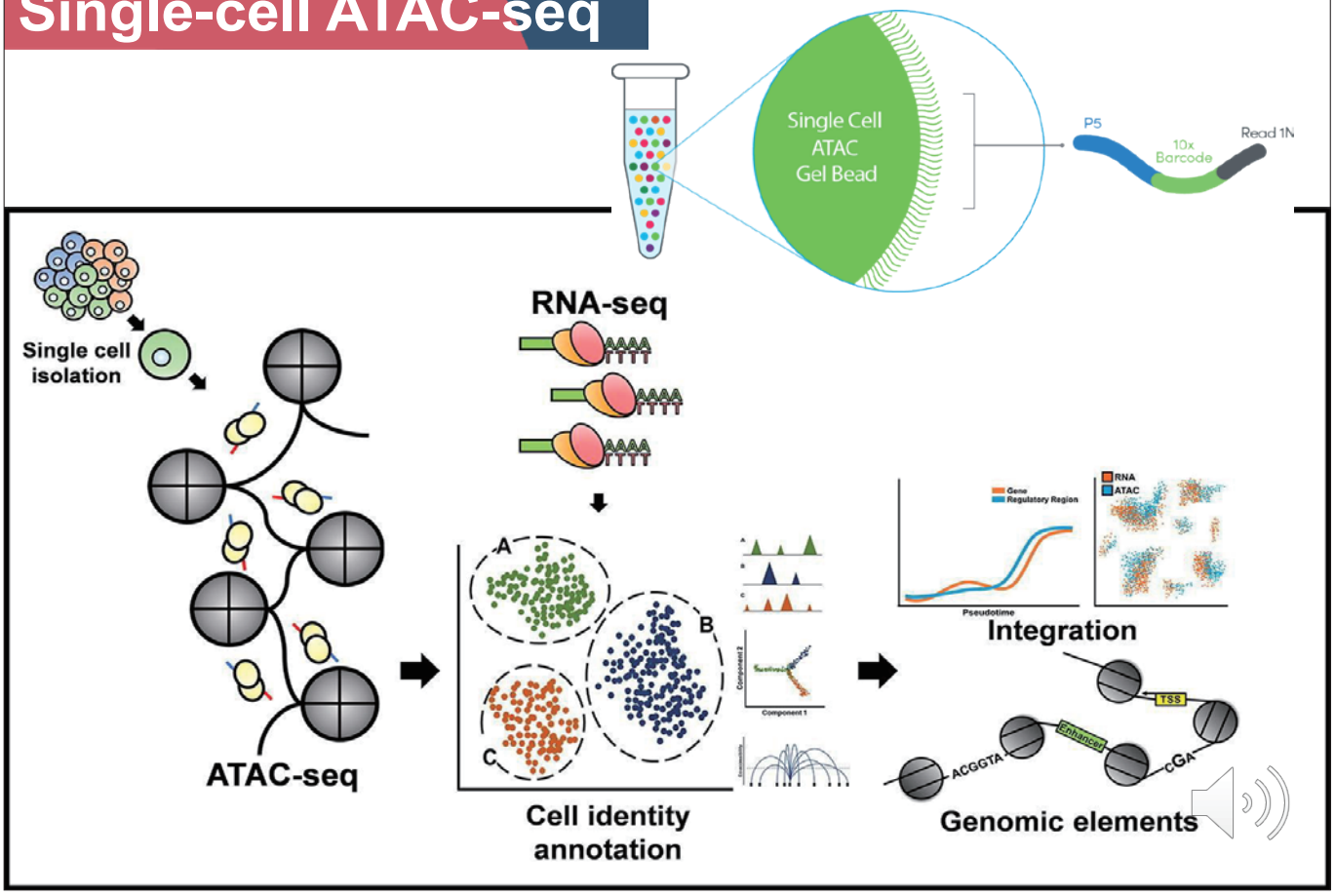
- Commercially available
- Robust
- scATAC-seq
- sc-RNA/ATAC Multiome

- Customizable
- Highly scalable (10^4-10^6)
- RNA, ATAC, HiC, mC, etc
- Mutil-omics
- No specialized Instrument

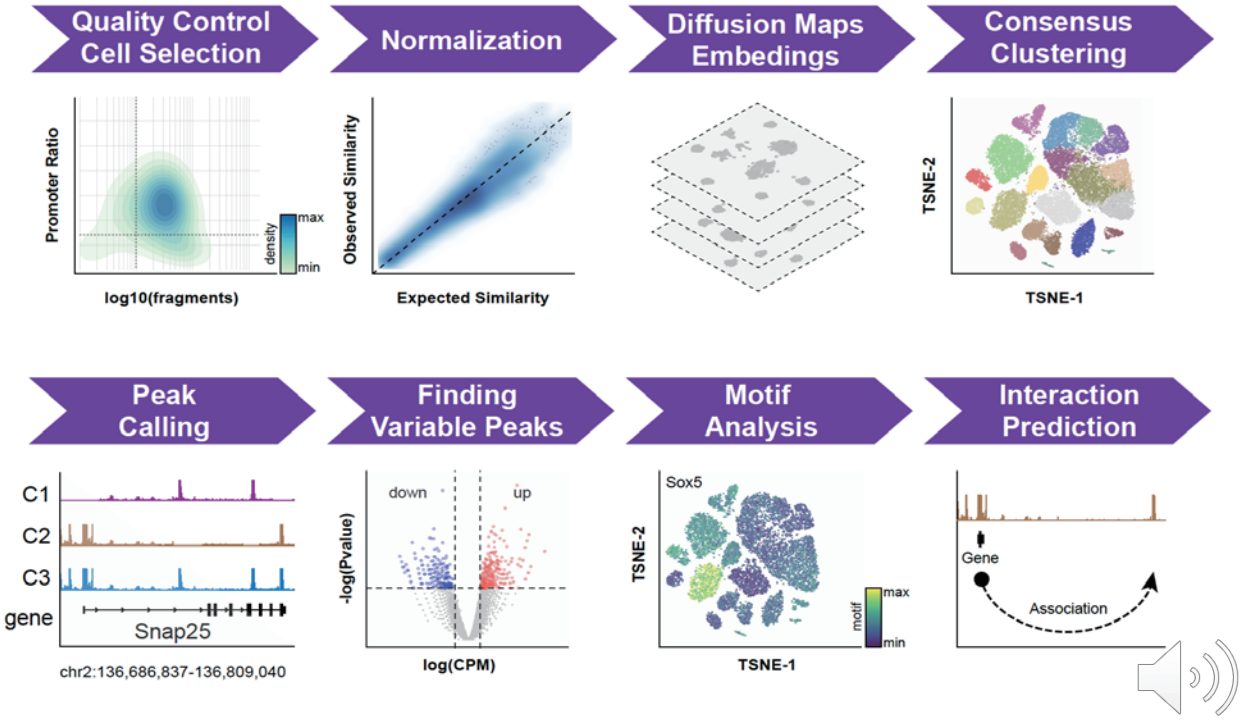
Modified from Kelsey, Stegle and Reik, *Science* 358, 69 (2017)



Single-cell ATAC-seq

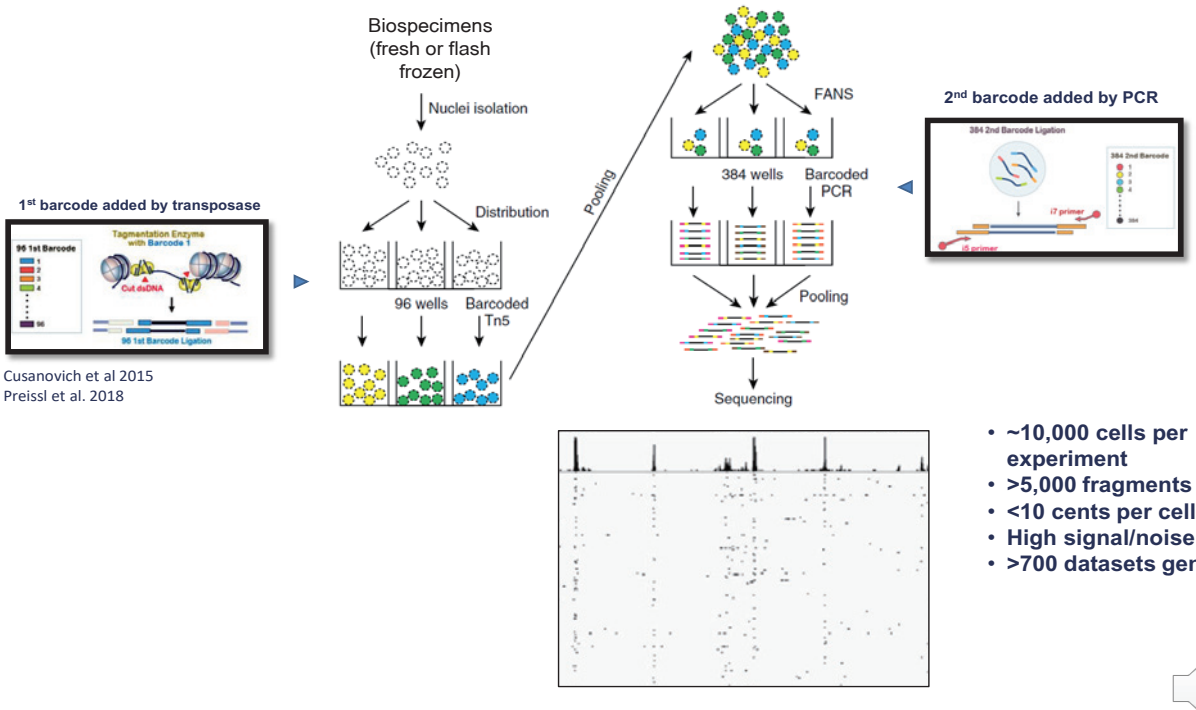


Single Nucleus Analysis Pipeline for ATAC-seq



Nature Communication (2021)

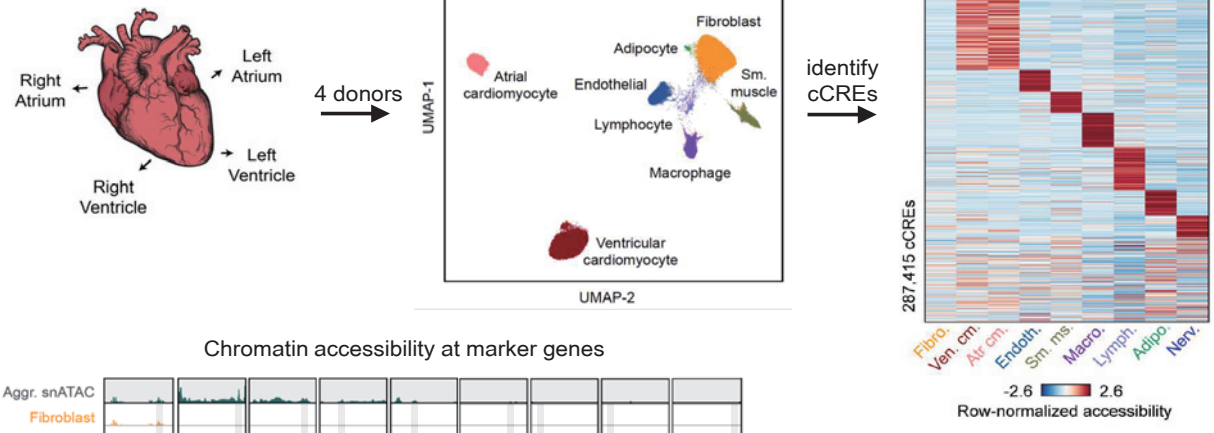
snATAC-seq by a combinatorial barcoding



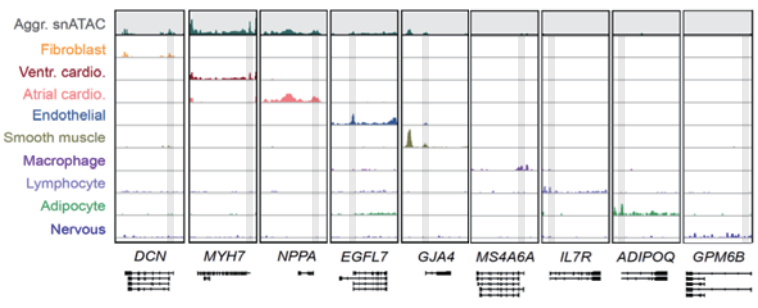
Cusanovich et al 2015
Preissl et al. 2018

- ~10,000 cells per experiment
- >5,000 fragments per cell
- <10 cents per cell
- High signal/noise
- >700 datasets generated

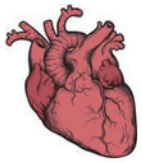
Single Nucleus ATAC-seq Atlas of the Human

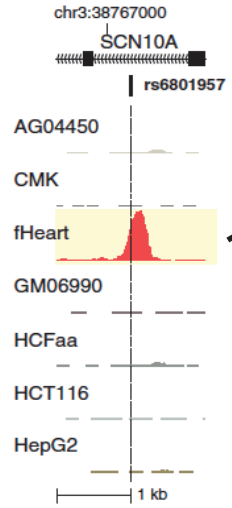


Chromatin accessibility at marker genes

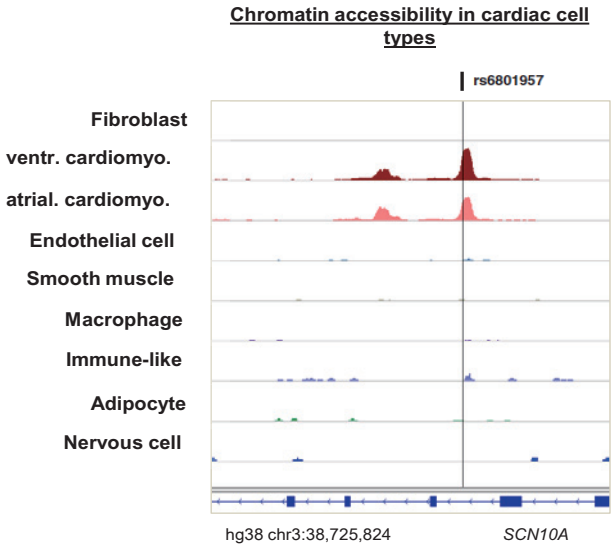


Single nucleus chromatin accessibility maps identify the likely cardiac cell types affected by risk variants

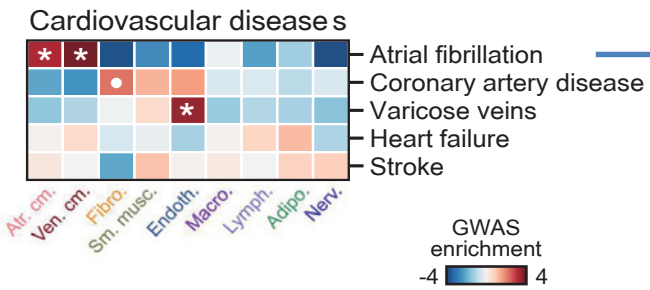




Maurano et al., Science 2012



Risk variant enrichment links cardiovascular disease risk variants to cardiac cell types



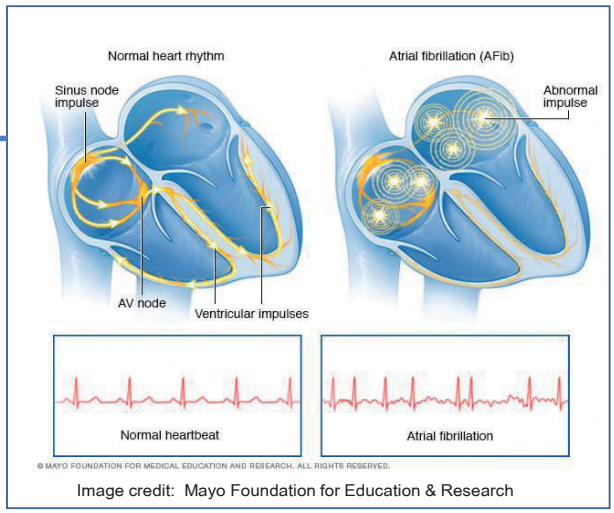
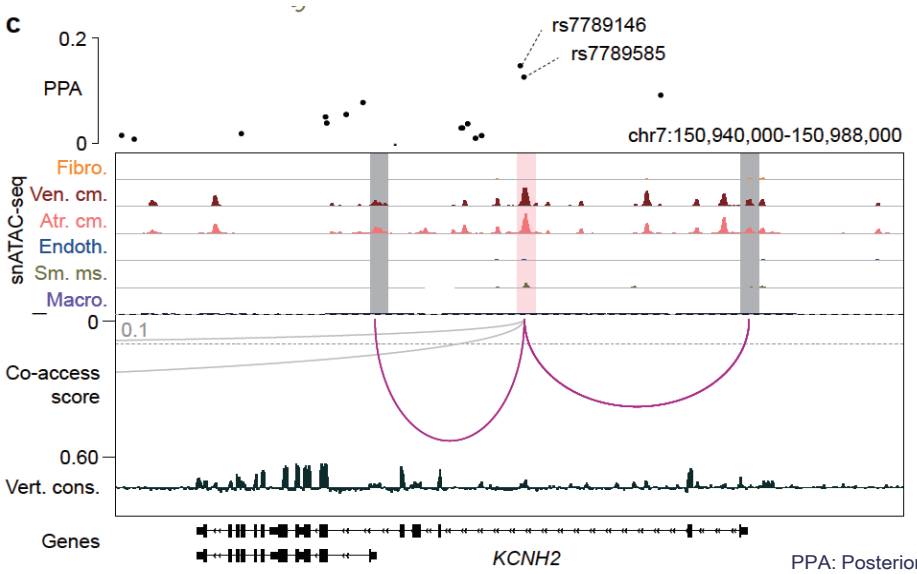


Image credit: Mayo Foundation for Education & Research

Atrial fibrillation (심방세동): 심방에서 비정상적인 전기신호가 나타나 심방이 제대로 수축하지 못하고 빠르고 불규칙하게 떠는 상태를 의미

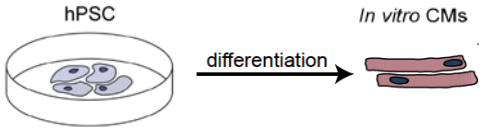
A cardiomyocyte cCRE at the *KCNH2* locus contains fine-mapped AF risk variants



Hocker et al. Science Advances, 2021

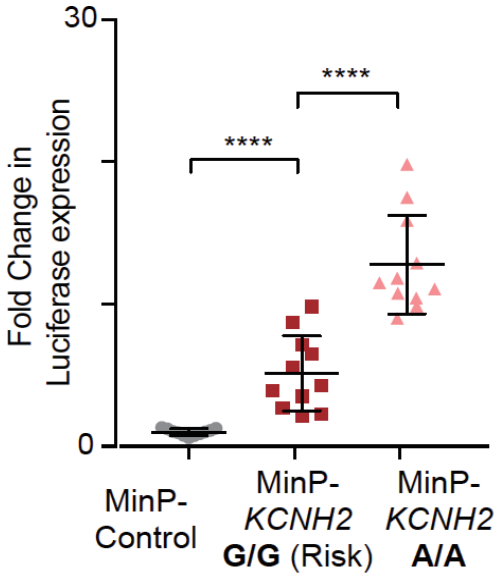
PPA: Posterior probability of association
AF: Atrial Fibrillation

AF risk alleles reduce enhancer activity of *KCNH2* cCRE in hPSC-cardiomyocytes

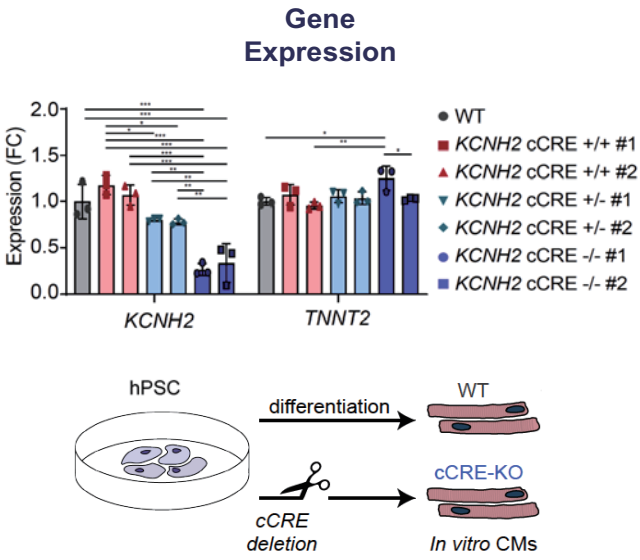


hPSC: human pluripotent stem cell
CM: cardiomyocytes
MinP: Minimal promoter

Luciferase reporter assay



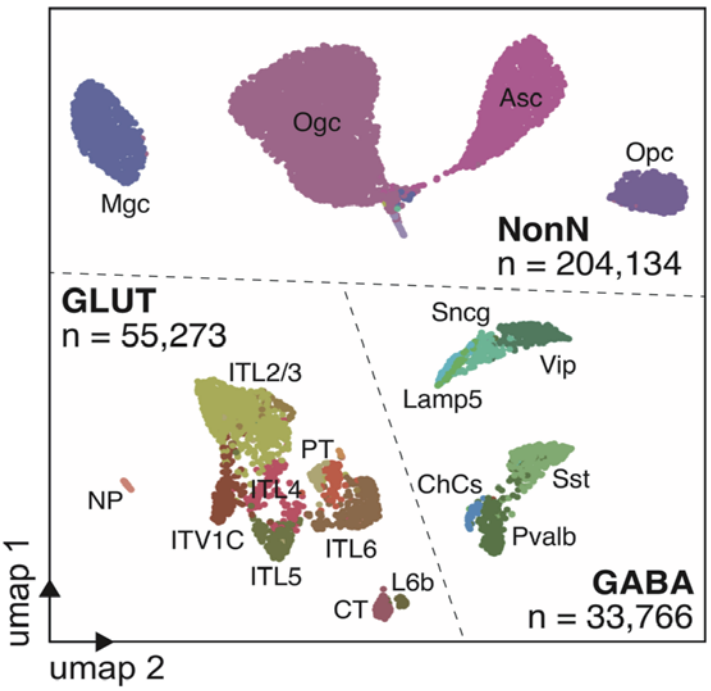
KCNH2 CRE is required for KCNH2 expression in hPSC-CMs

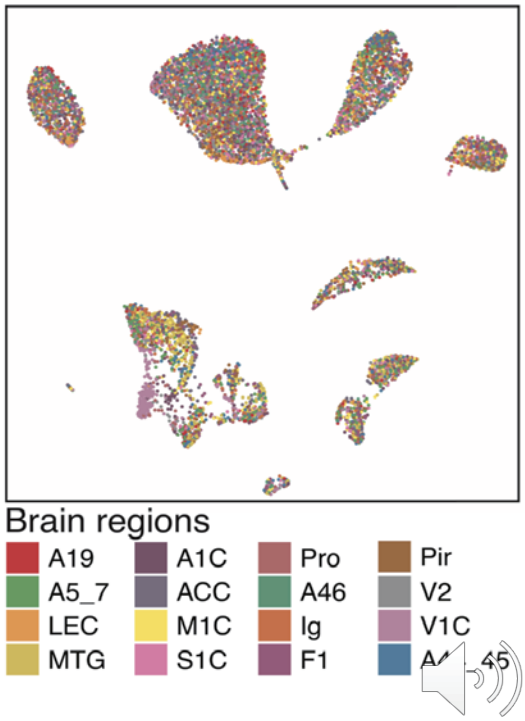


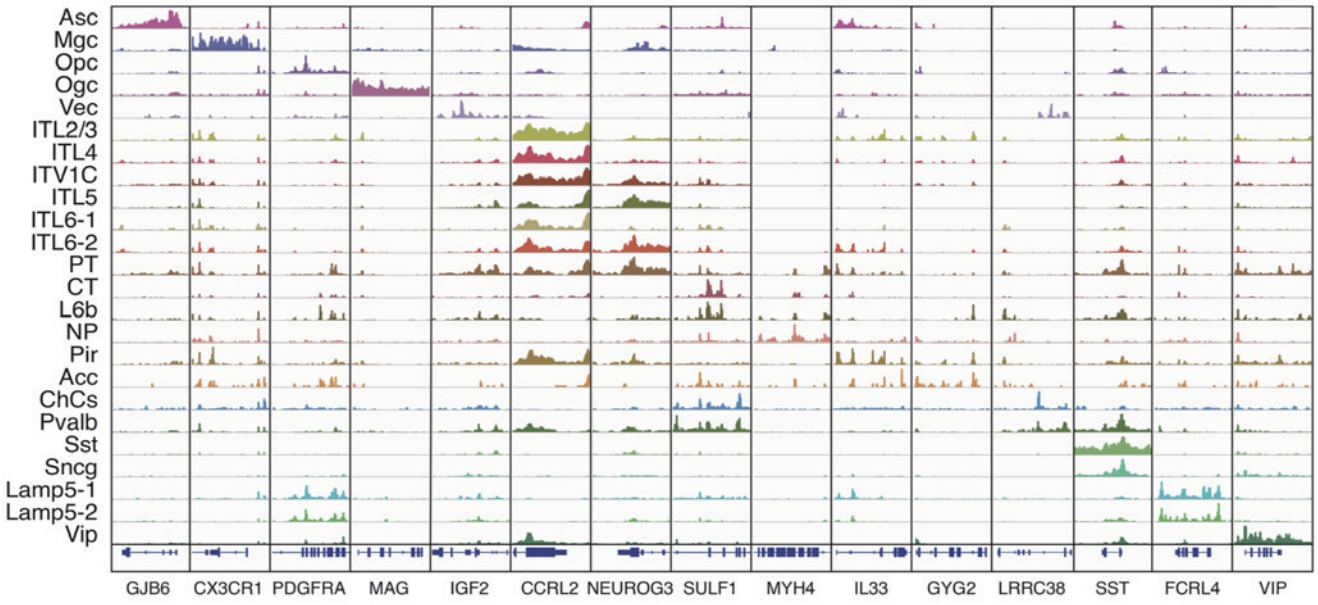
Fugui Zhu, Chi lab (UCSD)

hPSC: human pluripotent stem cell
CM: cardiomyocytes

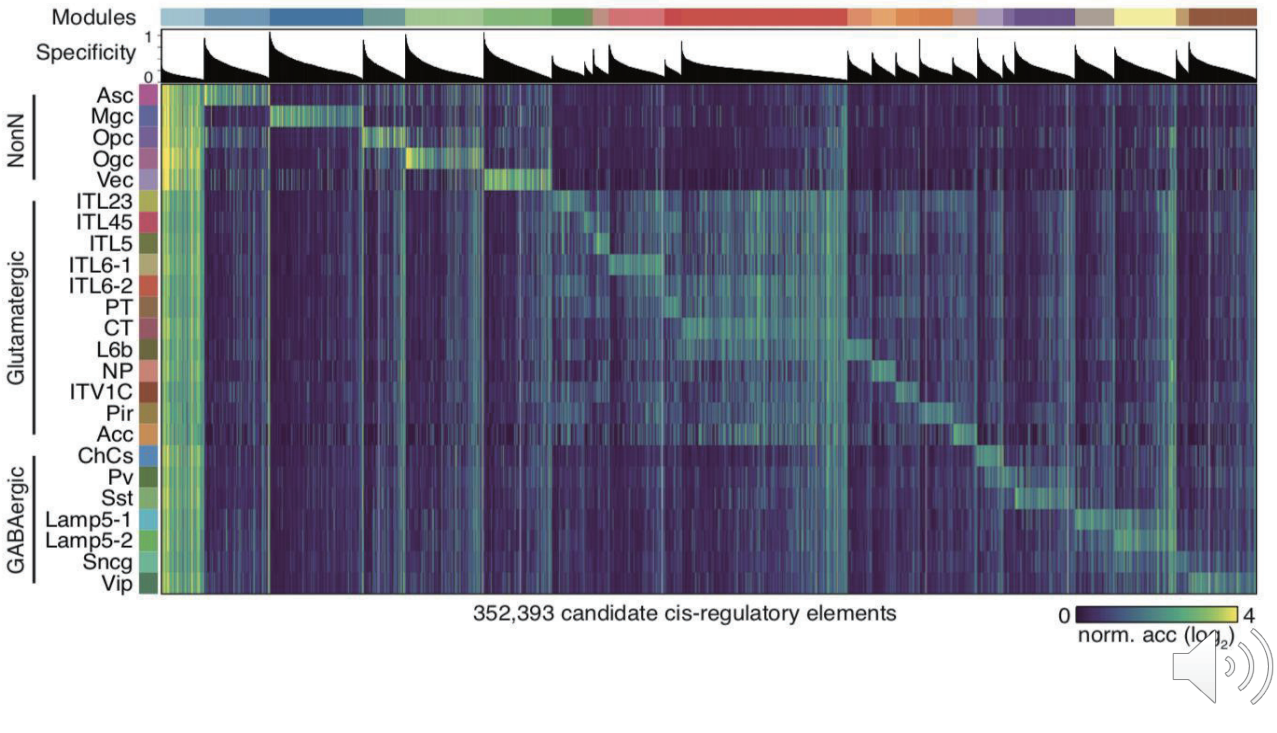
snATAC-seq in Human Cortex

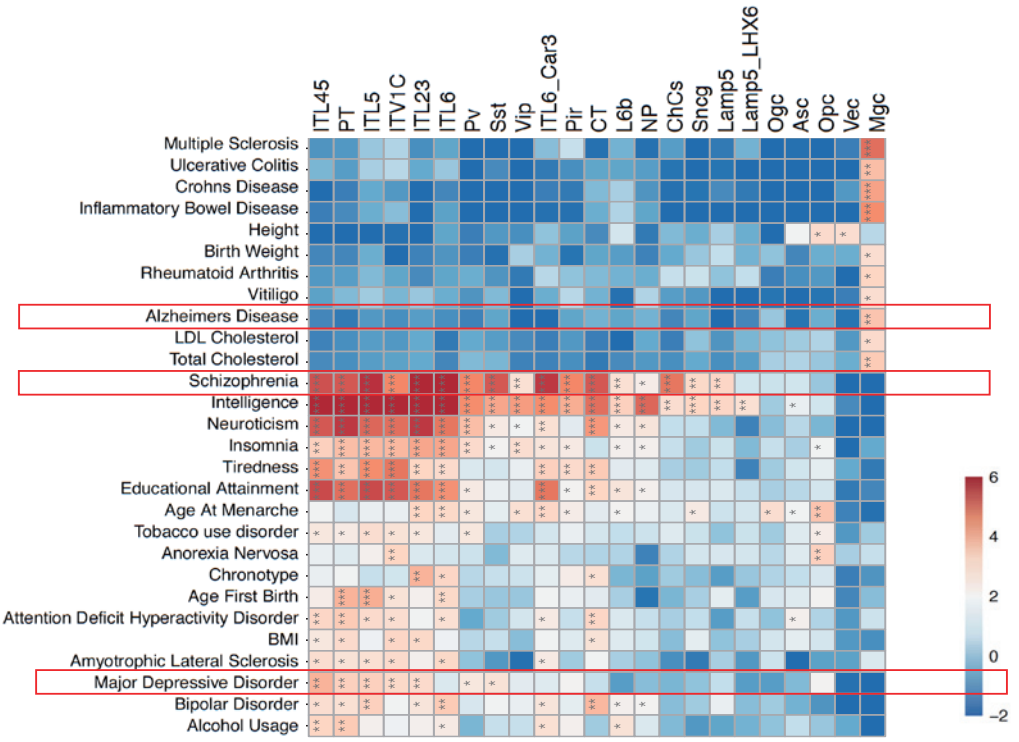




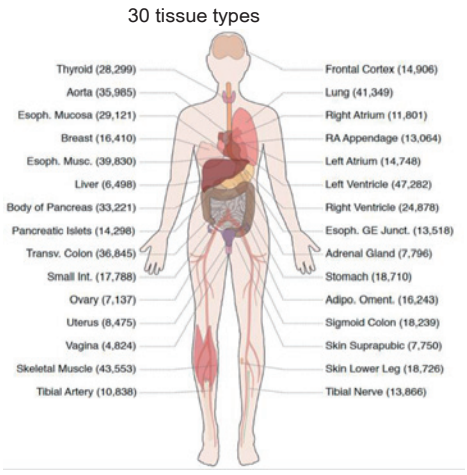


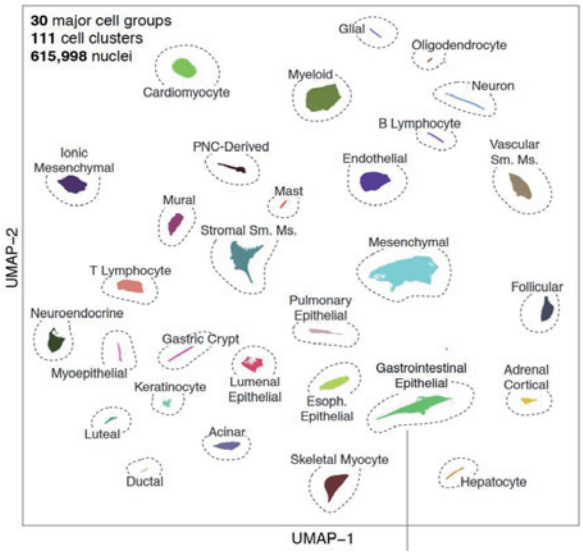
Annotation of CREs across human cortical cell types



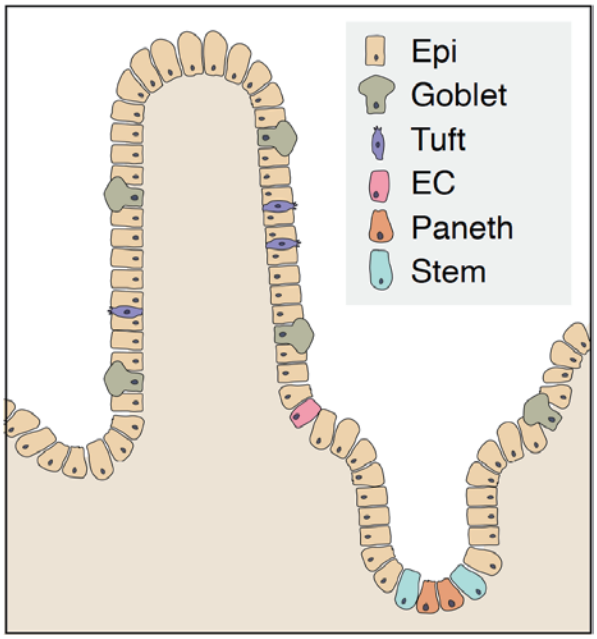


snATAC-seq in 30 Adult Human Tissues

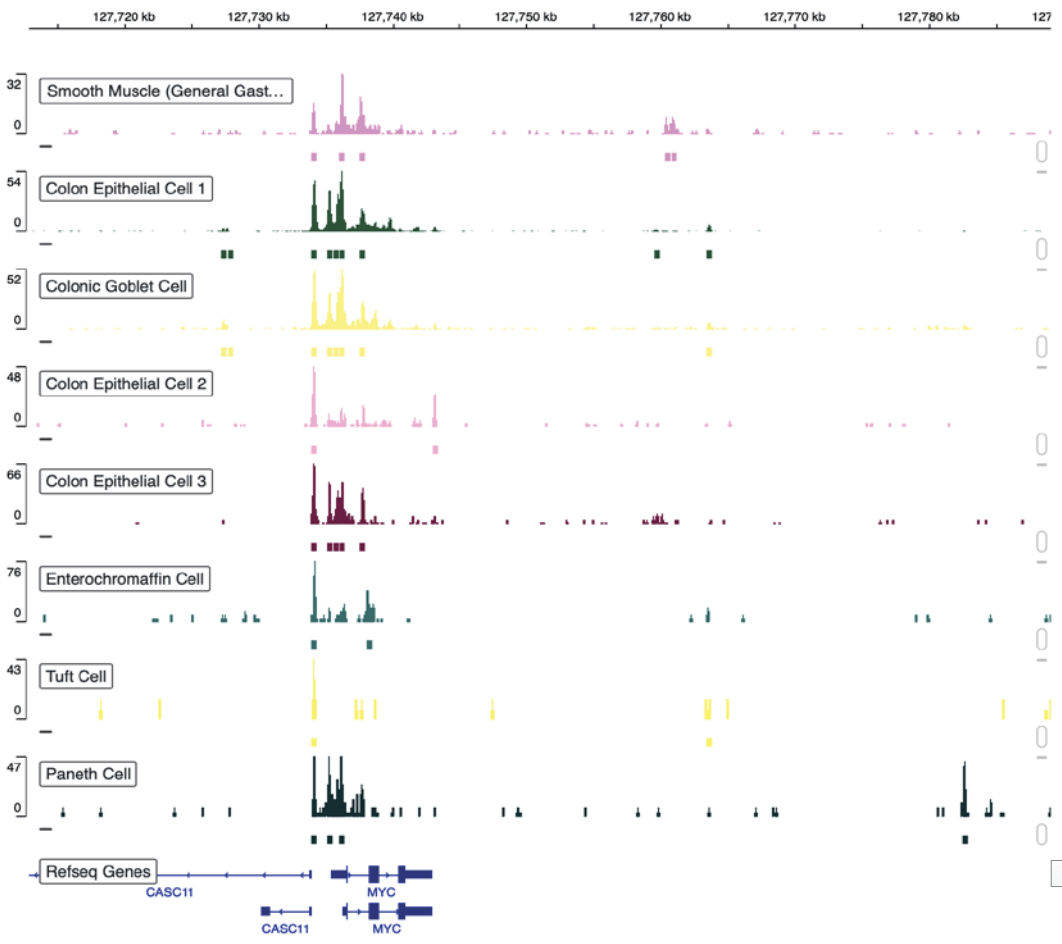


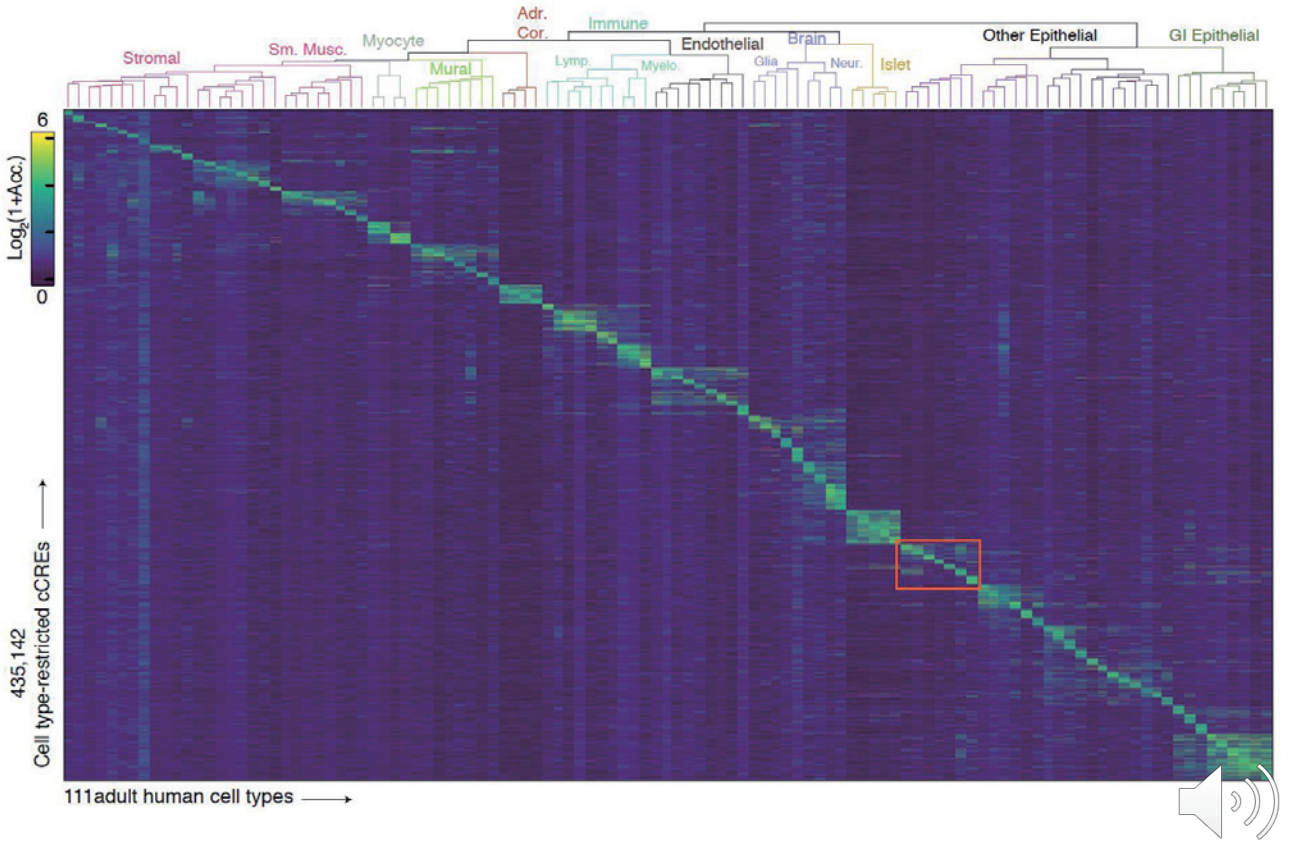


Iterative clustering

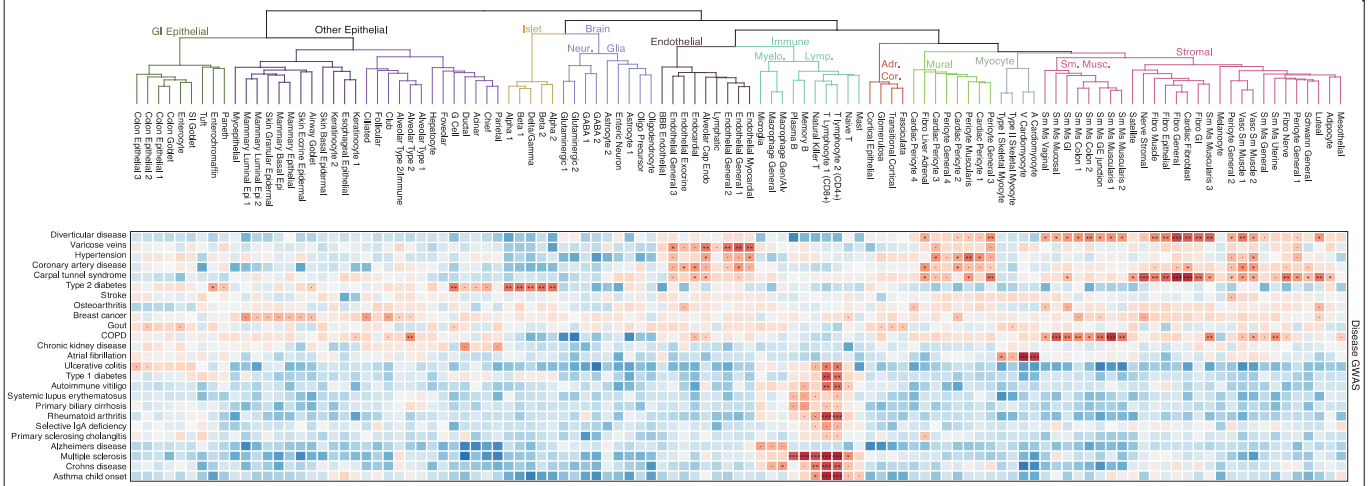


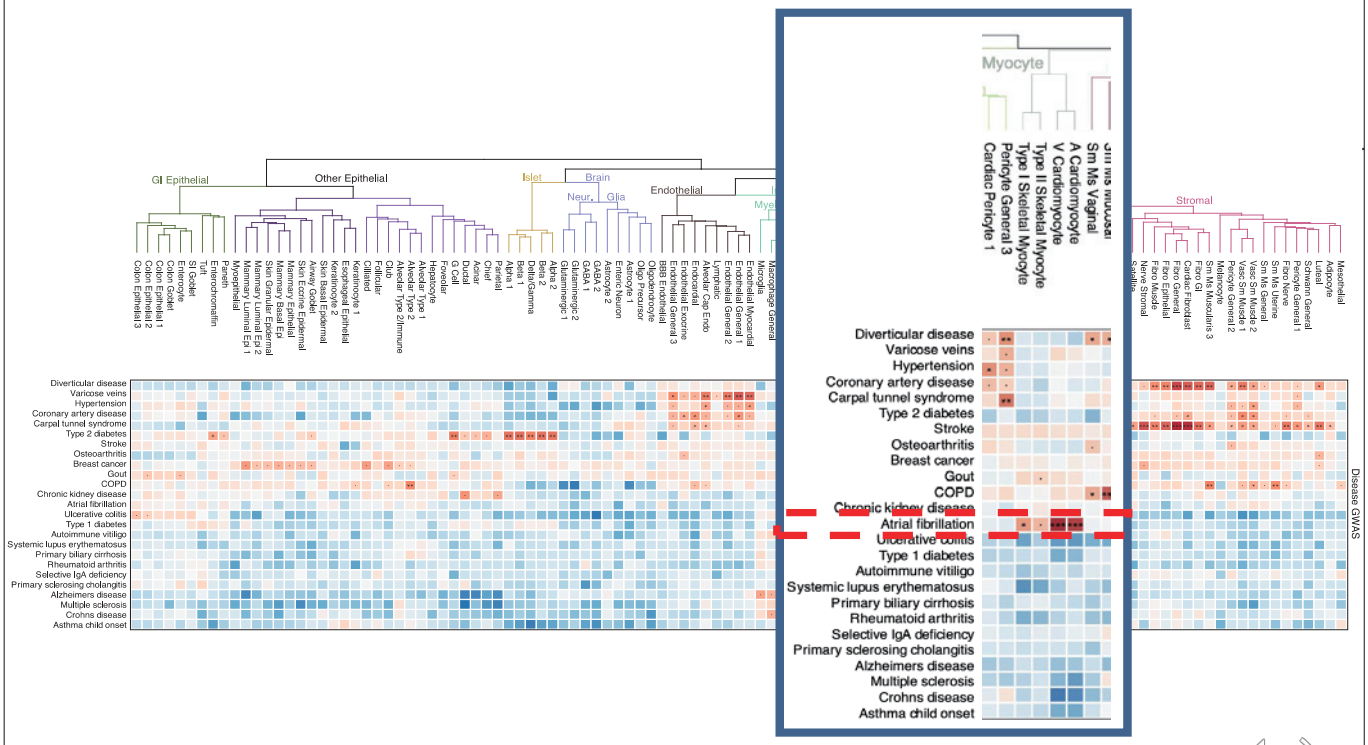
Enteroctye	44.32%
Fibroblast	12.24%
Goblet cells	11.55%
C1.1	8.21%
T cells	6.39%
Endothelial cells	4.63%
Macrophage	3.50%
B cells	2.96%
Paneth cells	1.51%
Enterochromaffin cells	0.71%
Smooth muscle	0.71%
Myofibroblast	0.70%
C9.6	0.64%
Schwann cells	0.32%
Tuft cells	0.19%
Mast cells	0.12%



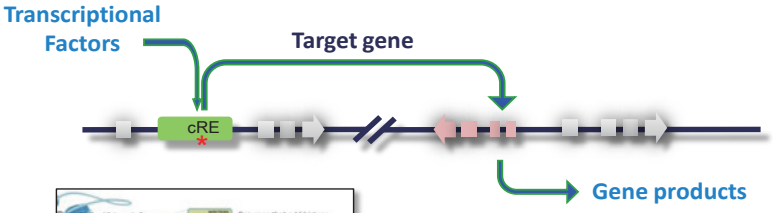


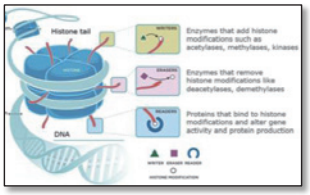
Enrichment of non-coding risk variants in candidate cis regulatory elements in different human cell types

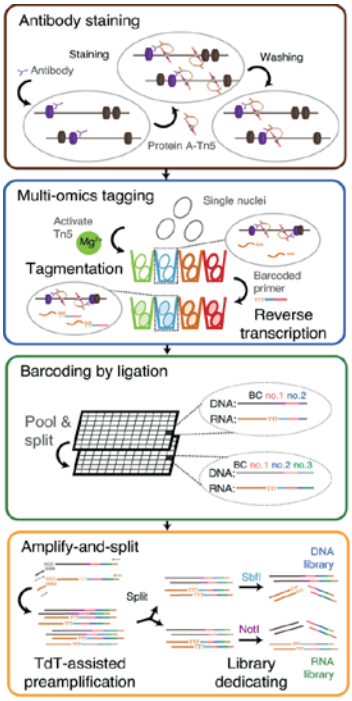




Paired-Tag: single cell co-assay of chromatin state or TF binding and transcriptome at large scale

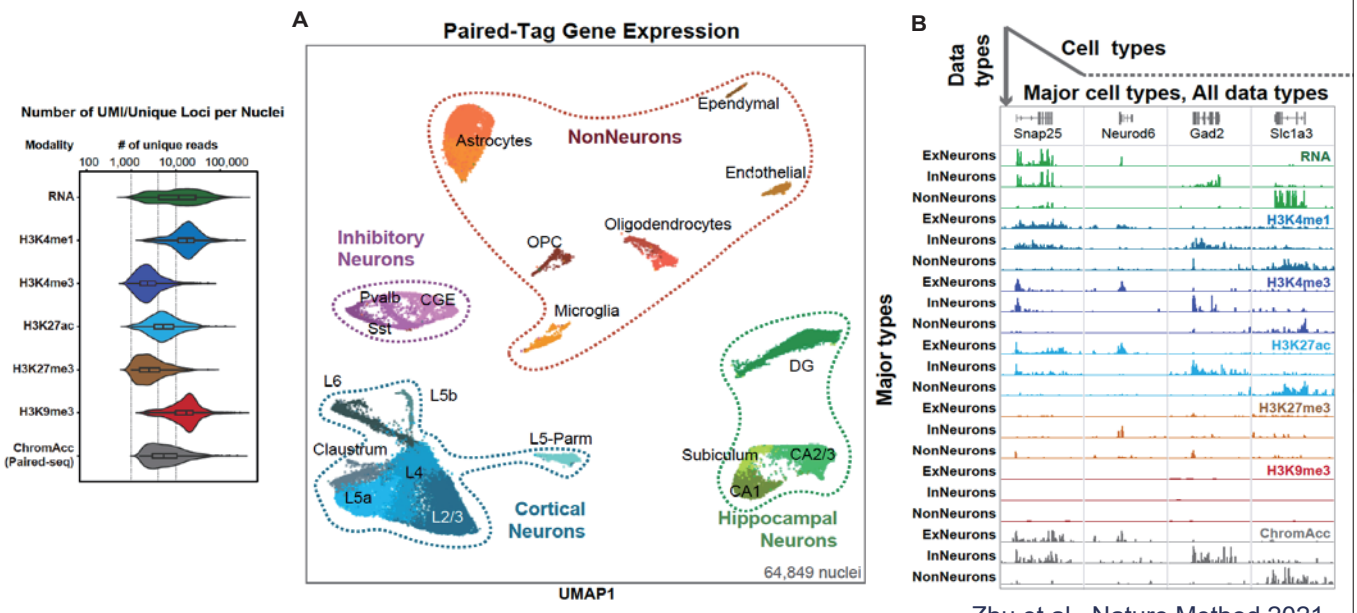






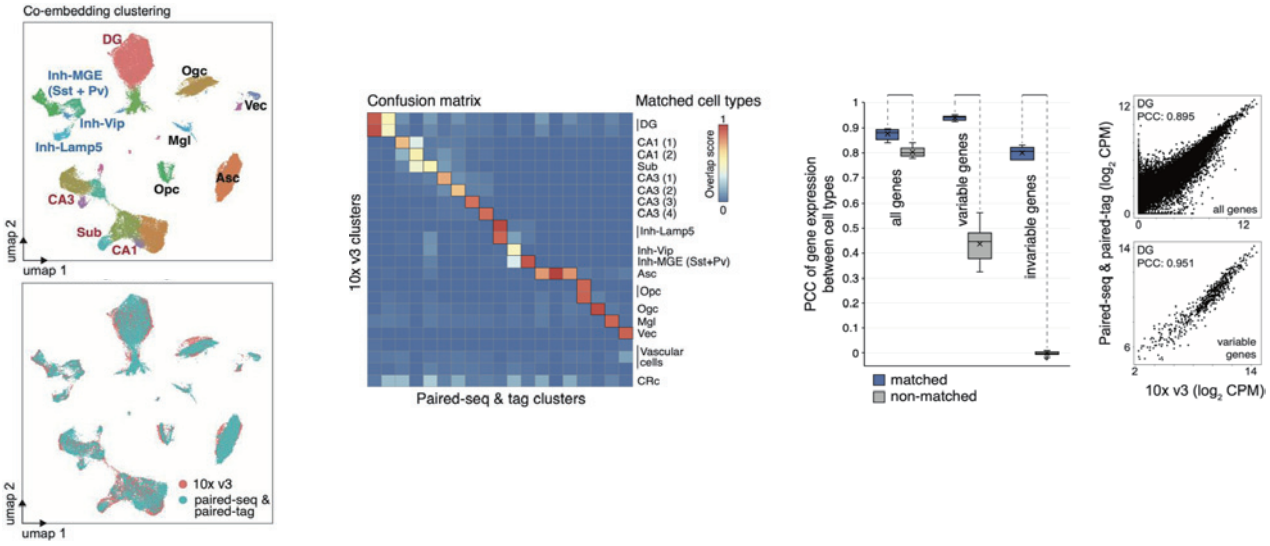
Zhu et al. Nature Method 2021

Paired-Tag analysis of adult mouse brain



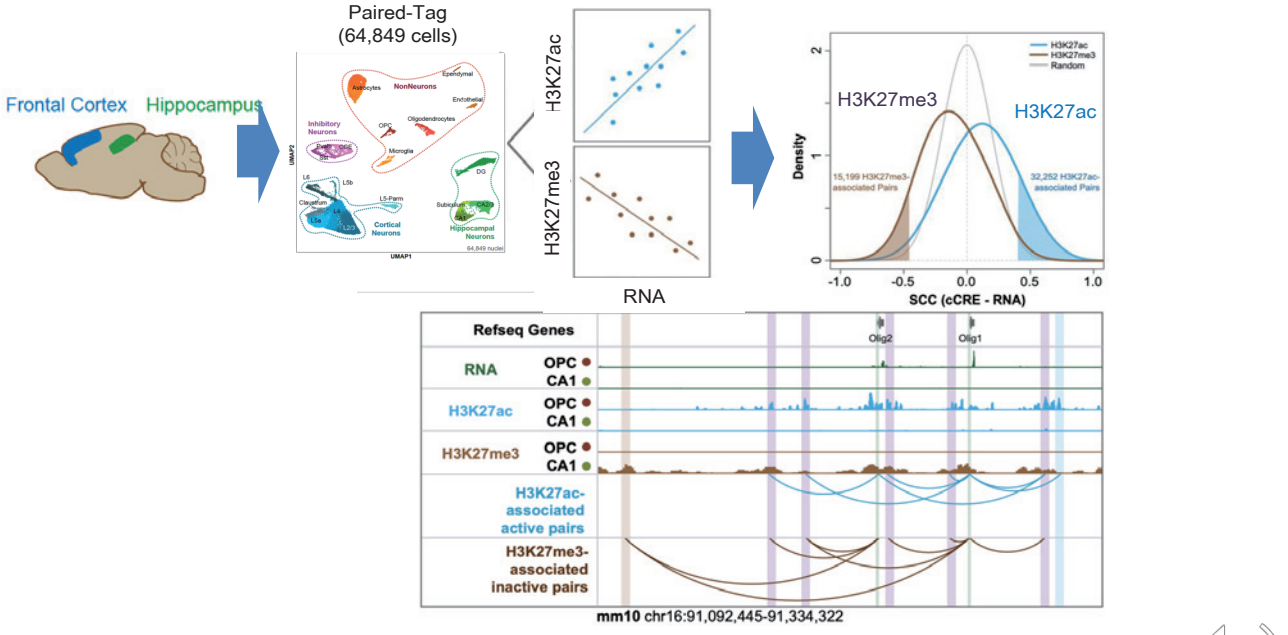
Zhu et al. Nature Method 2021

Comparison between Paired-Tag and 10x scRNA



Zhu et al. Nature Method 2021

Paired-Tag analysis links distal elements to putative target genes



How can we incorporate the 3D genome into single-cell multi-omics?

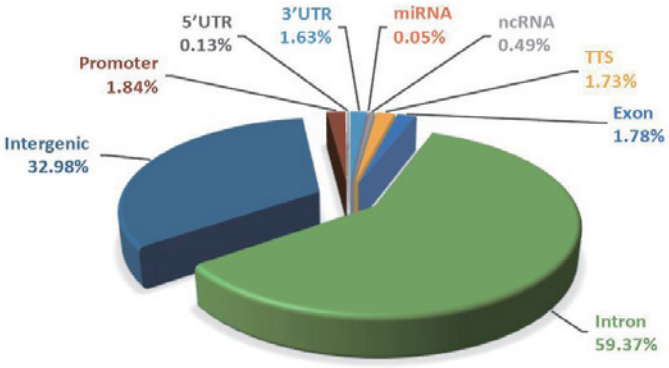
Mechanisms of noncoding located PD GWAS-SNPs

Various symptoms

Dementia
Hallucination
Delusion
Depression
Anxiety
Impulse control disorder
Sleep disturbance
Urinary frequency
Constipation
Anosmia
Orthostatic hypotension
Other non-motor symptoms

Complex molecular mechanisms

① α-synuclein proteostasis
② mitochondrial function
③ oxidative stress
④ calcium homeostasis
⑤ axonal transport
⑥ neuron-inflammation

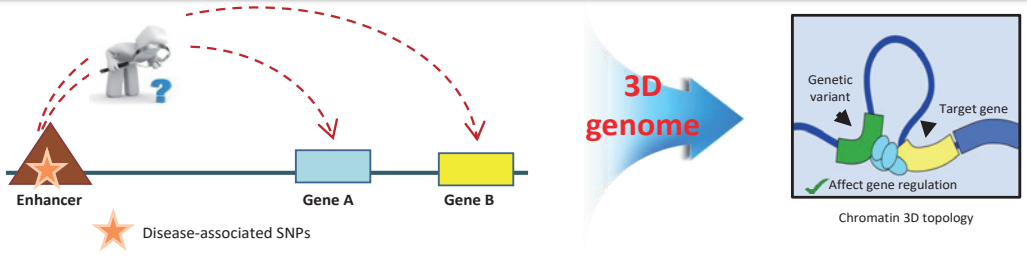


Number of sporadic PD GWAS SNPs: 3,869

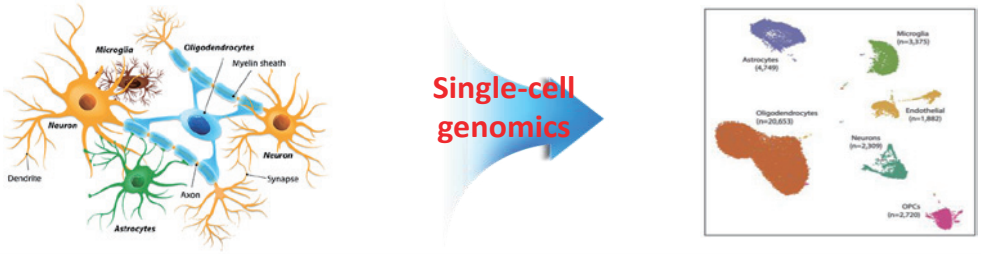
Foo et al., 2017; Nalls et al., 2019; Chang et al., 2017; Nalls et al., 2014

Single-cell 3D epigenomics in the human brain

Challenge 1. A lack of target gene information

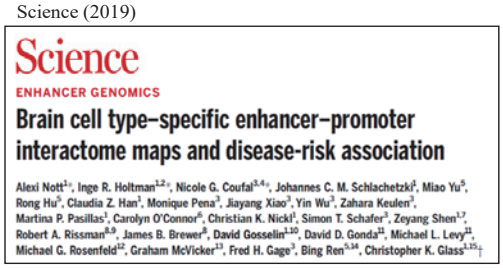


Challenge 2. Inability to identify disease relevent cell type

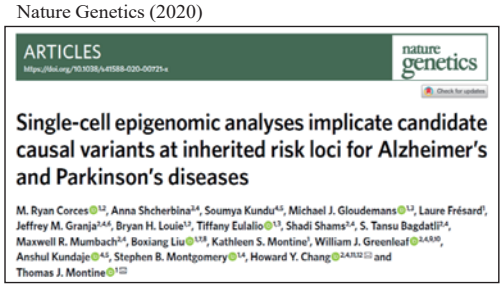


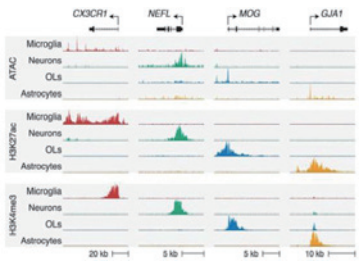
Efforts to dissect mechanisms of noncoding

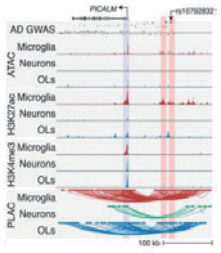
Science (2019)

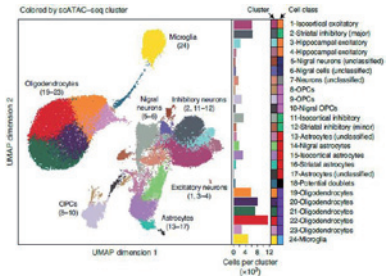

Science
ENHANCER GENOMICS
Brain cell type–specific enhancer–promoter interactome maps and disease-risk association
Alexi Nott[1†], Inge R. Holtman[1,2†], Nicole G. Coufal[3,4†], Johannes C. M. Schlachetzki[3], Miao Yu[5], Rong Hu[5], Claudia Z. Han[1], Monique Pena[3], Jiayang Xiao[3], Yin Wu[3], Zahara Keulen[3], Martina P. Pasillas[1], Carolyn O'Connor[6], Christian K. Nickl[3], Simon T. Schafer[3], Zeyang Shen[1,7], Robert A. Rissman[8,9], James B. Brewer[8], David Gosselin[1,10], David D. Gonda[11], Michael L. Levy[11], Michael G. Rosenfeld[12], Graham McVicker[13], Fred H. Gage[3], Bing Ren[5,14], Christopher K. Glass[1,15†]

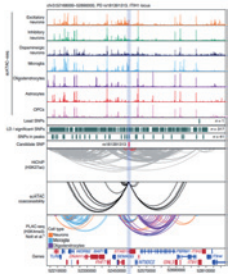
Nature Genetics (2020)


ARTICLES
nature genetics
Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases
M. Ryan Corces[1,2], Anna Shcherbina[3,4], Soumya Kundu[4,5], Michael J. Gloudemans[1,2], Laure Frésard[1], Jeffrey M. Granja[3,4,6], Bryan H. Louie[1,2], Tiffany Eulalio[1,2], Shadi Shams[3,4], S. Tansu Bagdatli[3,4], Maxwell R. Mumbach[3,4], Boxiang Liu[1,7,8], Kathleen S. Montine[1], William J. Greenleaf[3,4,9,10], Anshul Kundaje[4,5], Stephen B. Montgomery[1,4], Howard Y. Chang[2,4,11,12] ✉ and Thomas J. Montine[1] ✉



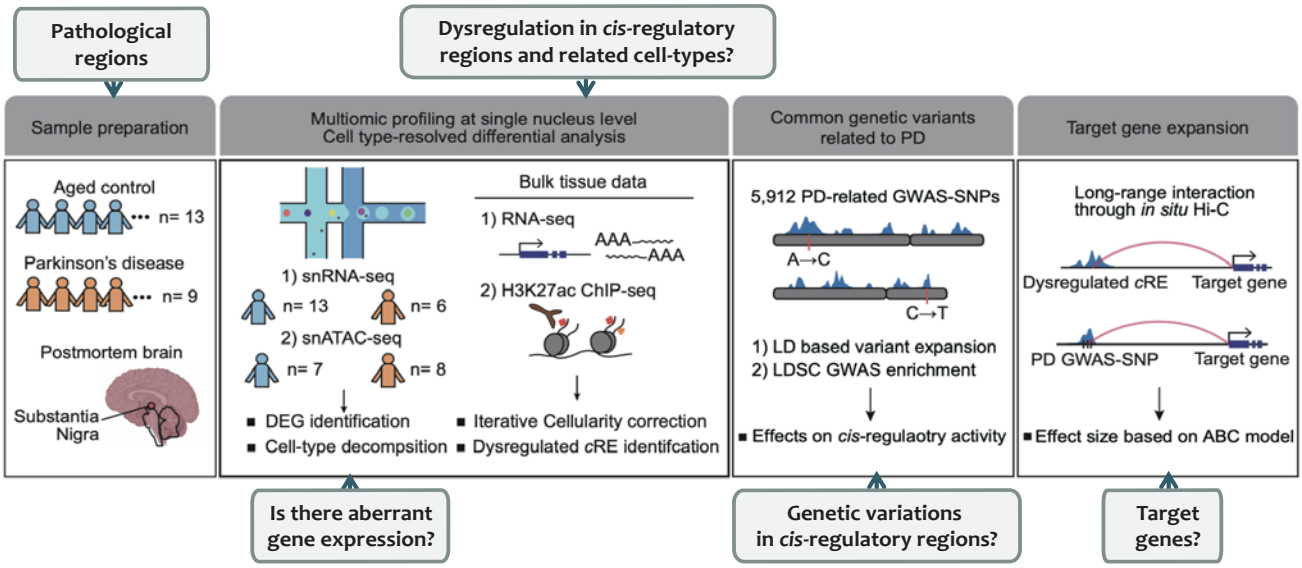


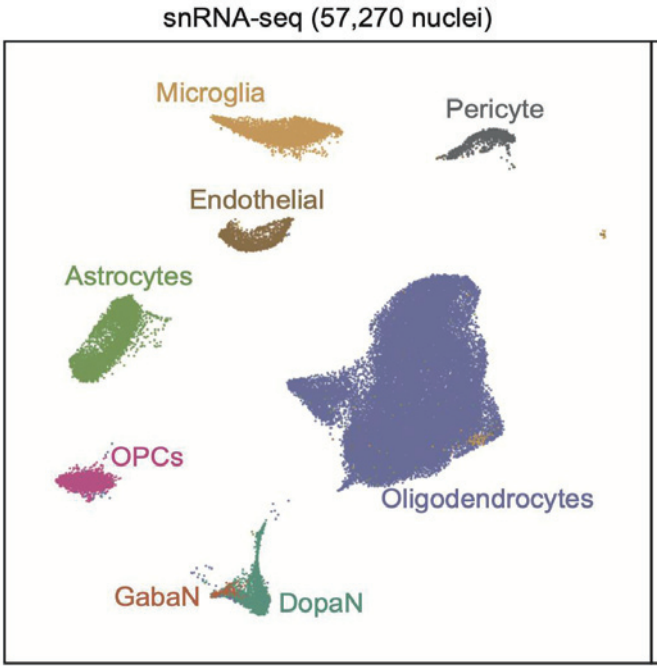


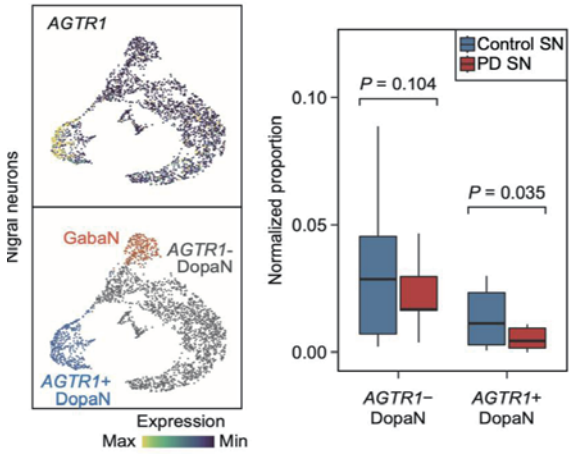


Limitations: a lack of diseased cases and a lack of disease-specific regulatory elements

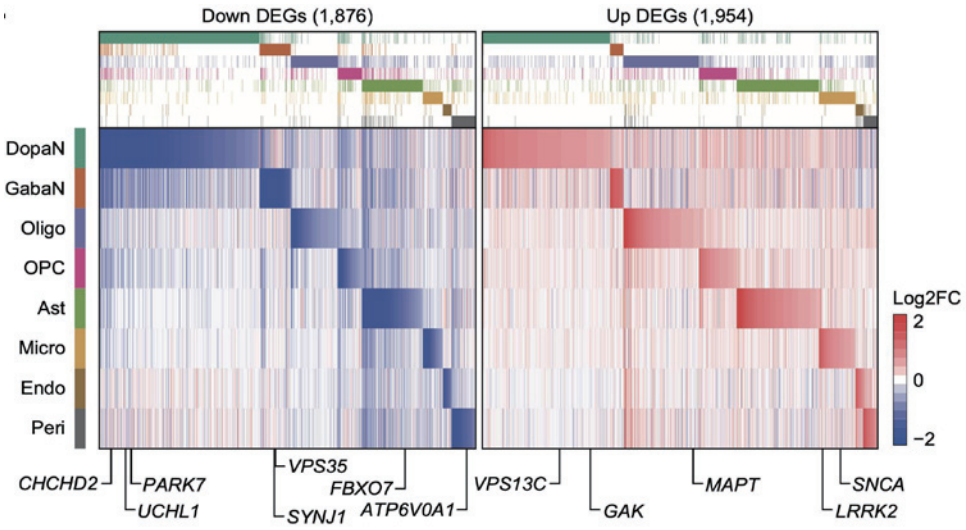
A single-cell 3D epigenomics







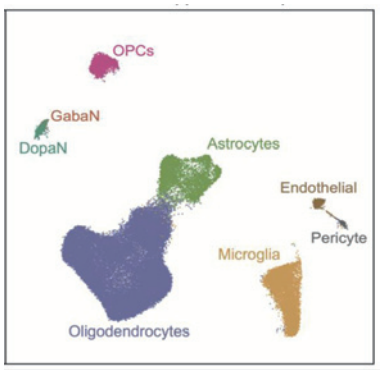
snRNA-seq reveals that aberrant PD genes are highly cell-type specific



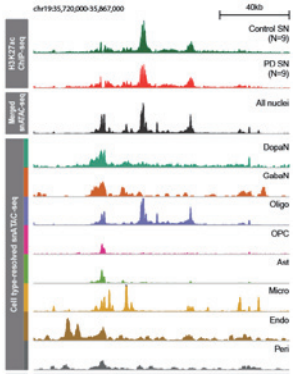
- Differentially expressed genes (DEGs) were identified in cell type specific manner.
- known PD genes were found in up-regulated DEGs, showing cell type-specificity.
- *SNCA* (Micro), *MAPT* (Oligo), *UCHL1* (DopaN), *LRRK2* (Micro), and *VPS13C* (DopaN)

How are aberrant PD genes regulated?
PD-specific enhancers are highly cell-type specific

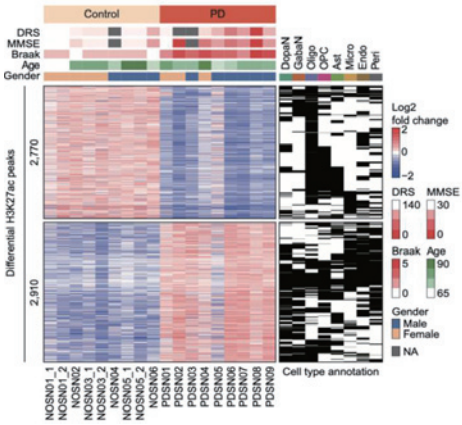
snATAC-seq (55,937 nuclei)



Cell type-specific epigenome

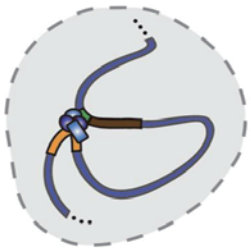


PD-specific enhancer dysregulation



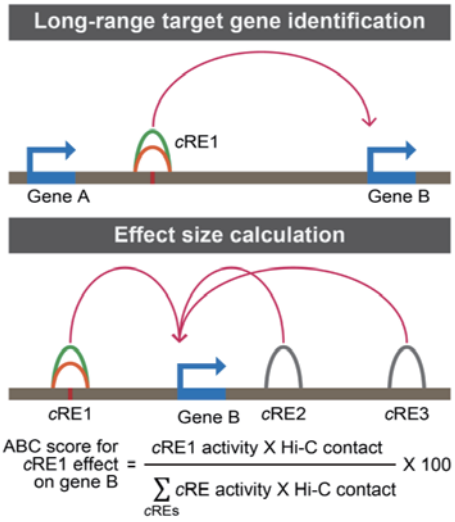
Target gene identification using 3D genome in PD brain

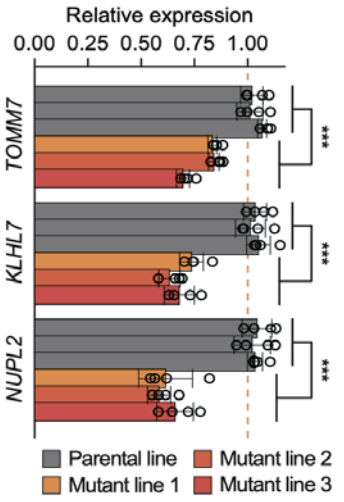
in situ Hi-C



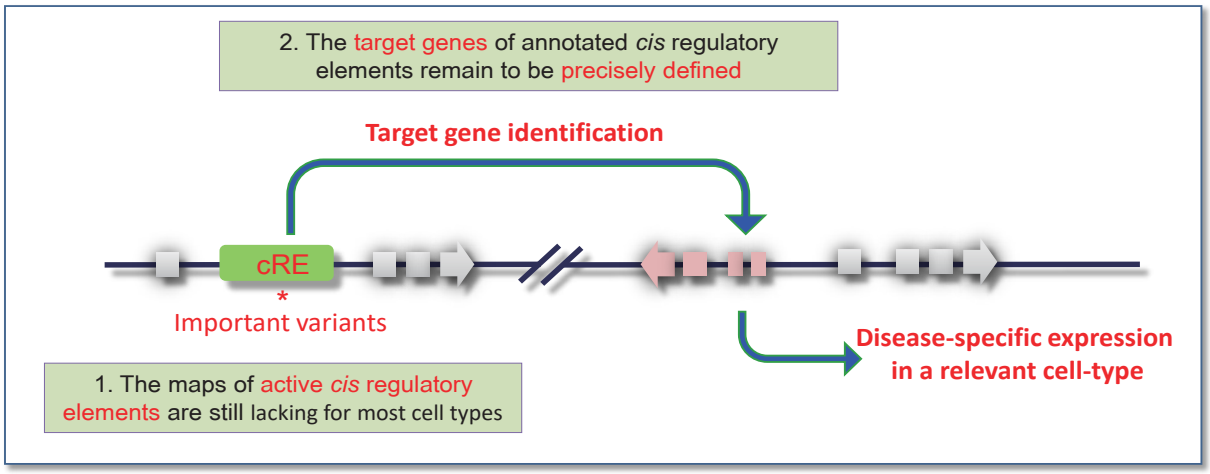
6 Control SN
(3 Male, 3 Female)
870,722,830 valid read-pairs
1,400,615 long-range interactions

5 PD SN
(2 Male, 3 Female)
744,285,811 valid read-pairs
1,031,124 long-range interactions



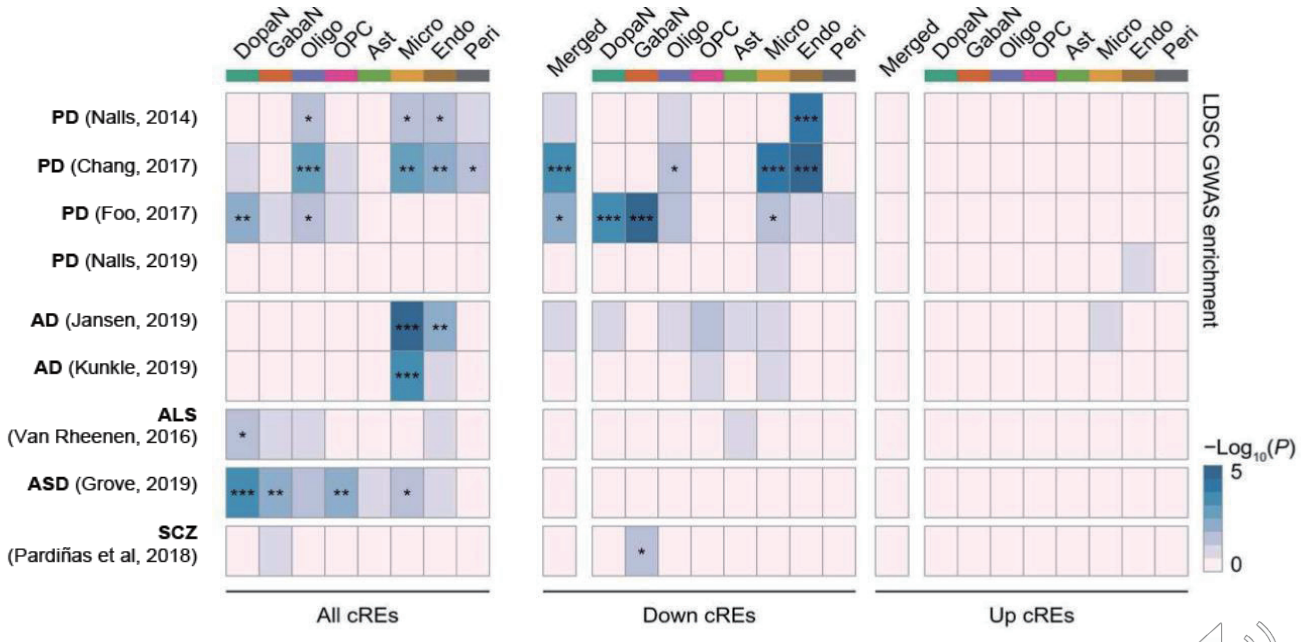


The first single-cell multi-omics and 3D genome map in PD brain

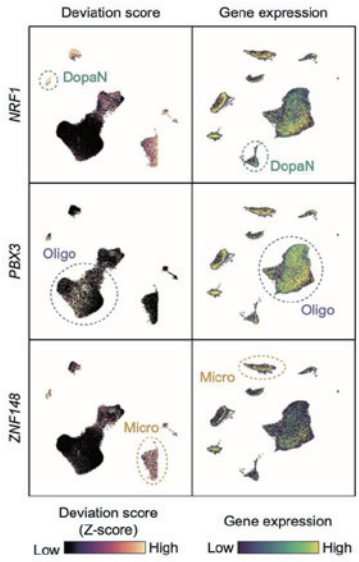


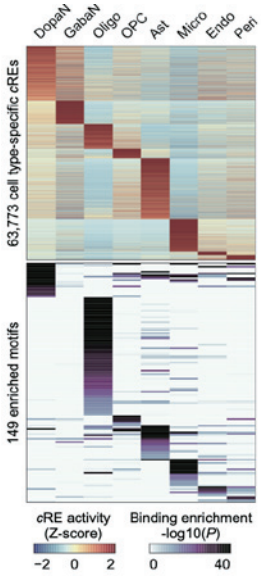
What are the **target genes** of PD GWAS variants and their **active cell type**?

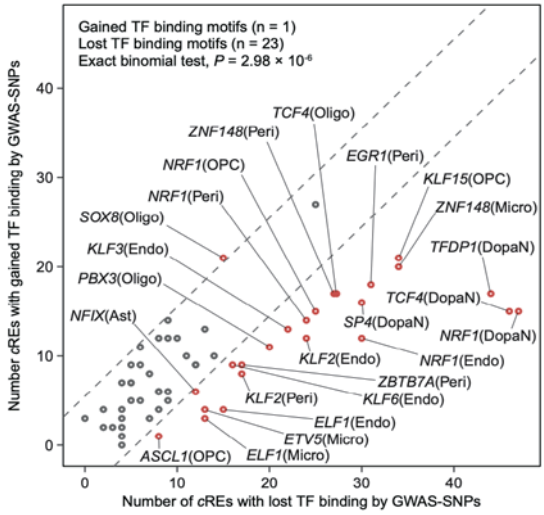
PD-GWAS SNPs are enriched at down-regulated enhancers



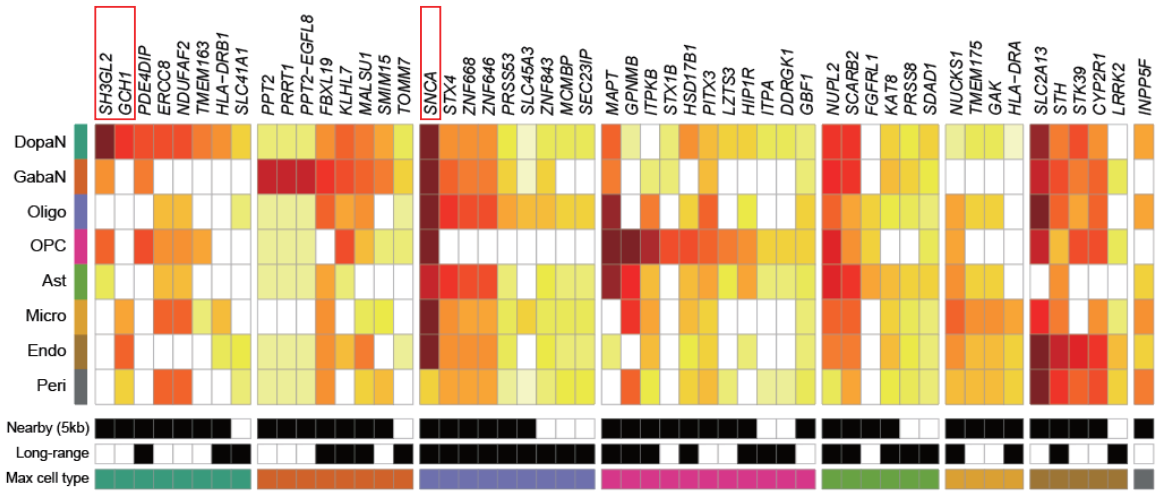
Why do GWAS-SNPs downregulate enhancer activity?

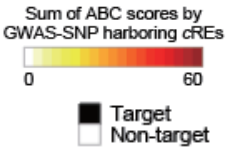






Identify active cell types for PD GWAS target genes

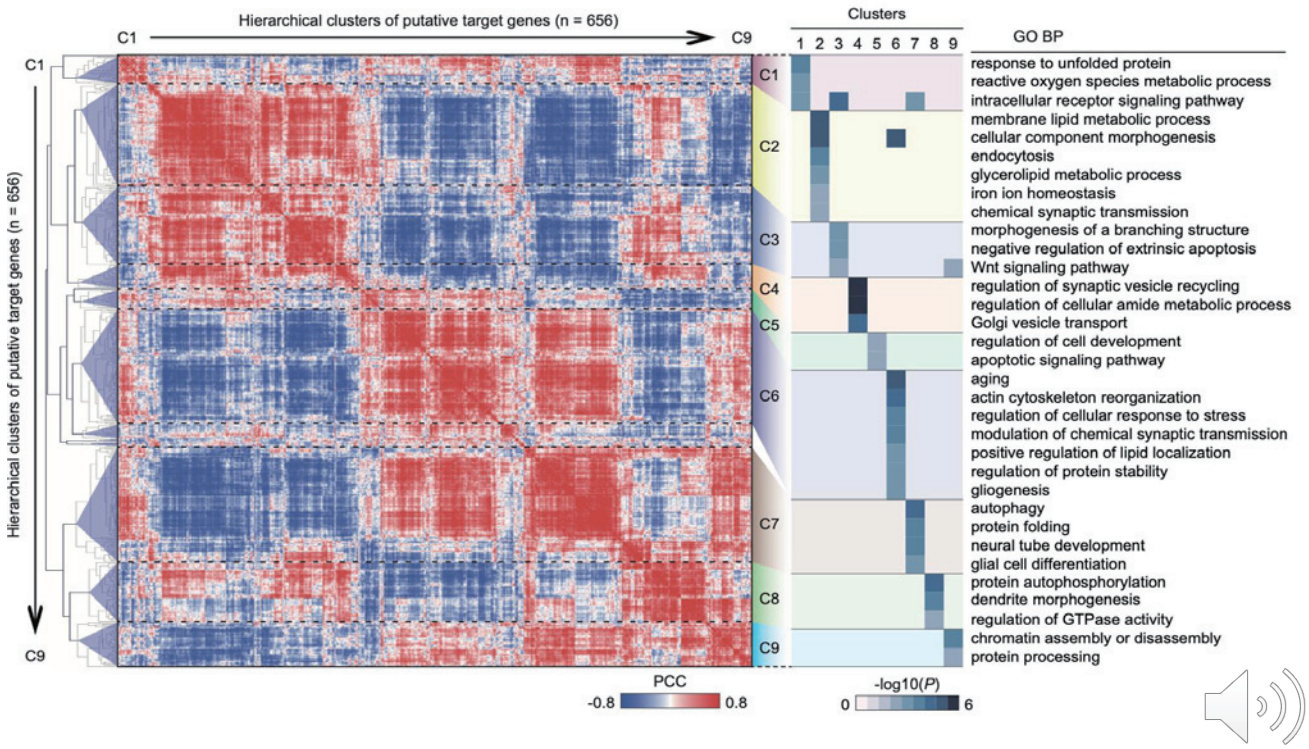




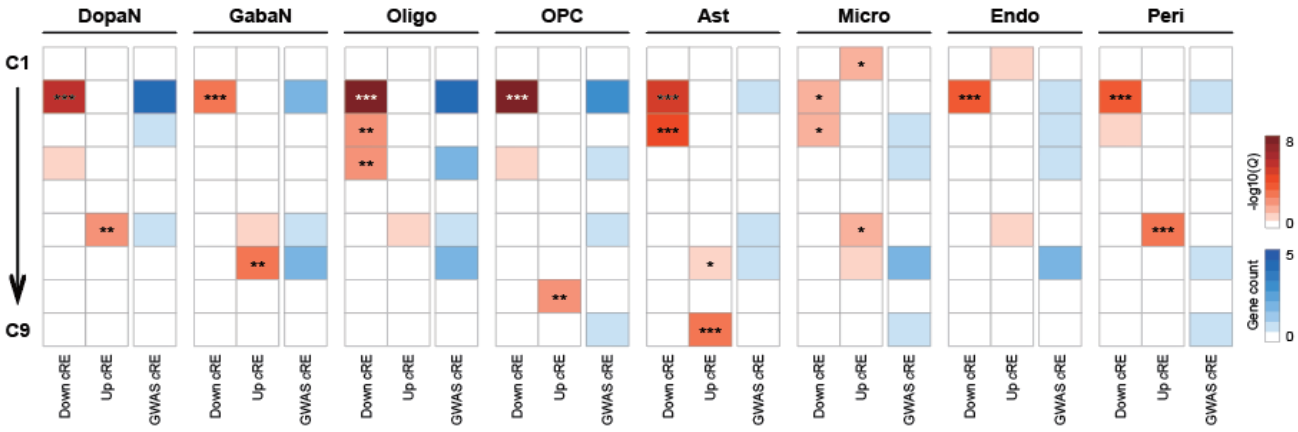
Target genes of PD GWAS-SNPs show active cell types for known PD genes.

1. *SH3HL2* / *GCH1*: highly specific to dopaminergic neurons
2. *SNCA*: significant in most cell types
3. *MAPT*: affected dominantly in oligodendrocytes, OPCs, and astrocytes
4. *GPNMB*: affected dominantly in OPCs, astrocytes, and microglia
5. *SCARB2*: most significant in neurons, and astrocytes

Can we modularize the functions of target genes?

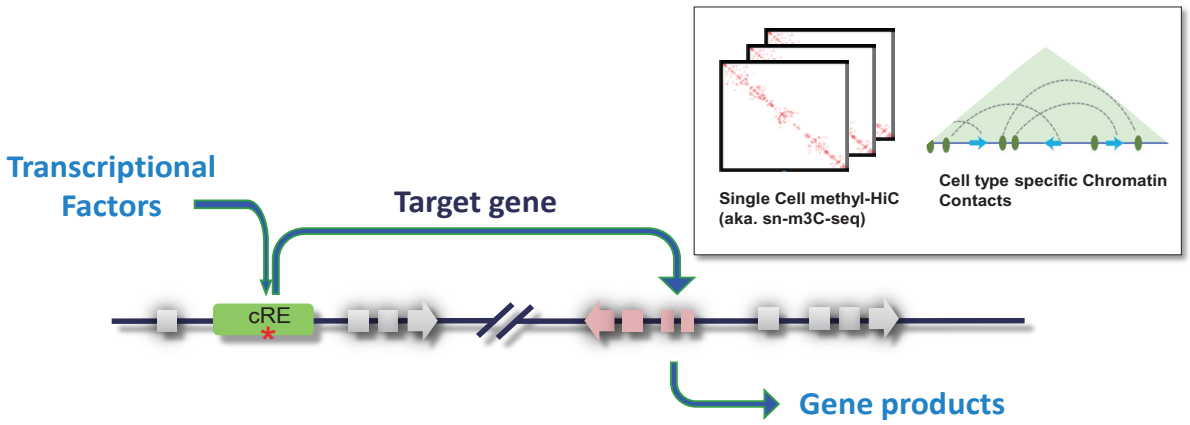


Cell-type resolved modular functions

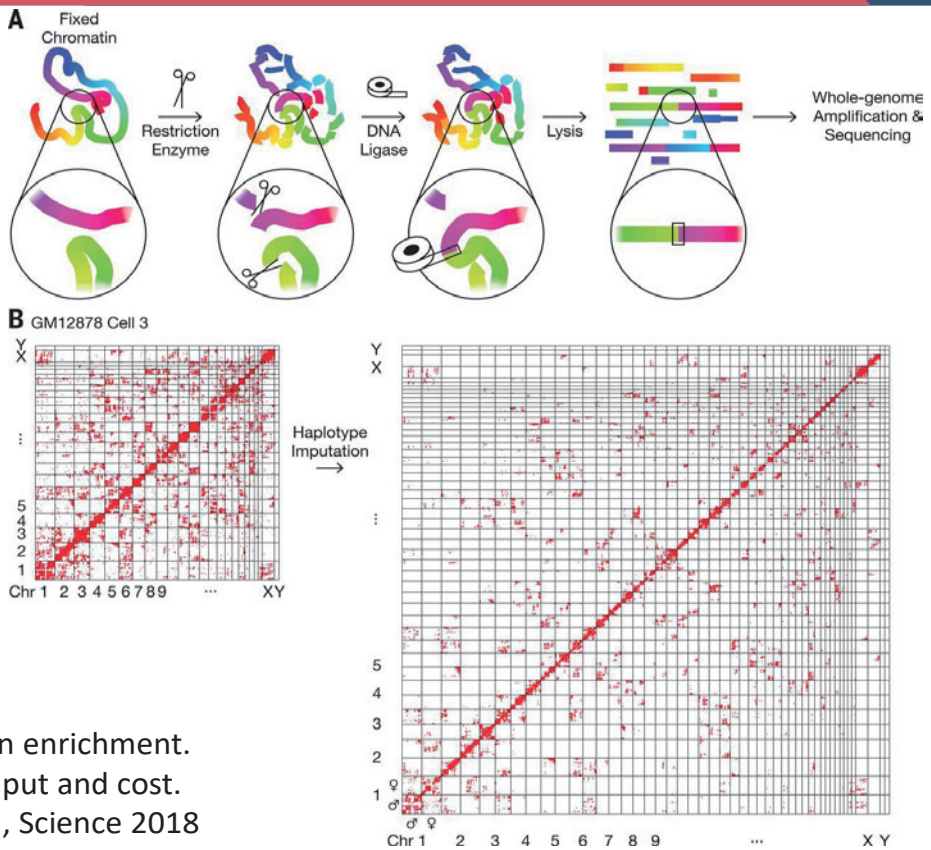


C1 (response to unfolded proteins, ROS metabolism) : Microglia
C2 (endocytosis, lipid metabolism, synaptic function) : All cell types
C3 (Negative regulation of apoptosis) : Oligodendrocytes, and Astrocytes
C6 (Cytoskeleton organization, stress, protein stability) : Dopaminergic neurons, Microglia
C7 (Glial cell differentiation) : Astrocytes
C8 (Dendrite morphogenesis) : OPC

Profiling the Higher-order Chromatin Structure with Single Cell Multi-omics



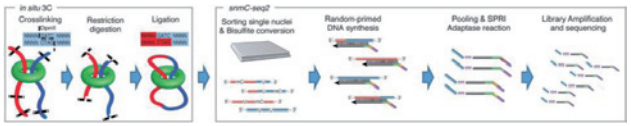
Single cell Hi-C (Dip-C, 3D genome)



WGA,
No biotin enrichment.
Throughput and cost.
Tan et al, Science 2018

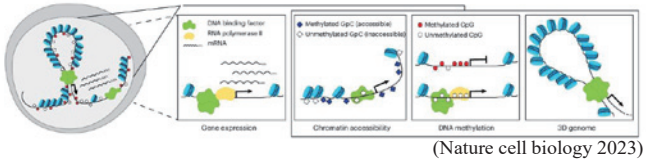
Single cell 3D multi-omics technologies

sn-m3C-seq (3C & DNA methylation)



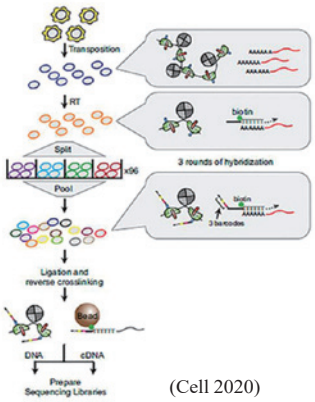
(Nature Methods 2019)

3DRAM-seq (HiC & RNA & accessibility & methylation)



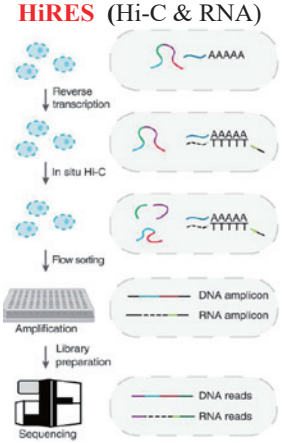
(Nature cell biology 2023)

SHARE-seq (RNA & ATAC)



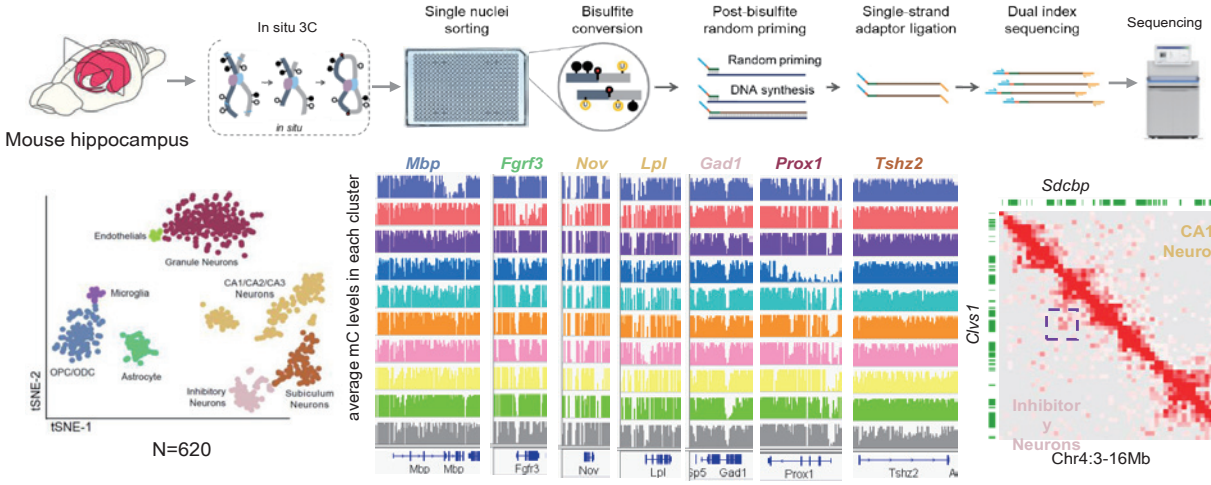
(Cell 2020)

HiRES (Hi-C & RNA)



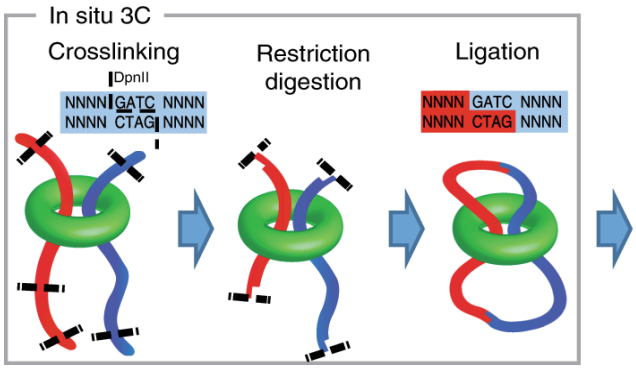
(Science 2023)

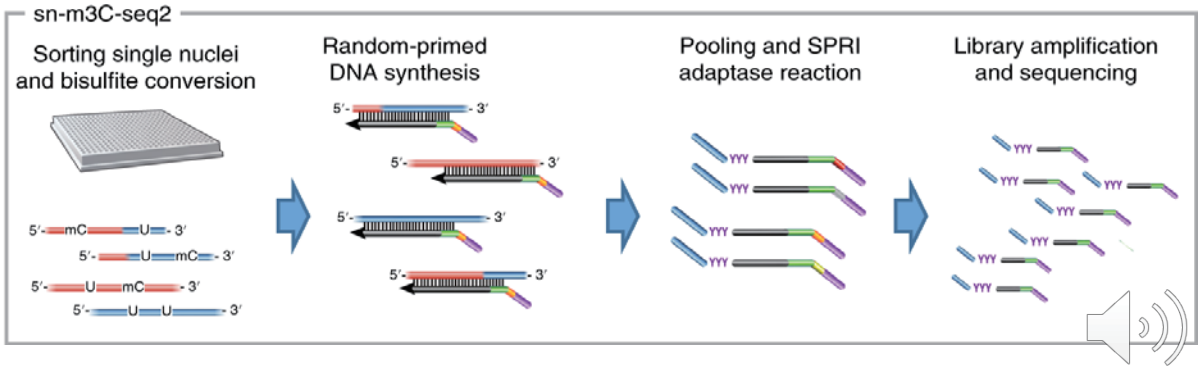
Joint Analysis of DNA Methylome and Chromatin Organization in Single Cells



Li et al., Nature Methods 2019

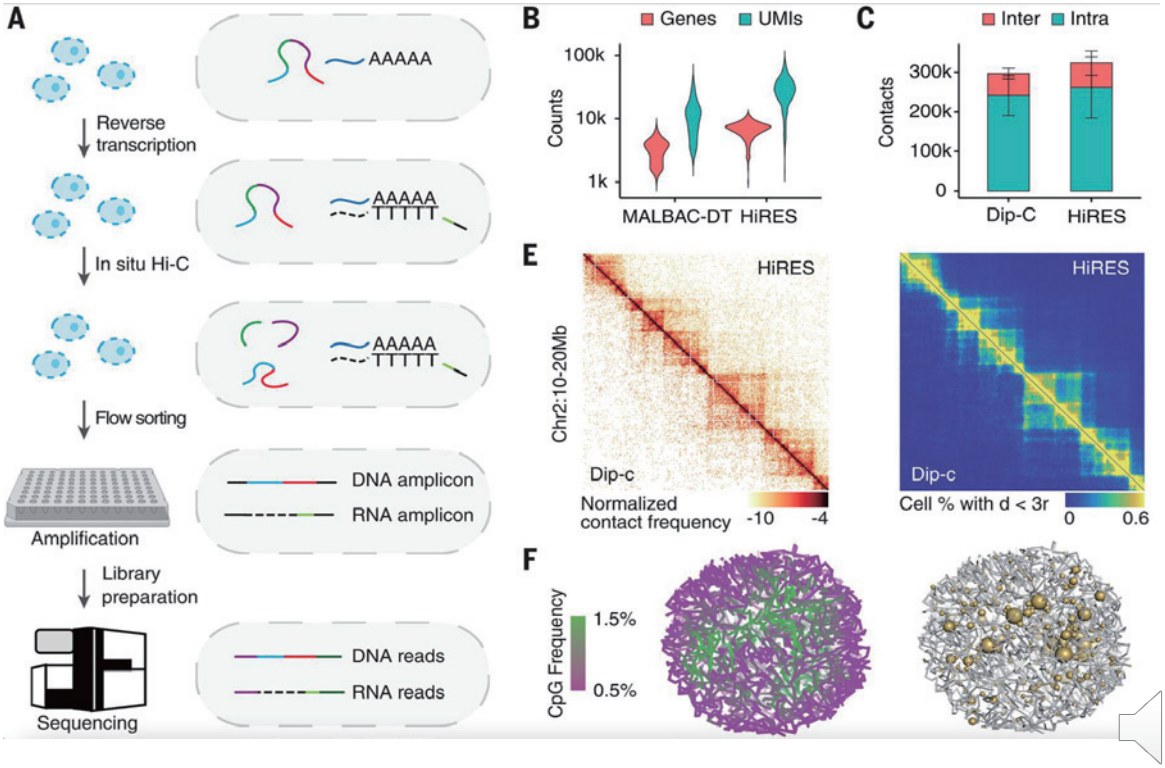
snm3C (DNA methylation and 3D genome)





Throughput, cost, Lee at al., Nature Methods, 2019

HiRES (RNA and 3D genome)

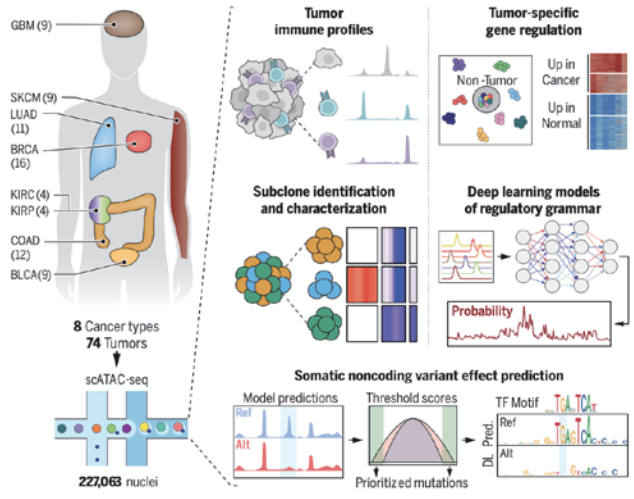


Liu et al., Science, 2023

A next challenge

How can we study oncogenic 3D gene regulation in single-cell resolution?

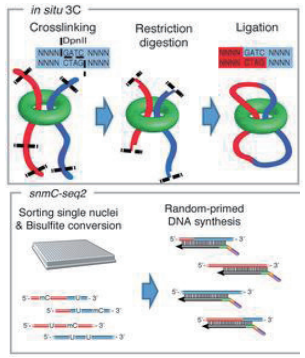
Single-nucleus pan-cancer epigenome ATLAS



Sundaram et al., Science 2024

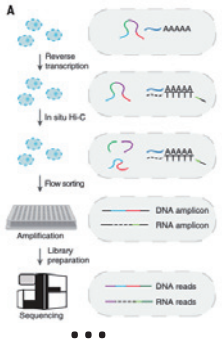
A lack of method profiling transcriptome, epigenome, and 3D genome simultaneously at single-cell resolution

sn-m3C-seq (Nat Methods 2019)



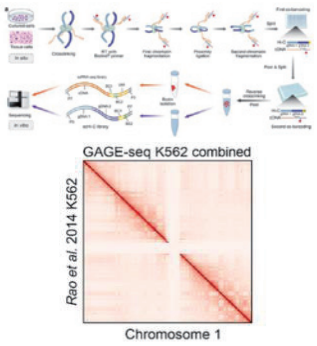
3D genome & DNA methylation

HiRES (Science 2023)



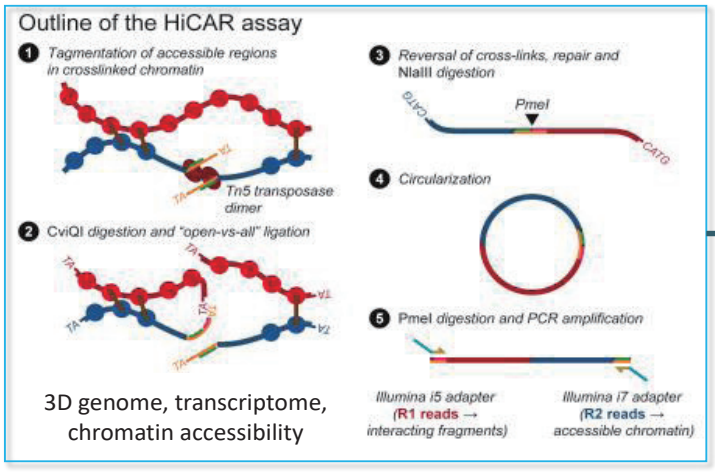
3D genome & Transcriptome

GAGE-seq (Nat Genet 2024)



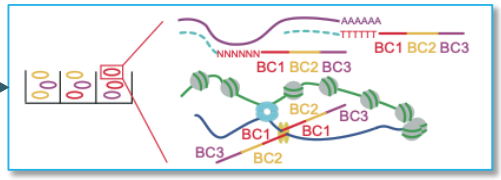
3D genome & Transcriptome

scHiCAR: Transcriptome, epigenome, and 3D genome in single-cell resolution

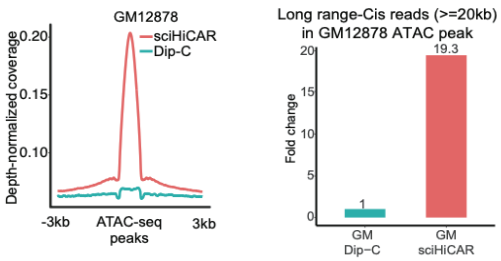


3D genome, transcriptome, chromatin accessibility

HiCAR (Hi-C on accessible regulatory DNA, Mol Cell, 2022)

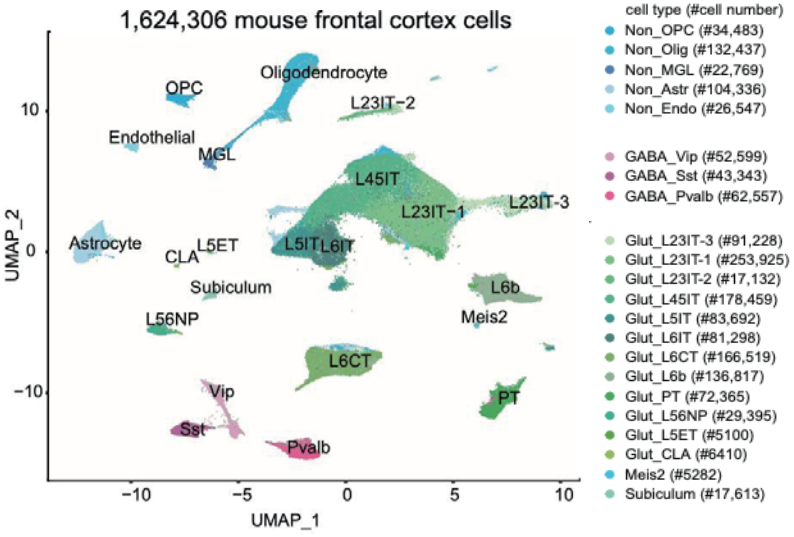


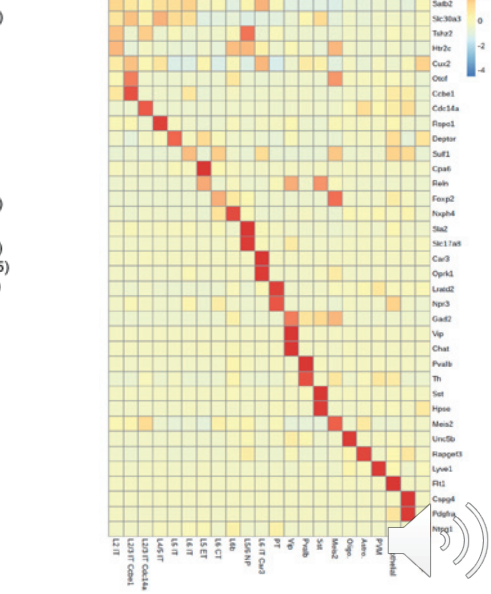
Single-cell HiCAR using split and pool



scHiCAR effectively captures
1) regulatory interactions 2) long-range interactions

scHiCAR is ultra high-throughput and low-cost profiling 1.6M cells in mouse brain

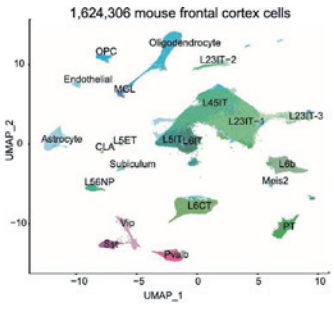




Gene expression of cell type markers

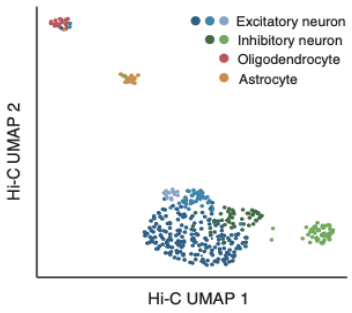
scHiCAR is ultra high-throughput and low-cost

Mouse Brain scHiCAR (1.6M cells)
scHiC +scRNA + epigenome



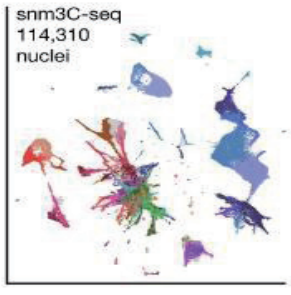
Unpublished

Mouse Brain HiRES (399 cells)
scHiC +scRNA, no epigenome



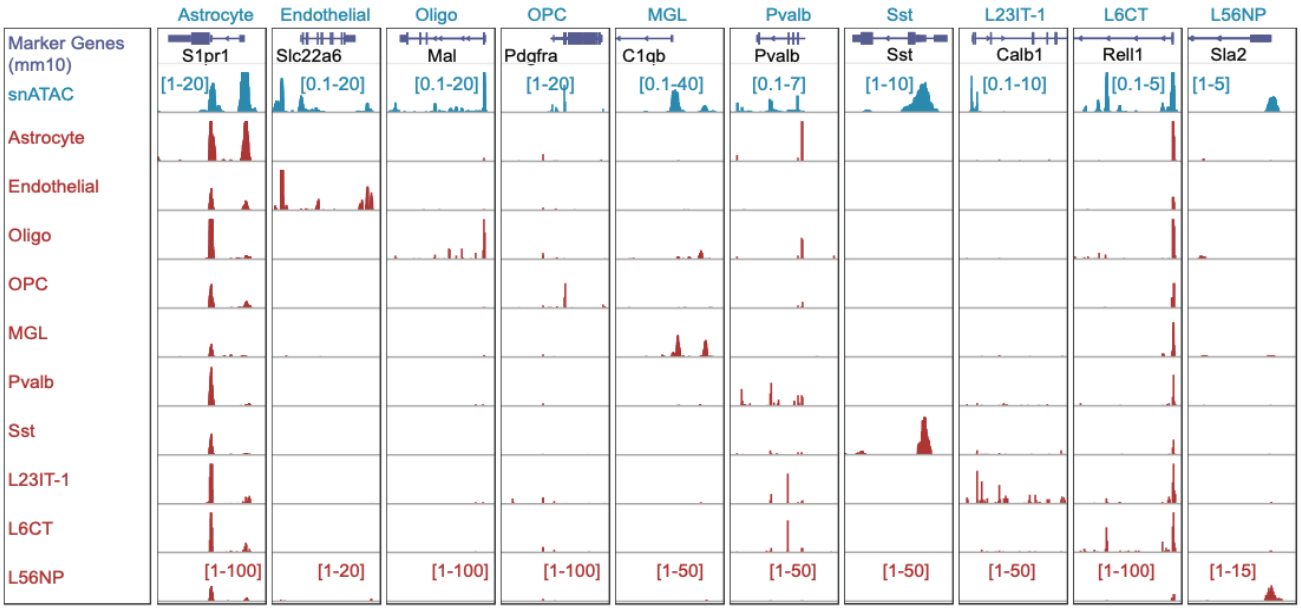
Science 2023

Mouse Brain snm3C (114k cells)
scHiC +epigenome, no RNA

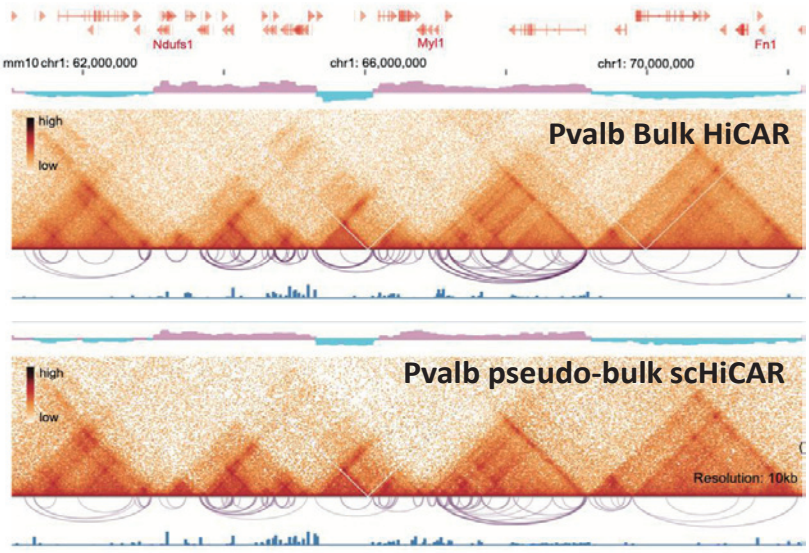


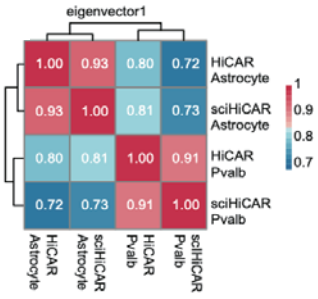
Nature 2023

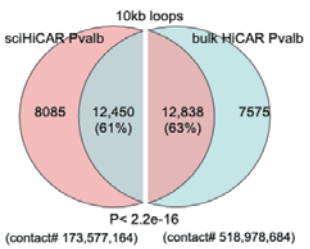
scHiCAR well recapitulates cell-type specific CREs



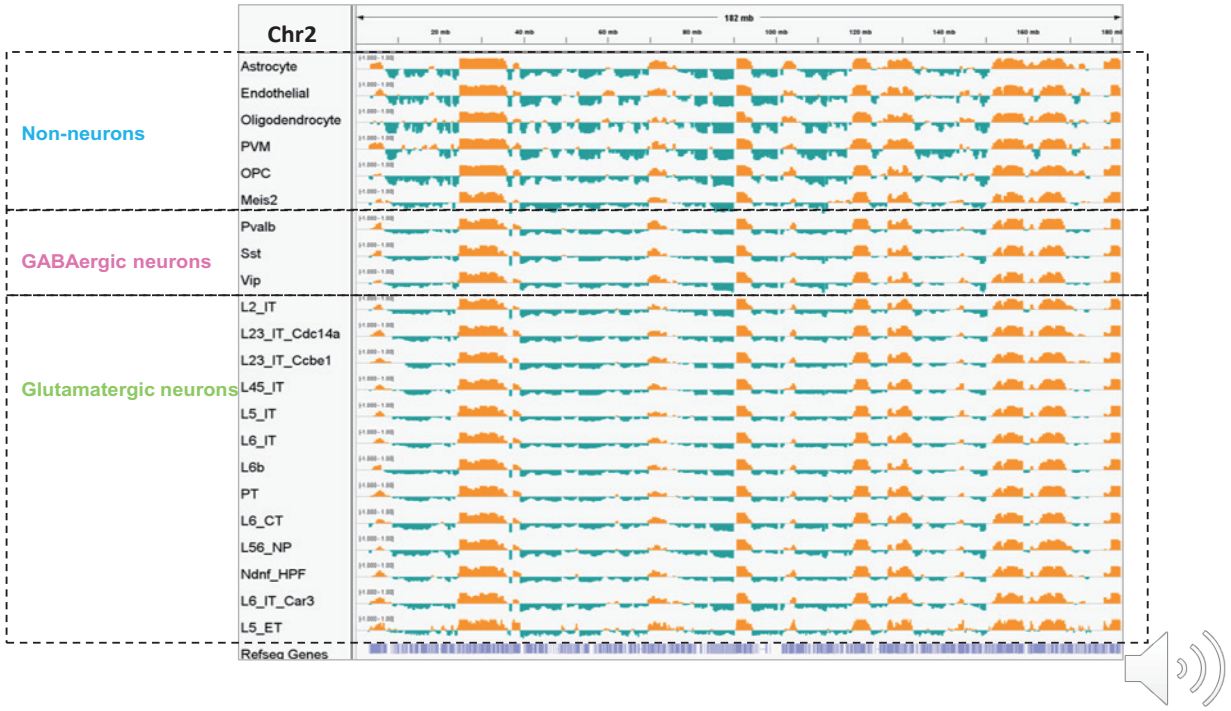
scHiCAR well recapitulates 3D genome in each cell-type



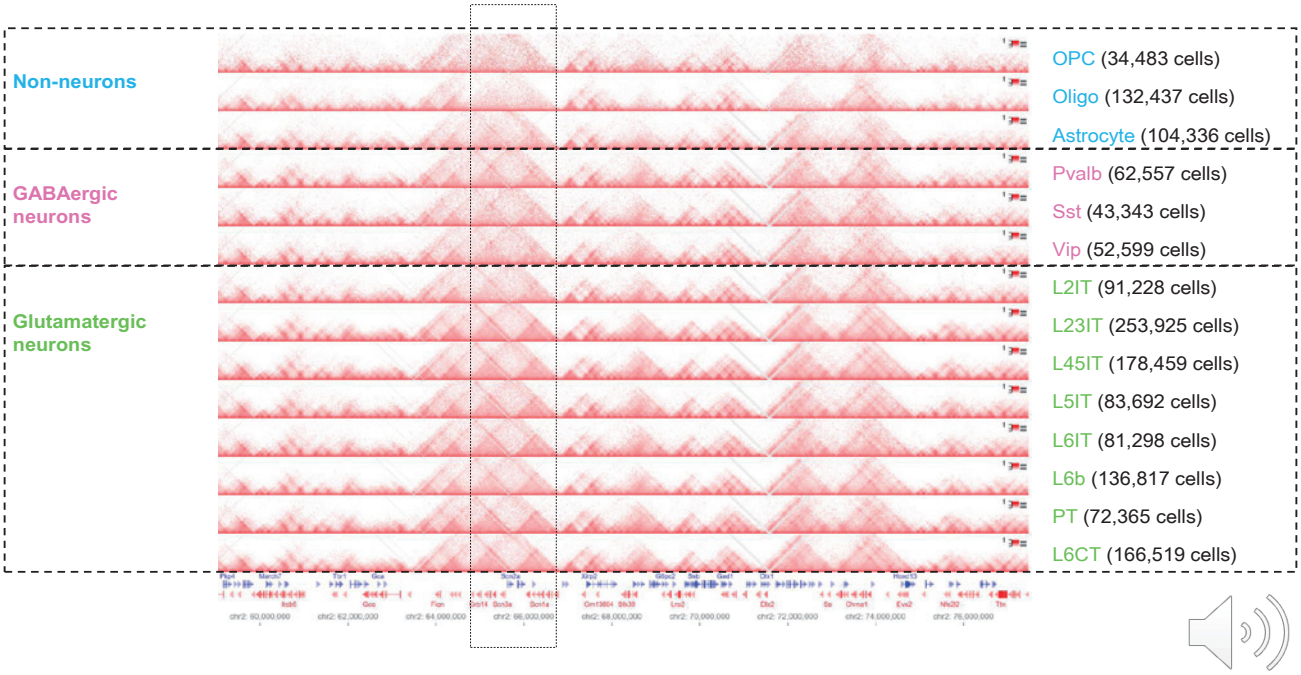




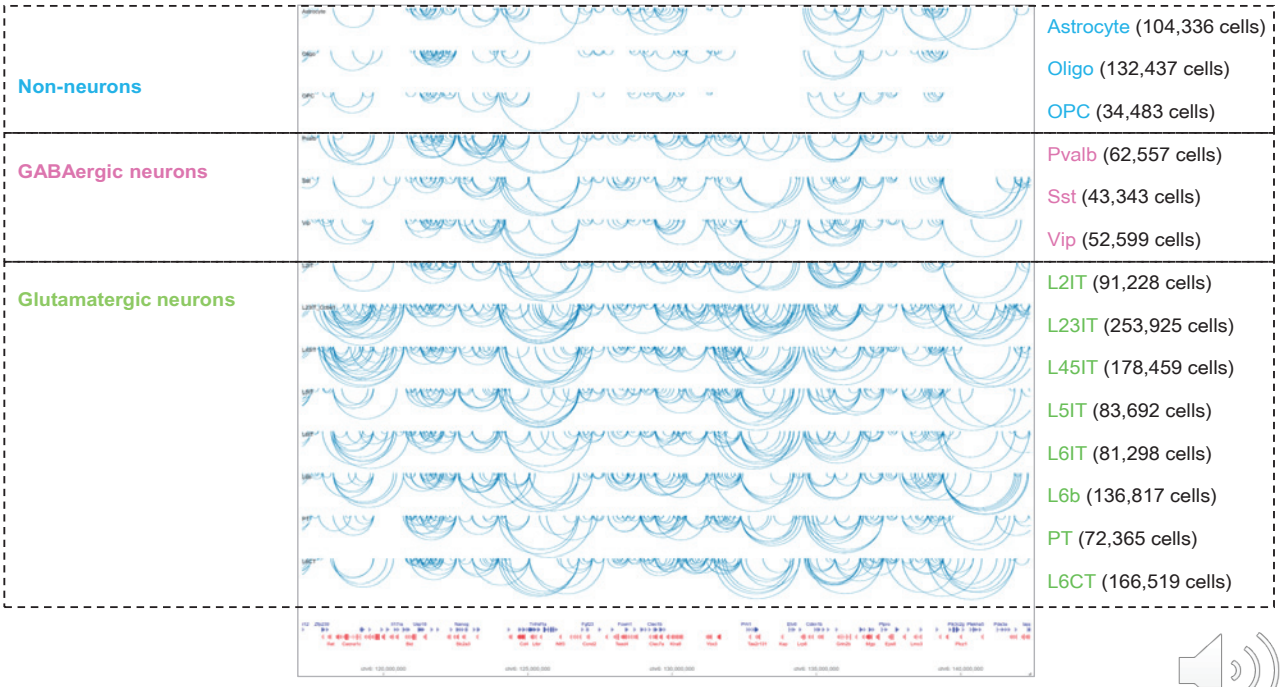
Dynamic compartment A/B patterns across cell-types



Cell-type independent TAD boundary conservation

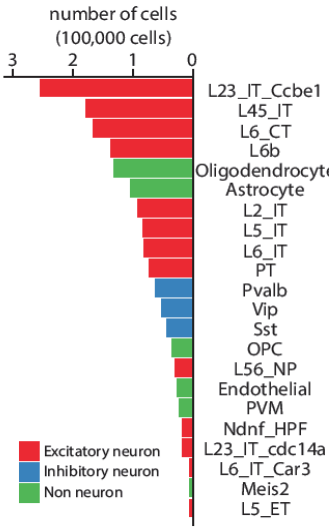


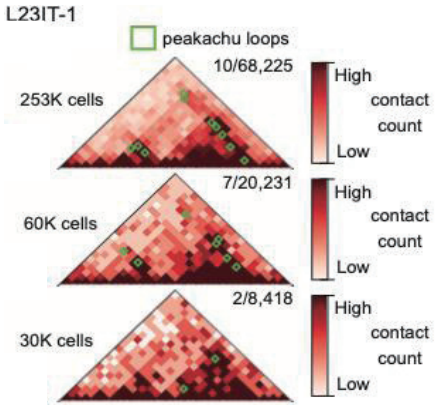
Dynamic E-P pairs across cell-types

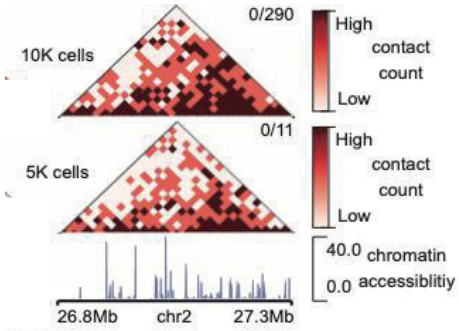


A challenge

Identifying E-P interactions depends on varying cell proportions



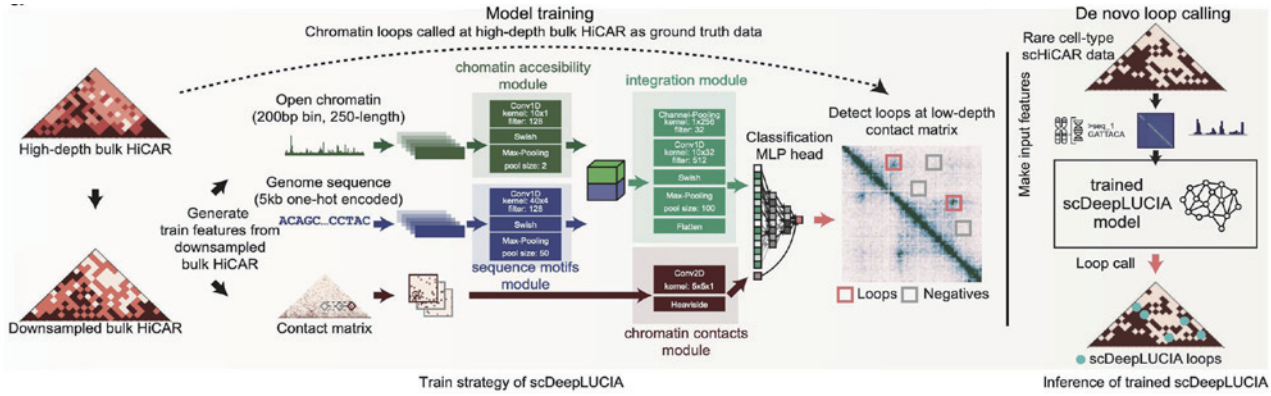




Conventional loop calling approach

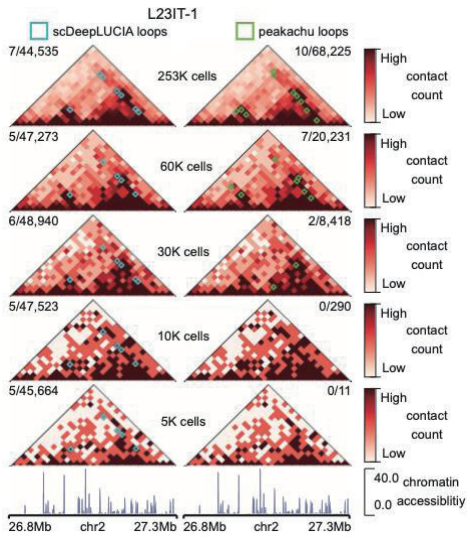
Model goal: reproduce loop callings from high-depth using low-depth scHiCAR

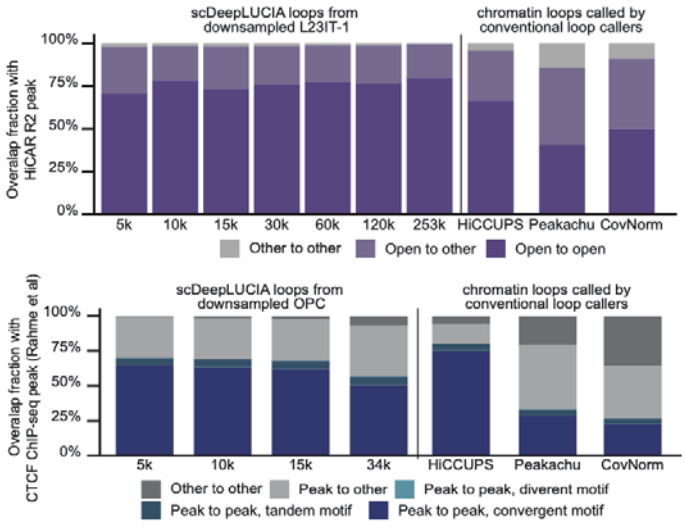
Robust identification of E-P interactions with scDeepLUCIA



scDeepLUICA (single-cell Deep Learning-based Universal Chromatin Interaction Annotator)

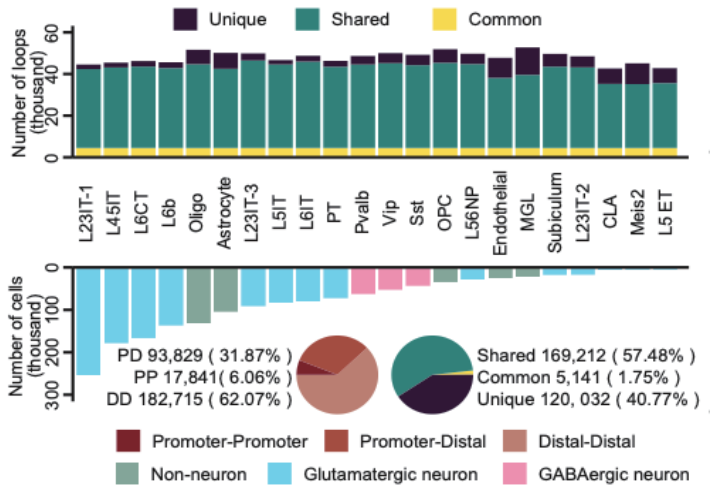
scDeepLUCIA is robust in defining CRE-interactions from at least 5,000 cells

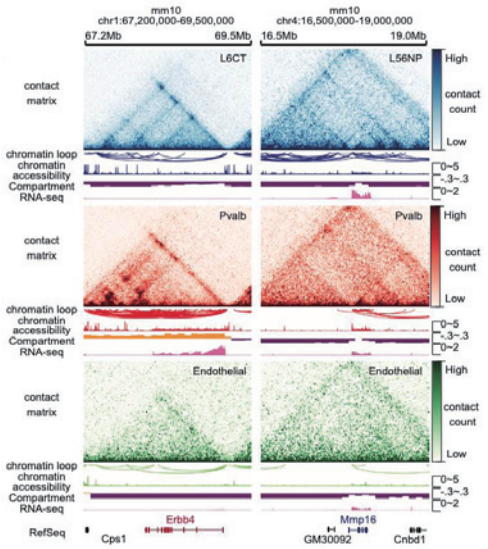




scDeepLUCIA can identify CRE-interactions

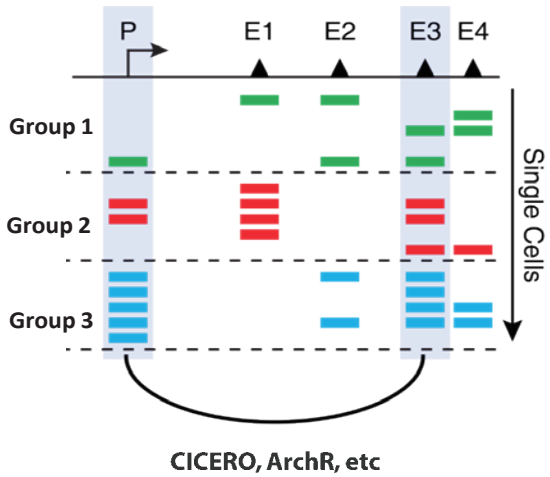
from major to rare brain cell types

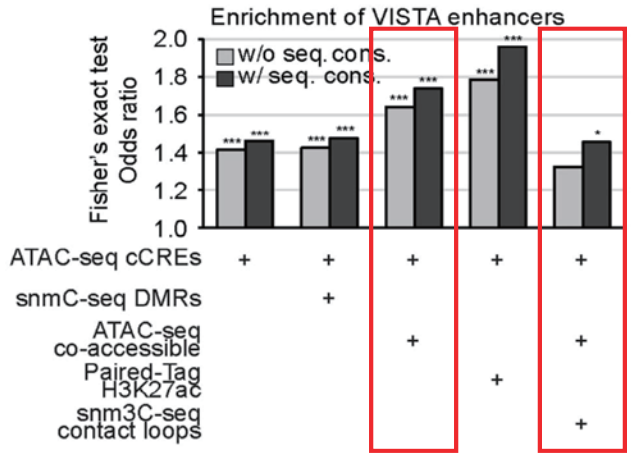




Defining functional CRE-interactions:

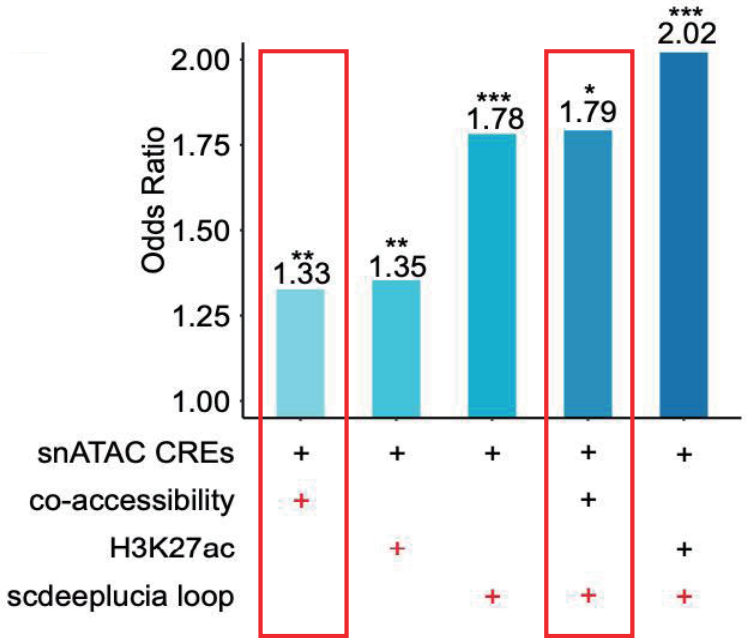
Co-accessibility vs physical interactions



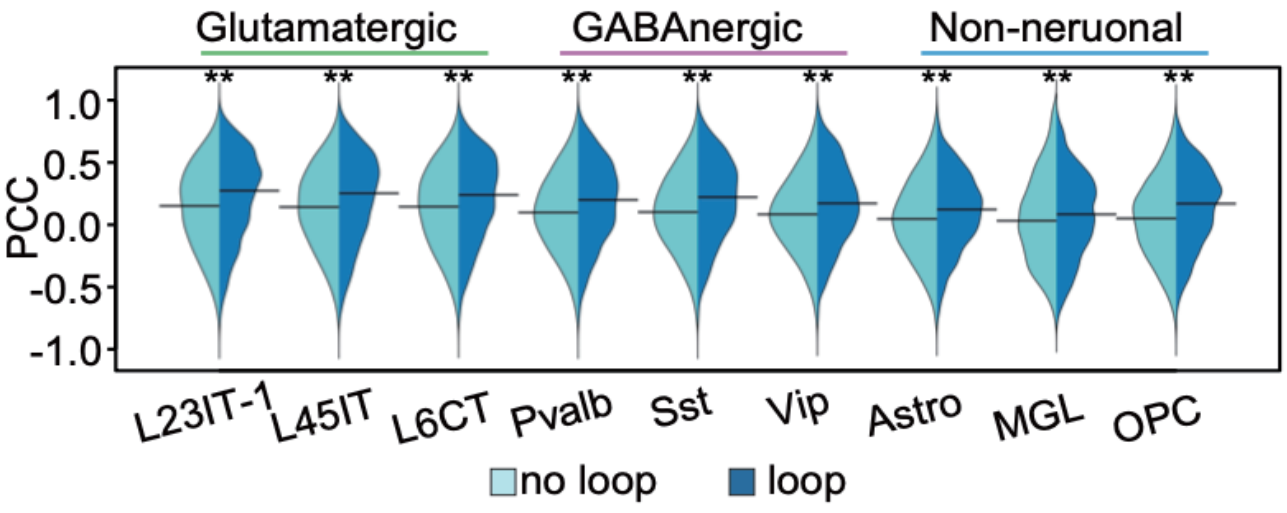


Li et al., Science 2023

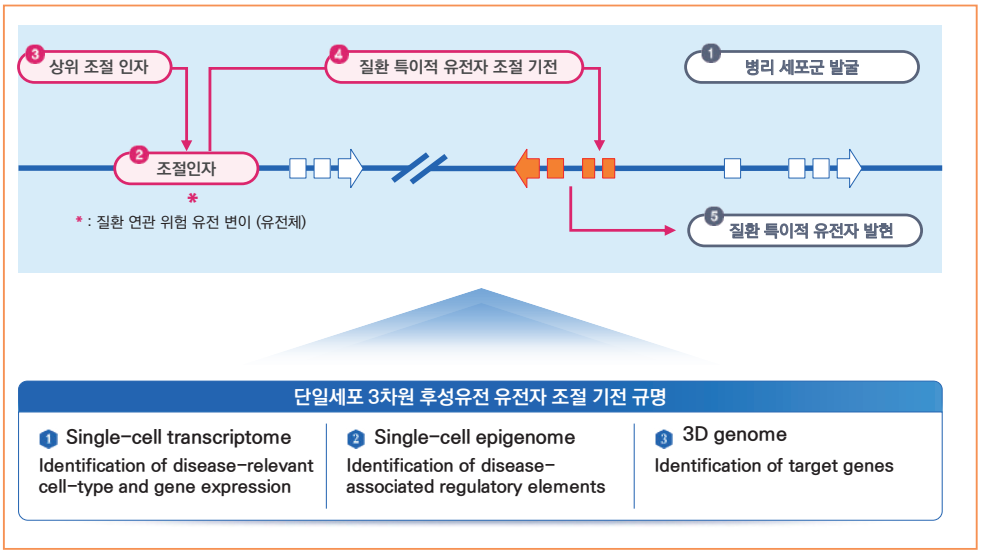
scDeepLUCIA accurately identifies functional enhancers



CRE-interactions are required for gene regulation



Delineation of 3D gene regulation



KSBI | 한국생명정보학회

KSBi-BIML 2026

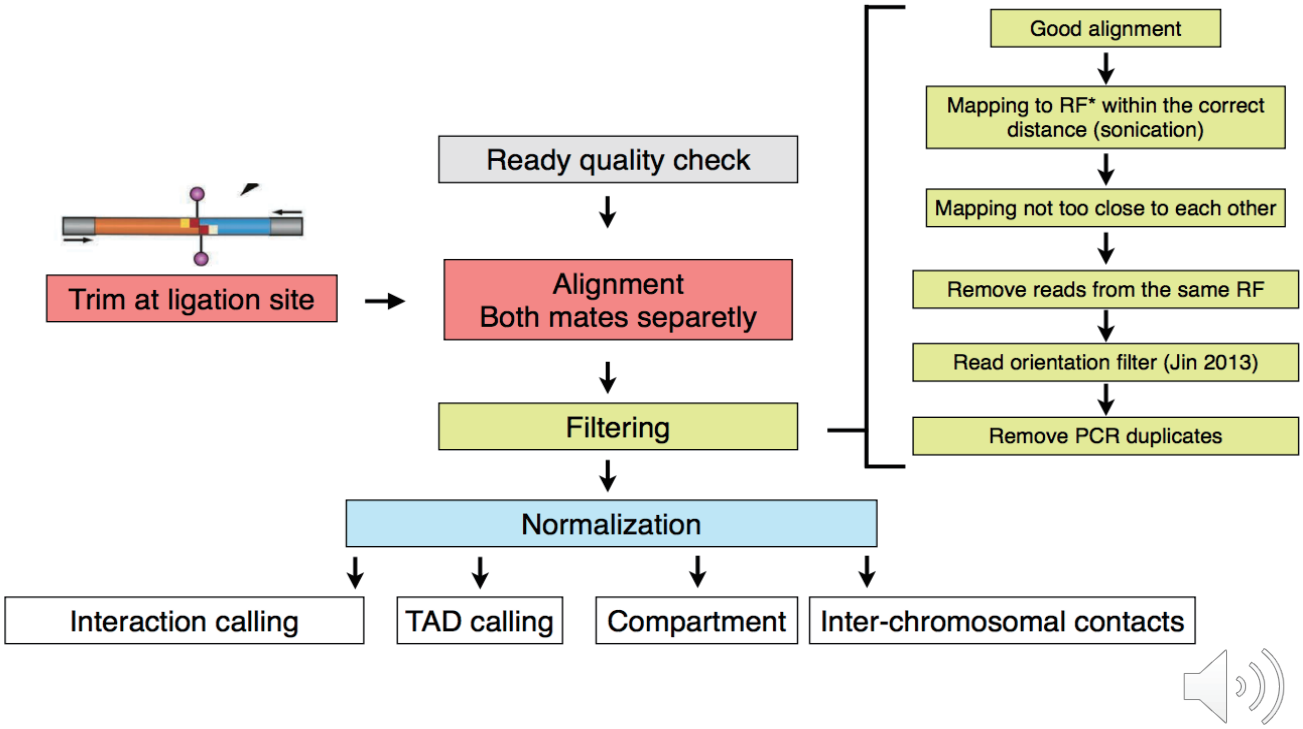
(Single-cell) 3D Epigenome Data Analysis

정인경(KAIST)

Contents

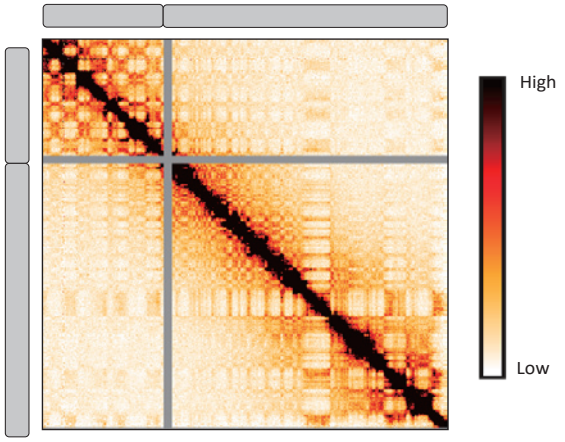
1. 후성유전학/염색질 3차구조 개요
2. 염색질 3차구조 중심의 단일세포 multi-omics 개요
3. **염색질 3차구조 데이터 분석 방법**
4. 3DIV 기반 Hi-C 데이터 분석 실습

How to define interactions? - Overall Workflow



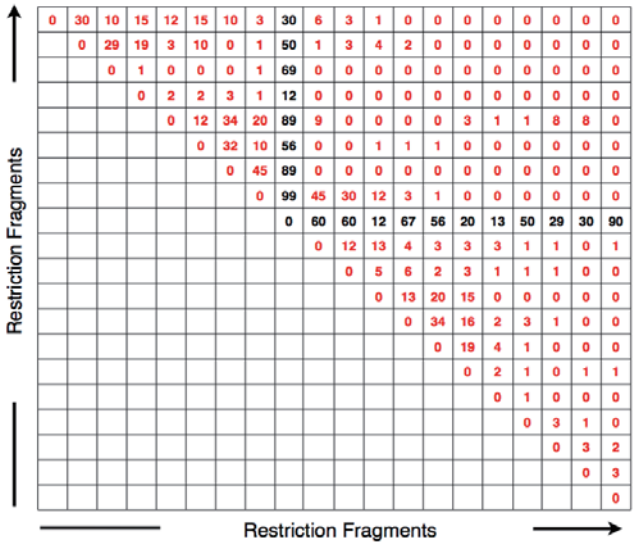
Sequencing Hi-C Library and then What?

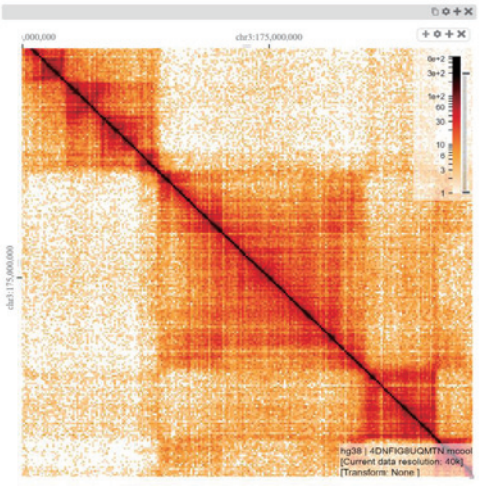
- Put the ends back together
  - Map to a reference genome
  - Determine ligation frequency
- Binning → Matrix
  - Sizes (resolution)



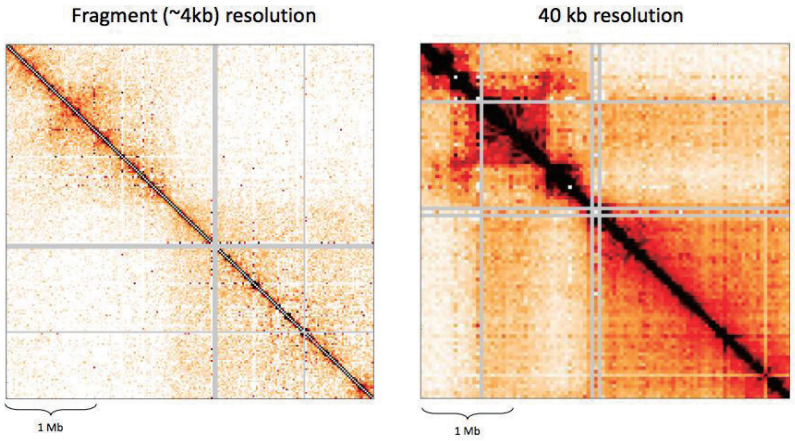
- Unless single cell, ensemble average (over 10^6 cells)
- Unless phased, averages over homologs.
- Unless synchronized, averages over the cell cycle.

Ligation Frequency Matrix (Interaction Map)





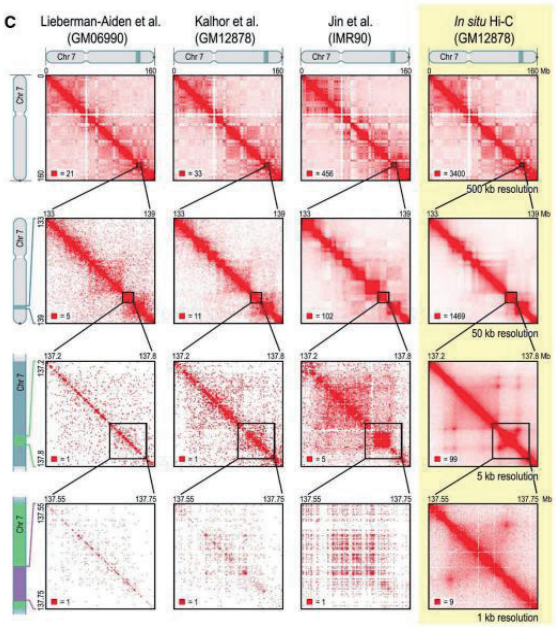
Resolution of Hi-C Interaction Map



- For the same Hi-C result, but different visualization depending on the resolution (bin size)
  - At 40kb resolution, TAD looks clear
  - At 4kb resolution location interactions look clear
- Determining optimal resolution is critical to precisely interpret Hi-C result

No clear definition to determine Hi-C resolution so far since there is no clear resolution dependent properties. However one paper proposed that **bin size can be determined as at least 80 % of all possible bins having more than 1,000 contact**

Hi-C Resolution



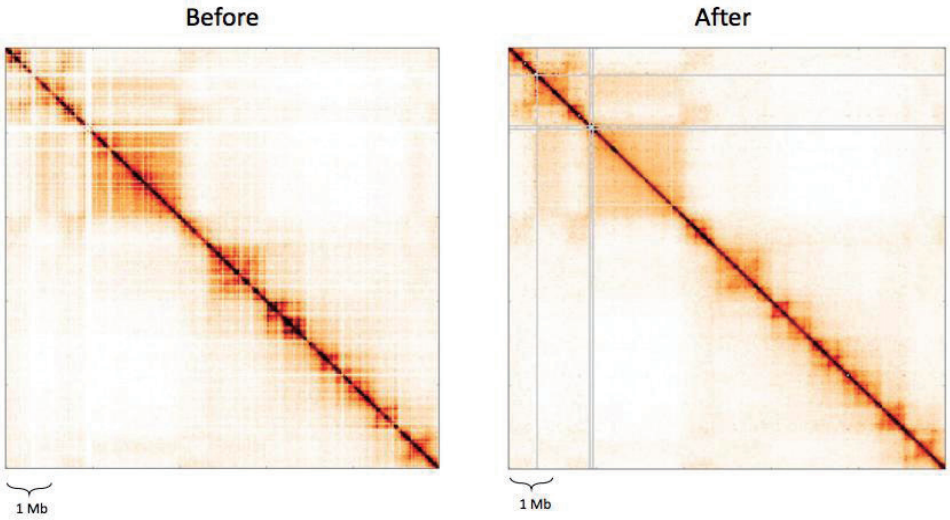
Some numbers in Hi-C resolution

- Human genome size: 3*10^9bp
- Restriction fragment length: ~4kb (6bp cutter)
- Number of fragments: 7.5*10^5
- Total interaction space: 5.6*10^11
- Number of cells per experiment: 10^6
- When we have 200M usable reads (2*10^8), 200 interactions were measured per cell
- Interaction space is under-sampled

How can we increase Hi-C resolution?

- Reduced interaction space (targeted approaches)
- Increased sequencing depth
- Reduced fragment size

Bias Correction (Normalization)



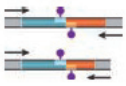
Uncorrected

Corrected (balanced)

Two major approaches for bias correction

Digestion efficiency as a function of sequence composition (and DNA compaction)

Ligation efficiency as a function of fragment length



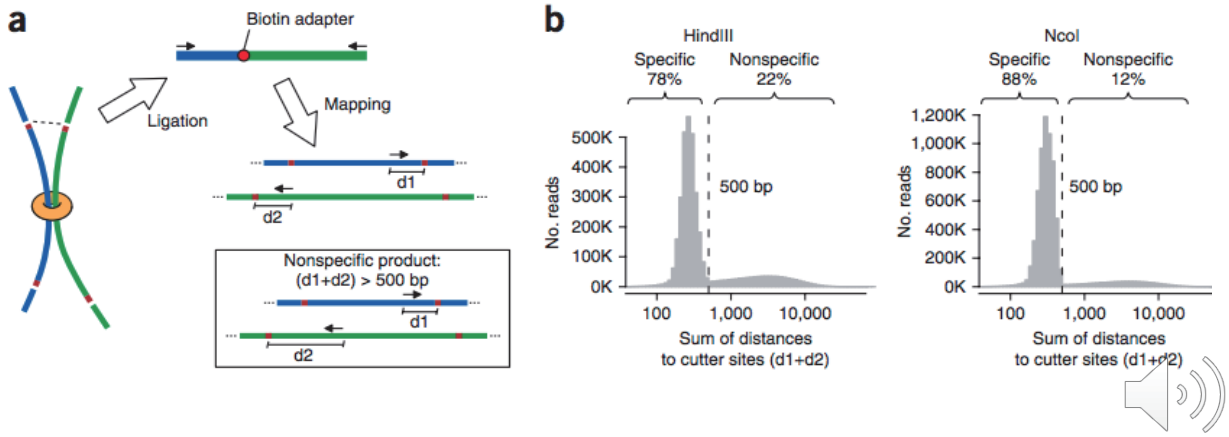
Sequencing efficiency as a function of sequence composition

1. Explicit factor methods (ex: HiCNorm)
   - Model bias due to GC content, fragment lengths, etc.
2. Coverage based methods (ex: ICE)
   - Don't model explicit sources of bias. Only assumes factorizable biases

Bias correction with explicit factor methods

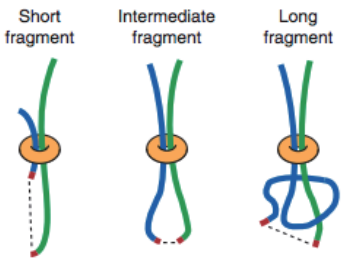
- Yaffe and Tanay or HiCNorm used explicit bias model based on 3 features that cause biases in Hi-C result
  - GC content, mappability, and effective length (or fragment length).
  - External information dependent
  - Effect of GC content on Hi-C library is enzyme-dependent.

Filtering for random ligation events

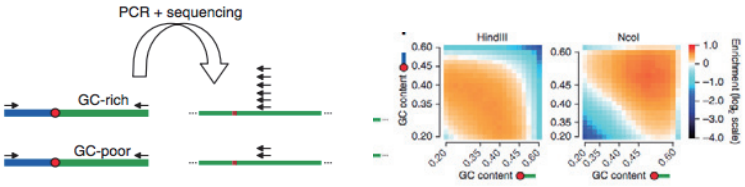


Bias correction with explicit factor methods

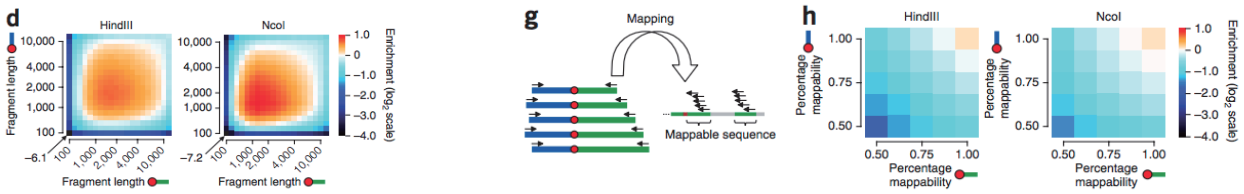
Fragment length



GC bias



Mappability bias



Bias correction with explicit factor methods

Let $\boldsymbol{U^i} = \{u^i_{jk}\}_{1\le j,k\le n_i}$ represent the $n_i \times n_i$ Hi-C *cis* contact map for chromosome i, where n_i is the number of consecutive, disjoint 1 MB bins in chromosome i. Each entry u^i_{jk} represent the number of paired-end reads spanning two bins L^i_j and L^i_k. Let x^i_j, y^i_j and z^i_j represent the effective length feature, the GC content feature and the mappability feature at locus j for chromosome i, respectively. Similarly, let x^i_k, y^i_k and z^i_k represent the effective length feature, the GC content feature and the mappability feature at locus k for chromosome i, respectively. We assume that u^i_{jk} follows Poisson distribution with rate θ^i_{jk}:

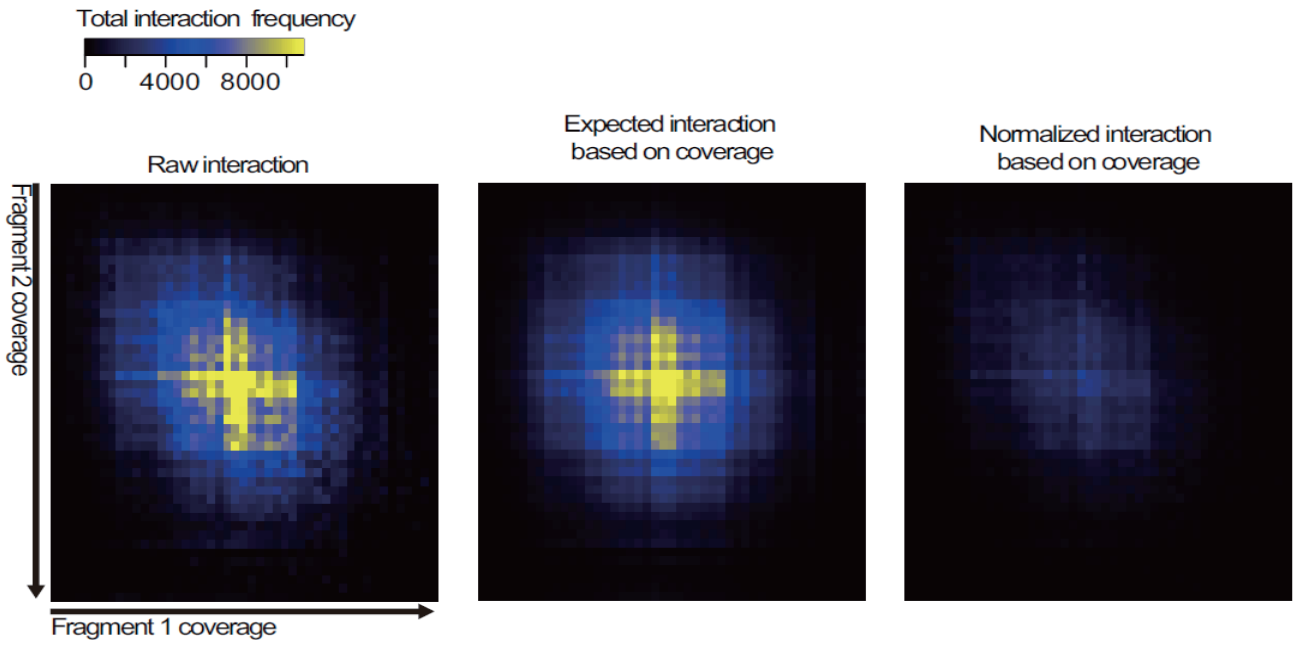
$$\log(\theta^i_{jk}) = \beta^i_0 + \beta^i_{len}\log(x^i_j x^i_k) + \beta^i_{gcc}\log(y^i_j y^i_k) + \log(z^i_j z^i_k).$$

Here β^i_0 is the intercept term. β^i_{len} and β^i_{gcc} represent the effective length bias and the GC content bias, respectively. $\log(z^i_j z^i_k)$ is the Poisson offset term of the mappability bias. We fit this Poisson regression model, and let $\hat{\beta}^i_0$, $\hat{\beta}^i_{len}$ and $\hat{\beta}^i_{gcc}$ represent the corresponding parameter estimates. We further define the estimated Poisson rate $\hat{\theta}^i_{jk}$ as following:

$$\hat{\theta}^i_{jk} = exp\{\hat{\beta}^i_0 + \hat{\beta}^i_{len}\log(x^i_j x^i_k) + \hat{\beta}^i_{gcc}\log(y^i_j y^i_k) + \log(z^i_j z^i_k)\}.$$

The residual $e^i_{jk} = u^i_{jk}/\hat{\theta}^i_{jk}$ is the normalized *cis* interaction between two bins L^i_j and L^i_k.

HiCNorm / covNorm



Bias correction with coverage based methods

- Do not try to identify sources of biases but learn their effect from data

number of reads between segments *i* and *j*

normalized ligation frequency between segments *i* and *j*

$$f_{ij} = \frac{c_{ij}\left(\sum_{k=1}^{K-1}\sum_{l=k+1}^{K} c_{kl}\right)}{\left(\sum_{k=1}^{K} c_{ik}\right)\left(\sum_{k=1}^{K} c_{kj}\right)}$$

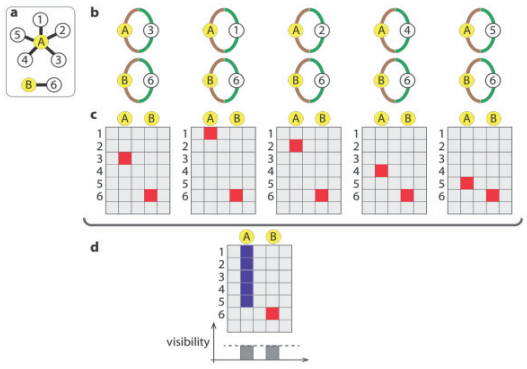
Total number of reads

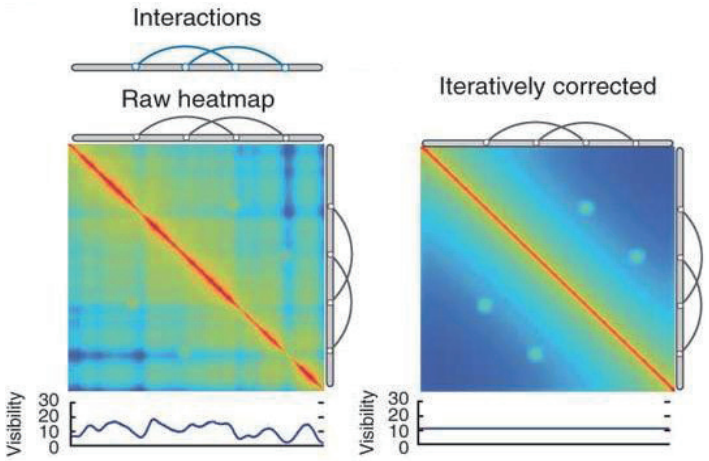
Total number of for segment *i*

Total number of for segment *j*

Bias correction with coverage based methods - equalized visibility

ICE equalized visibility: Each bin has equal coverage





Imakaev *et al* (2012)

Bias correction with coverage based methods - equalized visibility

Biases

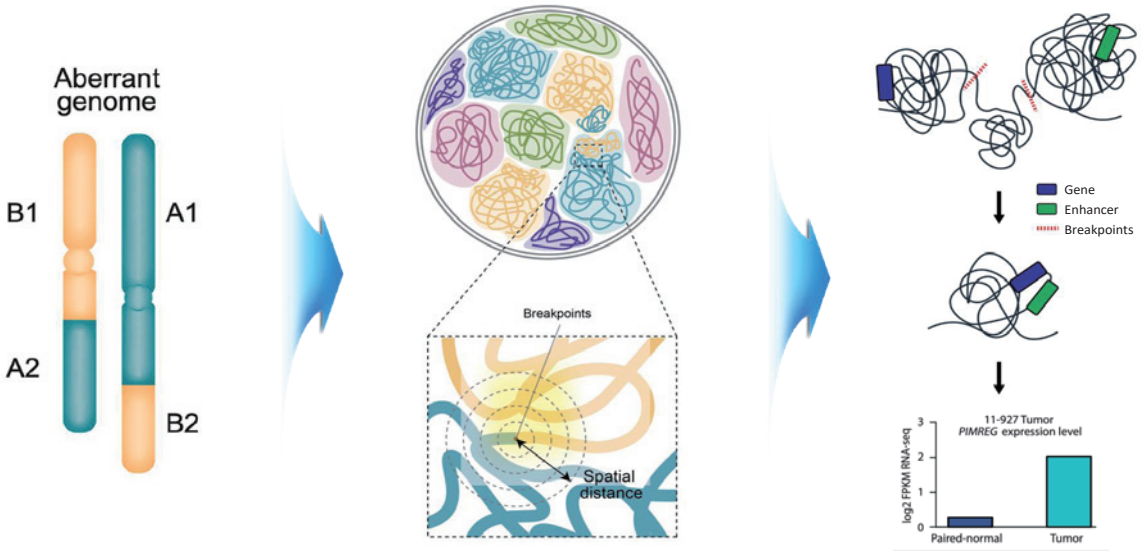
Observed $O_{ij} = B_i B_j T_{ij}$ True

1. Start from Wij (=Oij) as the iterative process gradually changes this matrix to Tij
2. Calculate coverage of i as s_i=sum_j(Wij)
3. Additional biases delta Bi are calculated by renomralizitng s_i to have the unit mean as delta Bi = S_i / mean(S_i)
4. New Wij = Wij / (delta Bi * delta Bj)
5. Iterate step 2-4 until the variance of the additional biases becomes negligible

Issues:
- local signals can be removed
- interaction hubs such as transcriptional factory are not covered by ICE normalization

Advanced Topics (enhancer-hijacking)

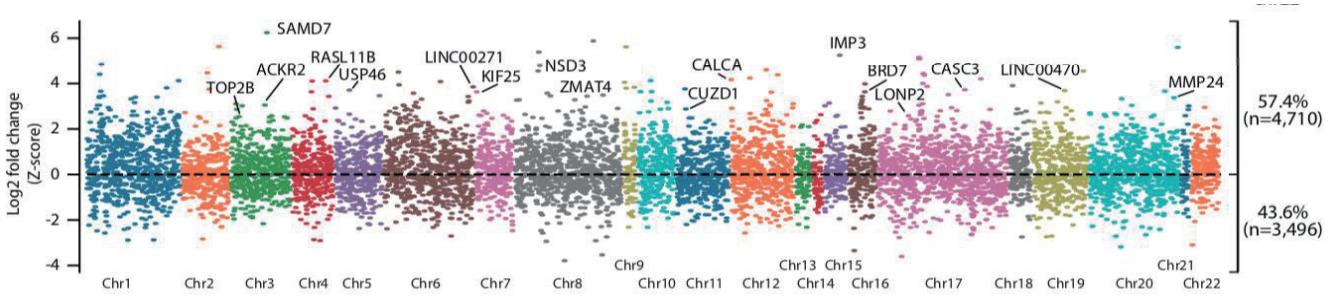
Structural variations induce oncogenic 3D genome

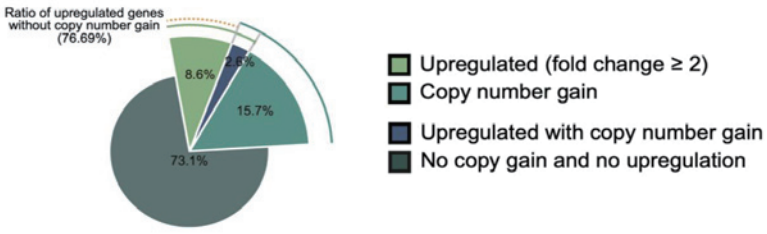


1D-genomic rearrangement **3D genome reorganization** **Aberrant gene expression**

Kim et al., Mol Cells (2019), Kim et al., Semin Cell Dev Biol. (2021)

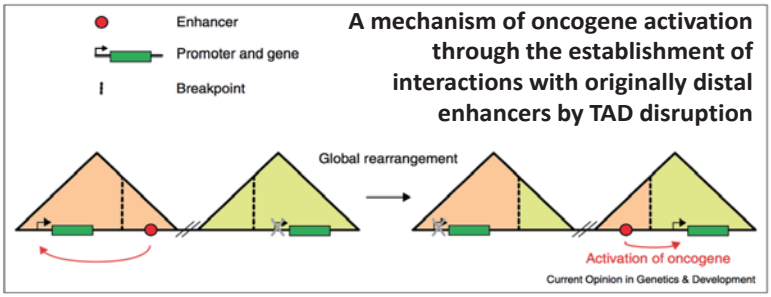
Copy number variations are not sufficient to explain oncogene activation nearby SVs





A model of Enhancer Hijacking

Enhancer Hijacking



A mechanism of oncogene activation through the establishment of interactions with originally distal enhancers by TAD disruption

Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer

Kadir C. Akdemir[1], Victoria T. Le[2], Sahaana Chandran[2], Yilong Li[3], Roel G. Verhaak[4], Rameen Beroukhim[5,6,7,8], Peter J. Campbell[3,9], Lynda Chin[10], Jesse R. Dixon[2], P. Andrew Futreal[1]✉, PCAWG Structural Variation Working Group[11] and PCAWG Consortium[12]

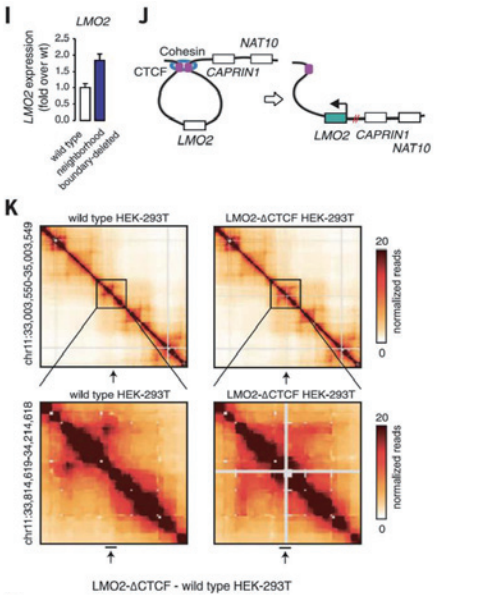
Nat Genet 2020

Conclusion:

Most domain disruptions do not result in marked gene expression changes

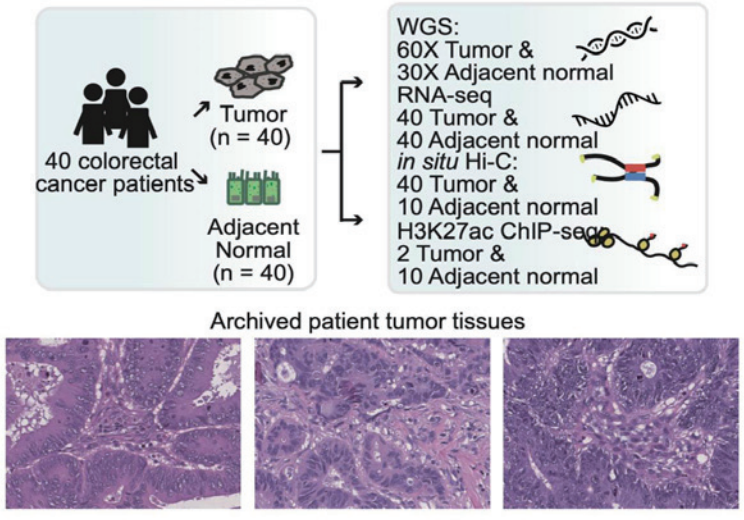
Limitation:

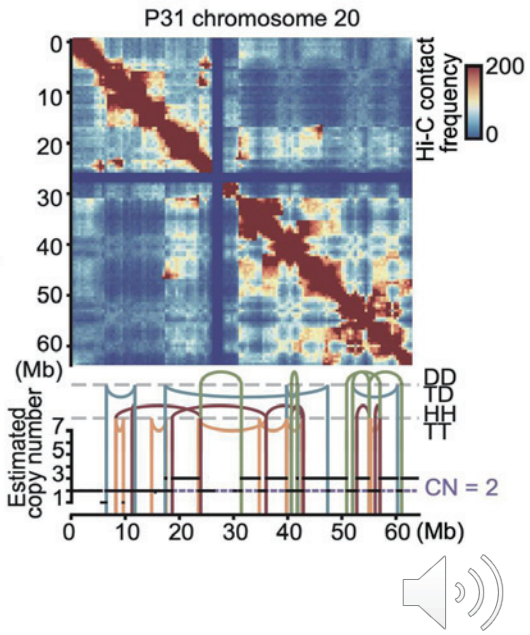
A lack of 3D genome in patients' cancer genome



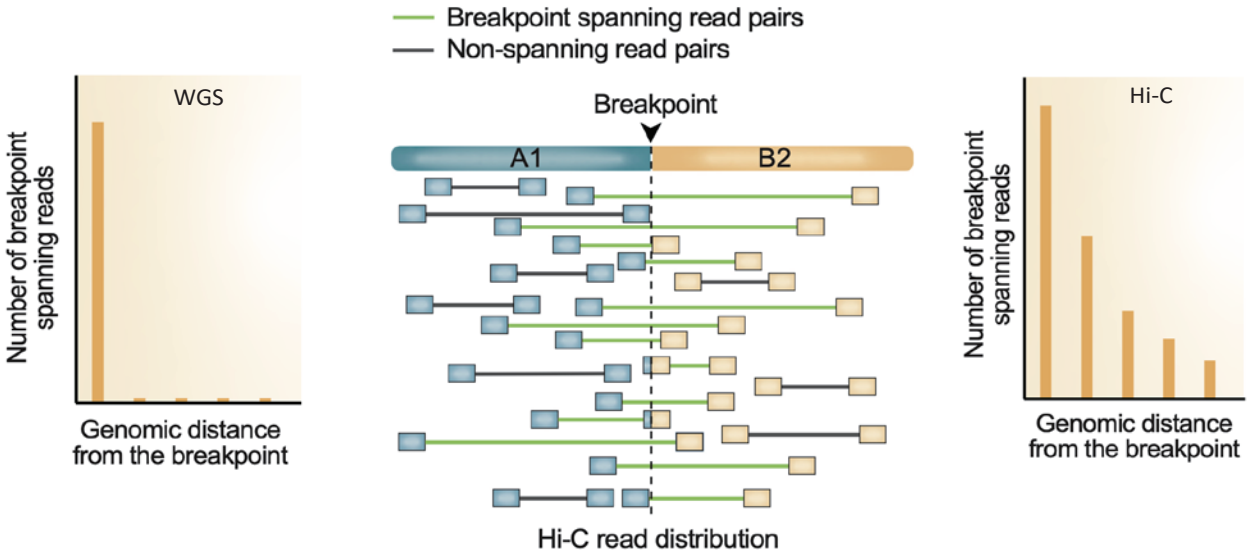
Science 2016

Disorganized patient–specific 3D genome can be oncogenic by rewiring enhancer-promoter interactions



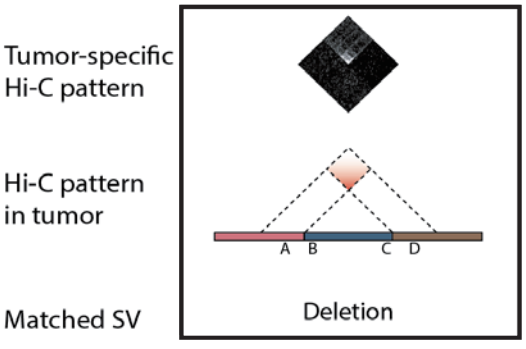


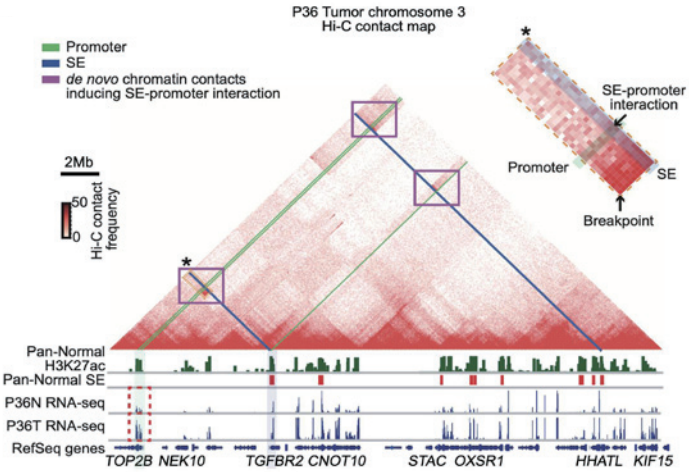
Hi-C is a molecular rangefinder for detection of SVs



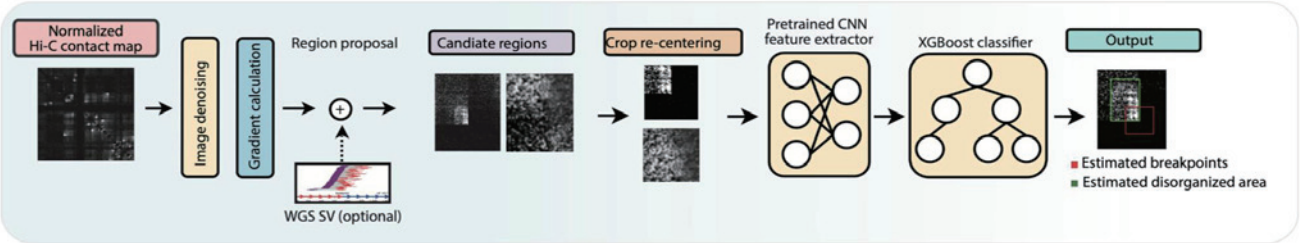
Kim et al., 2019; Kim et al., 2021

A unique contact signature of 3D cancer genome

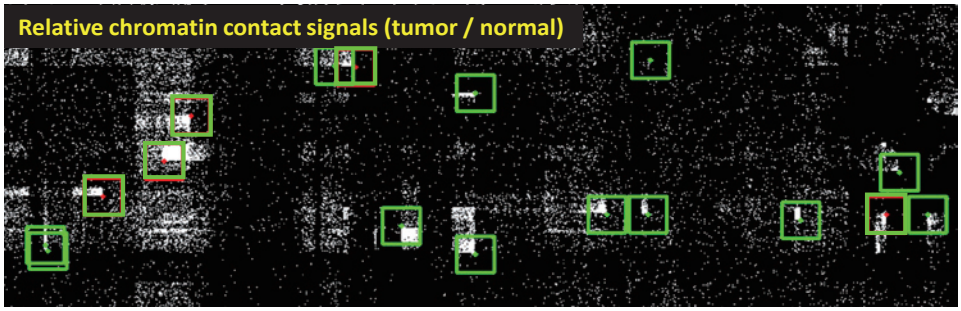




→ Defining cancer-specific de novo chromatin contacts
→ Identify genes and SEs involving the de novo contact
→ Test if SE-hijacking genes are over-expressed
→ Determine if the genes are recurrently affected



Problem: tumor purity and tissue heterogeneity

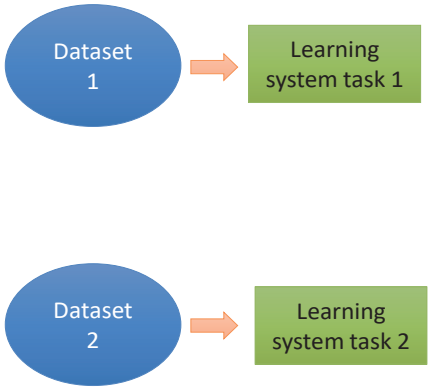


Signal patterns to be identified

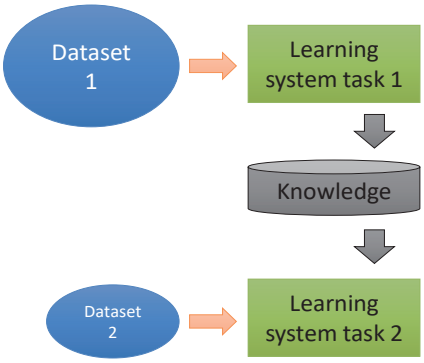
Background noise patterns to be removed

Solution: A pre-trained model for 3D cancer genome

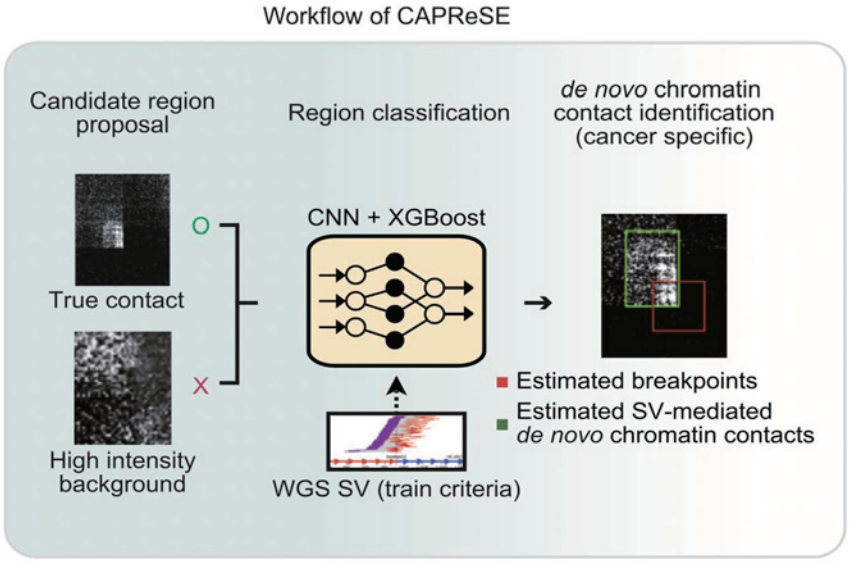
Traditional machine learning
Isolated & single task learning



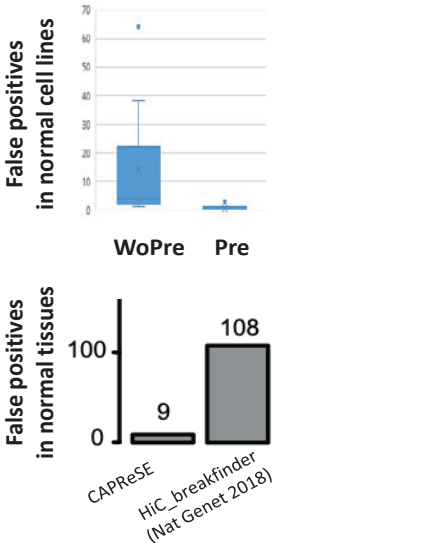
Pre-trained model construction
Learning of a new tasks relies on the previous learned tasks



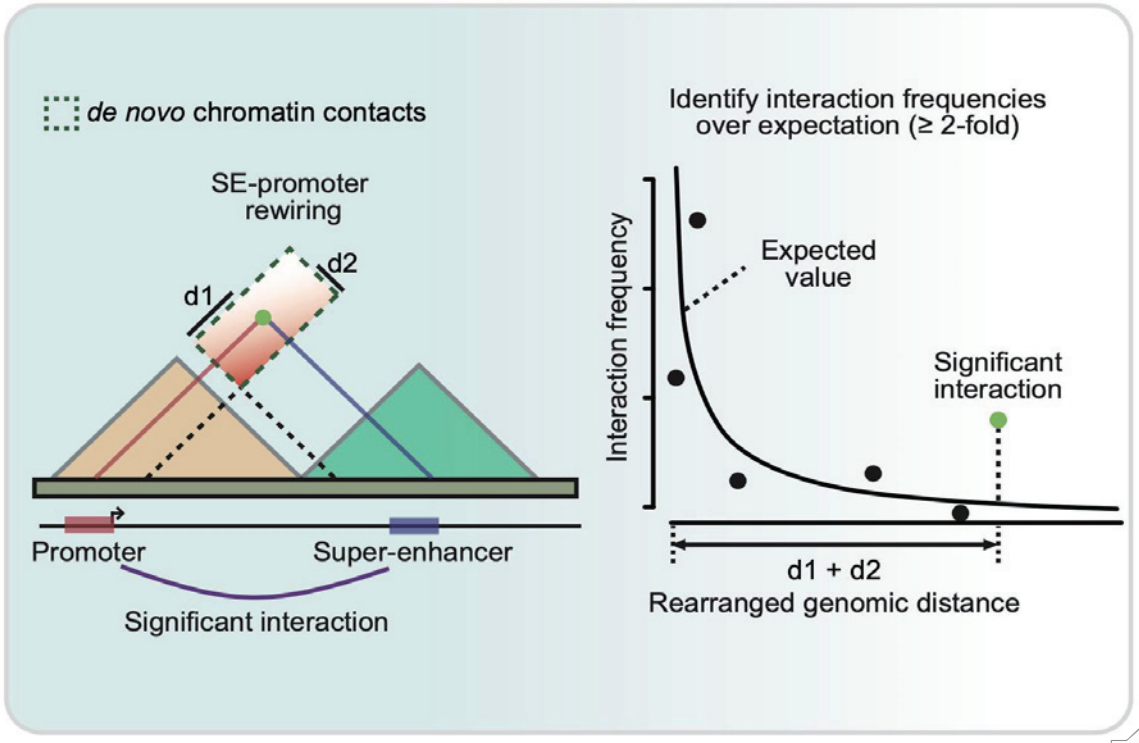
CAPReSE (chromatin anomaly pattern recognition and size)

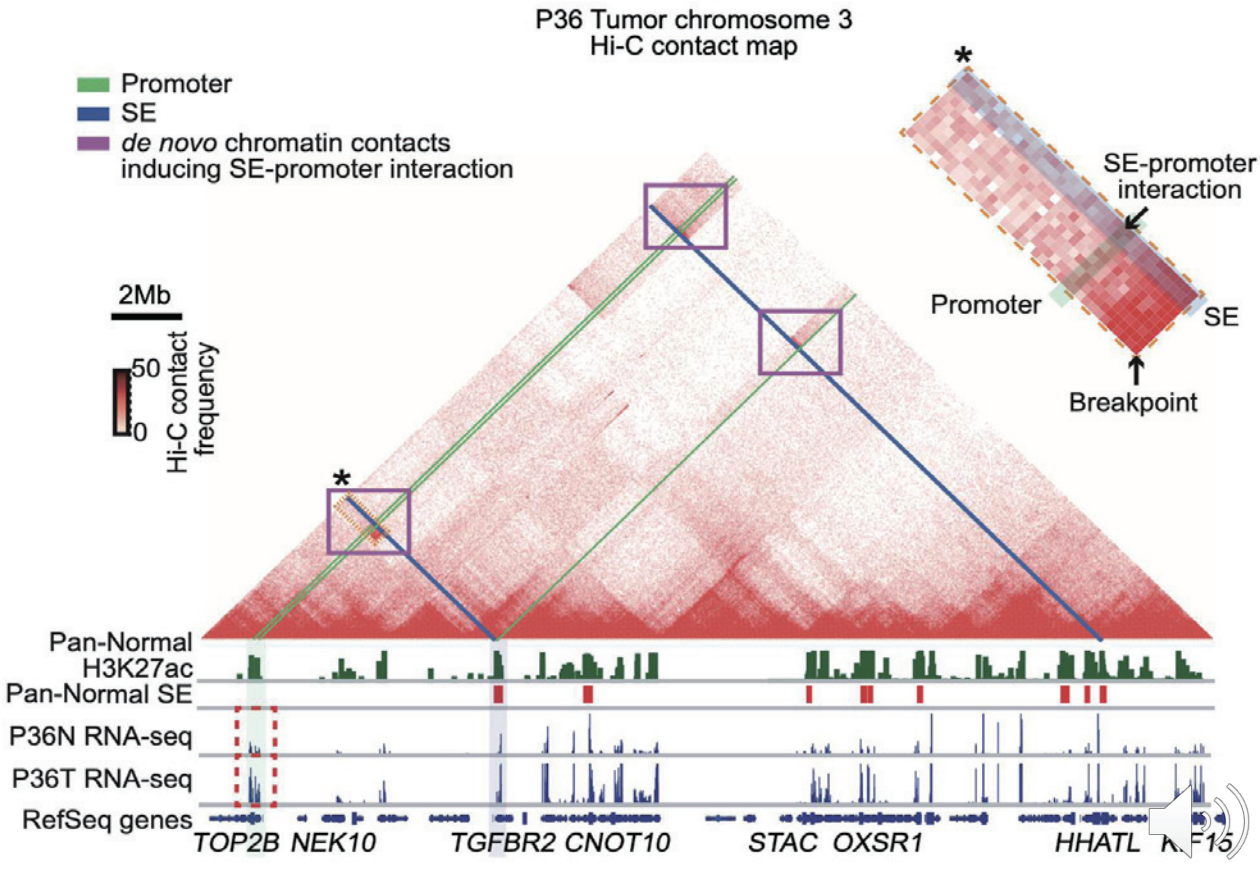


Performance evaluation

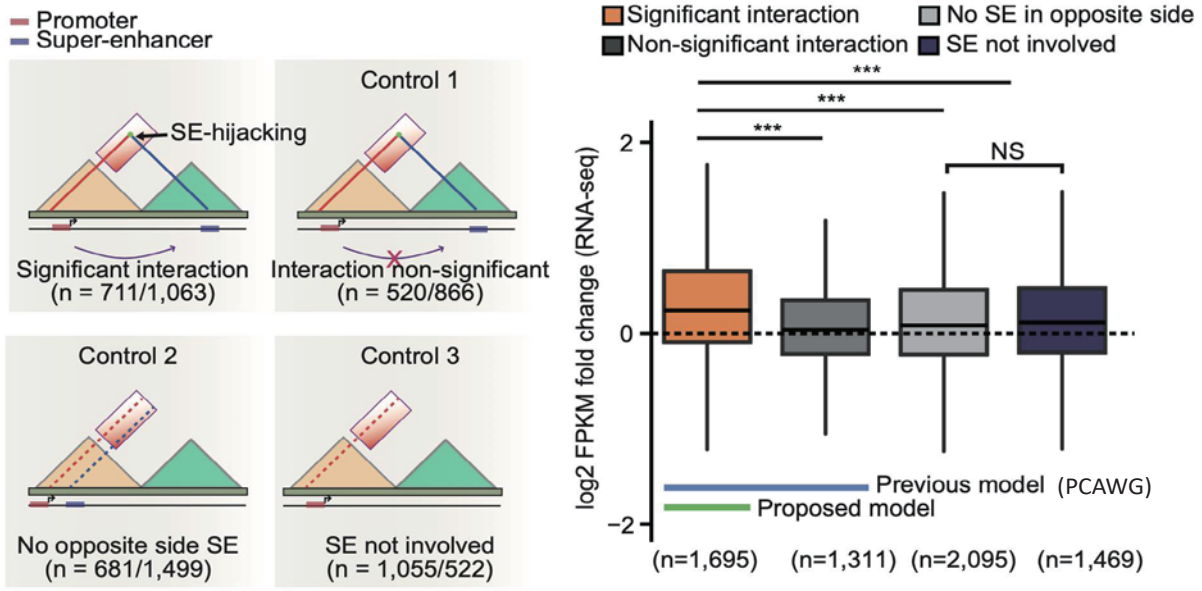


Kim et al., Cell Reports 2023

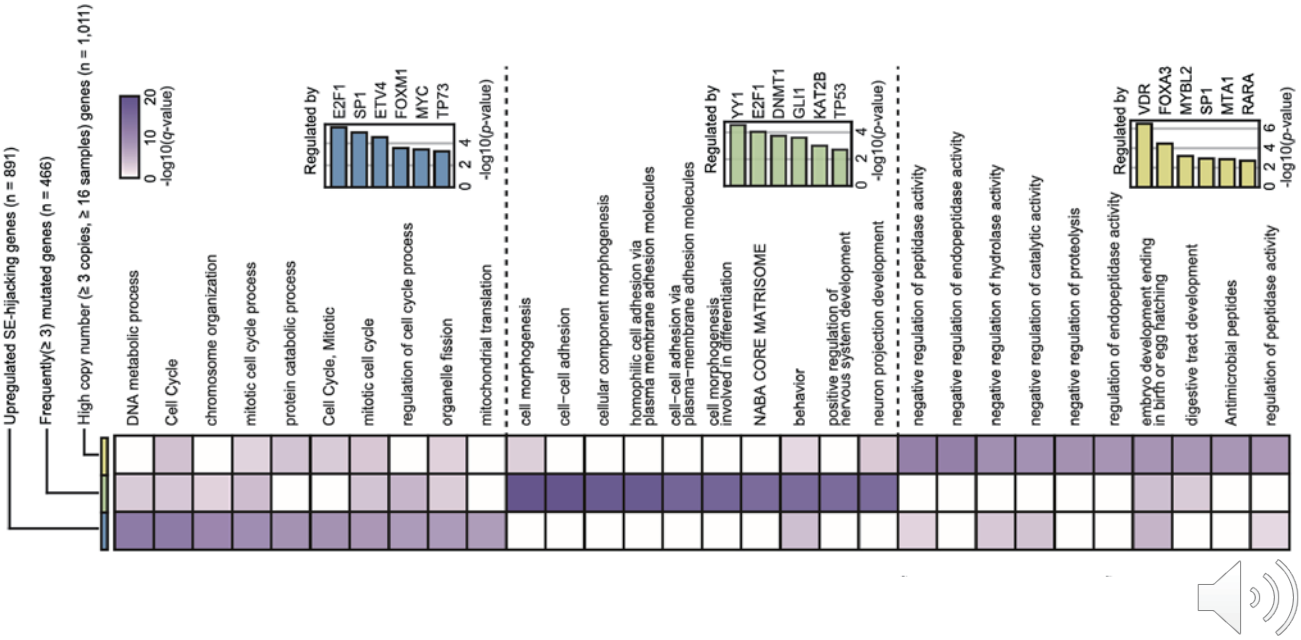

de novo chromatin contacts
SE-promoter rewiring
d1
d2
Promoter
Super-enhancer
Significant interaction
Identify interaction frequencies over expectation (≥ 2-fold)
Interaction frequency
Expected value
Significant interaction
d1 + d2
Rearranged genomic distance


P36 Tumor chromosome 3
Hi-C contact map
Promoter
SE
de novo chromatin contacts inducing SE-promoter interaction
2Mb
50
0
Hi-C contact frequency
SE-promoter interaction
Promoter
SE
Breakpoint
Pan-Normal H3K27ac
Pan-Normal SE
P36N RNA-seq
P36T RNA-seq
RefSeq genes
TOP2B NEK10
TGFBR2 CNOT10
STAC OXSR1
HHATL

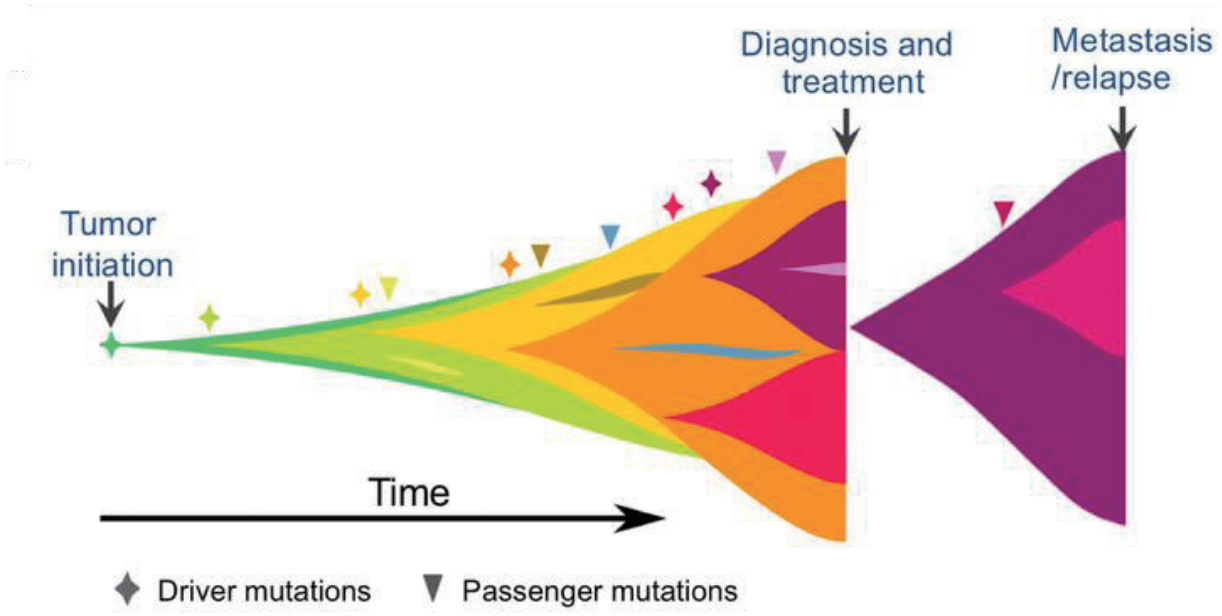
Rewired SE-promoter induces oncogene activation

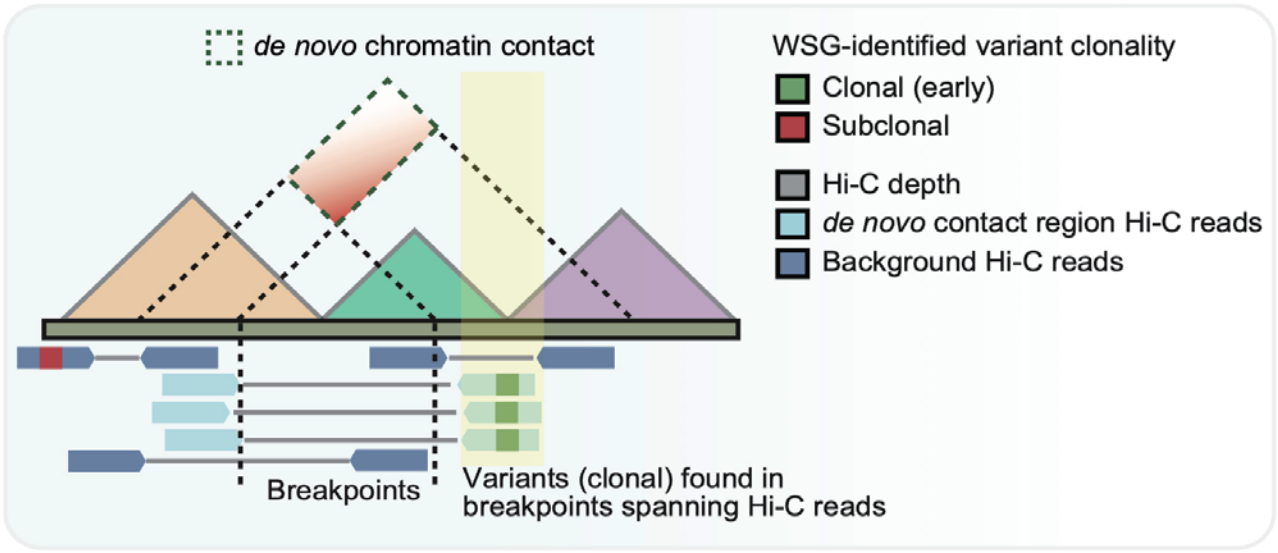


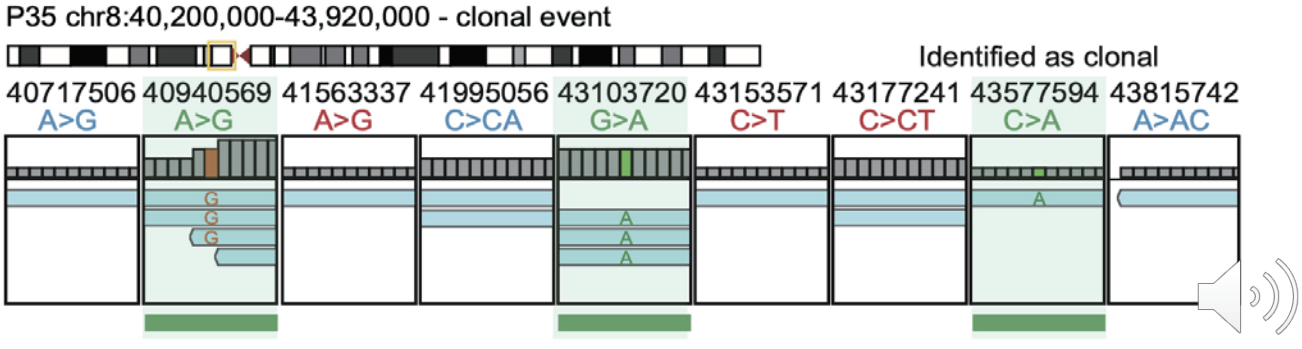
SE-hijacking genes are enriched by unique biological functions



Clonal vs sub-clonal Enhancer-hijacking genes

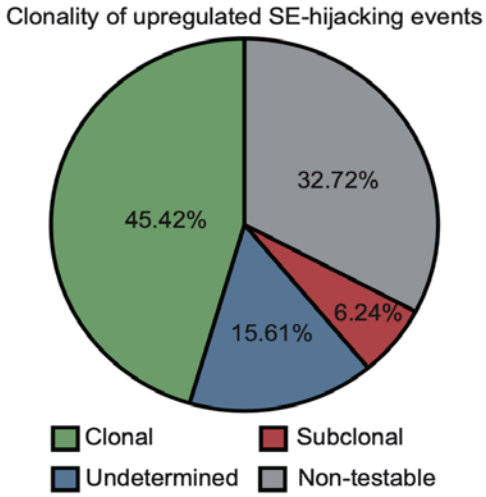


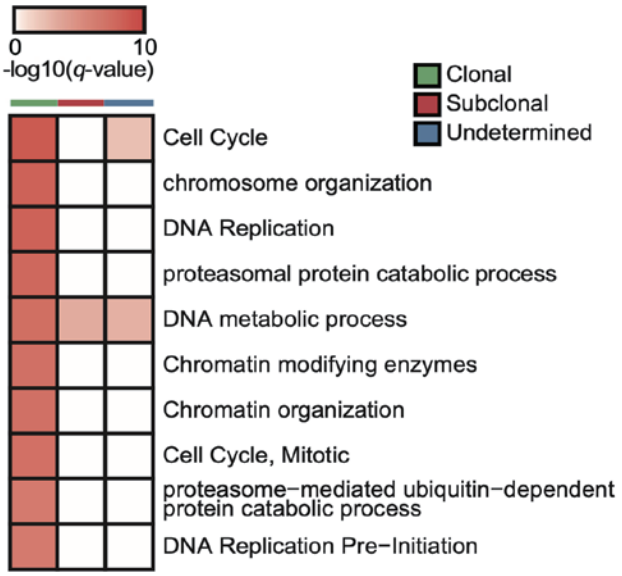




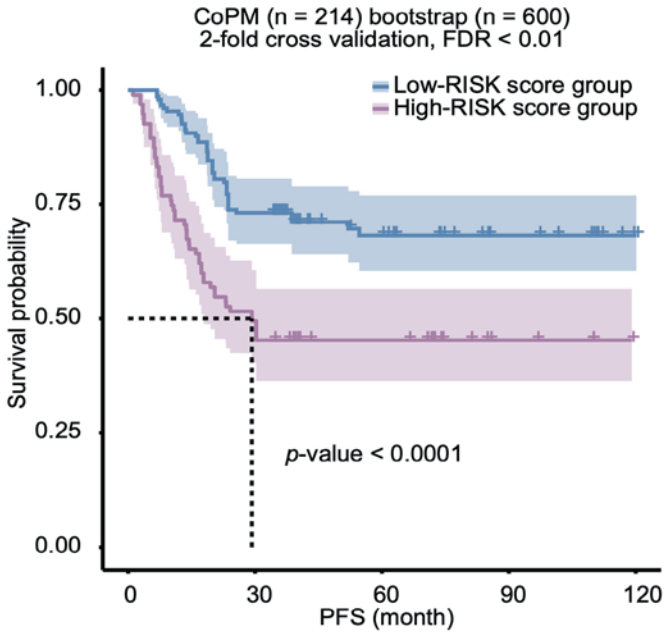
Gustung et al., 2020

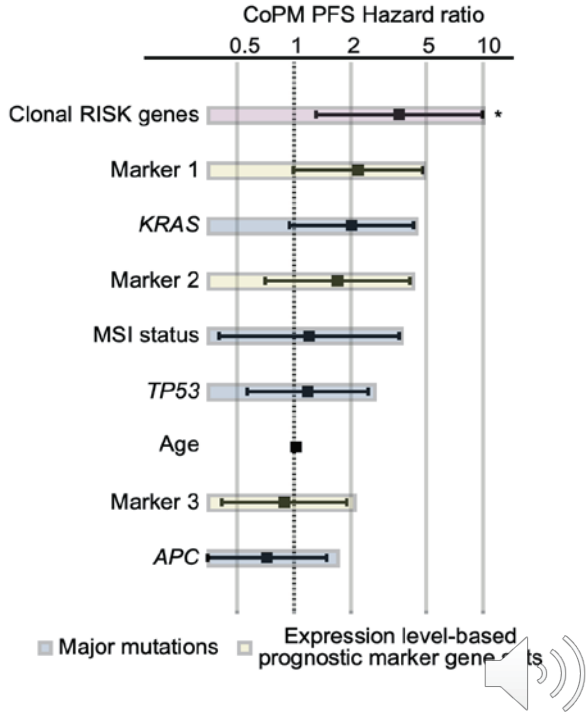
Clonal enhancer-hijacking genes are functionally important





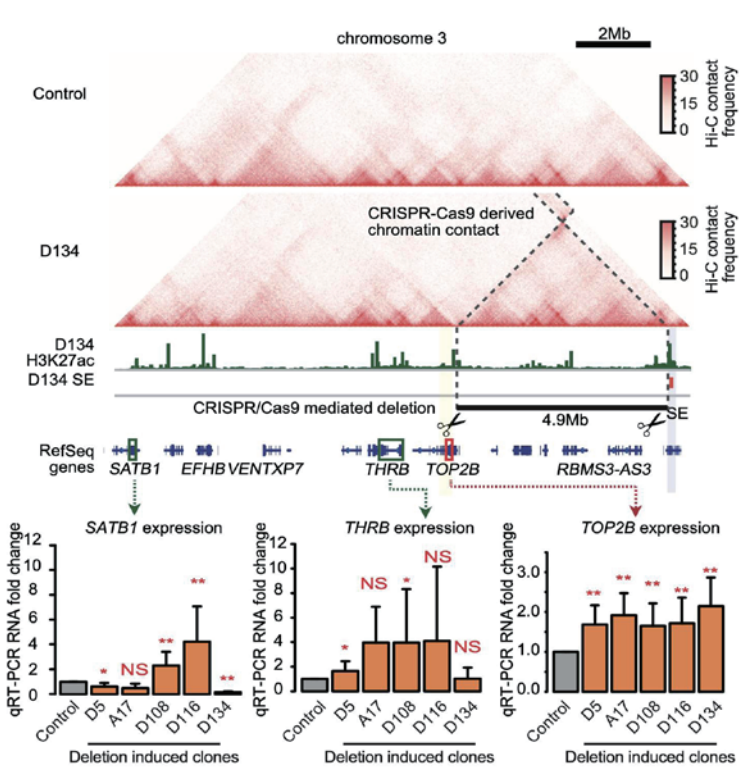
The prognostic potential of clonal hijacking genes

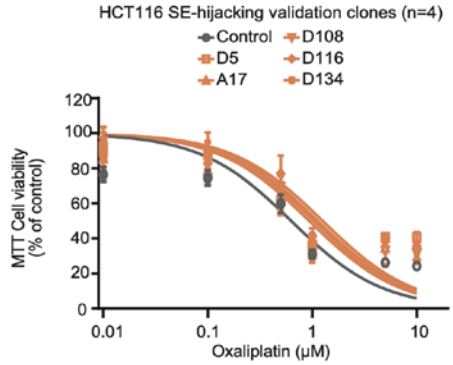


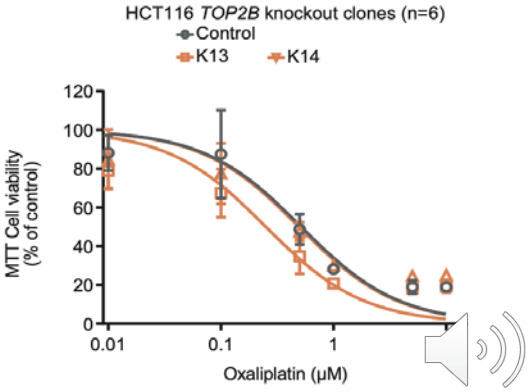


Marker 1, 2, and 3 gene sets were obtained from Koncina et al., 2020

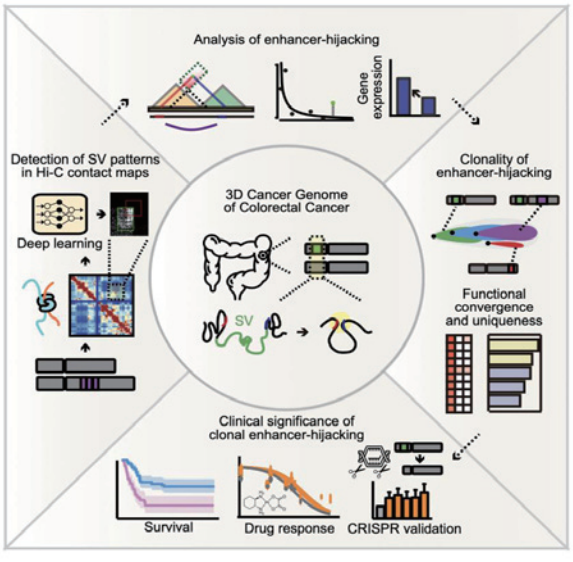
Clinical implication of clonal hijacking genes







Clonal enhancer-hijacking can be oncogenic and may serve as an 'Active supporter' during cancer progression



- **Patients' specific 3D genome** requires to precise prediction of enhancer-hijacking
- **Clonal enhancer-hijacking genes** are enriched to **recurrent oncogenic functions**
- Clonal enhancer-hijacking genes may serve as **prognostic markers** for colorectal cancer

Kim et al., Cell Reports 2023

KSBI | 한국생명정보학회

KSBi-BIML 2026

(Single-cell) 3D Epigenome Data Analysis

정인경(KAIST)

Contents

1. 후성유전학/염색질 3차구조 개요
2. 염색질 3차구조 중심의 단일세포 multi-omics 개요
3. 염색질 3차구조 데이터 분석 방법
4. 3DIV 기반 Hi-C 데이터 분석 실습

3DIV: http://3div.kr

ABOUT 3DIV

3D genome organization is tightly coupled with gene regulation in various biological processes and diseases. 3D Interaction Viewer and Database (3DIV) is a database providing chromatin interaction visualization in a variety of options from one-to-all chromatin interaction with epigenetic annotation to unique dynamic browsing tools allowing examination of large-scale genomic rearrangement mediated impacts in cancer 3D genome. 3DIV will be the most comprehensive resource to explore gene regulatory effects of both normal and cancer 3D genome.

Hi-C

3DIV provides querying list of significant interacting partner locus, visualization, and comparative analysis of 3D chromatin interaction across about 400 samples.

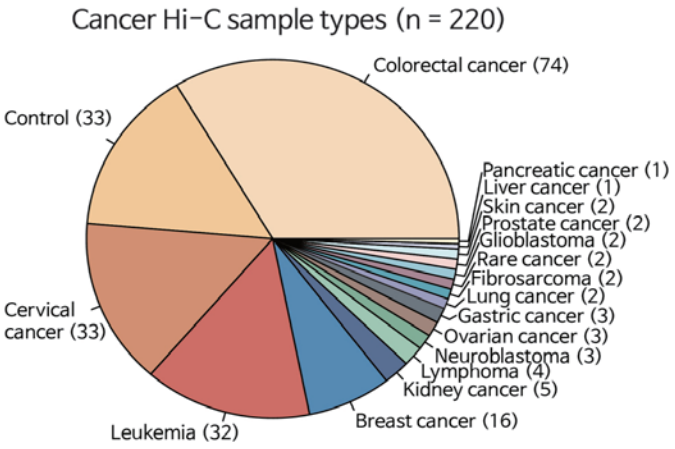
Capture Hi-C

3DIV provides promoter capture Hi-C (pcHi-C) results across 28 normal human cell/tissue types, a great resource in identifying target genes of disease-associated genetic variants.

Cancer Hi-C

3DIV provides unique visualization and manipulation tools that allows user to generate rearranged 3D genome by selecting listed SVs, creating own SVs, and providing order of rearranged chromosomes.

Hi-C data collection in 3DIV



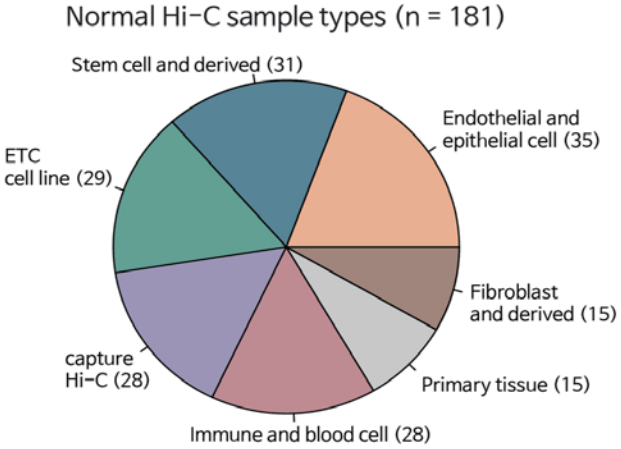


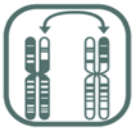
Table 1. Comparison of the updated 3DIV and other 3D genome databases as of October 2020

Software	Number of samples[a]	Hi-C contact map	TAD annotation	One-to-all interaction	Interaction table	Distance normalization	Interactive Hi-C contact map browsing	Live manipulation of genomic rearrangement	Structural variation annotation
3DIV 2021 Update	401	✓	✓	✓	✓	✓	✓	✓	✓
3DIV	80	✓	✓	✓	✓	✓	✓		
4D Nucleome	337[b]	✓					✓		
Nucleome Browser	138[c]	✓					✓		
WashU Epigenome Browser	36[c,d]	✓					✓		
HiView	2		✓	✓	✓	✓	✓		
HUGIn2	83	✓	✓	✓	✓	✓			
3D Genome Browser	113	✓	✓						
GITAR	20[c]	✓	✓						
Hi-C Data Browser	69	✓		✓					

Unique functionalities of 3DIV

Features of 3DIV

187 cancer/tumor tissue samples
with 33 control samples



Pan-cancer SV data
for corresponding cancer type

153 cell line/tissue Hi-C and
28 promoter capture Hi-C data

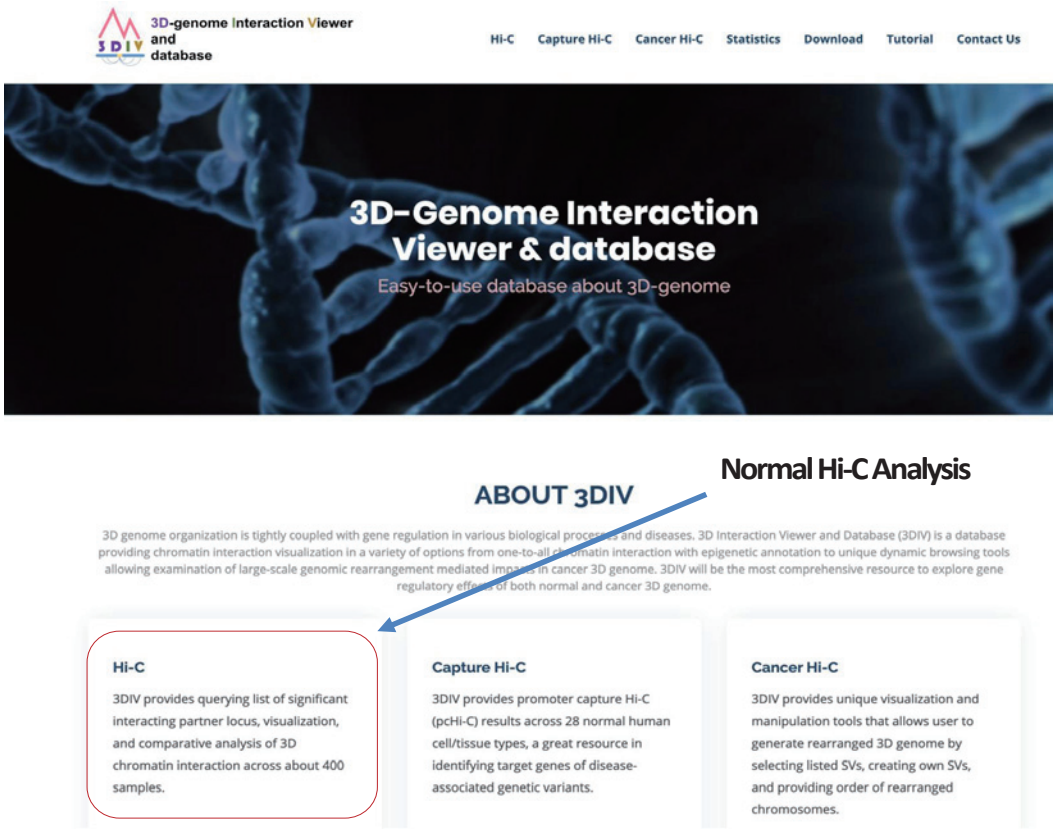
MySQL + Java Spring + HTML5
based webserver implementation

230 billion reads processed
and normalized Hi-C contact maps

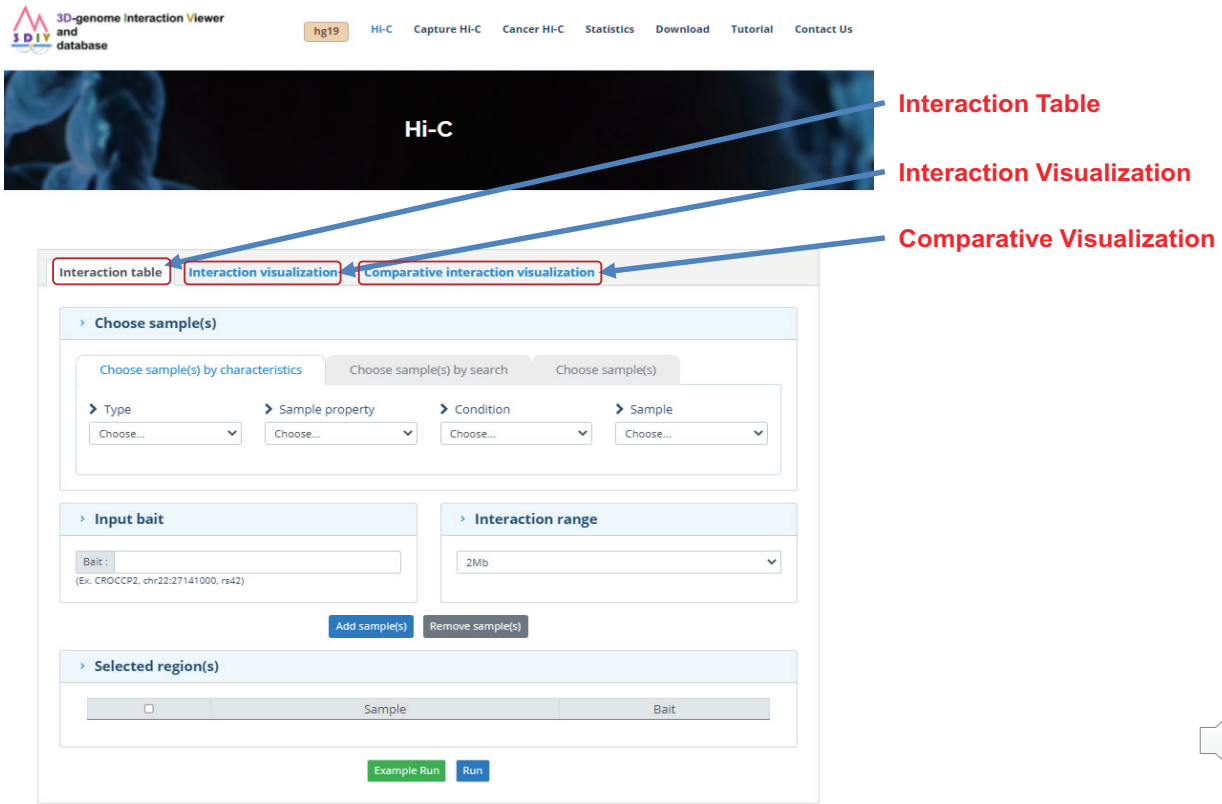
Interactive visualization function
on web page

Normal Hi-C Analysis

Normal Hi-C Analysis



Normal Hi-C Analysis



Functionalities of normal Hi-C in 3DIV

Interaction Table

- Bias-removed/distance-normalized Interaction frequency
- Disease-associated GWAS SNPs
- Promoter/Enhancer/super-enhancer annotation
- Histone ChIP-seq signal

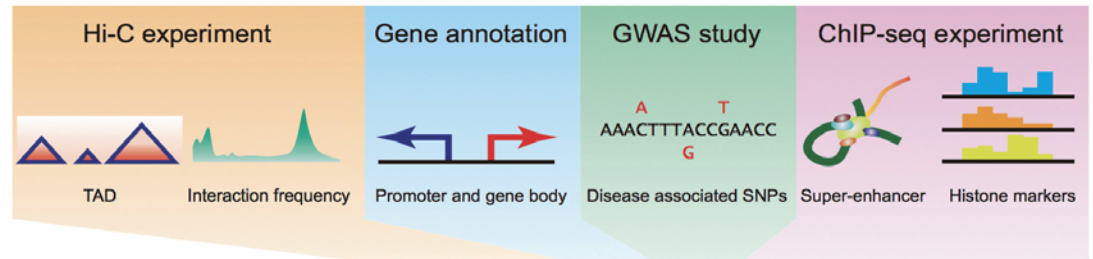
Interaction Visualization

- Interaction frequency heatmap
- Topologically associating domains
- One-to-all interaction plot
- Arc-representation of significant interactions

Comparative Visualization

- Comparative interaction frequency heatmap
- Synchronized interaction visualization

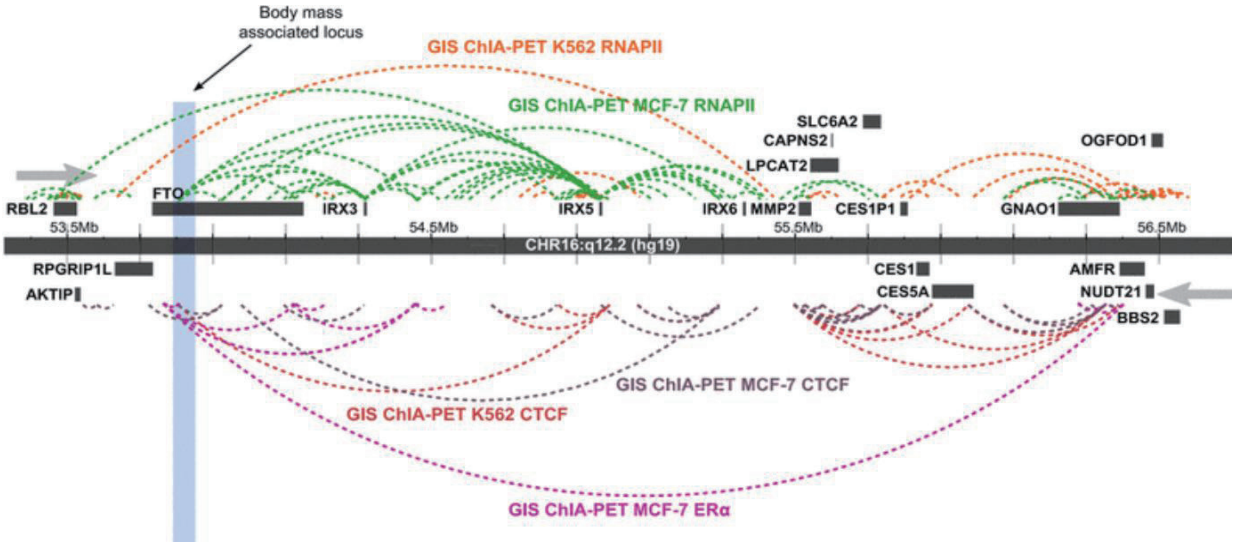
Module 1 : Interaction Table



Sample	Locus (unit : kb)	Bias-removed interaction frequency	Distance-normalized interaction frequency	Gene Name	GWAS SNP ID	Enhancer or Super-enhancer	H3K27ac Fold change	H3K4me1 Fold change	H3K4me3 Fold change
Mesenchymal Stem Cell	chr16:54965-54970	3.16	7.07	IRX5			2.6	4.15	23.9
Mesenchymal Stem Cell	chr16:55505-55510	2.25	6.99			Enhancer	8.75	3.61	2.27
Mesenchymal Stem Cell	chr16:55500-55505	2.01	6.24			Enhancer	6.41	4.15	3.86
Mesenchymal Stem Cell	chr16:55540-55545	1.29	4.11	LPCAT2			1.54	5.61	11.77
Mesenchymal Stem Cell	chr16:55510-55515	1.29	4.03	MMP2			2.07	5.61	23.5
Mesenchymal Stem Cell	chr16:55535-55540	1.02	3.22			Enhancer	6.1	5.98	2.14
Mesenchymal Stem Cell	chr16:55355-55360	1.06	3.01	IRX6			2.47	3.04	11.24
Mesenchymal Stem Cell	chr16:54320-54325	2.56	2.88	IRX3			3.87	6.16	18.62
Mesenchymal Stem Cell	chr16:55530-55535	0.86	2.73			Enhancer	6.63	5.39	2.8
Mesenchymal Stem Cell	chr16:55515-55520	0.74	2.32	MMP2			1.34	3.73	1.75
Mesenchymal Stem Cell	chr16:55310-55315	0.83	2.3			Enhancer	3.36	3.59	2.14
Mesenchymal Stem Cell	chr16:52115-52120	0.73	2.24	LINC00919			1.33	1.24	1.22
Mesenchymal Stem Cell	chr16:55705-55710	0.62	2.16	SLC6A2			1.12	1.79	2.27
Mesenchymal Stem Cell	chr16:54375-54380	1.74	2.14			Enhancer	3.12	5.96	2.8
Mesenchymal Stem Cell	chr16:55600-55605	0.62	2.05	CAPNS2			1.65	2.33	1.75
Mesenchymal Stem Cell	chr16:55315-55320	0.74	2.05			Enhancer	6.31	4.15	1.75
Mesenchymal Stem Cell	chr16:54490-54495	0.89	1.29		rs9921518		1.54	2.88	1.75
		6.22							1.33

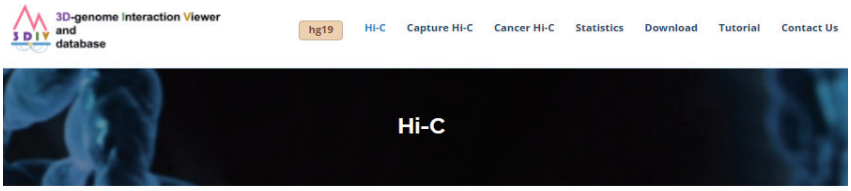
Example : Interaction profile of rs1421085

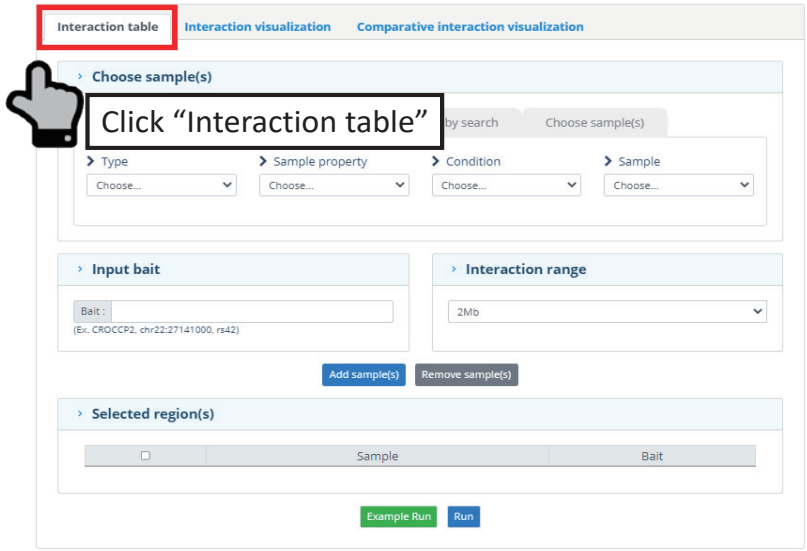
rs1421085 : an obesity variant in FTO gene intron region.
It is well characterized by significant interactions with IRX3 and IRX5 promoters.



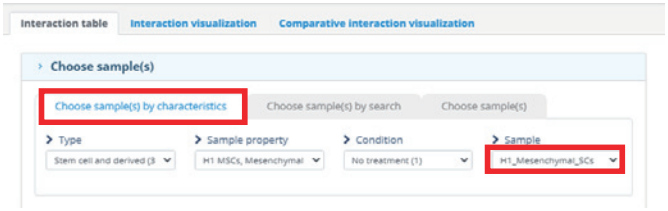
Rask-Andersen et al, Hum. Genet. (2015)

Step 1 : Open Interaction Table Module

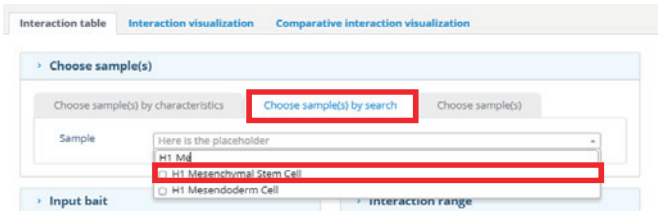




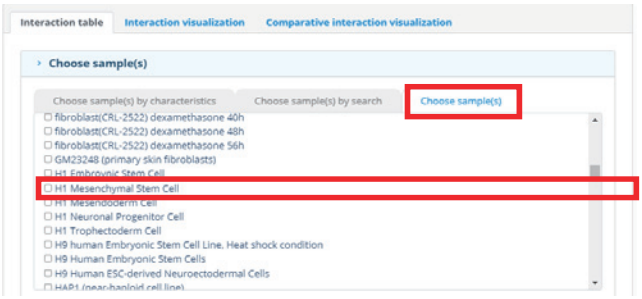
Step 2 : Choose a sample



1) Choose samples with condition

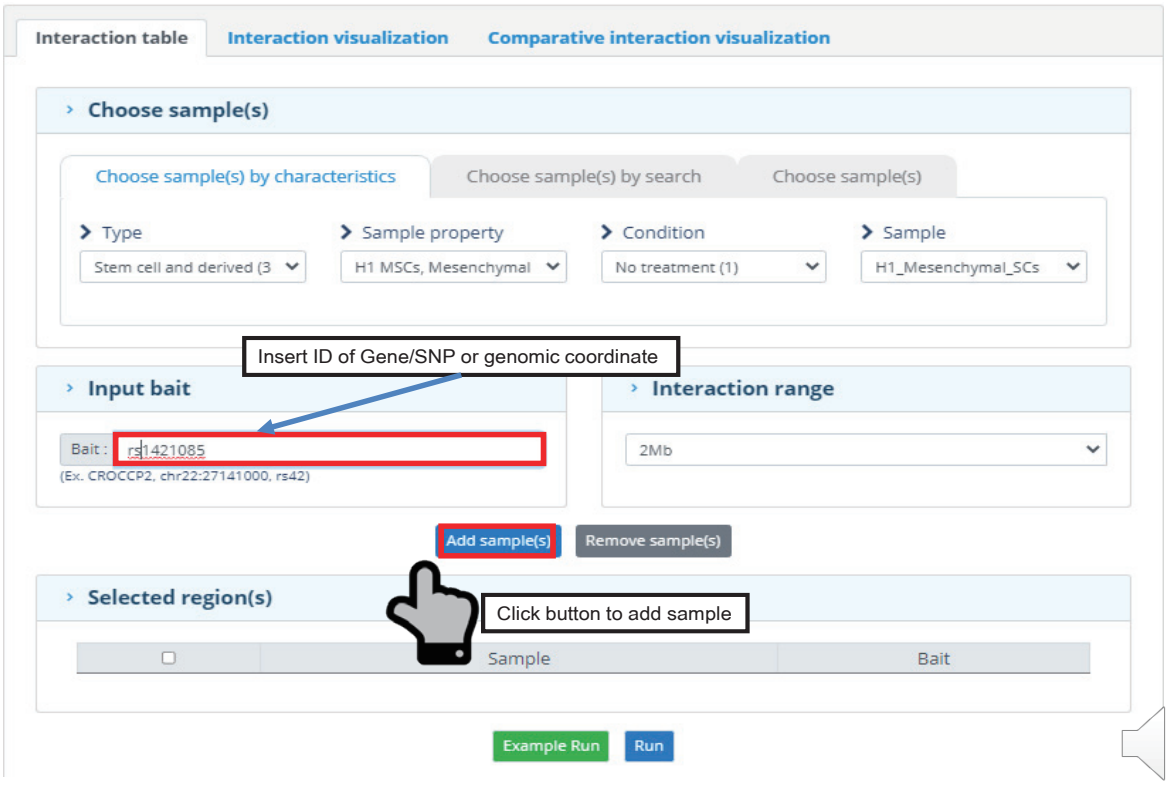


2) Choose samples with searching window

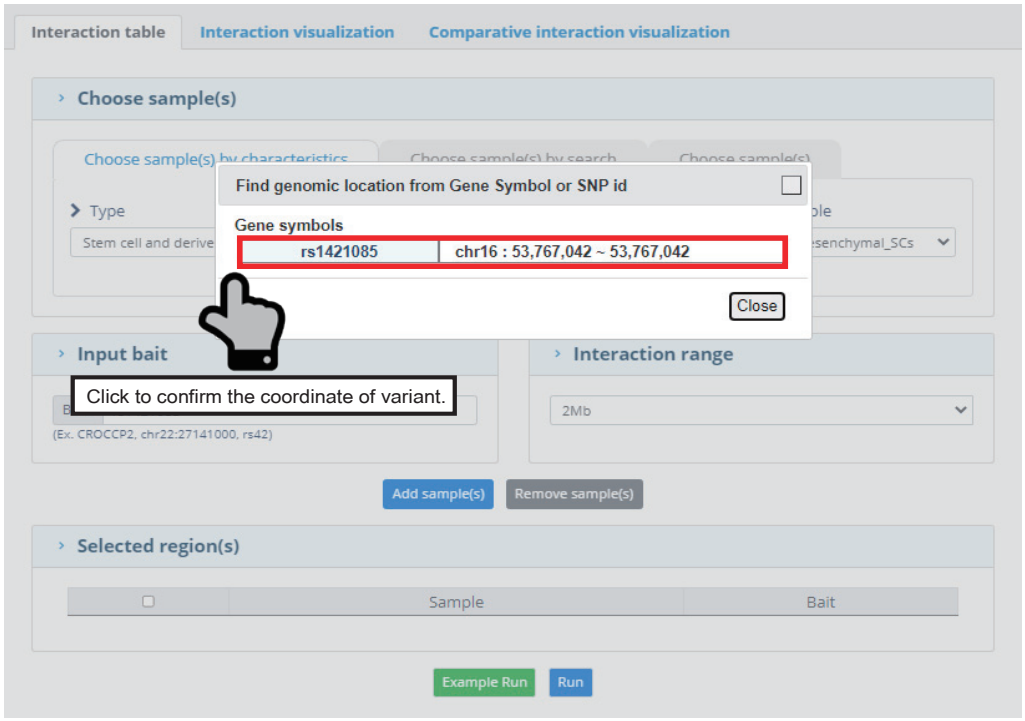


3) Choose samples from the list directly

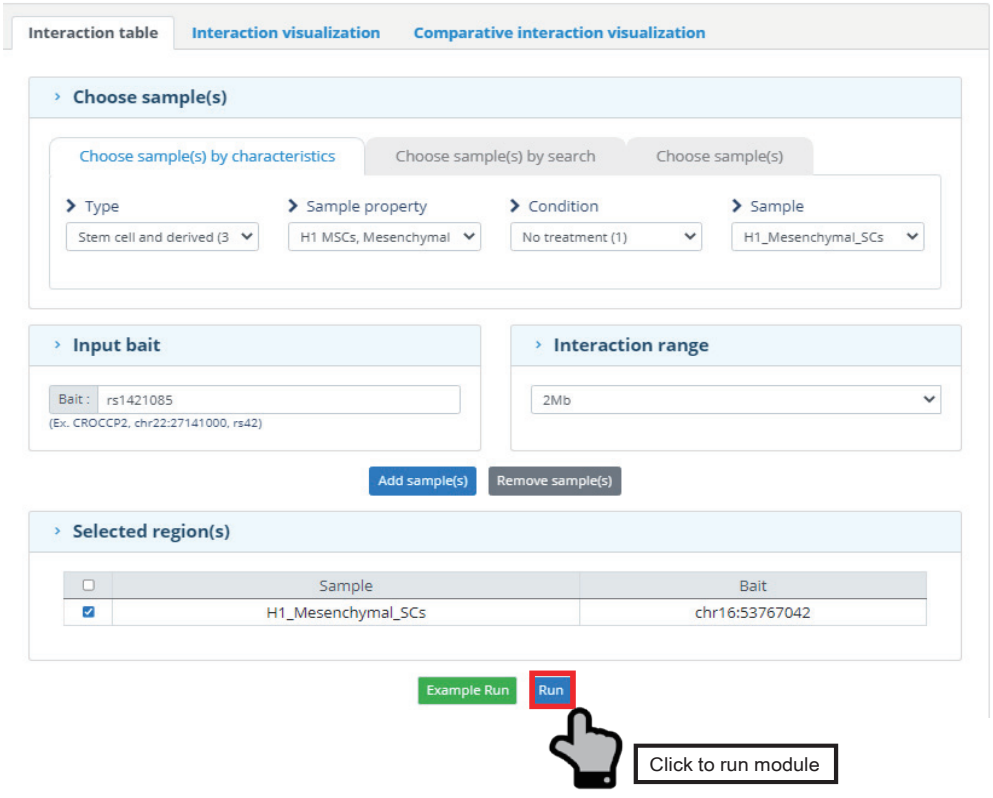
Step 3 : Choose a bait



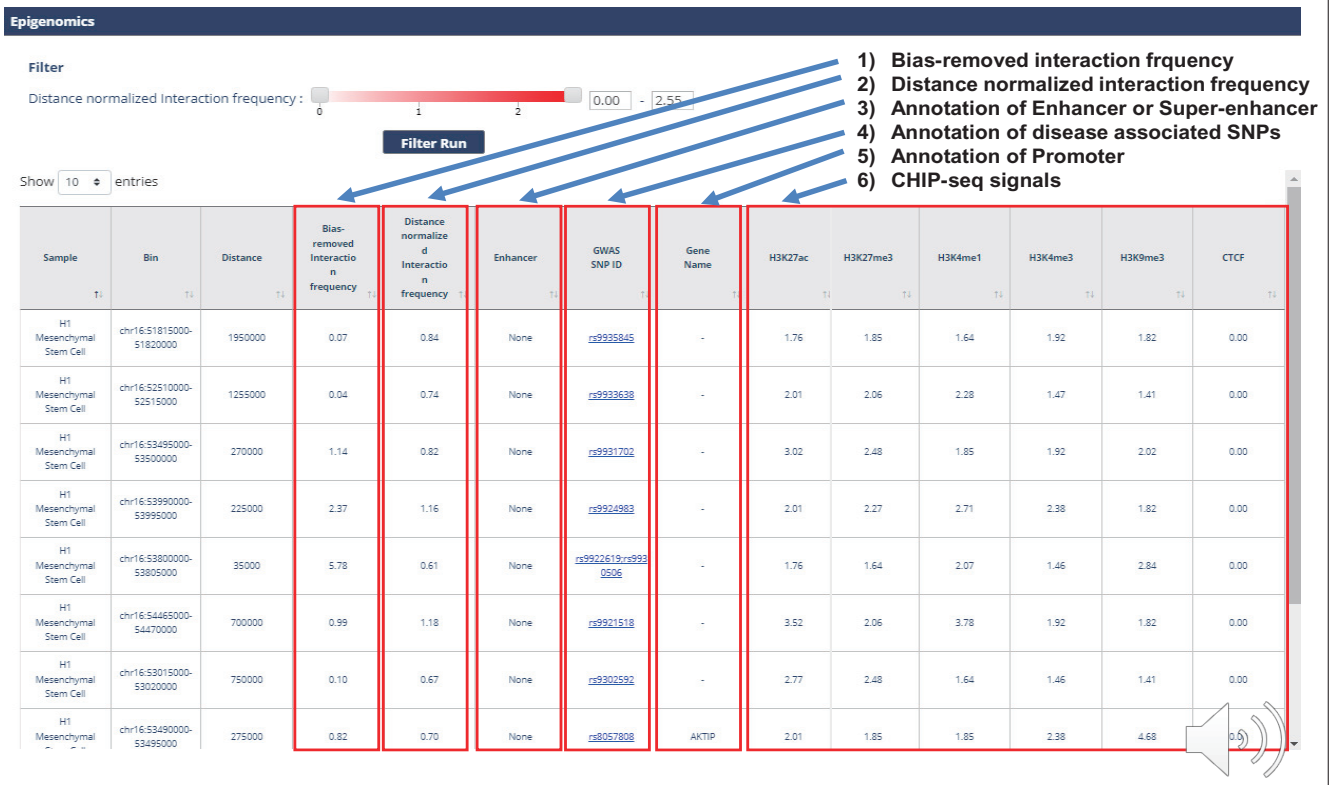
Step 3 : Choose a bait



Step 4 : Run Module

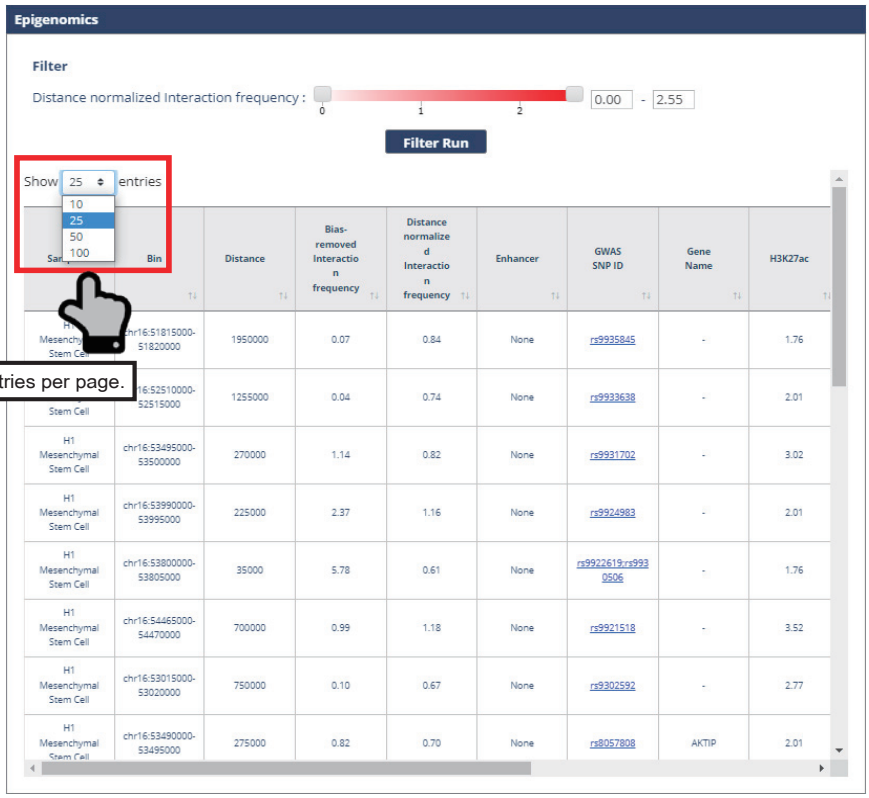


Step 5 : Browse the table



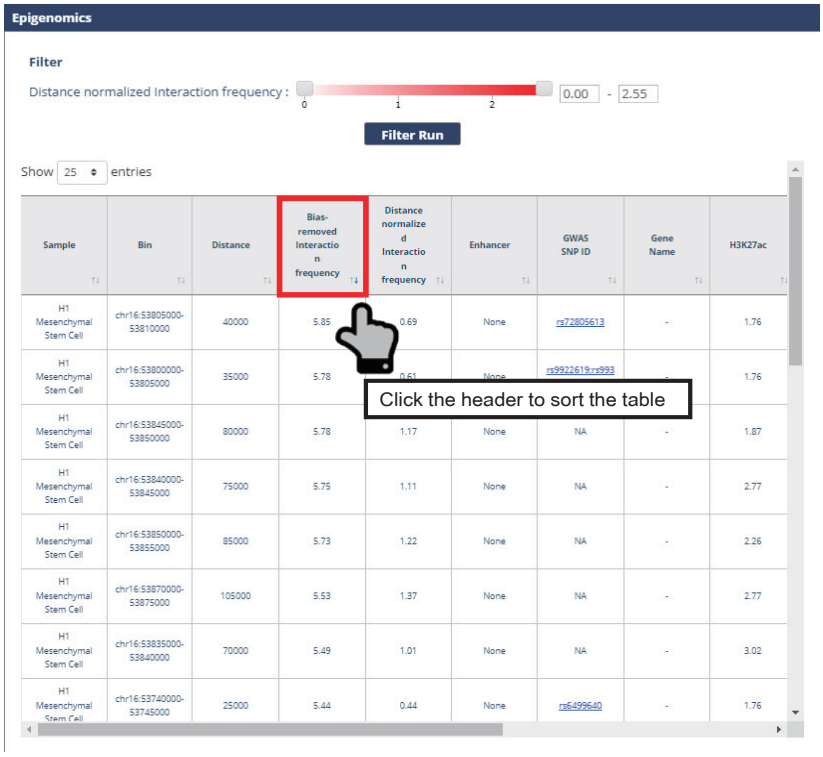
Sample	Bin	Distance	Bias-removed Interaction frequency	Distance normalized Interaction frequency	Enhancer	GWAS SNP ID	Gene Name	H3K27ac	H3K27me3	H3K4me1	H3K4me3	H3K9me3	CTCF
H1 Mesenchymal Stem Cell	chr16:51815000-51820000	1950000	0.07	0.84	None	rs9935845	-	1.76	1.85	1.64	1.92	1.82	0.00
H1 Mesenchymal Stem Cell	chr16:52510000-52515000	1255000	0.04	0.74	None	rs9933638	-	2.01	2.06	2.28	1.47	1.41	0.00
H1 Mesenchymal Stem Cell	chr16:53495000-53500000	270000	1.14	0.82	None	rs9931702	-	3.02	2.48	1.85	1.92	2.02	0.00
H1 Mesenchymal Stem Cell	chr16:53990000-53995000	225000	2.37	1.16	None	rs9924983	-	2.01	2.27	2.71	2.38	1.82	0.00
H1 Mesenchymal Stem Cell	chr16:53800000-53805000	35000	5.78	0.61	None	rs9922619:rs9930506	-	1.76	1.64	2.07	1.46	2.84	0.00
H1 Mesenchymal Stem Cell	chr16:54465000-54470000	700000	0.99	1.18	None	rs9921518	-	3.52	2.06	3.78	1.92	1.82	0.00
H1 Mesenchymal Stem Cell	chr16:53015000-53020000	750000	0.10	0.67	None	rs9302592	-	2.77	2.48	1.64	1.46	1.41	0.00
H1 Mesenchymal Stem Cell	chr16:53490000-53495000	275000	0.82	0.70	None	rs8057808	AKTIP	2.01	1.85	1.85	2.38	4.68	0.0

Step 5a : Adjust the table



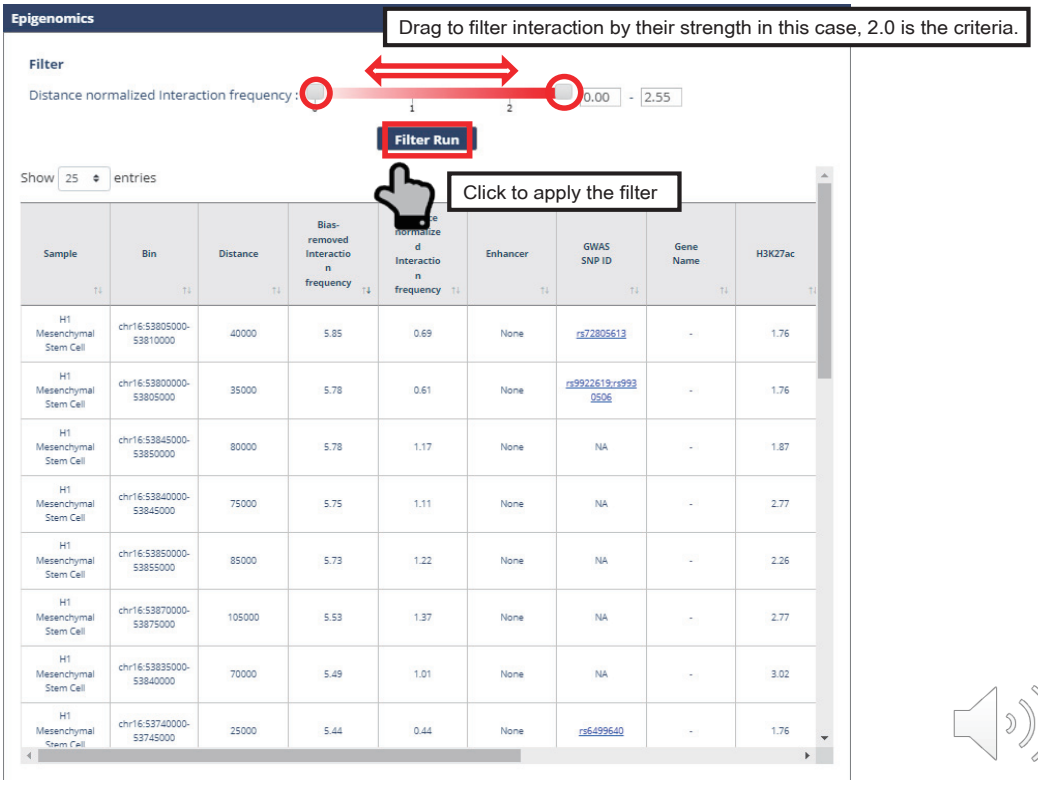
Adjust the number of entries per page.

Step 5b : Sort the interaction table



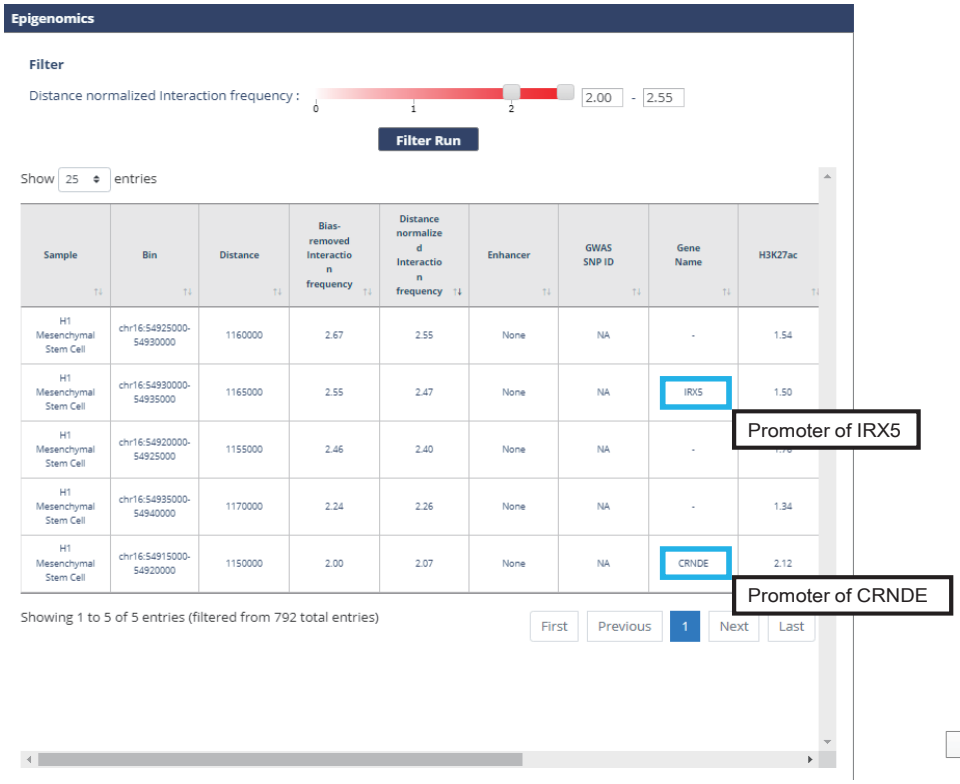
Sample	Bin	Distance	Bias-removed Interaction frequency	Distance normalized Interaction frequency	Enhancer	GWAS SNP ID	Gene Name	H3K27ac
H1 Mesenchymal Stem Cell	chr16:53805000-53810000	40000	5.85	0.69	None	rs72805613	-	1.76
H1 Mesenchymal Stem Cell	chr16:53800000-53805000	35000	5.78	0.61	None	rs9922619;rs993	-	1.76
H1 Mesenchymal Stem Cell	chr16:53845000-53850000	80000	5.78	1.17	None	NA	-	1.87
H1 Mesenchymal Stem Cell	chr16:53840000-53845000	75000	5.75	1.11	None	NA	-	2.77
H1 Mesenchymal Stem Cell	chr16:53850000-53855000	85000	5.73	1.22	None	NA	-	2.26
H1 Mesenchymal Stem Cell	chr16:53870000-53875000	105000	5.53	1.37	None	NA	-	2.77
H1 Mesenchymal Stem Cell	chr16:53835000-53840000	70000	5.49	1.01	None	NA	-	3.02
H1 Mesenchymal Stem Cell	chr16:53740000-53745000	25000	5.44	0.44	None	rs6499640	-	1.76

Step 5c : Filter interaction

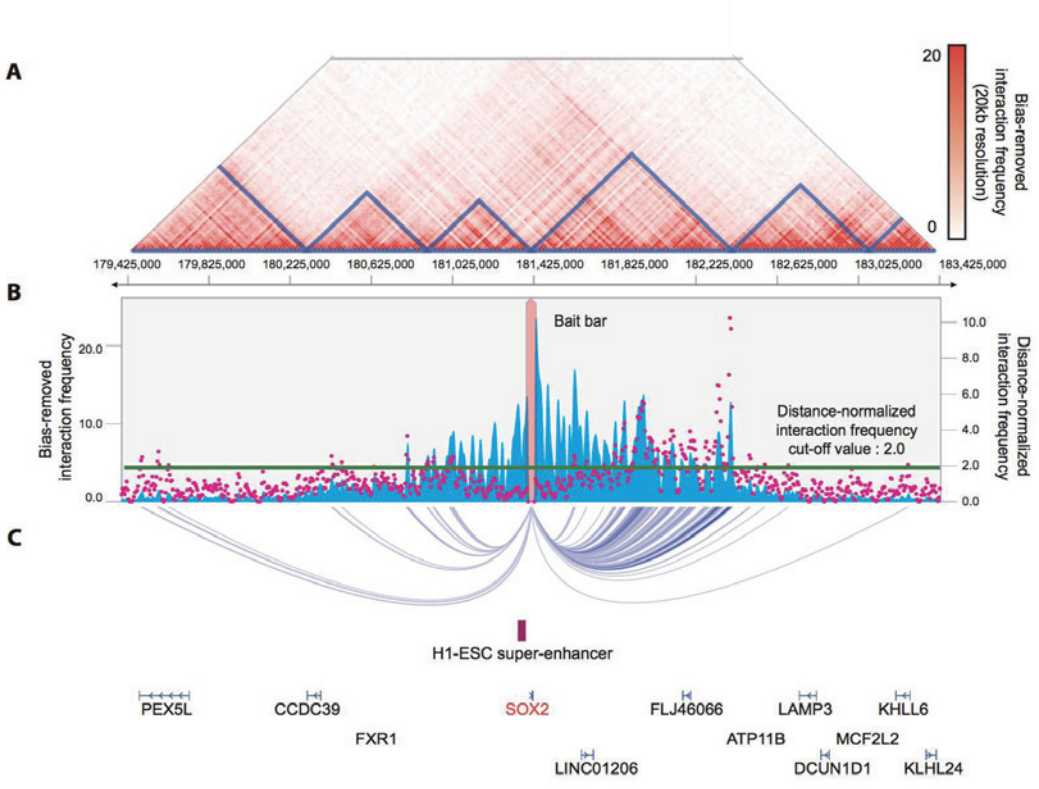


Sample	Bin	Distance	Bias-removed Interaction frequency	Distance normalized Interaction frequency	Enhancer	GWAS SNP ID	Gene Name	H3K27ac
H1 Mesenchymal Stem Cell	chr16:53805000-53810000	40000	5.85	0.69	None	rs72805613	-	1.76
H1 Mesenchymal Stem Cell	chr16:53800000-53805000	35000	5.78	0.61	None	rs9922619;rs9930506	-	1.76
H1 Mesenchymal Stem Cell	chr16:53845000-53850000	80000	5.78	1.17	None	NA	-	1.87
H1 Mesenchymal Stem Cell	chr16:53840000-53845000	75000	5.75	1.11	None	NA	-	2.77
H1 Mesenchymal Stem Cell	chr16:53850000-53855000	85000	5.73	1.22	None	NA	-	2.26
H1 Mesenchymal Stem Cell	chr16:53870000-53875000	105000	5.53	1.37	None	NA	-	2.77
H1 Mesenchymal Stem Cell	chr16:53835000-53840000	70000	5.49	1.01	None	NA	-	3.02
H1 Mesenchymal Stem Cell	chr16:53740000-53745000	25000	5.44	0.44	None	rs6499640	-	1.76

Step 5 : Browse the table

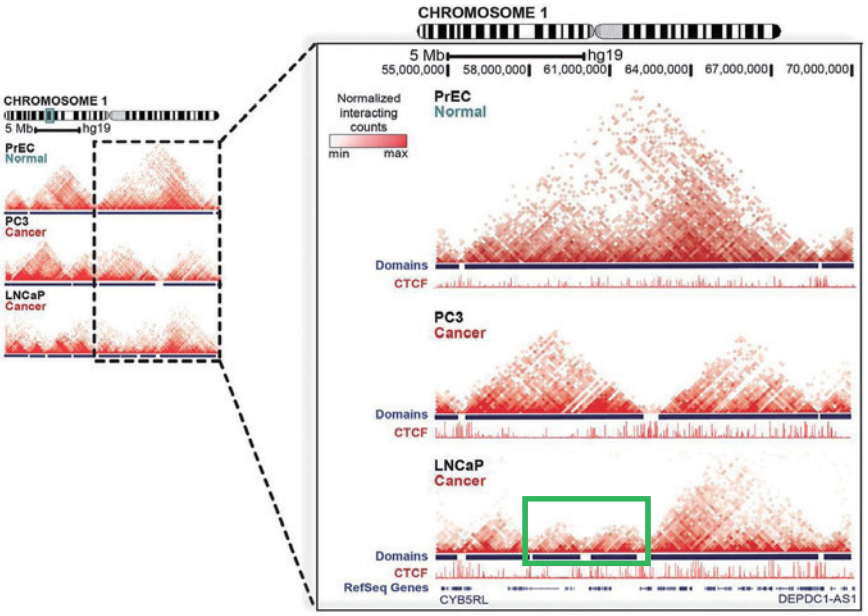


Module 2 : Interaction Visualization



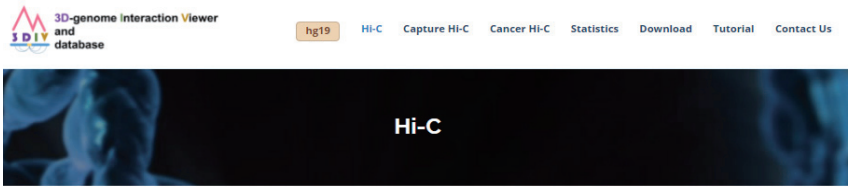
Example : Interaction profile of SOX2

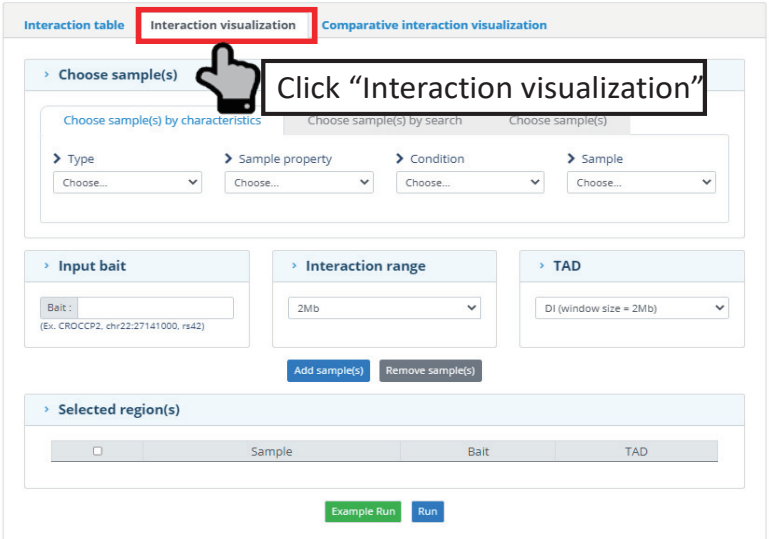
In cancer cells, the genomic structures are degraded into smaller sub-structures.
In this session, we will reproduce this result with 3DIV.



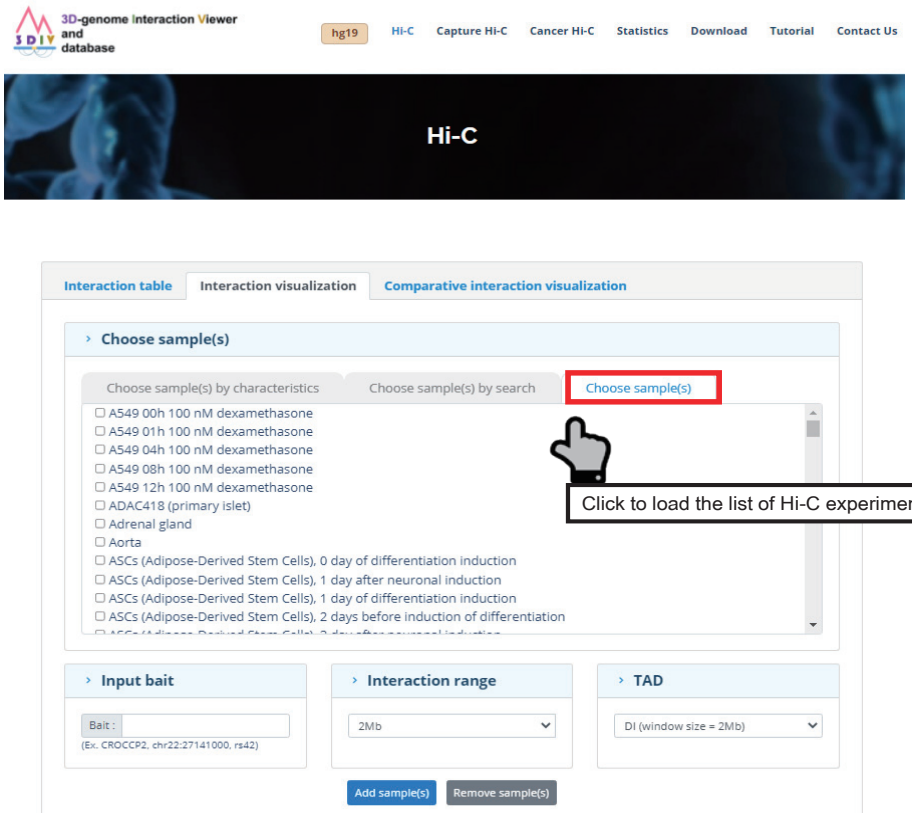
Taberlay et al, *Genome Res.* (2016)

Step 1 : Open Interaction Visualization Module

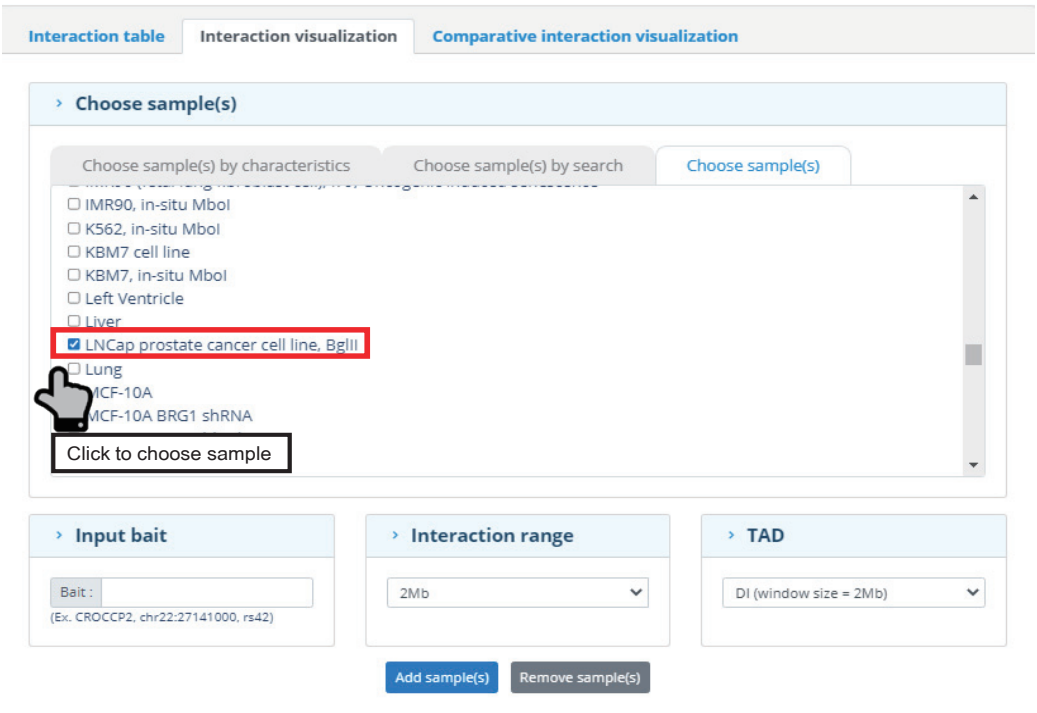




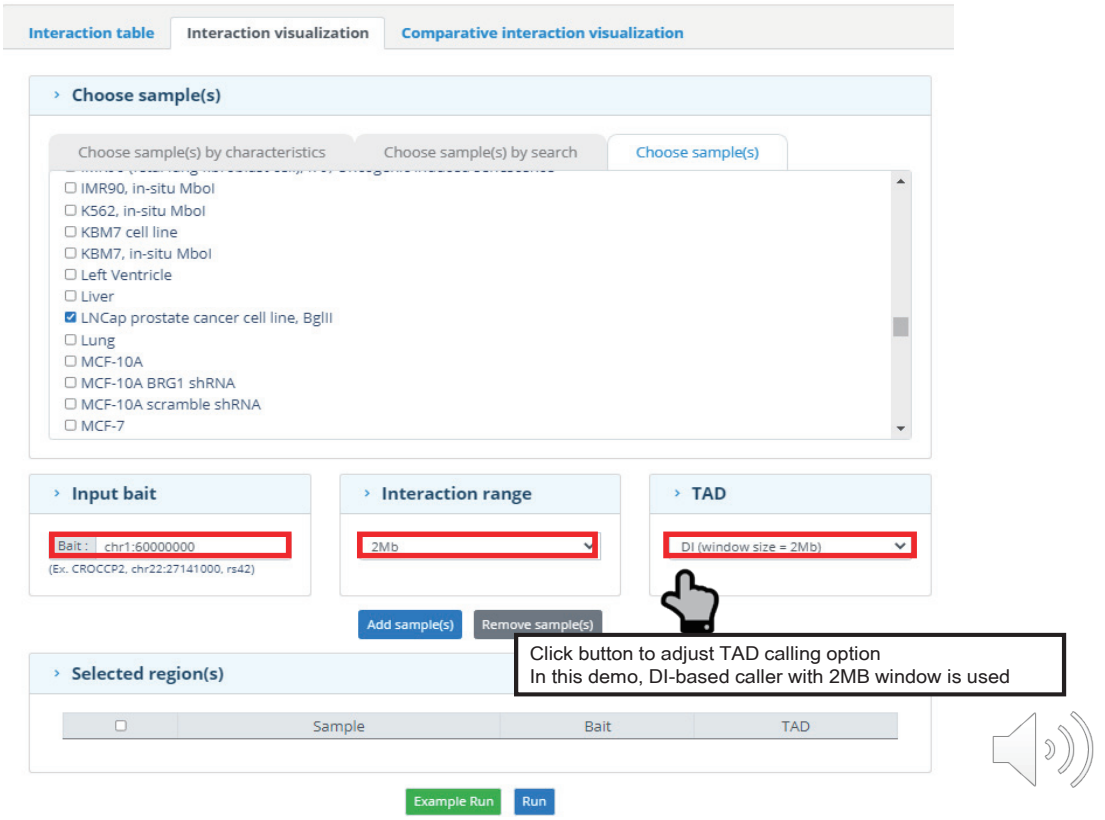
Step 2 : Choose a sample



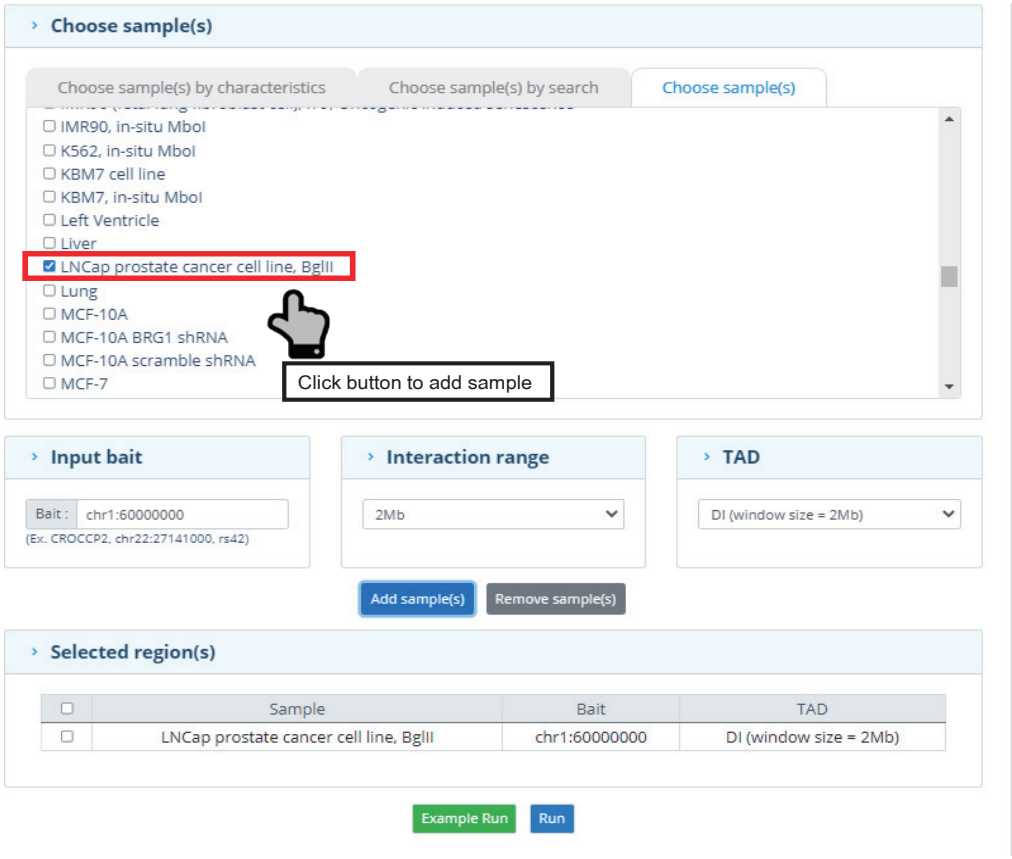
Step 2 : Choose a sample



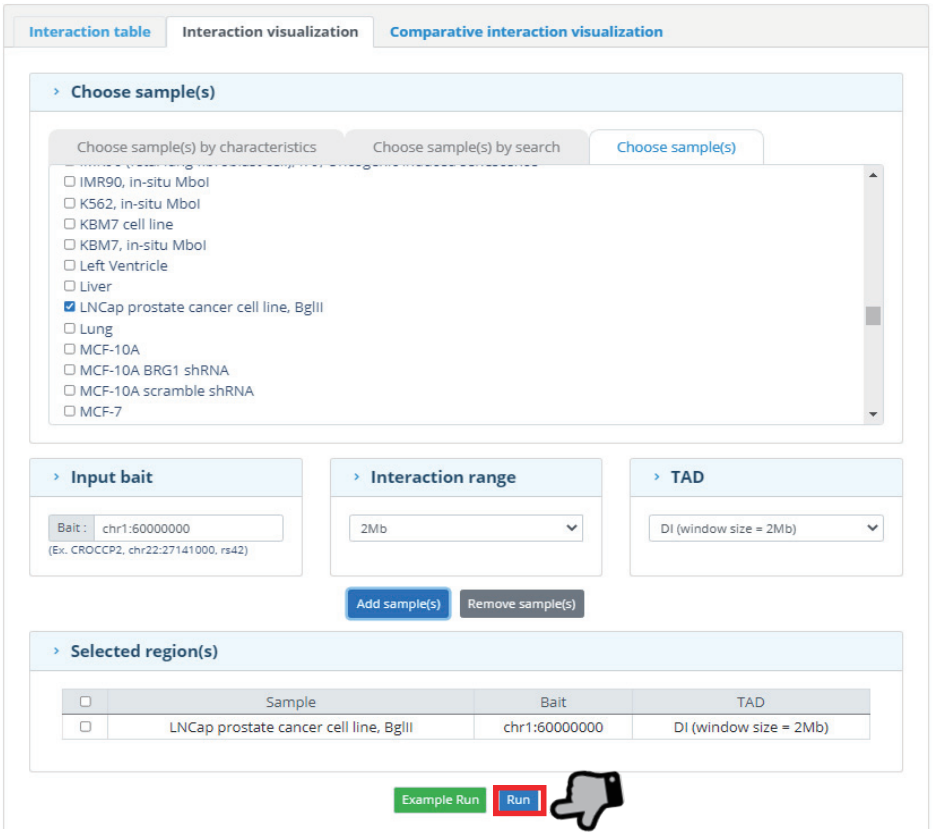
Step 3 : Choose a bait & TAD calling option



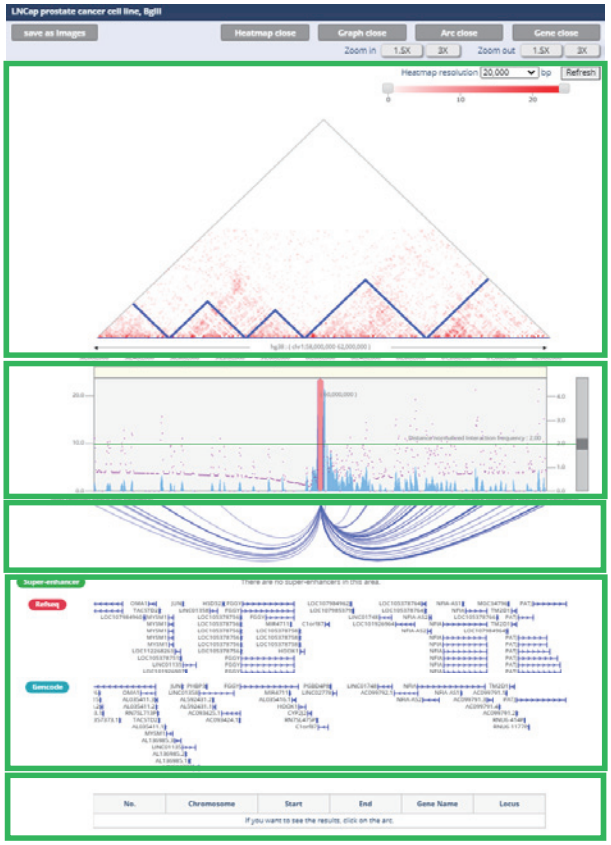
Step 3 : Choose a Bait & TAD calling option



Step 4 : Run Module



Step 5 : Adjust the interaction visualization



Interaction frequency heatmap with topologically associating domains(TAD) annotation.

One-to-all interaction plot

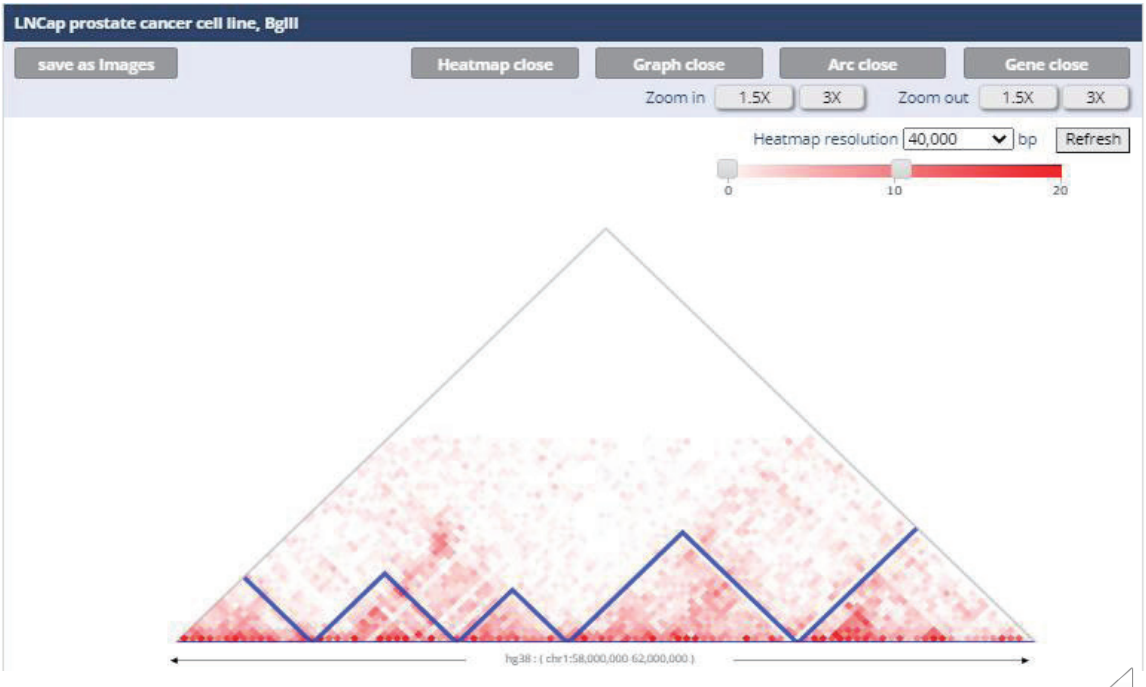
Arc-representation of significant interactions

Gene annotations

Description of selected interaction

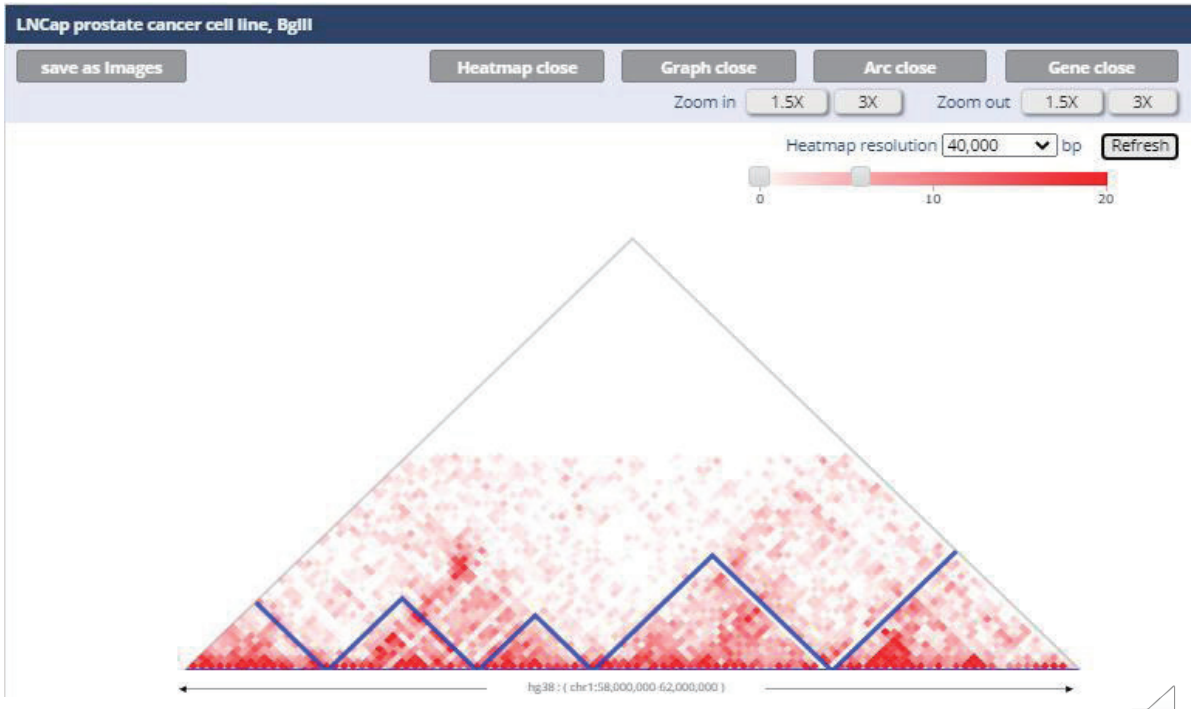
Step 5a : Adjust the heatmap resolution

Step 5a : Adjust the heatmap resolution



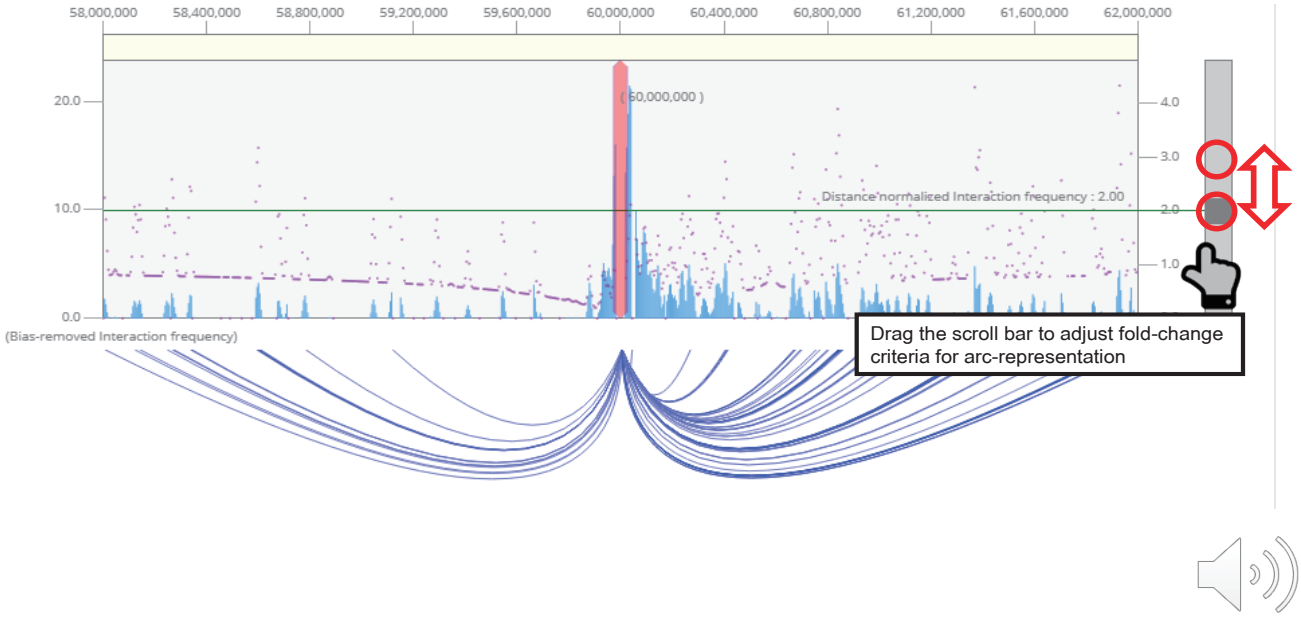
Step 5b : Adjust the heatmap color range

Step 5b : Adjust the heatmap color range

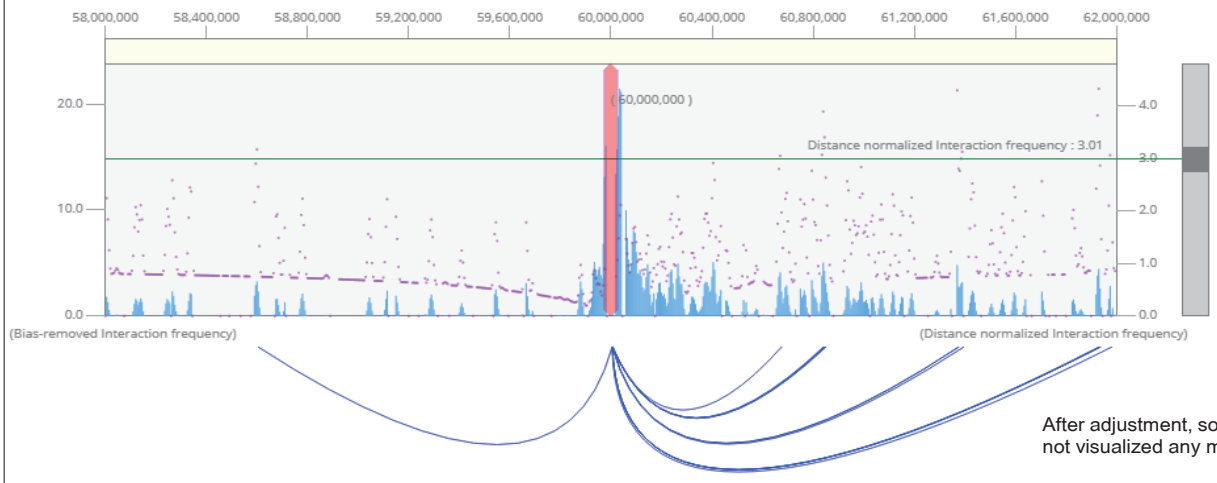


Step 5 : Adjust the interaction visualization

Adjust the fold-change criteria

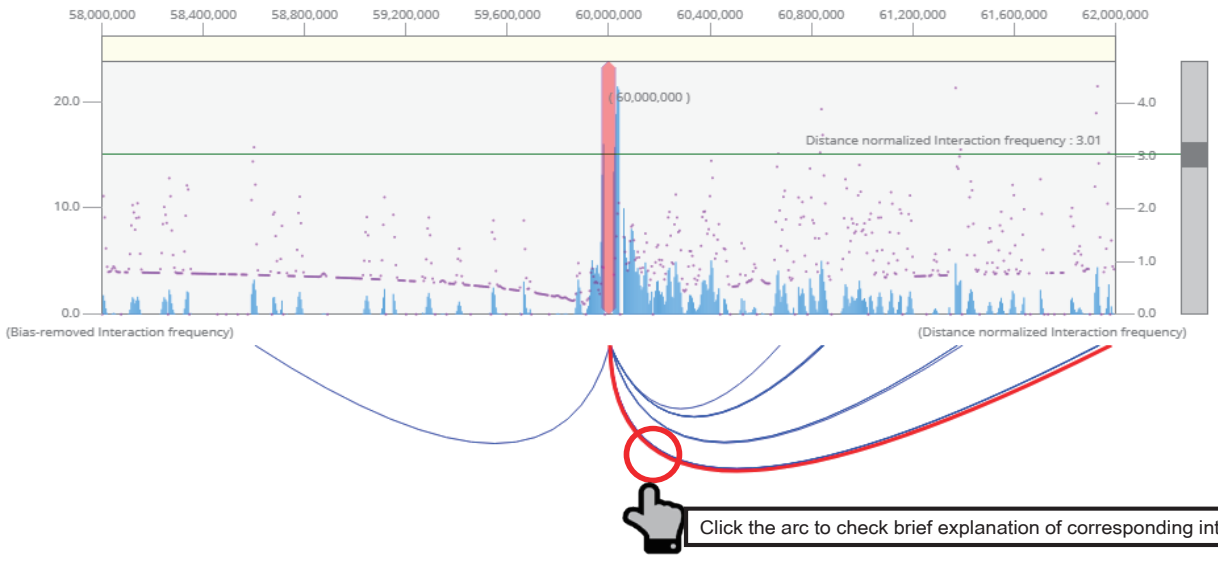


Adjust the fold-change criteria



After adjustment, some arcs are not visualized any more.

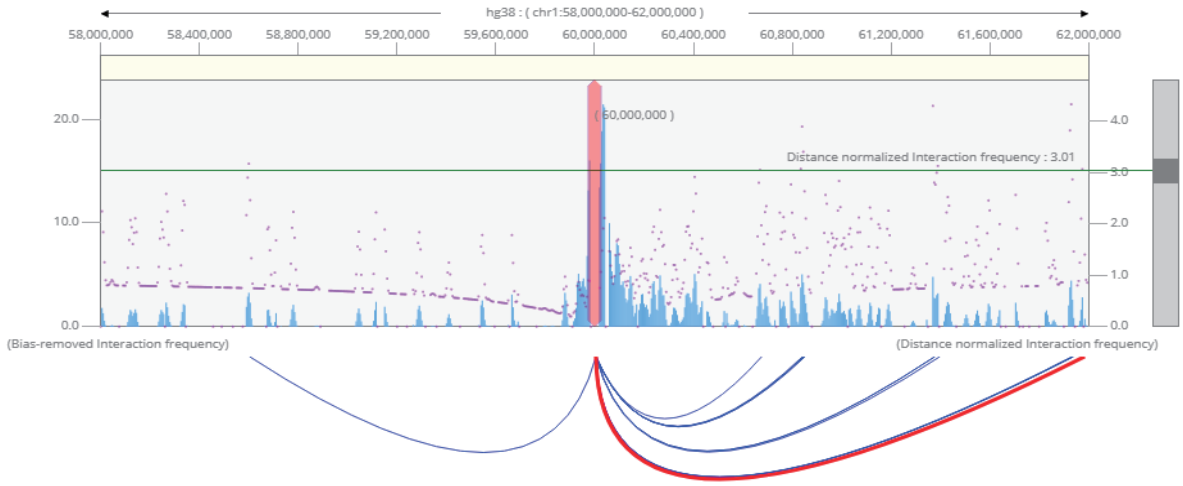
Description of identified interactions



Click the arc to check brief explanation of corresponding interaction

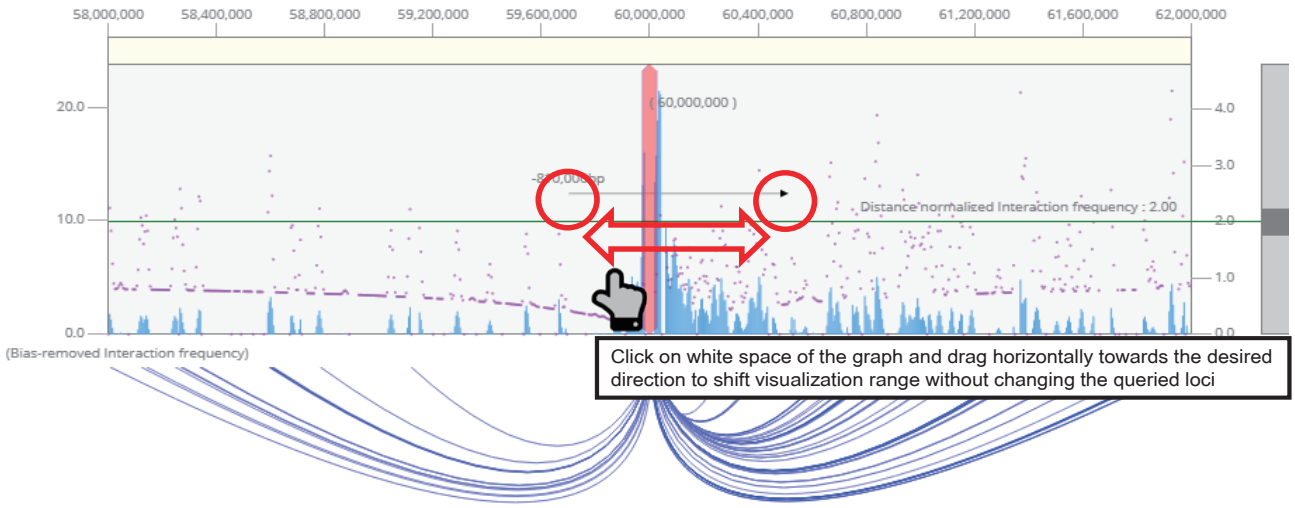
No.	Chromosome	Start	End	Gene Name	Locus
If you want to see the results, click on the arc.					

Description of identified interactions

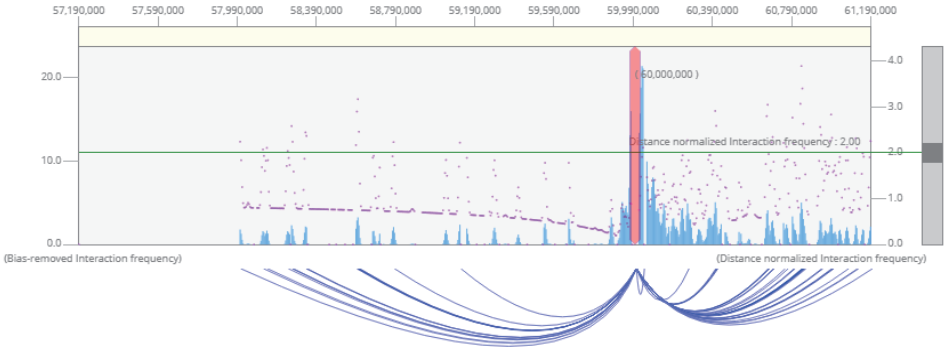


No.	Chromosome	Start	End	Gene Name	Locus
1	chr1	61975000	61980000		

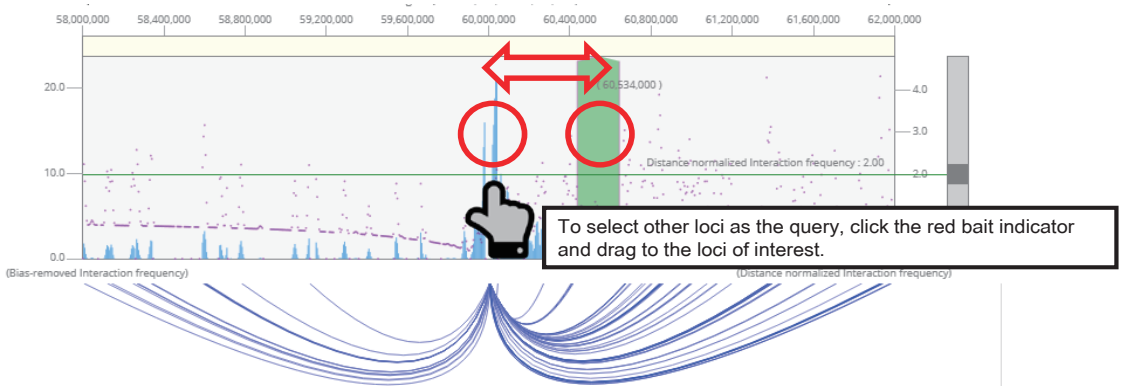
Browse interaction frequency w/o change the bait



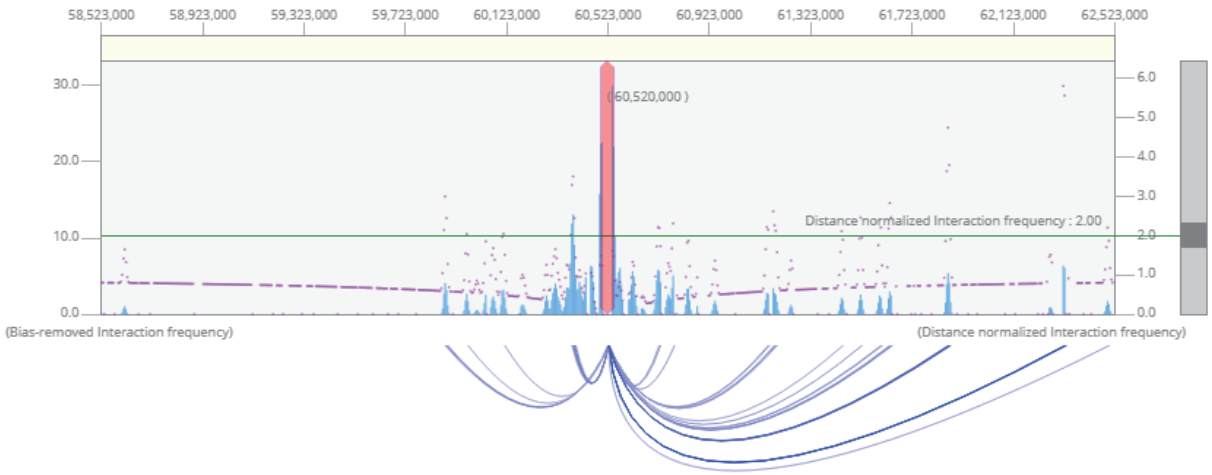
Browse interaction frequency w/o change the bait



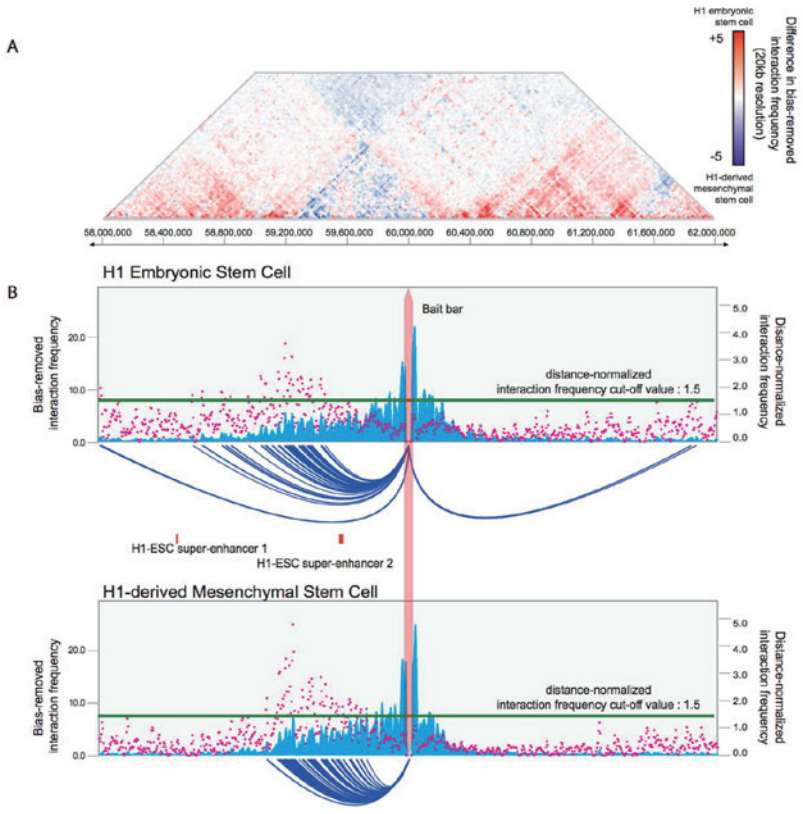
Adjust bait without resubmission



Adjust bait without resubmission

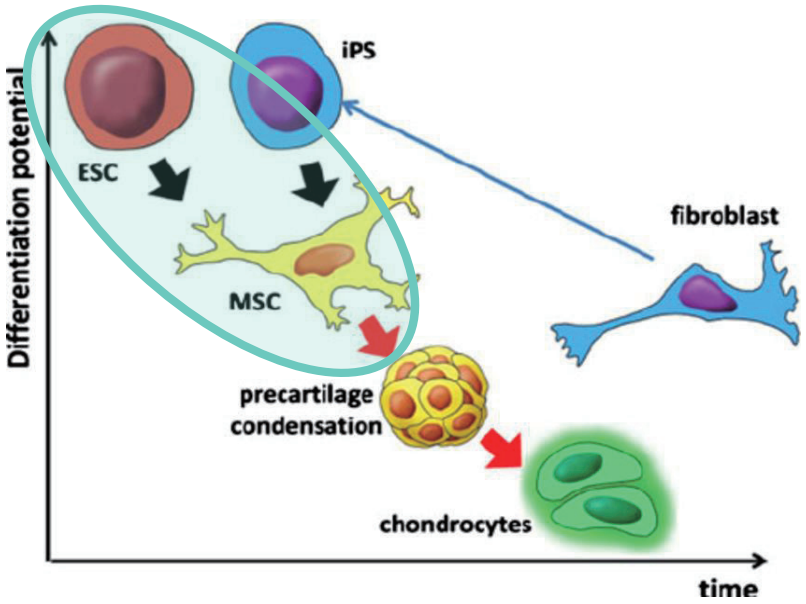


Module 3 : Comparative Visualization



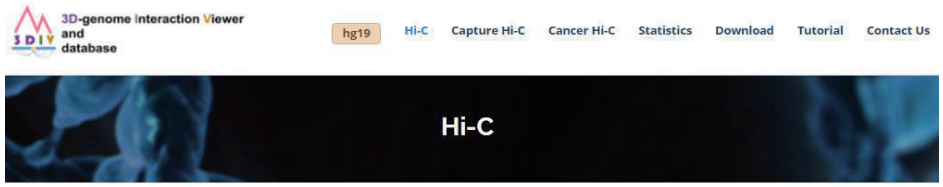
Example : Interaction change during differentiation

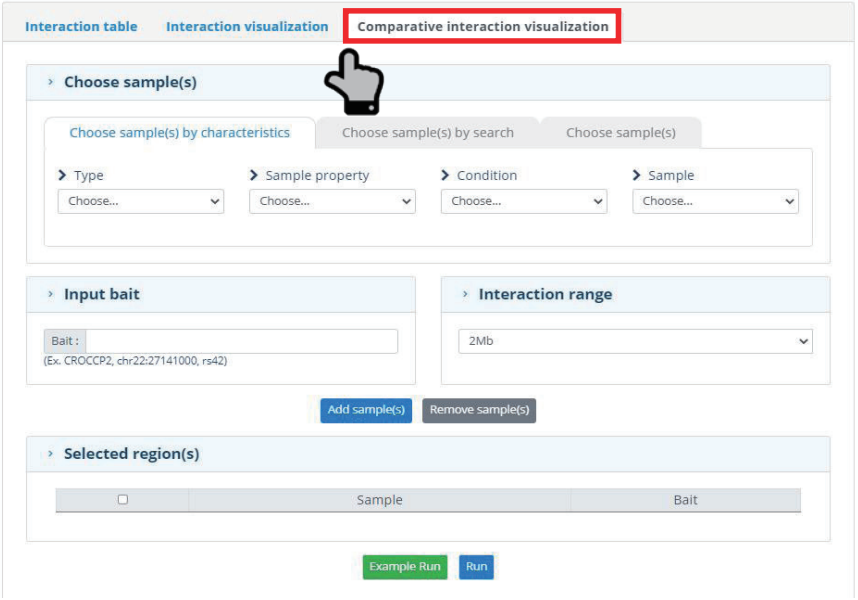
During the differentiation, the interaction profile is dramatically changed.
In this session, we will compare the interaction profile of ESC and MSC.
ESC : Embryonic Stem Cell, MSC : Mesenchymal Stem Cell



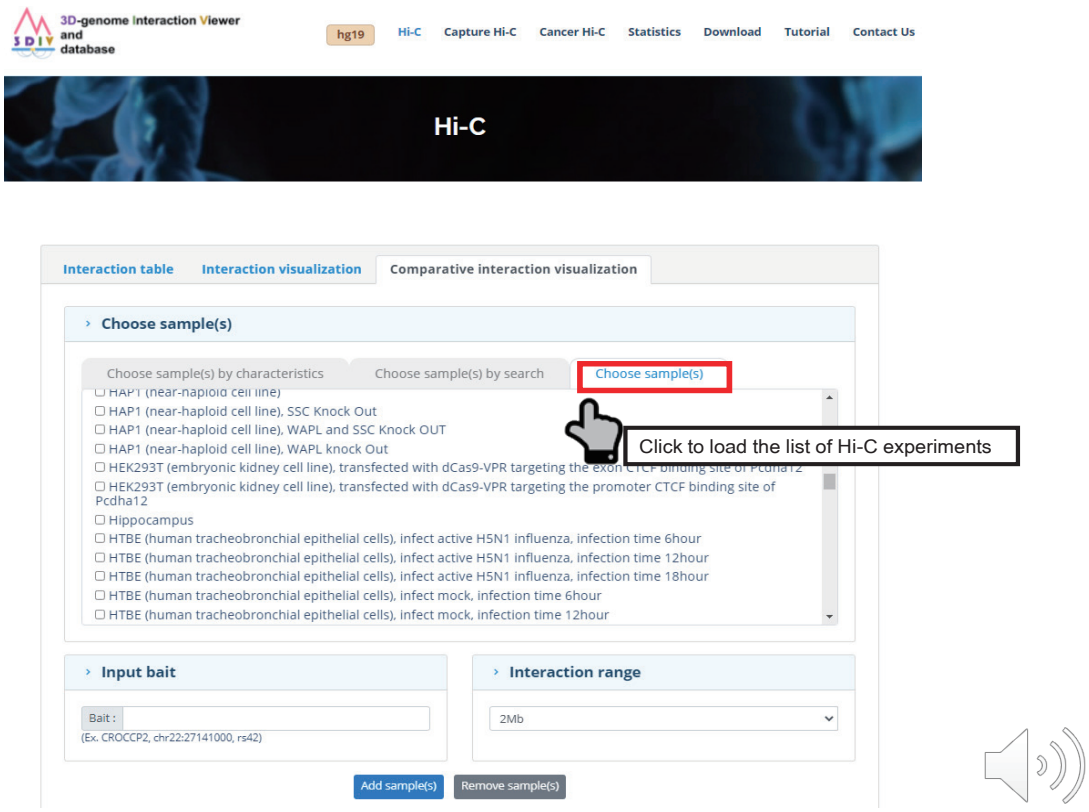
Gadjanski et al, Stem Cell Rev. Rep. (2012)

Step 1 : Open Comparative visualization Module

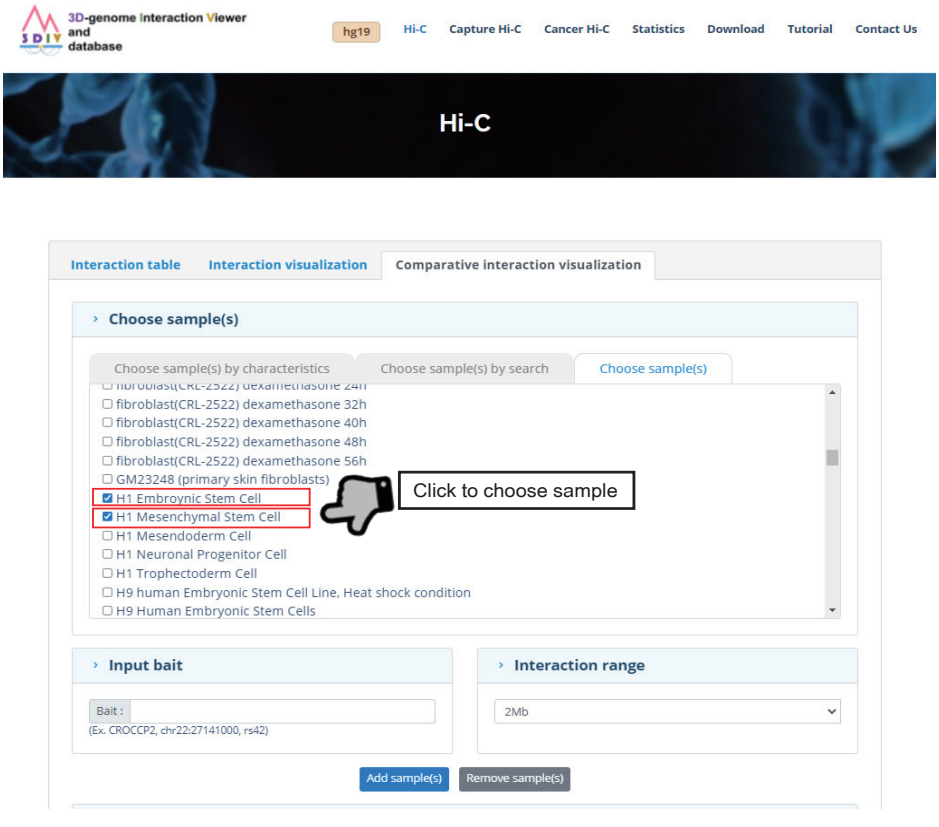




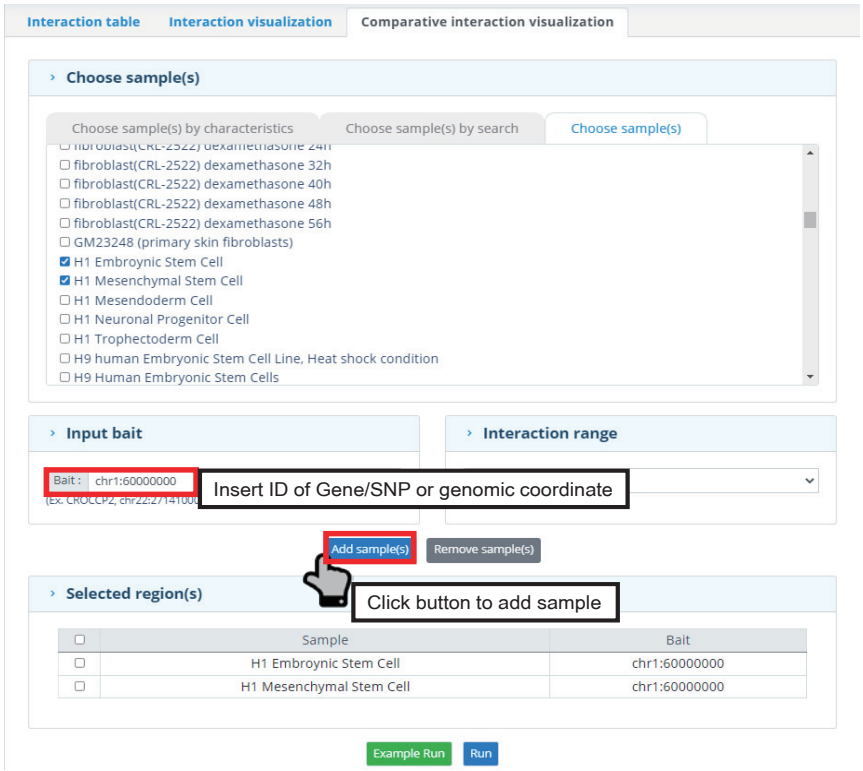
Step 2 : Choose a sample



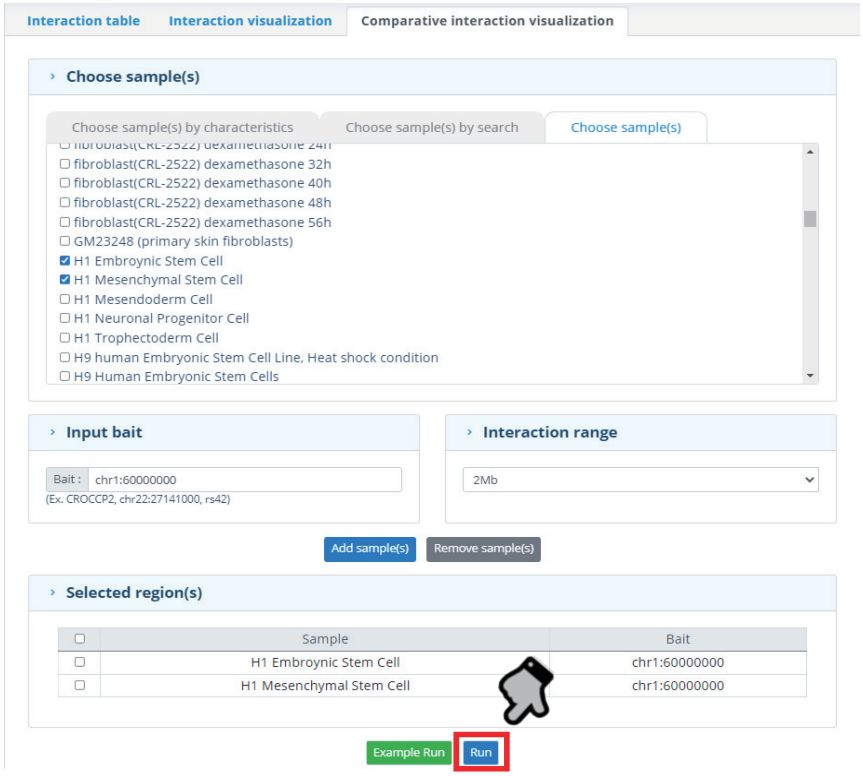
Step 2 : Choose a sample



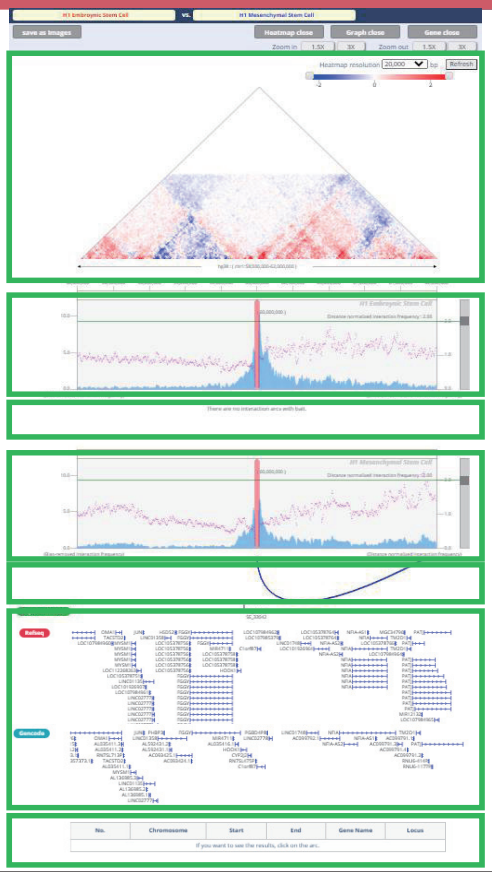
Step 3 : Choose a Bait



Step 4 : Run Module



Step 5 : Adjust comparative heatmap



Comparative heatmap of interaction frequency between 1st and 2nd samples.

Arc-representation of significant interactions in 1st sample

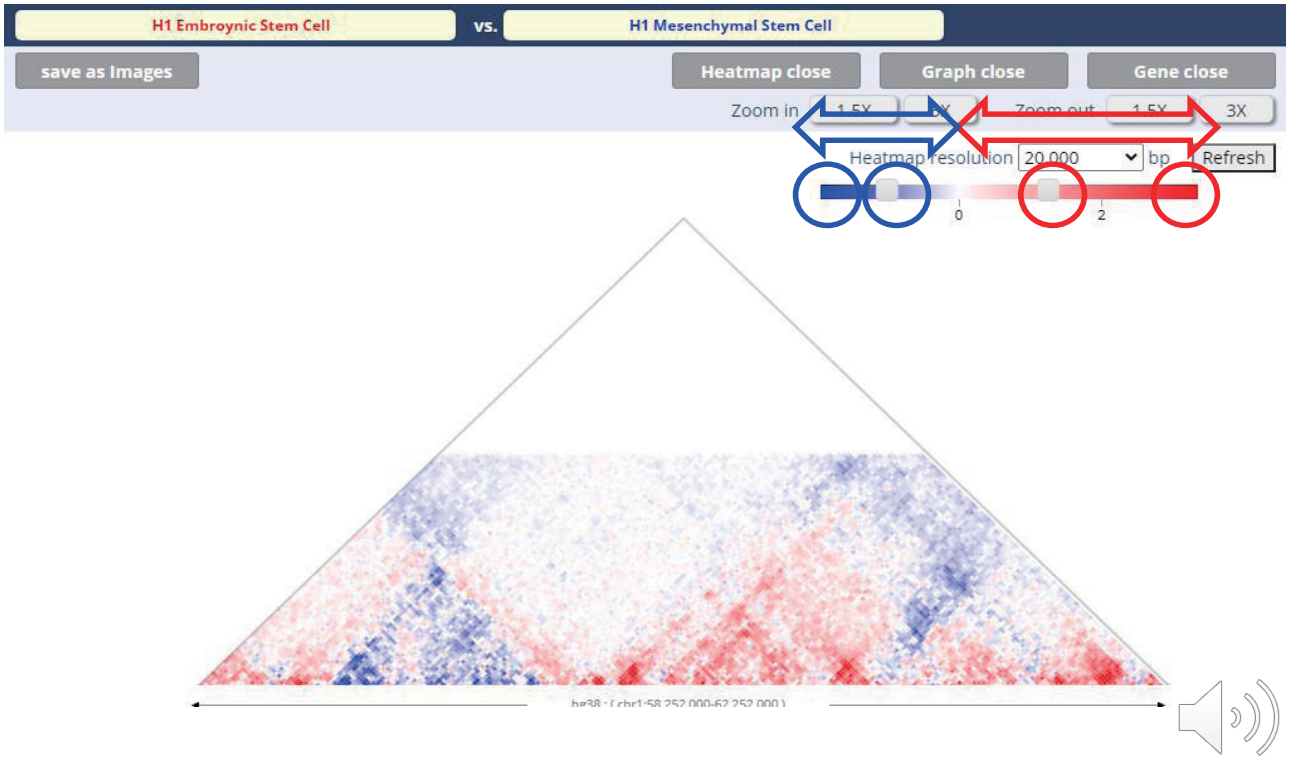
Arc-representation of significant interactions in 2nd sample

RefSeq Genes and super enhancer annotations

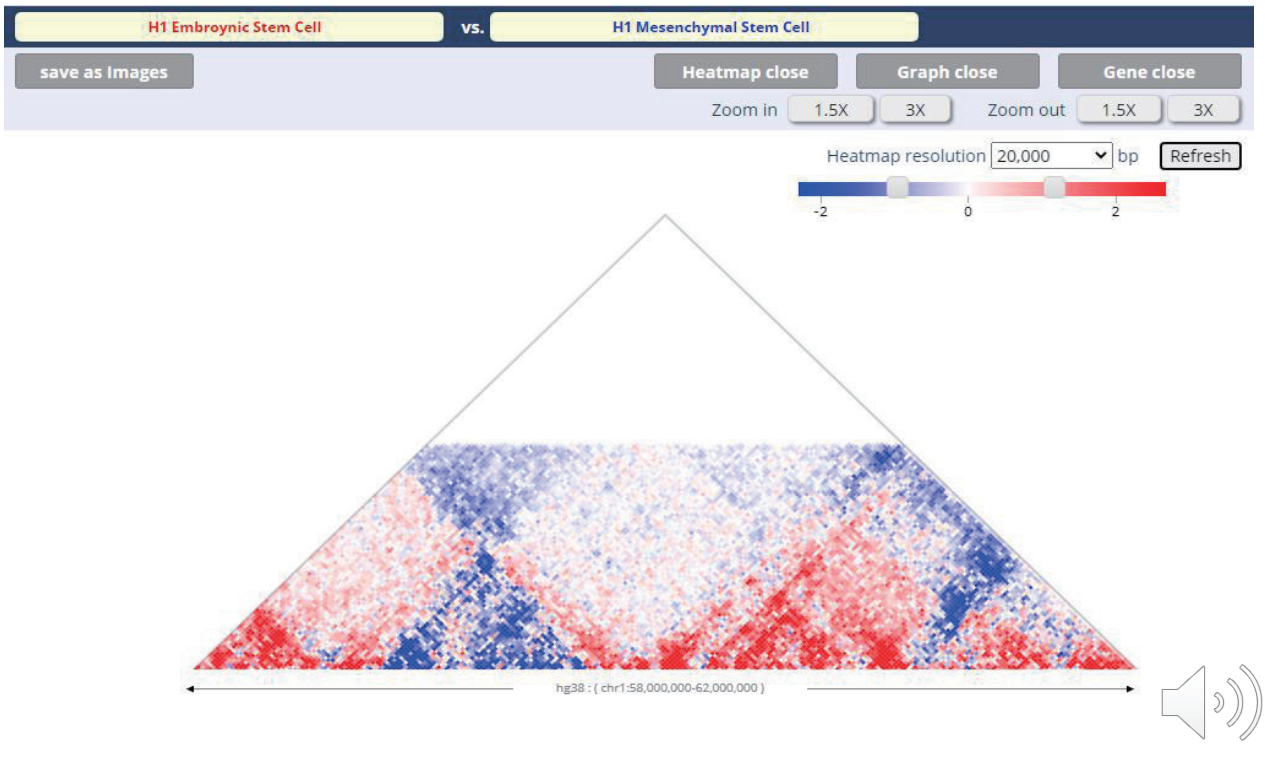
Description of selected interaction

Step 5a : Synchronized criteria change

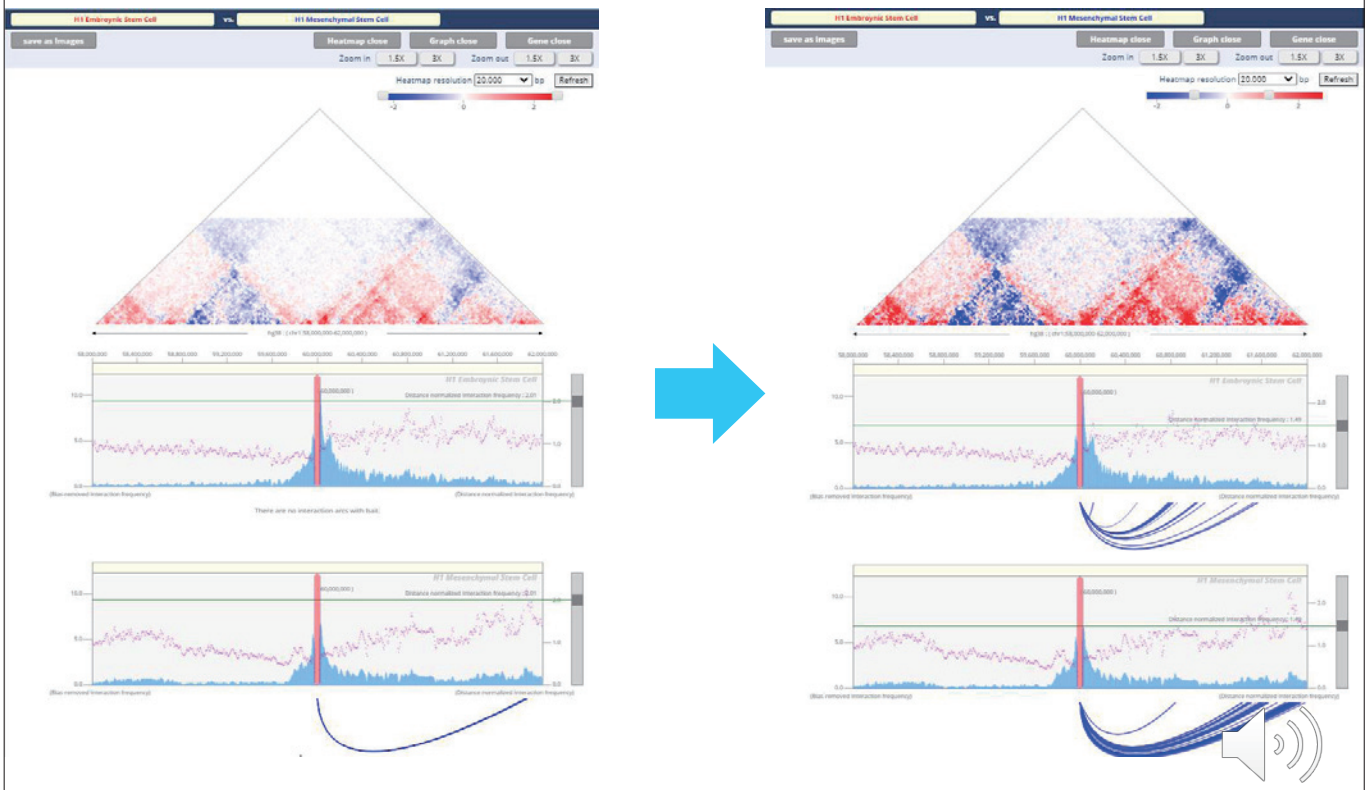
Step 5b : Adjust the heatmap color range



Step 5b : Adjust the heatmap color range

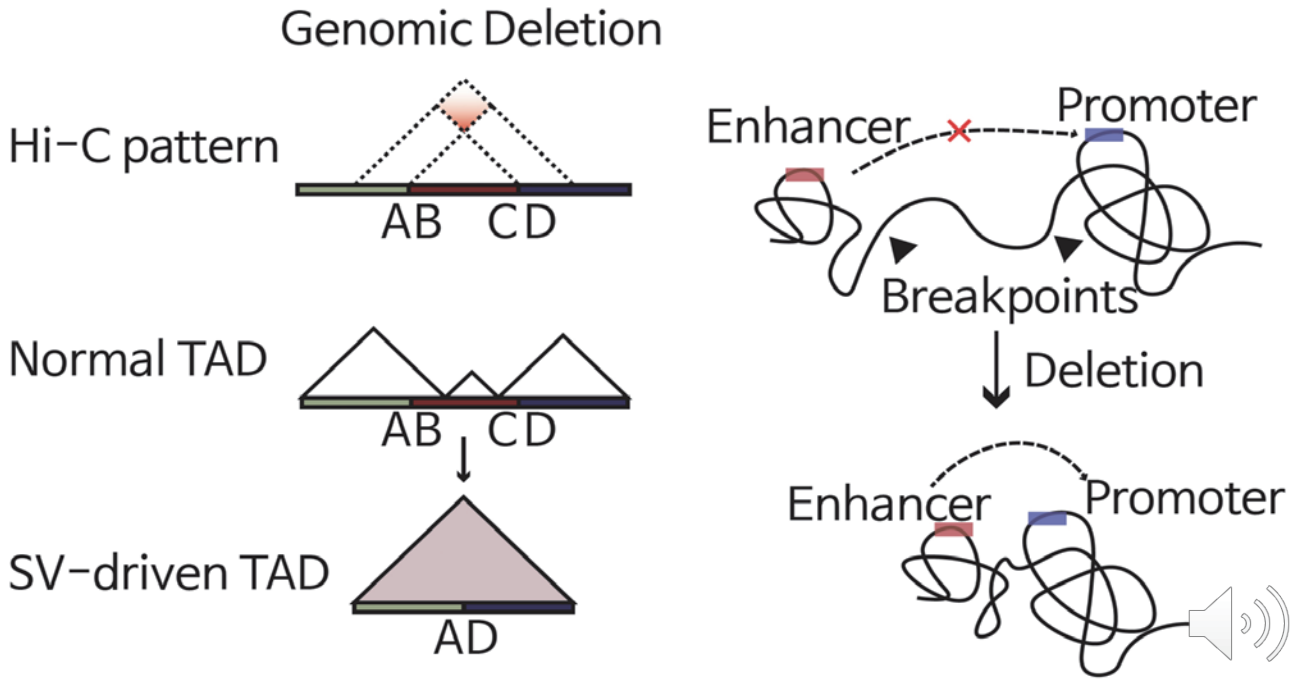


Step 5 : Adjust comparative heatmap



Cancer Hi-C Analysis

The impact of large scale structural variations to cancer 3D genome

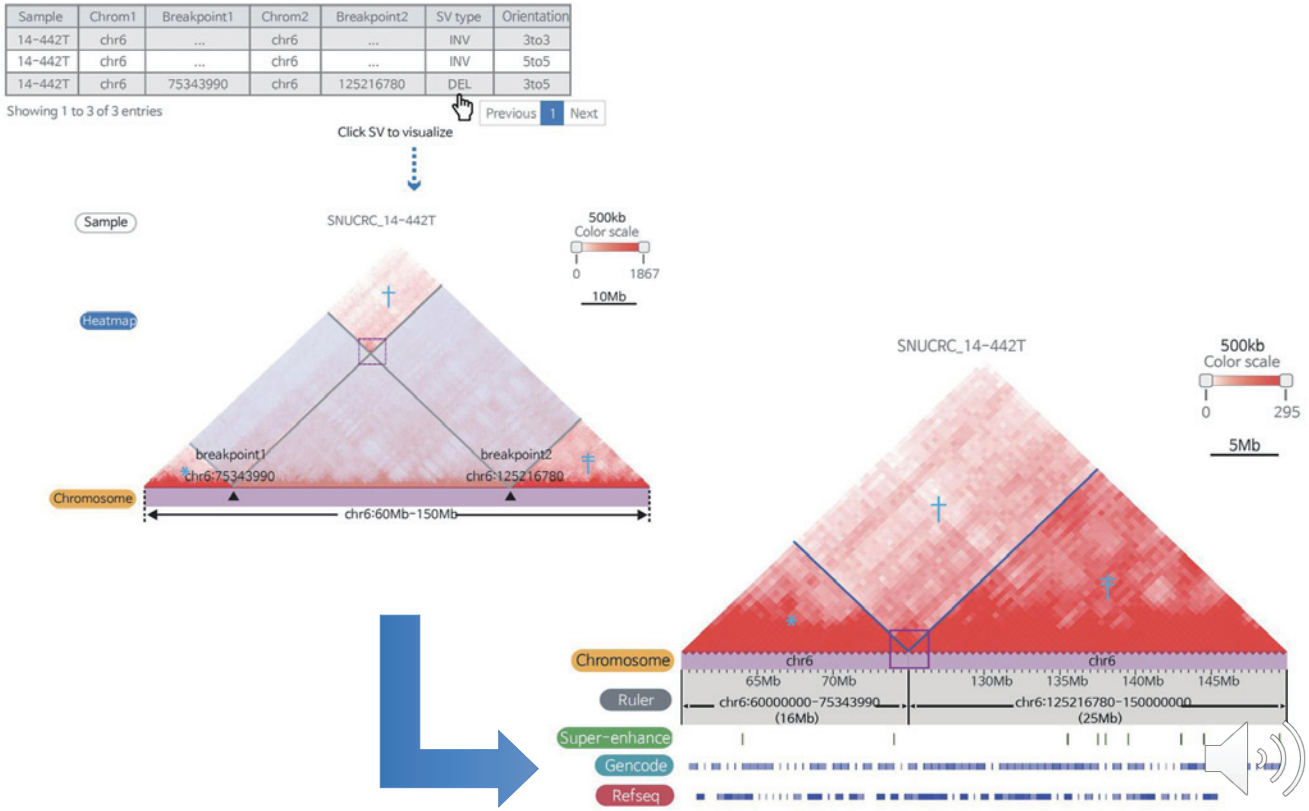


Interactively visualize and simulate the impact of structural variations to cancer 3D genome

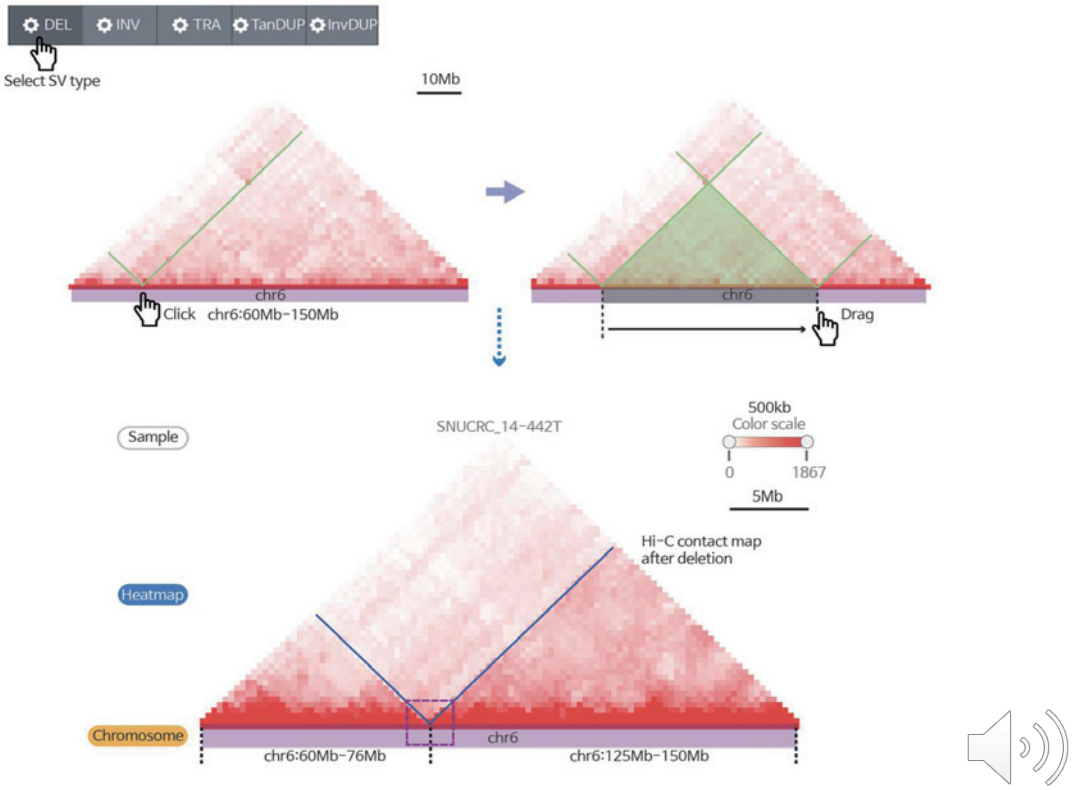
Problem statement
1. Frequent genomic rearrangements in cancer alters 3D genome
2. Abberant gene expression based on rewired regulatory elements
3. Requires appropriate visualization tools and processed data

Resolving issue
1. Collection of large cancer/normal Hi-C and pcHi-C data
2. Visualization of cancer 3D genome
3. Hi-C contact map manipulation to examine impact of SVs

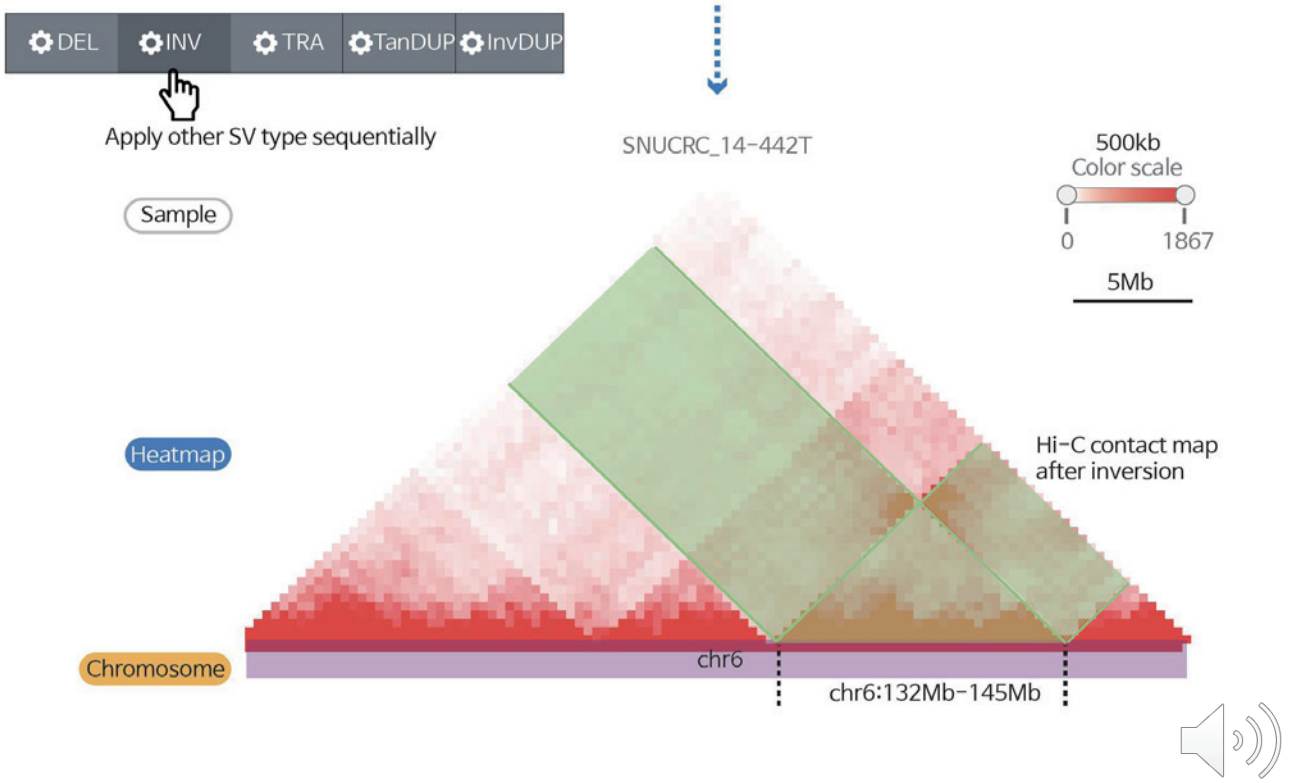
Module I. Pre-called SV and 3D genome



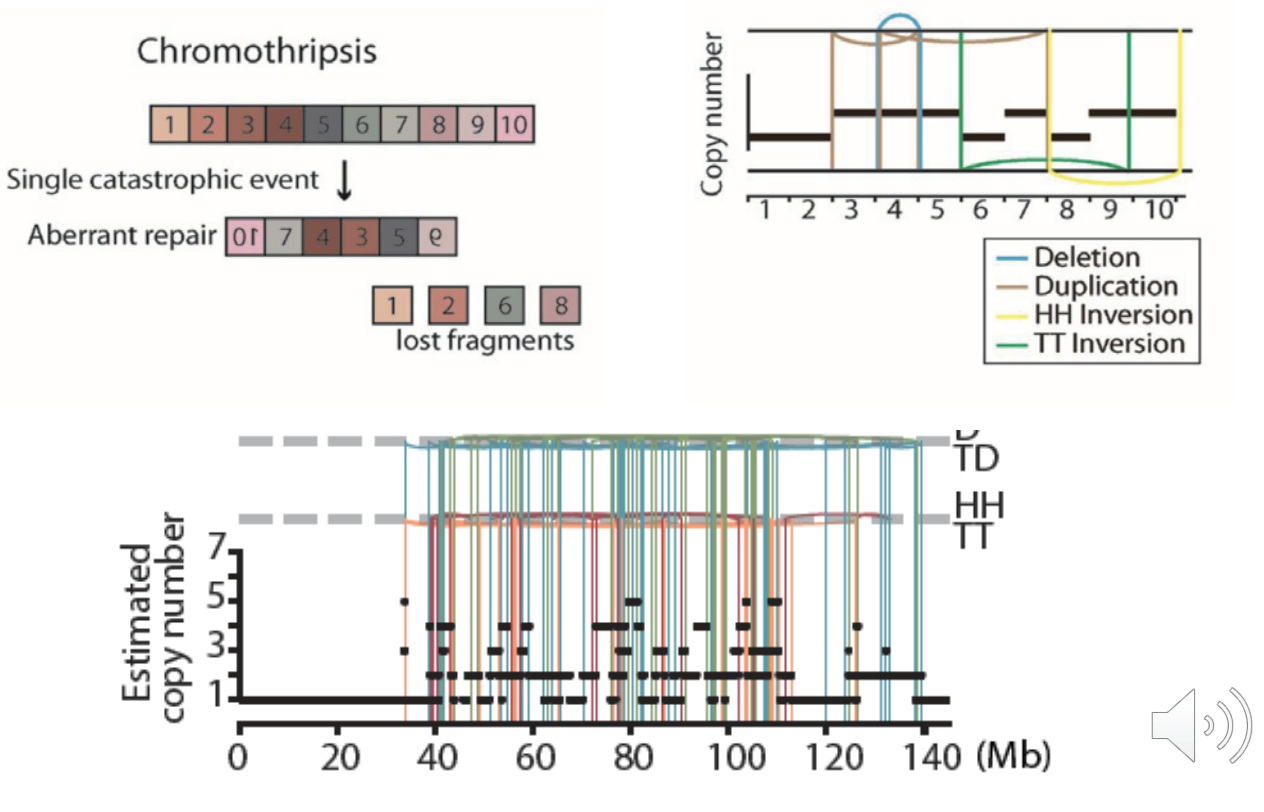
Module II. Interactive 3D genome manipulation



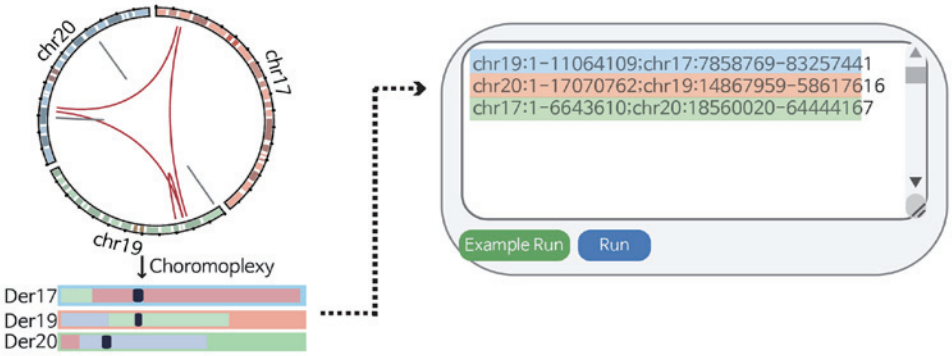
Module II. Interactive 3D genome manipulation

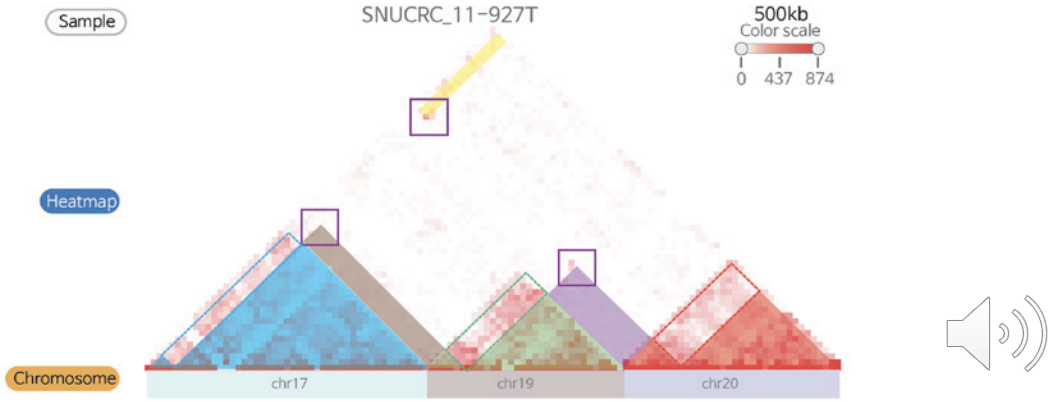


Complex forms of large-scale structural variations

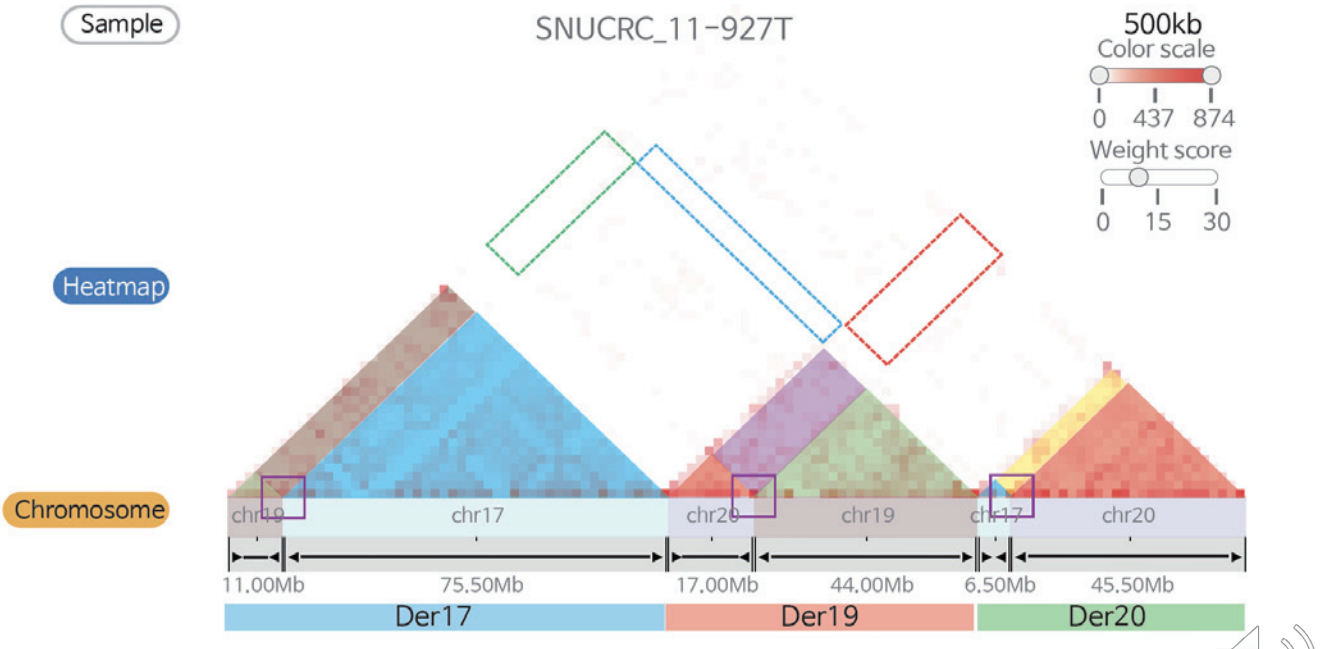


Module III. Complex SV and 3D genome





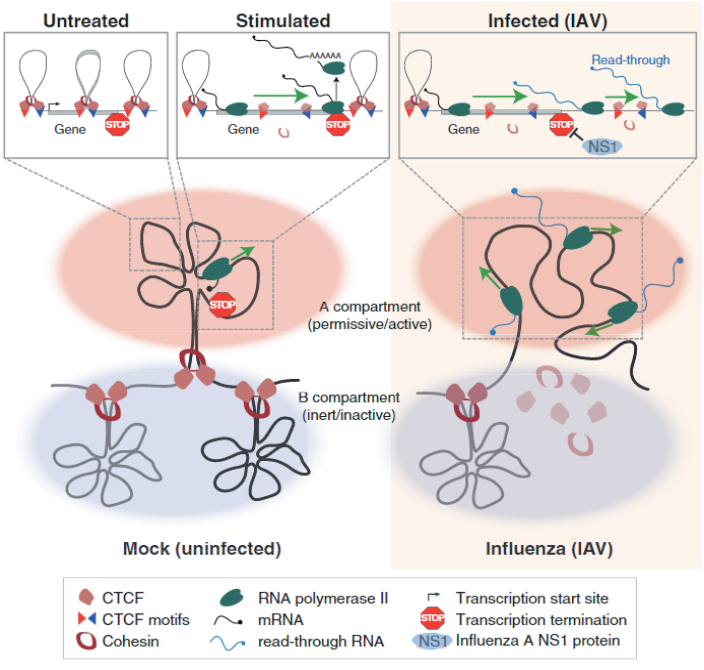
Module III. Complex SV and 3D genome



Question #1

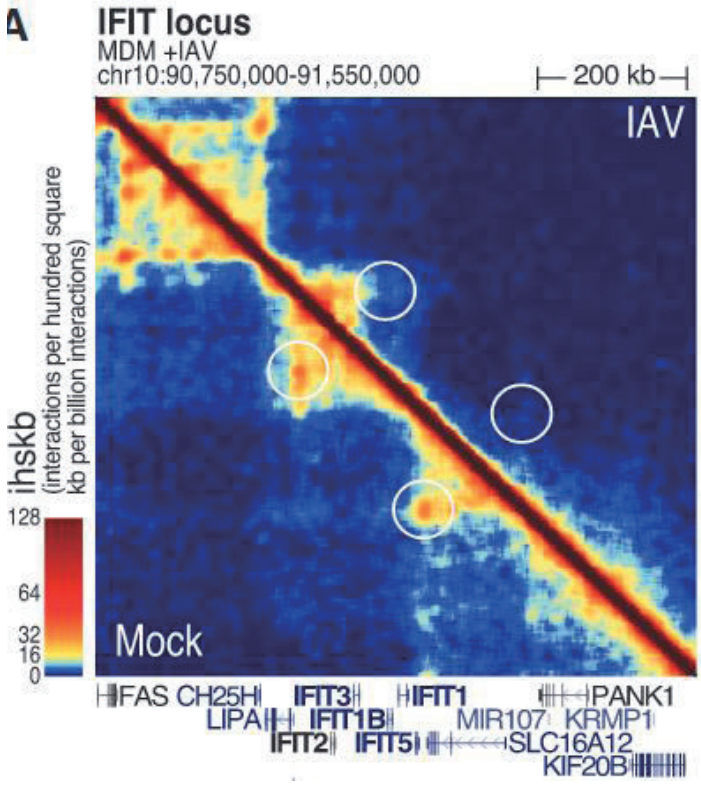
1. 3D genome organization in various cellular context

3D chromatin reorganization of macrophages after Flu infection



Heinz *et al.*, *Cell*. (2018)

Q: Does 3D chromatin structure change according to virus infection?



IAV: Influenza A virus

Sample: MDM (monocyte-derived macrophages)

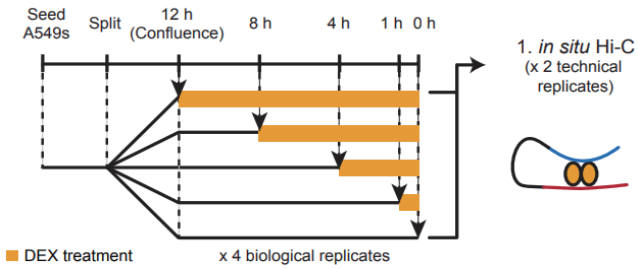
Target bait: IFIT locus

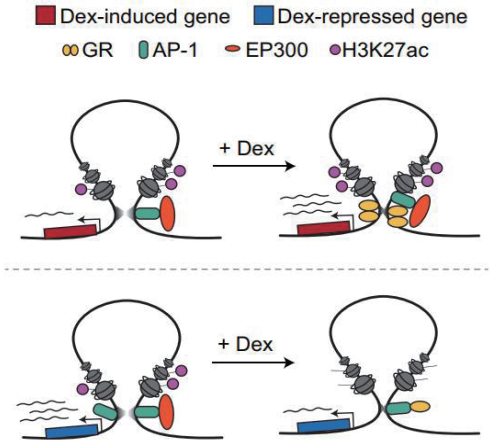
Heinz *et al.*, *Cell*. (2018)

Question #2

2. 3D Epigenome change by drug treatment

Pre-determined 3D chromatin structure strengthened
after drug treatment by recruited TFs

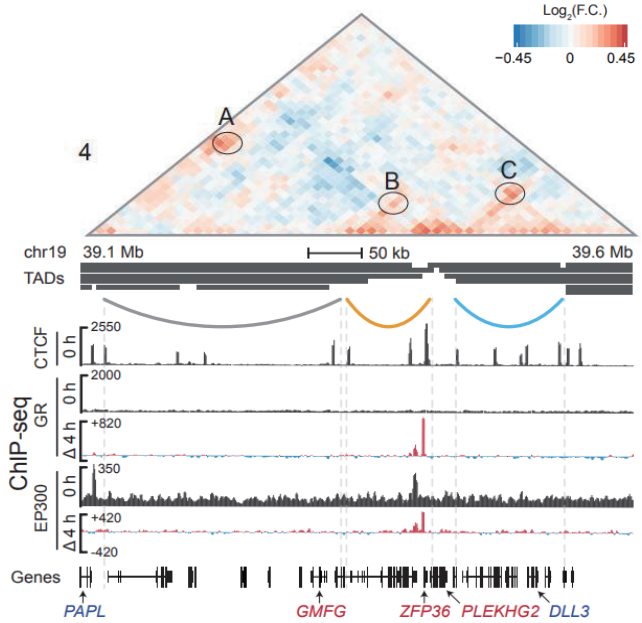




D'Ippolito *et al.*, *Cell Syste*. (2018)

69

Q: Does 3D chromatin structure & TF peaks change according to drug treatment?



Sample: A549 lung cancer cell line
-04h 100 nM dexamethasone
-00h 100 nM dexamethasone

+NR3C1 glucocorticoid receptor (GR) ChIP-seq

Target bait: *ZFP36* locus

Data: A549_Dex_practice

D'Ippolito *et al.*, *Cell Syste*. (2018)

https://dl.dropbox.com/s/bkudsuq5sexkzms/A549_Dex_practice.zip

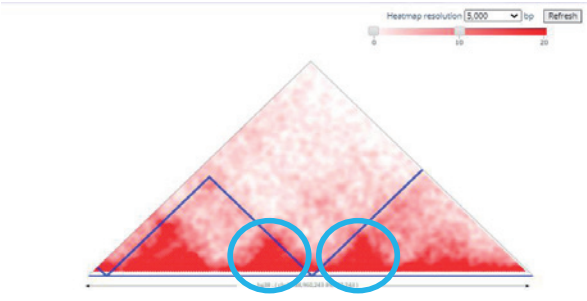
Example Answer for Q1

1. 3D genome organization in various cellular context

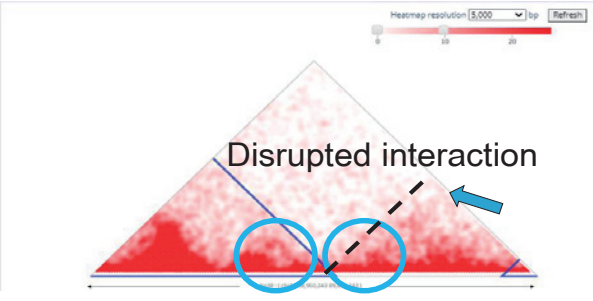
Example with visualization module)

Sample: MDM (monocyte-derived macrophages), infection time 12hour
Bait: SLC16A12
Genomic range: chr10:88,990,243-89,790,243
TAD: DI(window size=500kb)

Mock



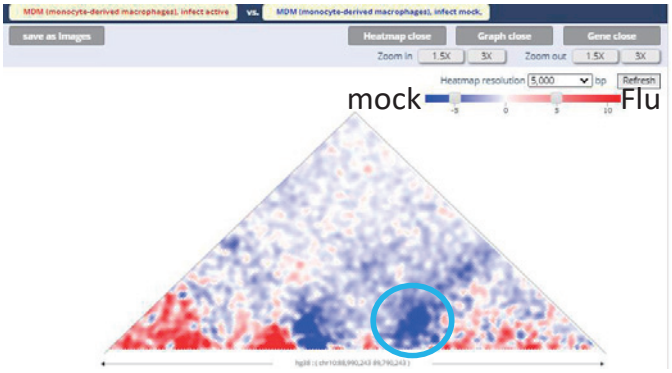
Flu infection



1. 3D genome organization in various cellular context

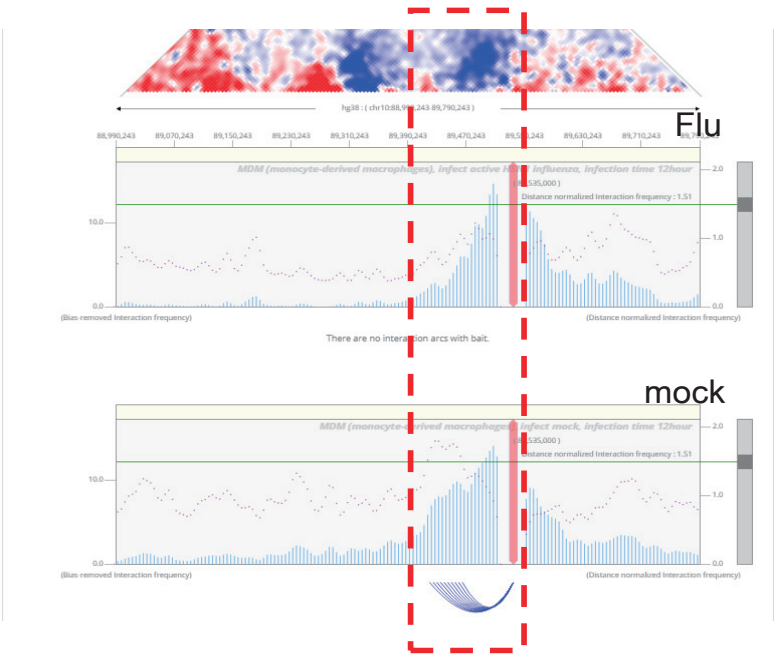
Example with Comparative module)

Sample: MDM (monocyte-derived macrophages), infection time 12hour
Mock vs Active Flu
Bait: SLC16A12
Genomic range: chr10:88,990,243-89,790,243
TAD: DI(window size=500kb)



1. 3D genome organization in various cellular context

Chromatin Interactions are disrupted when Flu infection occurs



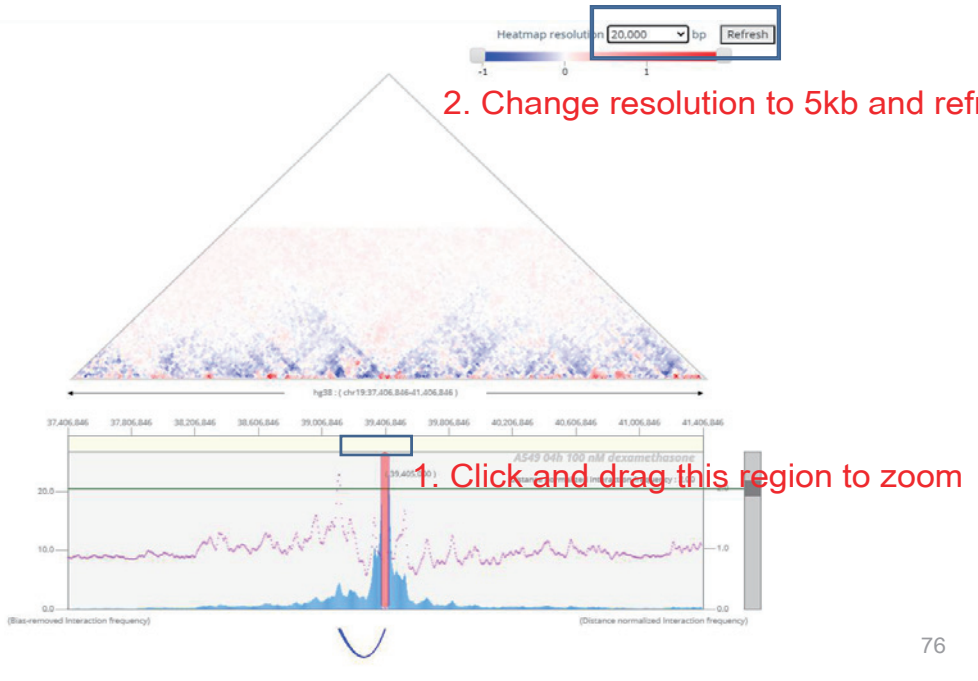
Example Answer for Q2

2. 3D Epigenome change by drug treatment

Sample: A549 00h 100 nM dexamethasone & A549 04h 100 nM dexamethasone

Bait: *ZFP36*

Comparative interaction module



76

2. 3D Epigenome change by drug treatment

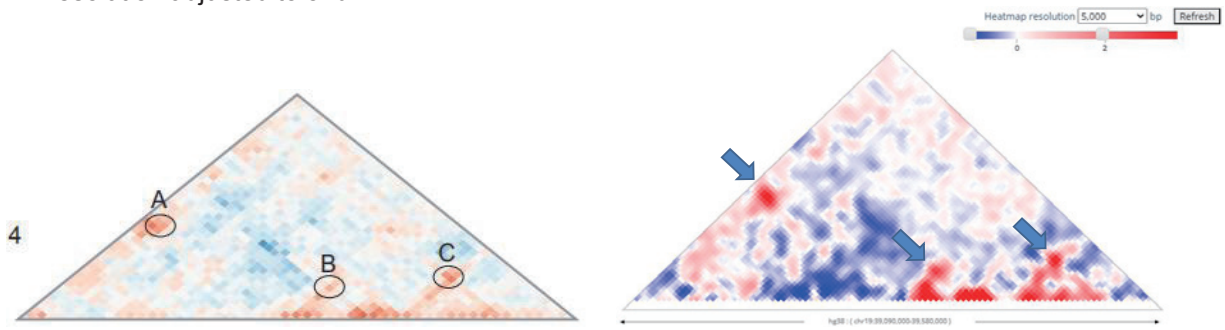
Sample: A549 00h 100 nM dexamethasone & A549 04h 100 nM dexamethasone

Bait: *ZFP36*

Comparative interaction module

Genomic range: chr19:39,090,000-39,580,000 (hg38)

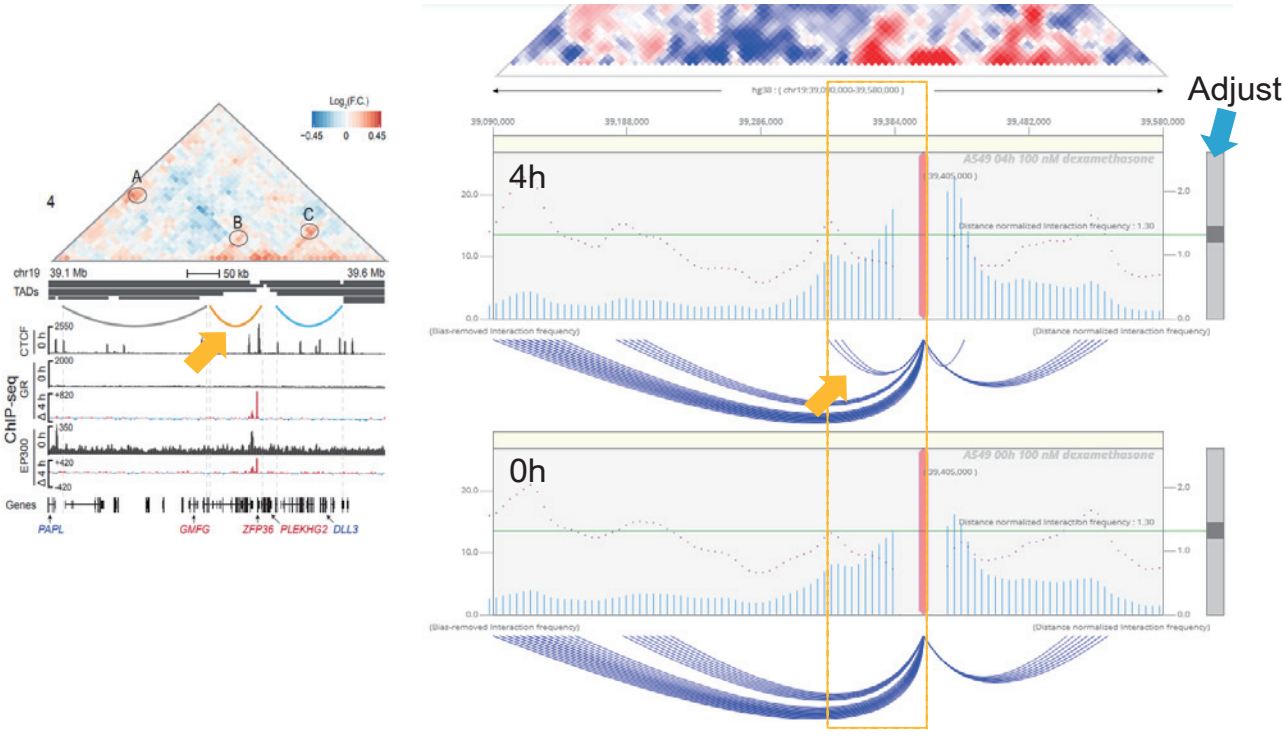
Resolution adjusted to 5kb



77

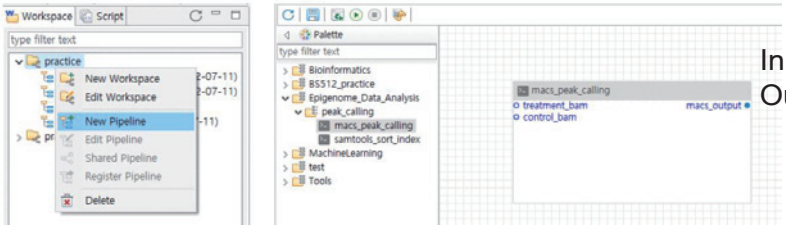
2. 3D Epigenome change by drug treatment

Enhanced chromatin interaction after 4hour from Dex treatment

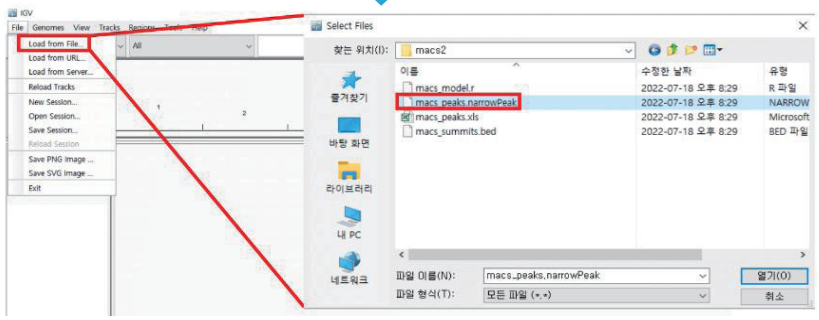


2. 3D Epigenome change by drug treatment

Run MACS scripts with treatement & control bam files as a pair (Watch out 0hr/4hr sample swap)
BAM files are provided by TAs
(1)Input 1: treatment BAM
(2)Input 2: control BAM
(3)Output: Set output folder to any exist folder in your workspace you want



Input : bam (treatment and control as pair)
Ouput: macs_peaks.narrowPeak



Load to IGV
(1)macs.peaks.narrowPeak
(2)bam
(3)bam index

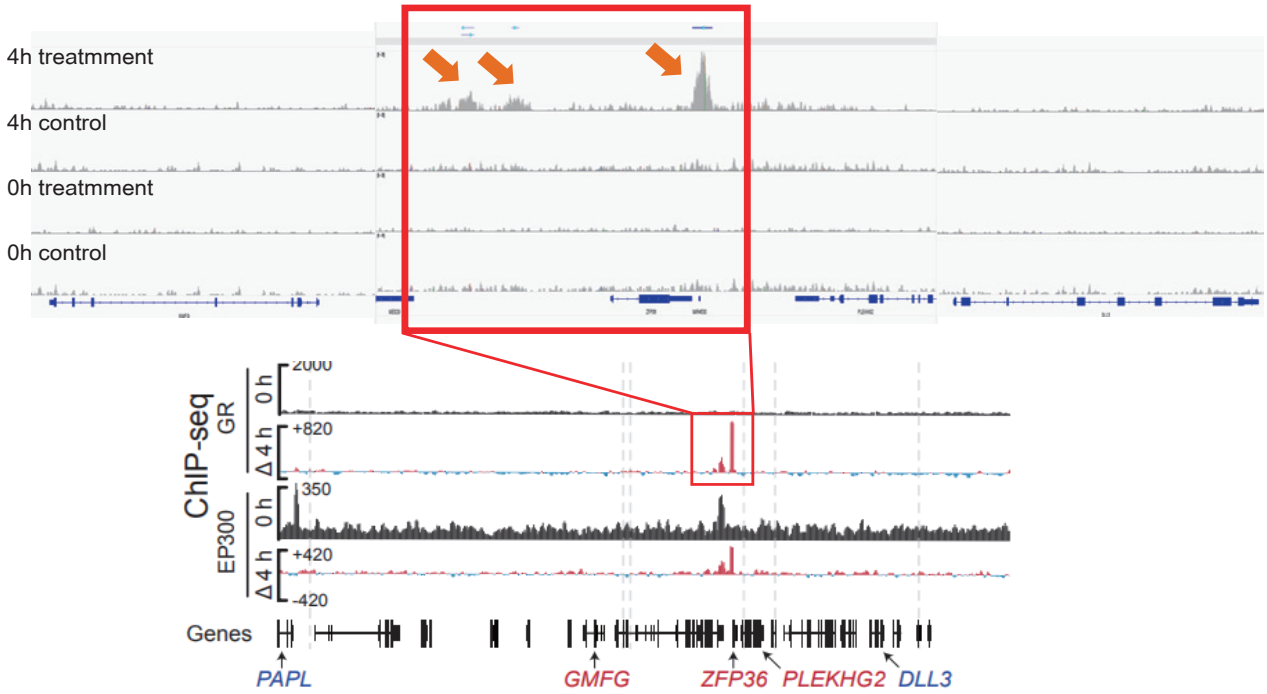
❖ IGV will automatically load BAM index file for selected BAM file

2. 3D Epigenome change by drug treatment

IGV showing 4 bam files and 4hr narrowpeaks (0hr has no peaks)
Adjust the scale to 0-35

2. 3D Epigenome change by drug treatment

Compare with the original figure.



81

2. 3D Epigenome change by drug treatment

Use genome track formats for better visualization