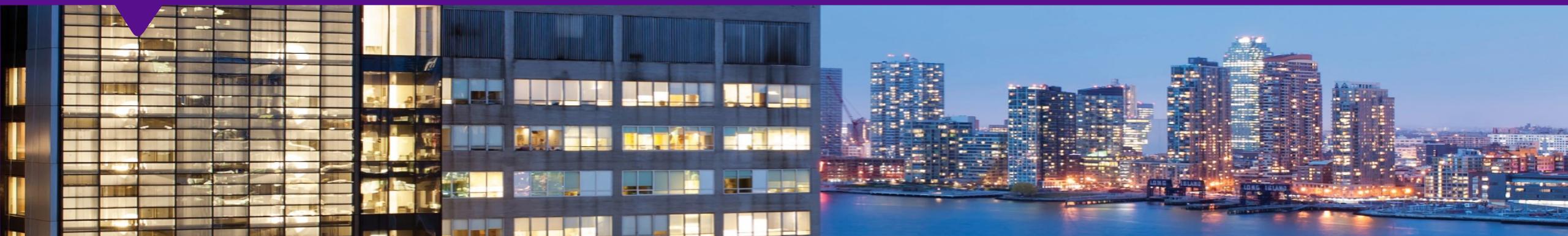


NYU School of Medicine

INTRODUCTION TO 3D CHROMATIN ORGANIZATION

Aristotelis Tsirigos, Professor, Pathology

Andreas Kloetgen, Sofia Nomikou, Sonali Narang

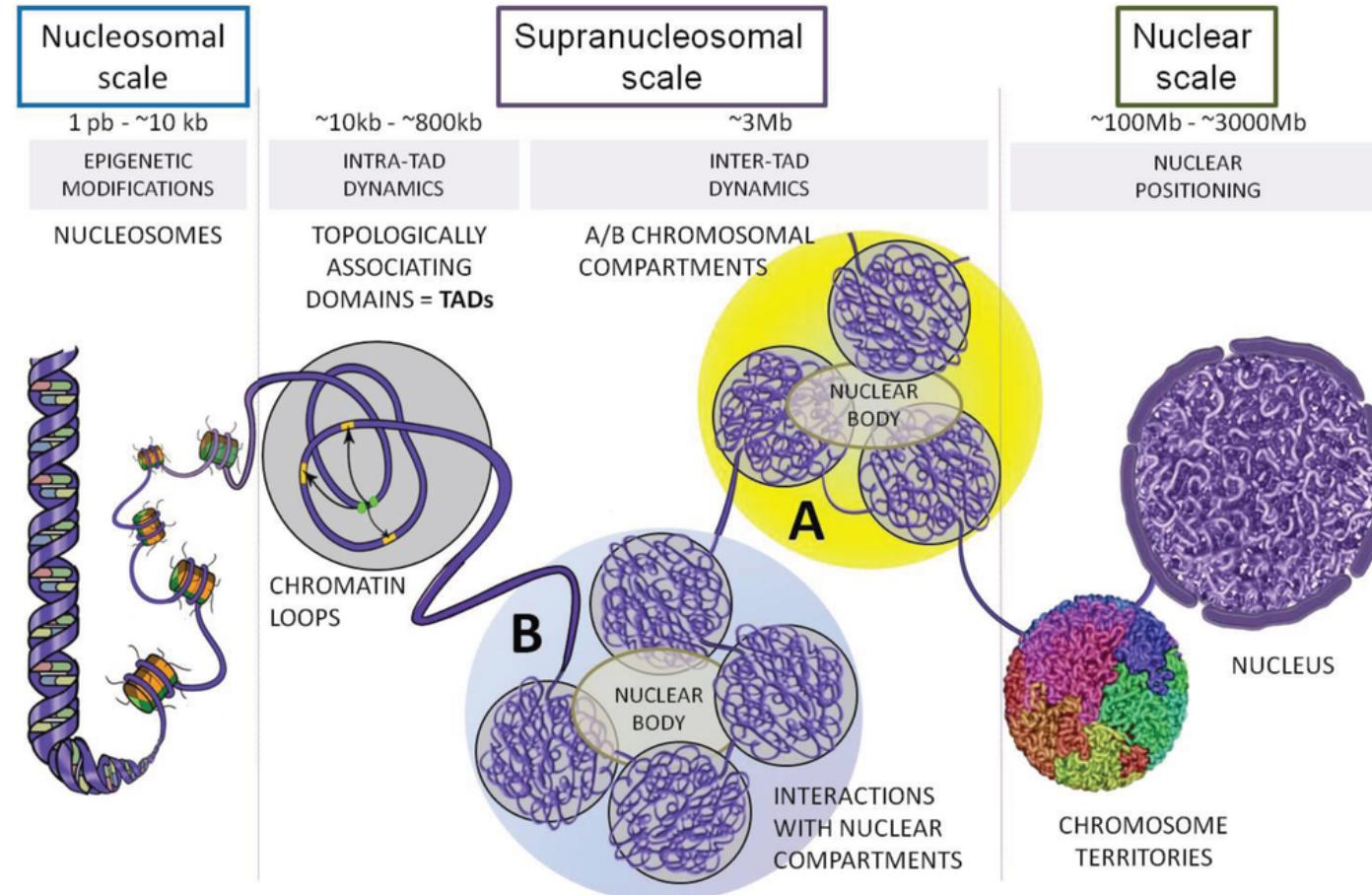


Outline

- What is 3D chromatin organization?
 - Why do we study it and why is it important for gene regulation?
 - What experimental approaches are used?
 - Disruption of chromatin organization in genetic diseases and in cancer
- Computational analysis of 3D data:
 - Data preprocessing
 - Compartment analysis
 - TAD calling
 - Loop calling

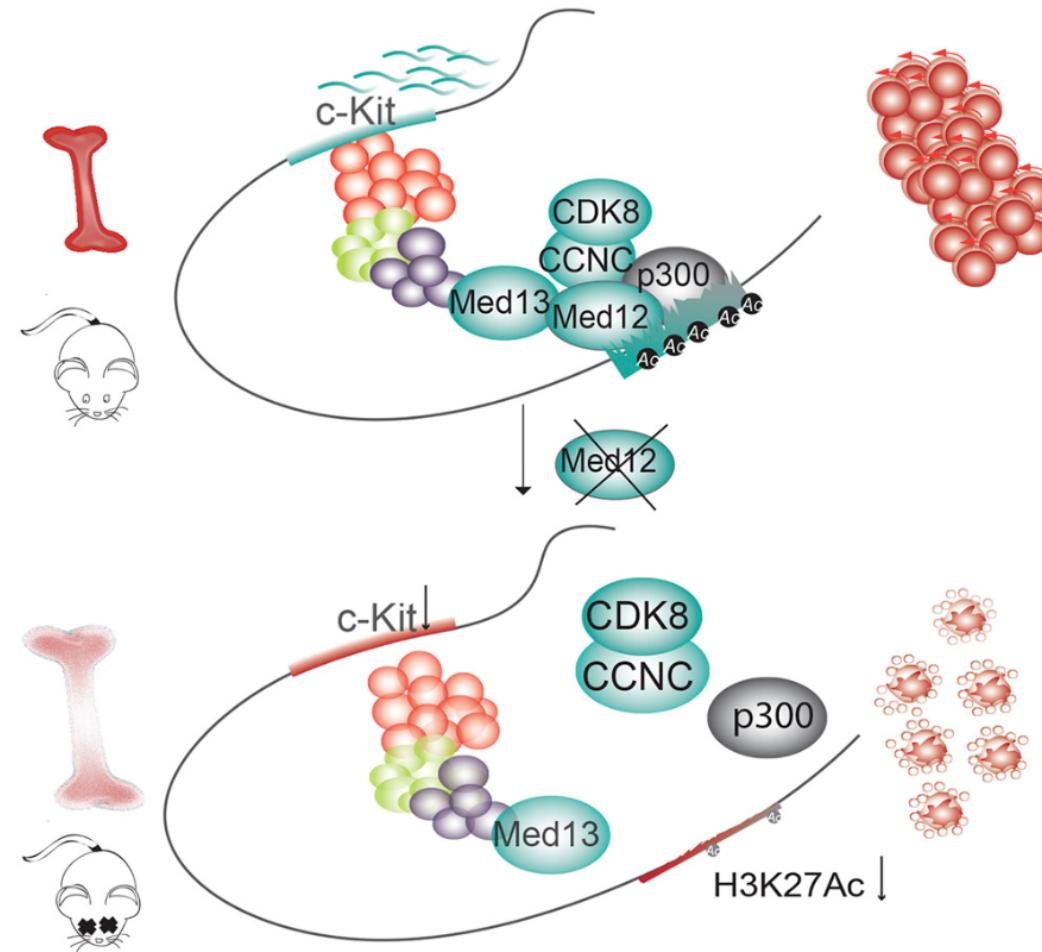
THE NUCLEAR ORGANIZATION OF CHROMATIN

Nuclear organization of chromatin



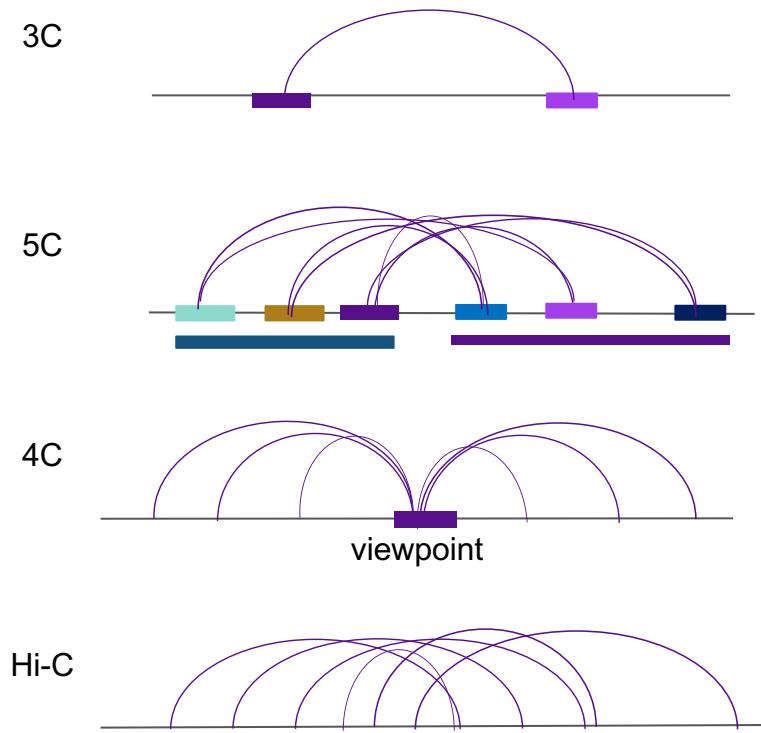
Importance of cis-regulatory elements for transcription

- Intergenic or non-promoter regions on the chromosome, referred to as “enhancers”
- Characterized by combinations of histone modification marks
- Strongly bound by cell-type specific transcription factors
- *Cis*-activation of gene transcription via formation of the Mediator-complex and the transcriptional machinery
- Example: c-Kit maintains self-renewal in hematopoietic progenitors → loss of enhancer results in loss of “stemness”



HIGH-THROUGHPUT ASSAYS TO STUDY CHROMATIN CONFORMATION

Chromatin Conformation Capture techniques

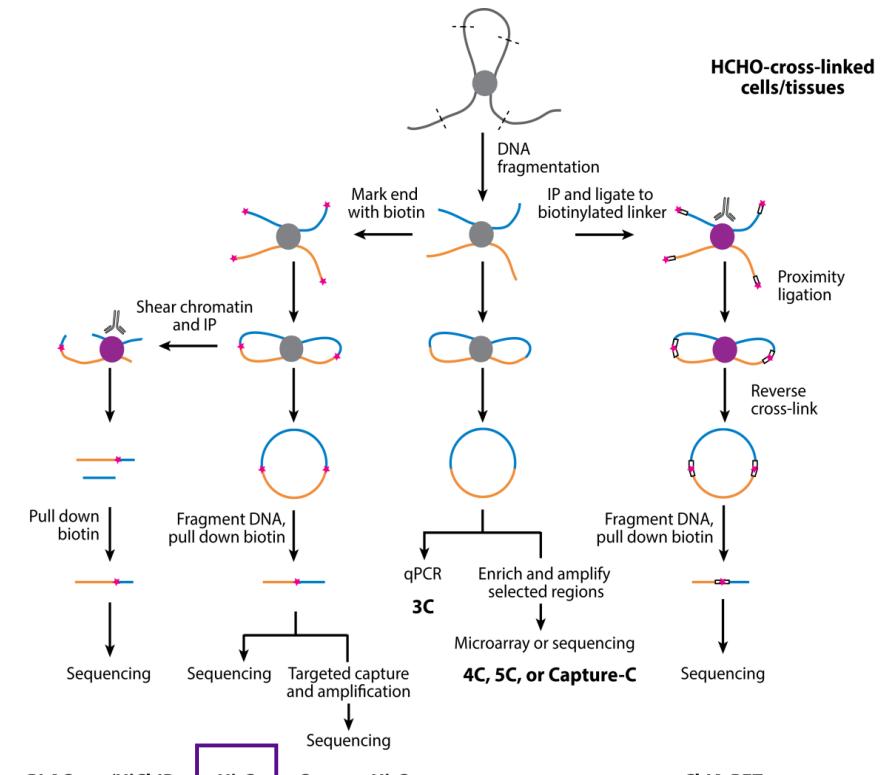


"one to one"

"many to many"

"one to all"

"all to all"



Yu M, Ren B. 2017.
Annu. Rev. Cell Dev. Biol. 33:265-89

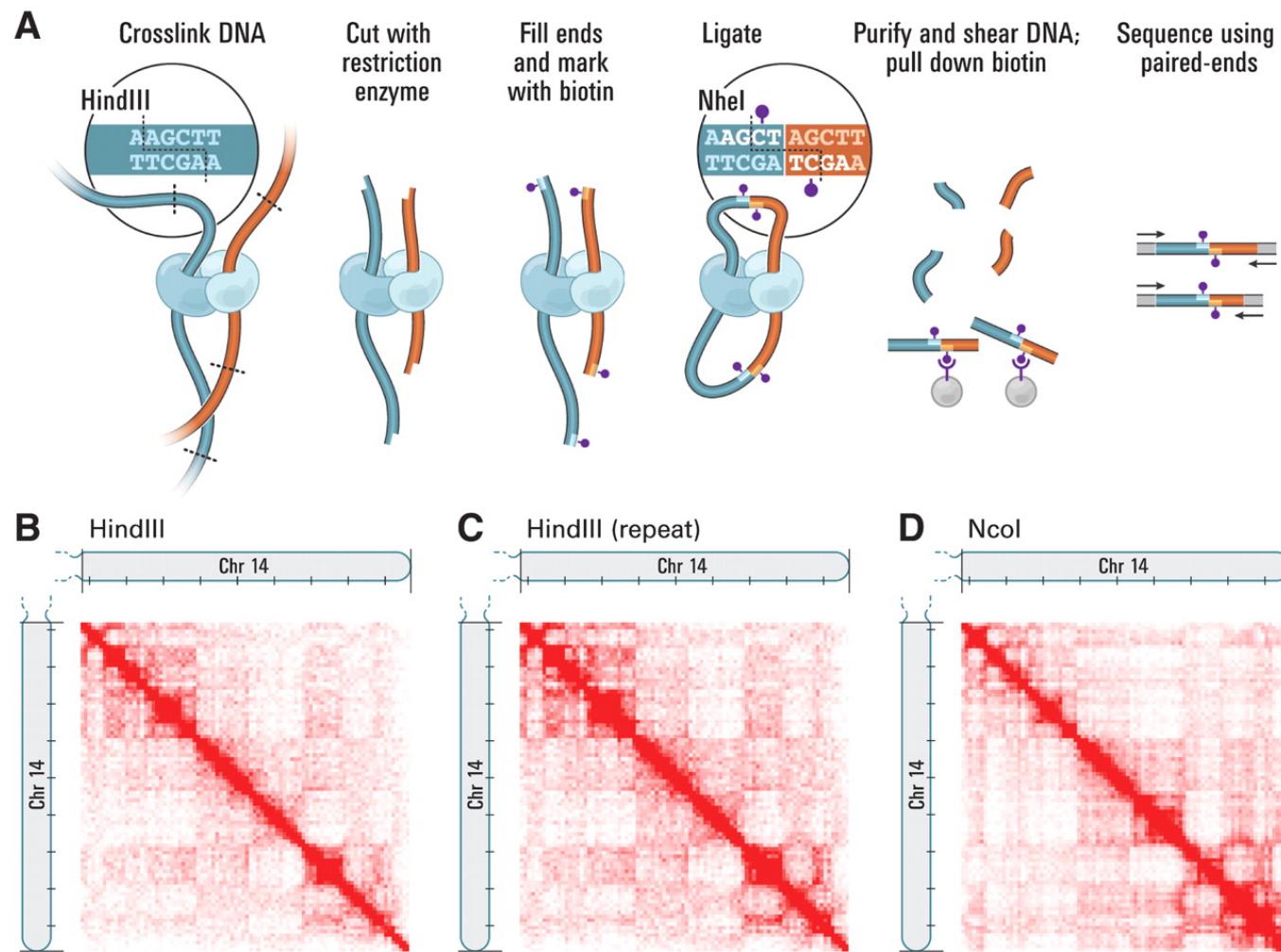
Introduction of the Hi-C-sequencing method

Science

Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

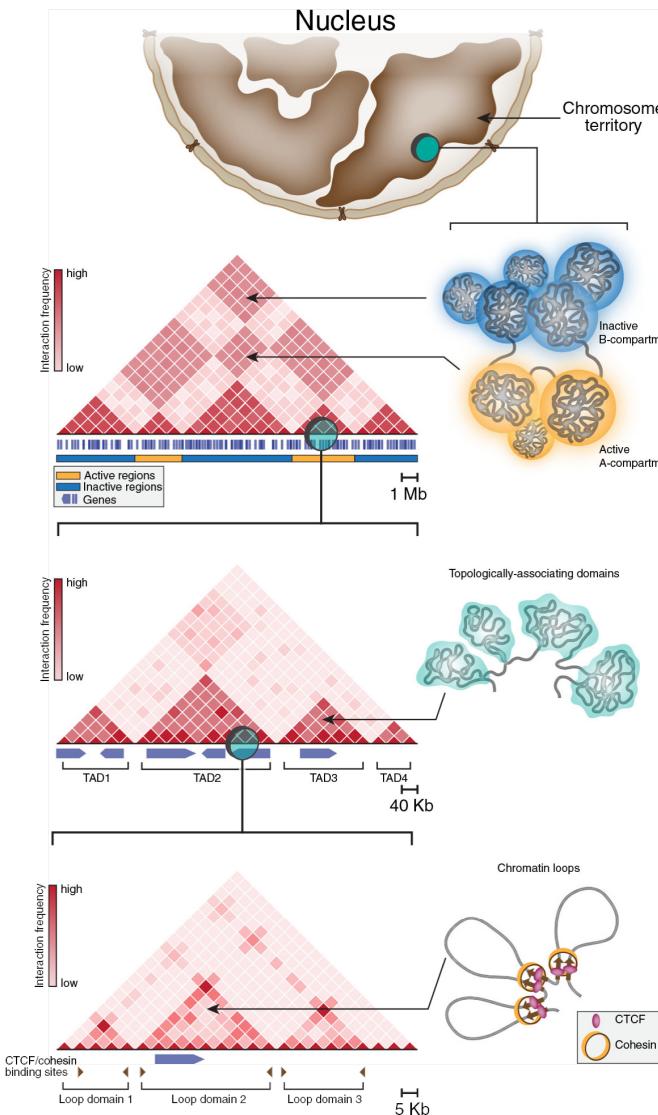
Erez Lieberman-Aiden^{1,2,3,4,*}, Nynke L. van Berkum^{5,*}, Louise Williams¹, Maxim Imakaev², Tobias Ragoczy^{6,7}, Agnes Telling^{6,7}, Ido Amit¹, Bryan R. Lajoie⁵, Peter J. Sabo⁸, Michael O. Dorschner⁸, Richard Sandstrom⁸, Bradley Bernstein^{1,9}, M. A. Bender¹⁰, Mark Groudine^{6,7}, Andreas Gnirke¹, John Stamatoyannopoulos⁸, Leonid A. Mirny^{2,11}, Eric S. Lander^{1,12,13,†}, Job Dekker^{5,†}

Hi-C outline



Hi-C data reveals most chromatin structures

Razin SV, Ulianov SV.
Cellular & Molecular Biology Letters. 2017;22:18



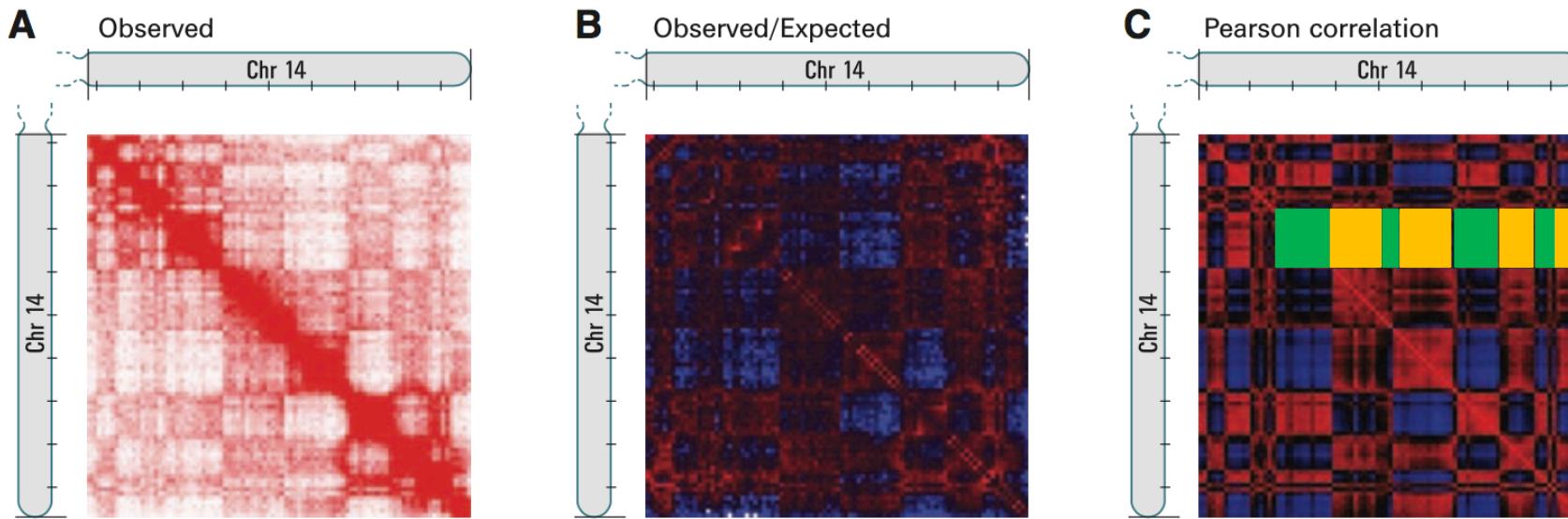
Compartments

Topological Domains (TADs)

3D chromatin loops

A/B COMPARTMENTS

Compartments (active vs. inactive chromatin)



- Compartment A :
 - gene rich, higher gene expression
 - more accessible chromatin
 - enriched for activating histone marks (H3K27ac, H3K4me3, H3K36me3)
 - Early DNA replication
 - Open, accessible and actively transcribed chromatin
- Compartment B:
 - Gene poor
 - Lower gene expression
 - Associated with nuclear lamina
 - Enriched for repressive histone marks (H3K27me3)
 - Late DNA replication
 - Closed, densely packed chromatin

■ A compartment
■ B compartment

TOPOLOGICALLY ASSOCIATING DOMAINS

Introduction of Topological Domains (TADs)

nature

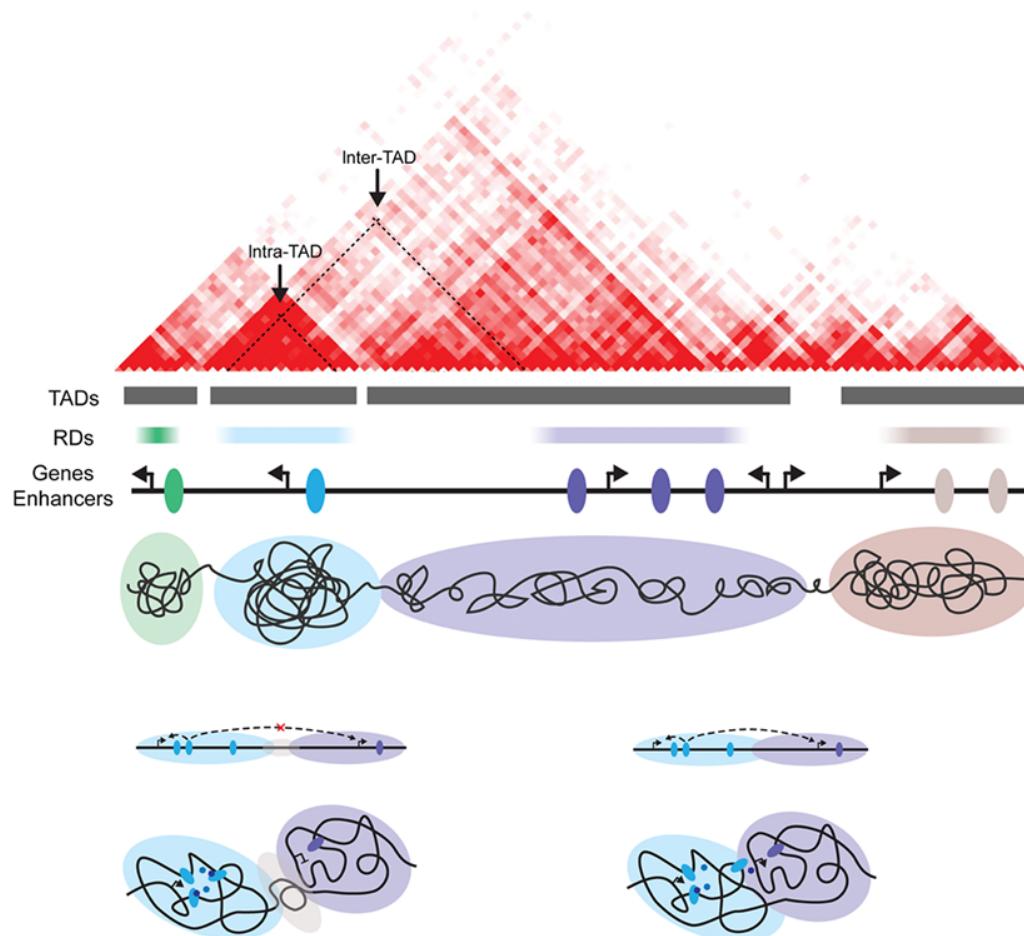
Letter | Open Access | Published: 19 November 2014

Topologically associating domains are stable units of replication-timing regulation

Benjamin D. Pope, Tyrone Ryba, Vishnu Dileep, Feng Yue, Weisheng Wu, Olgert Denas, Daniel L. Vera, Yanli Wang, R. Scott Hansen, Theresa K. Canfield, Robert E. Thurman, Yong Cheng, Günhan Gülsøy, Jonathan H. Dennis, Michael P. Snyder, John A. Stamatoyannopoulos, James Taylor, Ross C. Hardison, Tamer Kahveci, Bing Ren & David M. Gilbert✉

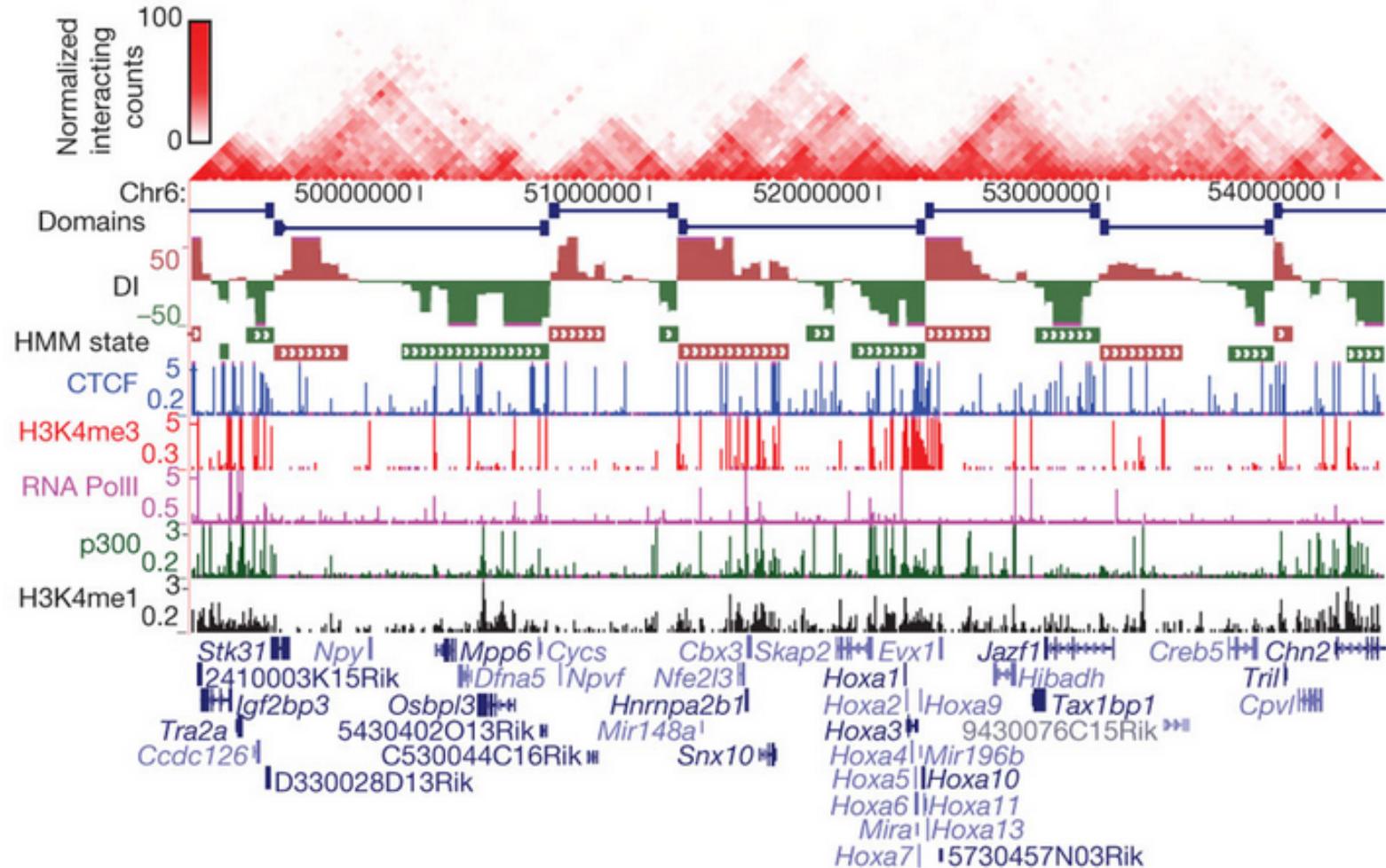
Topologically Associated Domains (TADs)

- Megabase-sized self-interacting regions
- Conserved across cell types and species
- Boundaries are enriched for the insulator binding protein CTCF, housekeeping genes, transfer RNAs and short interspersed element (SINE) retrotransposons
- Restrict interactions of regulatory elements to genes inside the TAD
- Prevent interactions of regulatory elements with genes outside the TAD

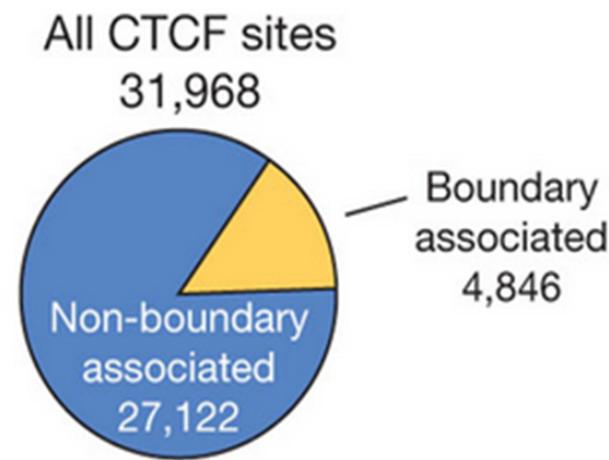
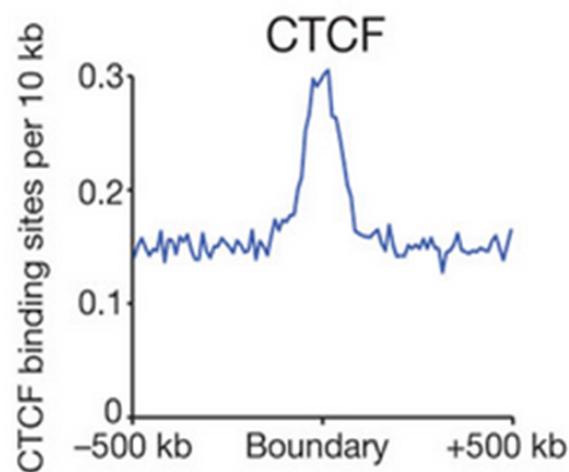


Remeseiro et al. *WIREs Dev Biol* 2015.

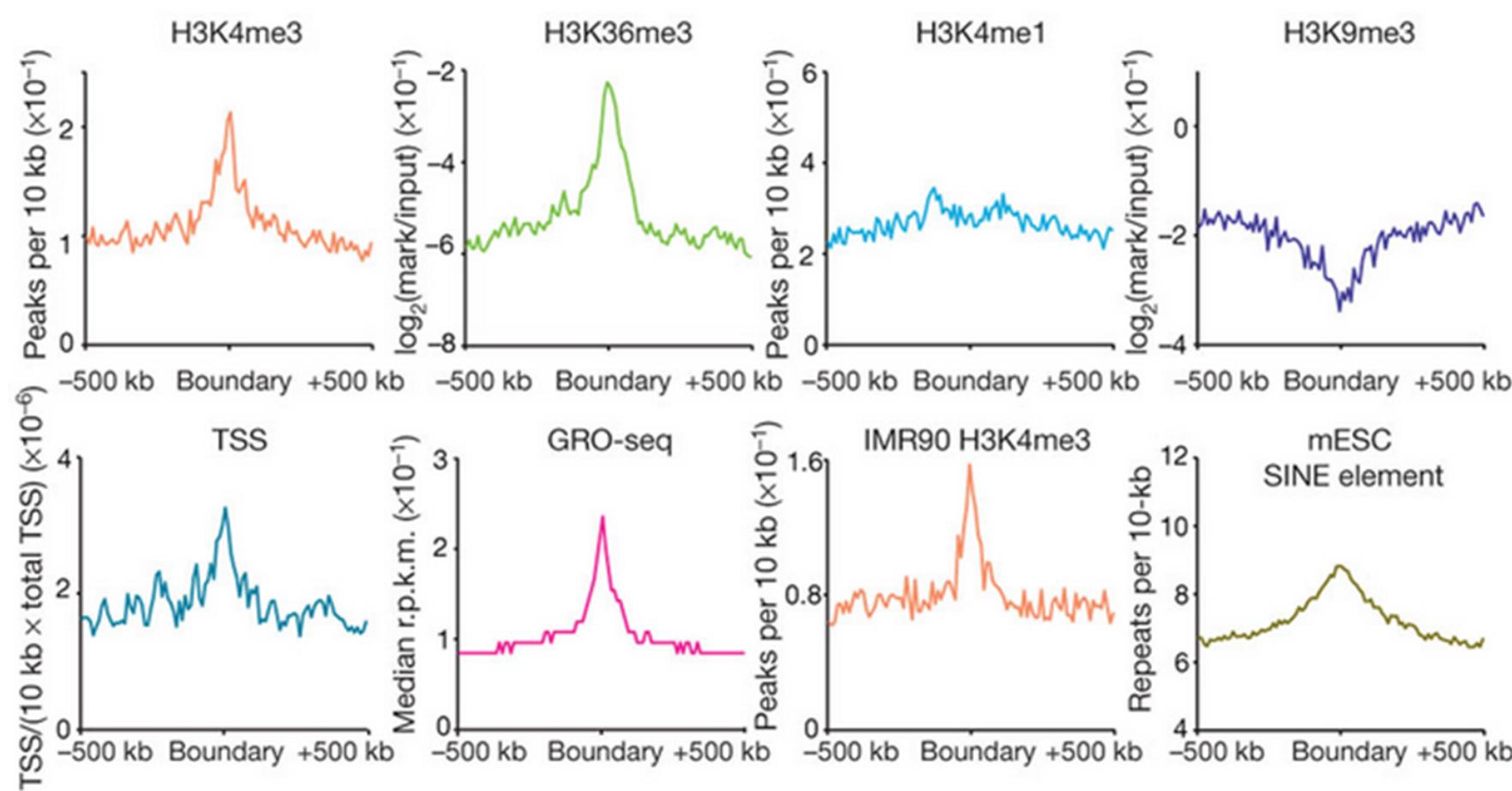
Topological domains in the mouse ES cell genome



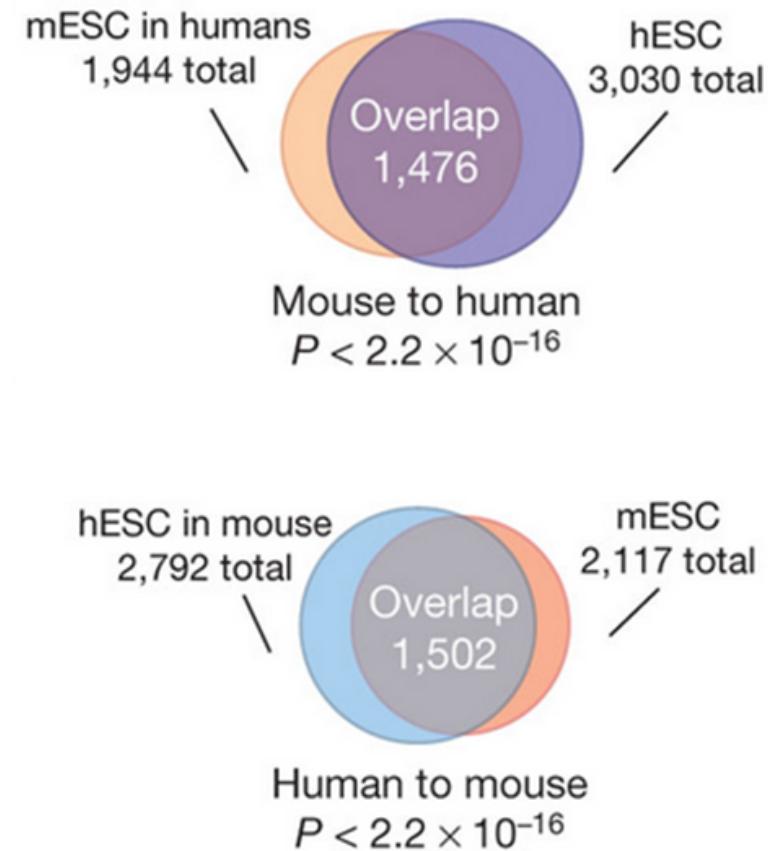
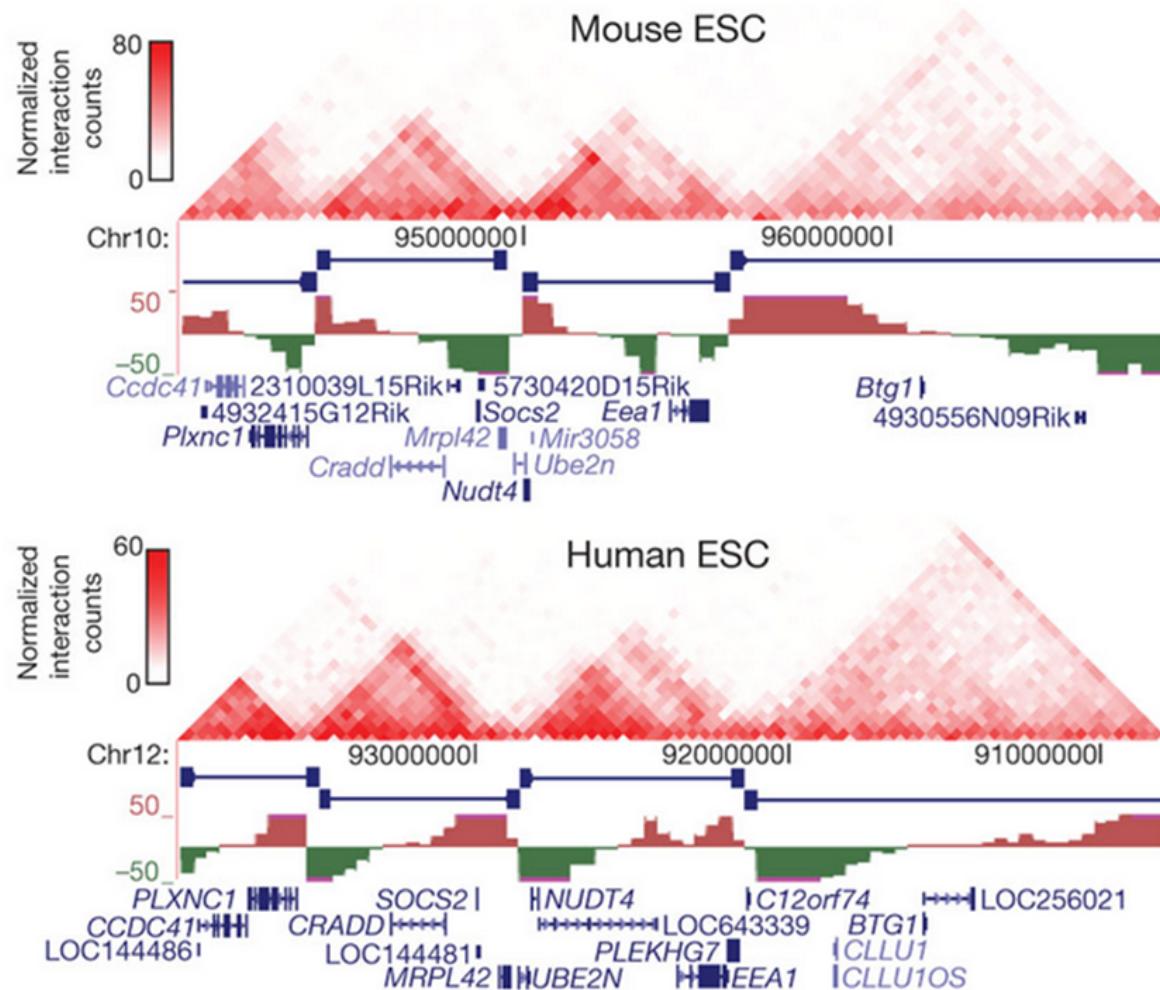
CTCF at TAD boundaries



Properties of domain boundaries



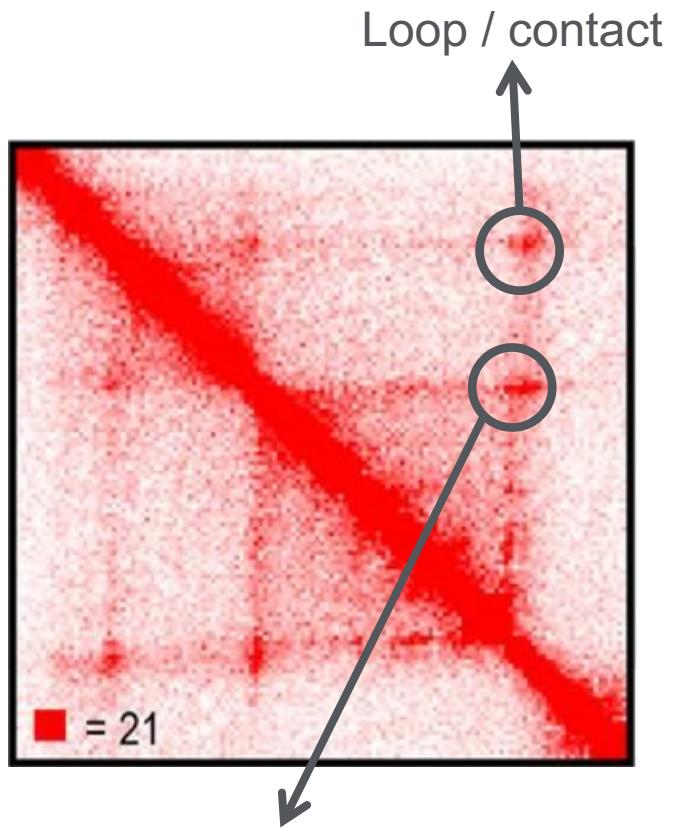
Common mESC/hESC boundaries



3D CHROMATIN LOOPS

Chromatin loops / contacts

- Usually <2Mb in size
 - Unclear to what degree long-range loops are of regulatory or structural function
- Can be found as “domain loops”, specifying contact points of TAD boundaries
 - Enriched in cohesin and CTCF binding
- Can connect promoters and cis-regulatory elements
 - Often associated with histone marks and other structural proteins, such as YY1

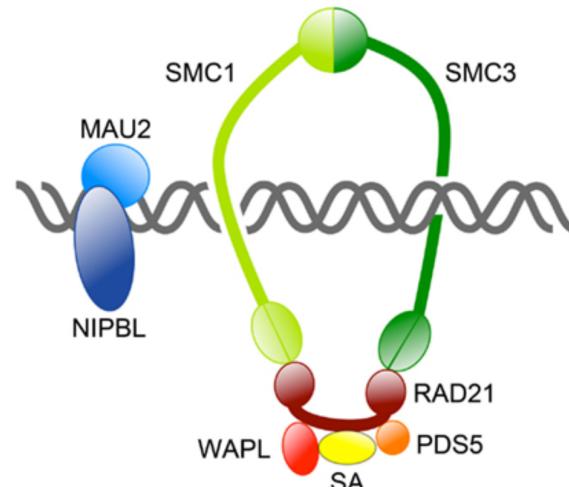


Rao S. et al., Cell, 2014
Weintraub A et al., Cell, 2017

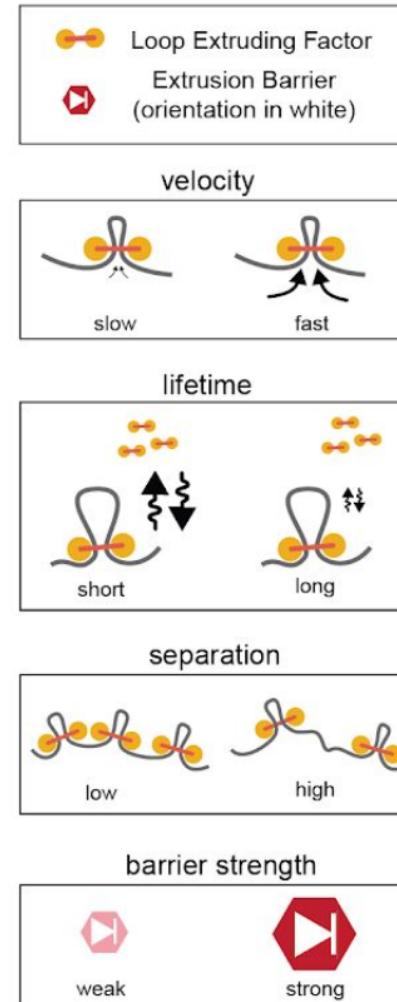
Cohesin mediates local looping but not compartmentalization

- Cohesin is loaded via NIPBL onto DNA
- It is a ring-like protein complex consisting of SMC1, SMC3, RAD21
- DNA is “extruded” through the cohesin rings until “blocked” by a barrier

→ Does not explain all features observed in 3D chromatin!



Fudenberg G et al., bioRxiv, 2018



PROBING CHROMATIN ORGANIZATION AT 1KB RESOLUTION

1-kb resolution using “4-cutter” enzymes

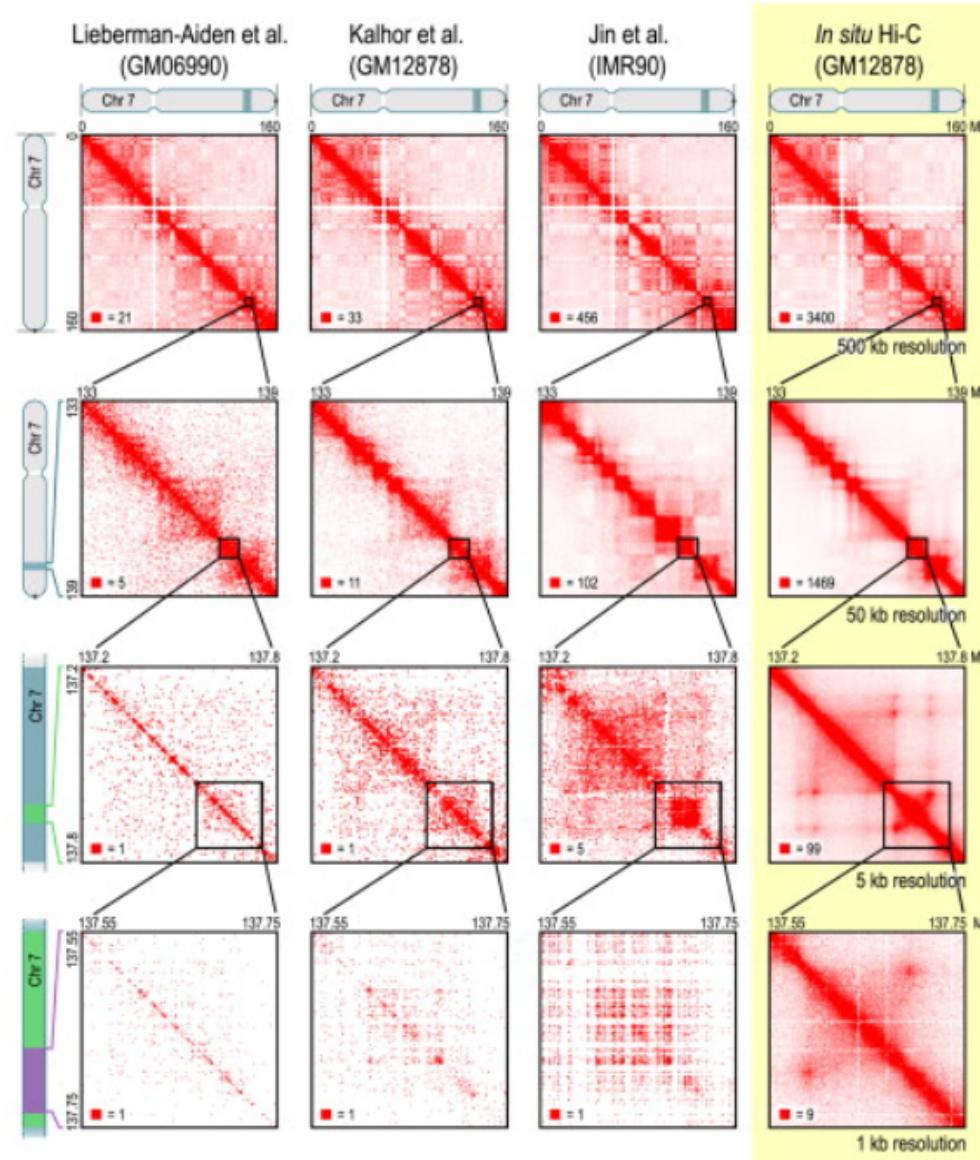
Cell

ARTICLE | VOLUME 159, ISSUE 7, P1665-1680, DECEMBER 18, 2014

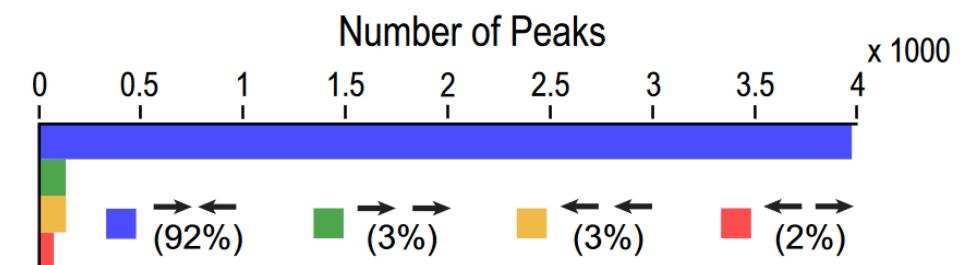
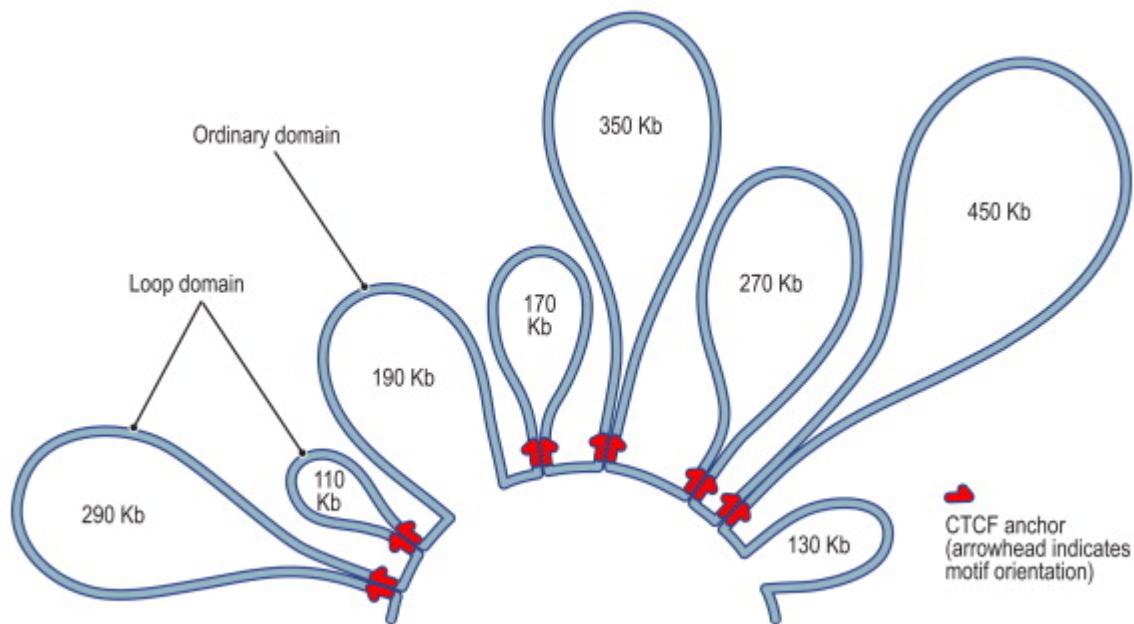
A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping

Suhas S.P. Rao ¹⁰ • Miriam H. Huntley ¹⁰ • Neva C. Durand • Elena K. Stamenova • Ivan D. Bochkov •
James T. Robinson • Adrian L. Sanborn • Ido Machol • Arina D. Omer • Eric S. Lander   •
Erez Lieberman Aiden   • Show less • Show footnotes

Hi-C at 1-kb resolution

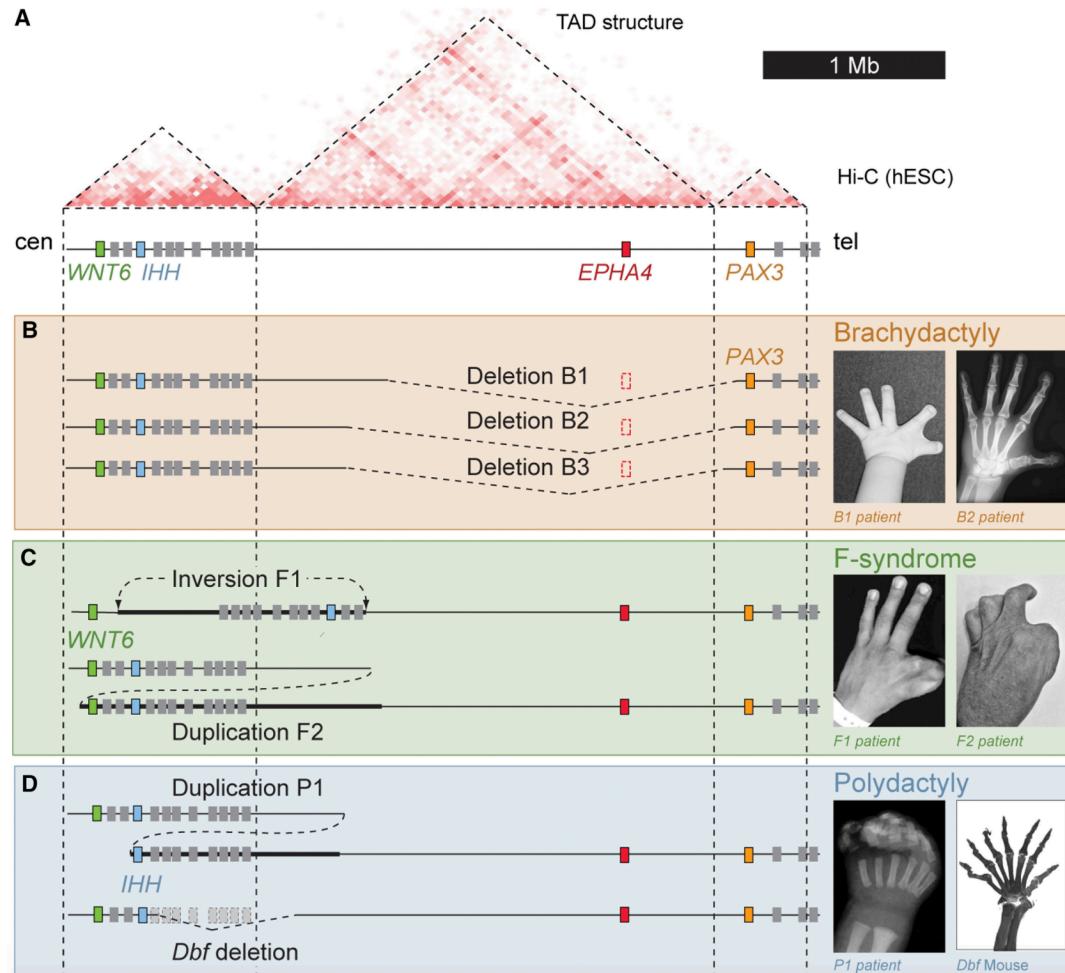


CTCF orientation at loop anchors



THE 3D GENOME TOPOLOGY IN DISEASE

Limb malformation due to rewiring of gene-enhancer interactions



Cell

Volume 161, Issue 5, 21 May 2015, Pages 1012-1025

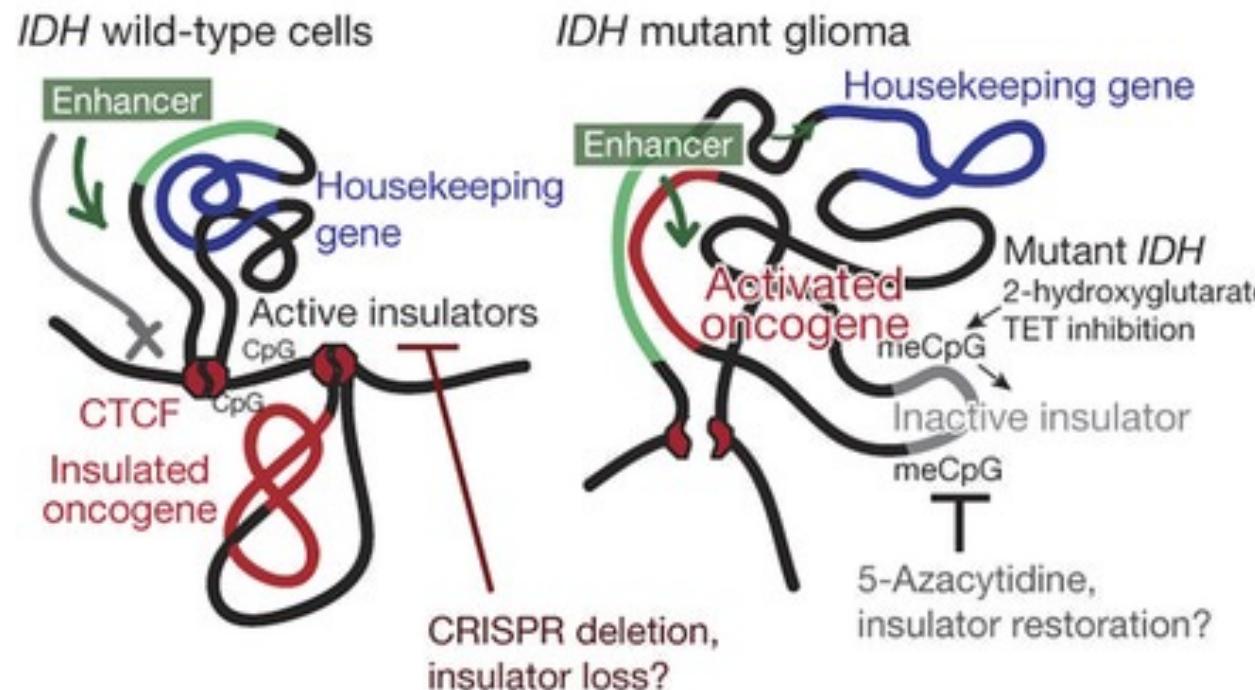


Article

Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

Darío G. Lupiáñez ^{1, 2}, Katerina Kraft ^{1, 2}, Verena Heinrich ², Peter Krawitz ^{1, 2}, Francesco Brancati ³, Eva Klopocki ⁴, Denise Horn ², Hülya Kayserili ⁵, John M. Opitz ⁶, Renata Laxova ⁶, Fernando Santos-Simarro ^{7, 8}, Brigitte Gilbert-Dussardier ⁹, Lars Wittler ¹⁰, Marina Borschüwer ¹, Stefan A. Haas ¹¹, Marco Osterwalder ¹², Martin Franke ^{1, 2}, Bernd Timmermann ¹³ ... Stefan Mundlos ^{1, 2, 14}

Insulator dysfunction and oncogene activation in IDH mutant gliomas



nature

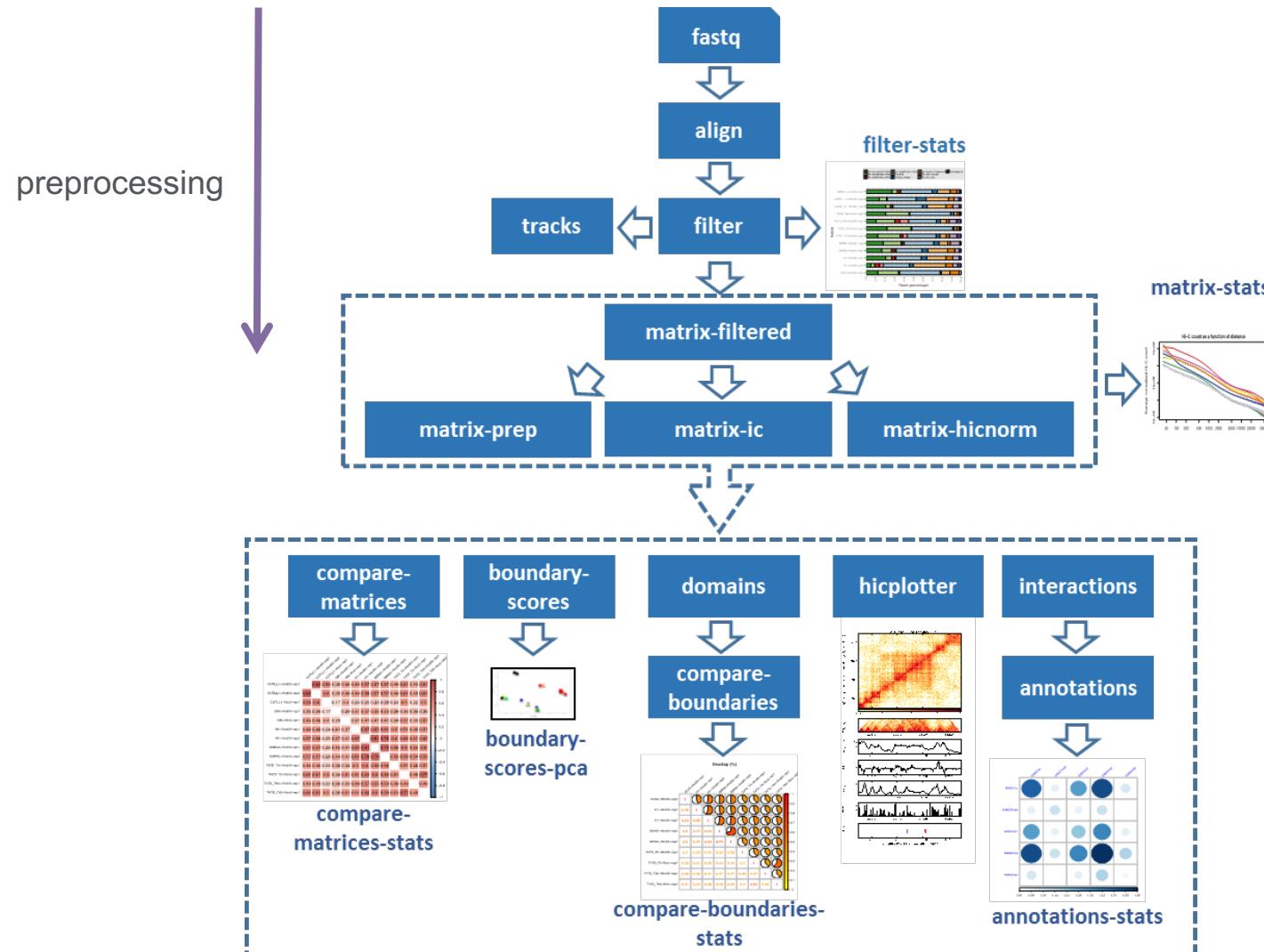
Letter | Published: 23 December 2015

Insulator dysfunction and oncogene activation in *IDH* mutant gliomas

William A. Flavahan, Yotam Drier, Brian B. Liau, Shawn M. Gillespie, Andrew S. Venteicher, Anat O. Stemmer-Rachamimov, Mario L. Suvà & Bradley E. Bernstein 

COMPUTATIONAL ANALYSIS OF HI-C DATA

Preprocessing of Hi-C matrix



HiC-bench (Lazaris et al., BMC Genomics 2017)

How does the data look like?

- Sequencing reads are usually stored in FASTQ files
 - Two files, one for each “mate” of the pair, ordered by pair
 - Each mate represents one end of the sequenced fragment

Illumina Sequencing example

@K00291:243:H575KBBXY:7:1101:2321:1121 1:N:0:NCAGTG

GATCCAAAGAAAGAGAGTTCAAAACTGCTCCATCAGCAGGATTGTTACCC

十

Header-line

Sequence

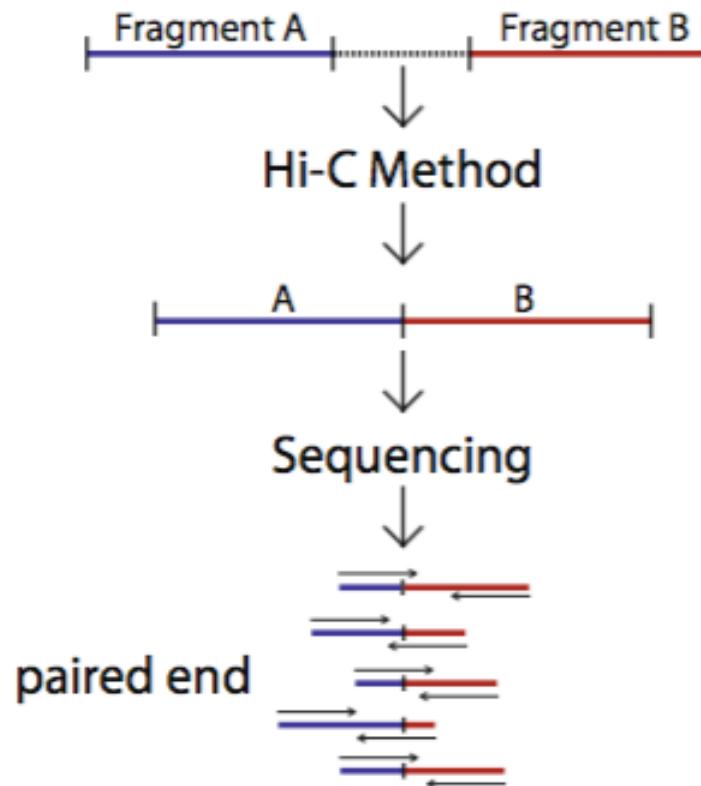
Quality score (Phred format)

Read-file 1

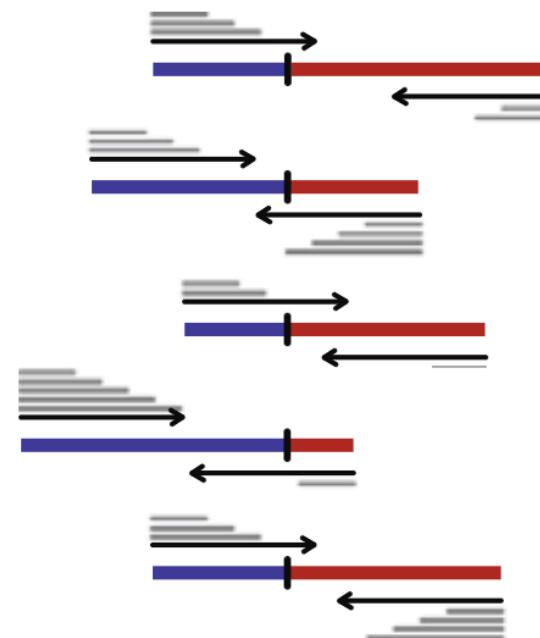
Read-file 2

@K00291:243:H575KBBXY:7:1101:2321:1121 2:N:0:NCAGTG
NAGCTGCANGTGGNTATTNGNCCNCTNTNANNATNTCGTTGGAAACGNGA
+
#A-<FFJJ#JJJF#FJJAJ#F#FJ#JF#<#F##F<#JFA-AAFJJFJJ#JJ
@K00291:243:H575KBBXY:7:1101:2341:1121 2:N:0:NCAGTG
NGGAGTGCNCTGNTATGTNTNAGNAANANCNGANACGAGTAACAAANGA
+
#<-<-77#<77F#111--#-#<F#F<#]#A##A7#-7-<<-<<AA<#F]

Alignment/Mapping

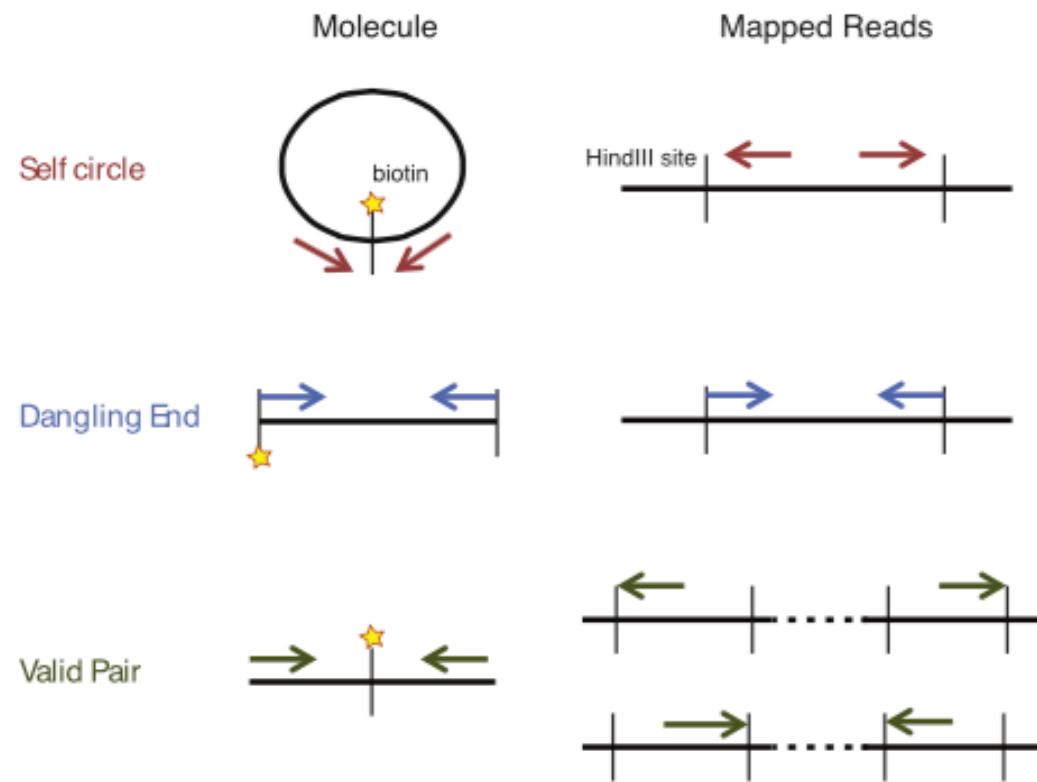


Iterative mapping



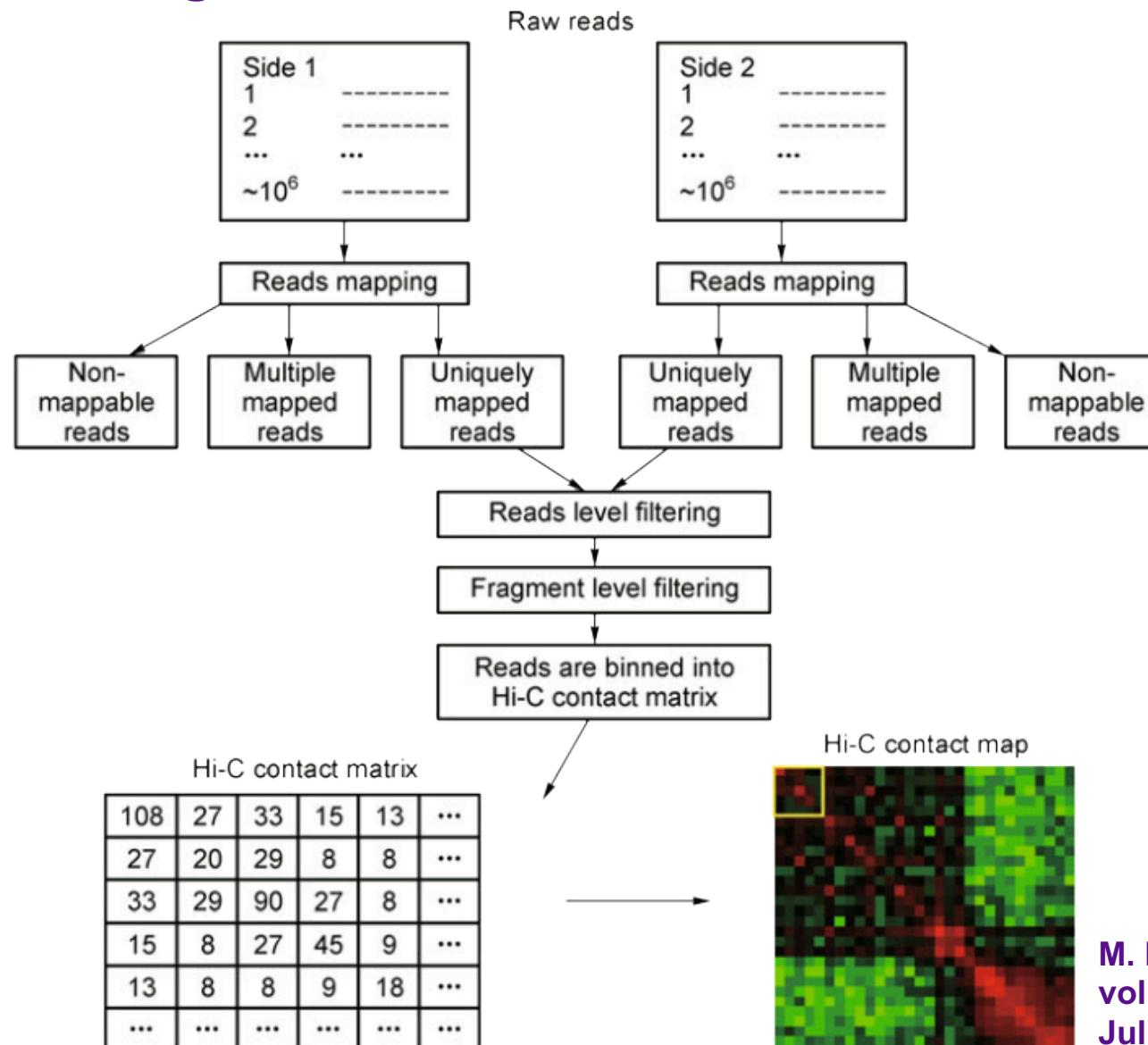
B. R. Lajoie et al. *Methods*, vol. 72, pp. 65–75, Jan. 2015.

Filtering to identify valid read-pairs

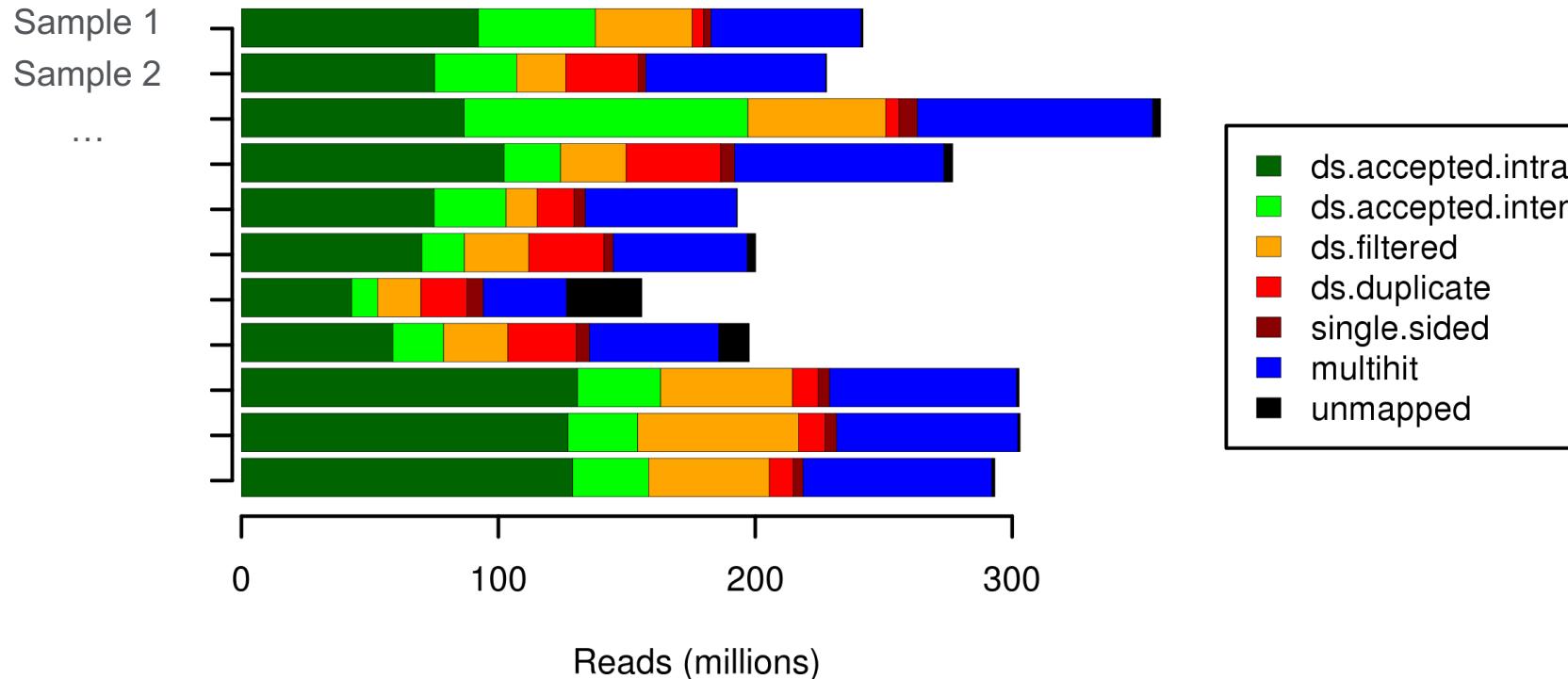


J-M. Belton et al. *Methods*, vol. 58, no. 3, pp. 268–276, Nov. 2012.

Hi-C contact matrix generation

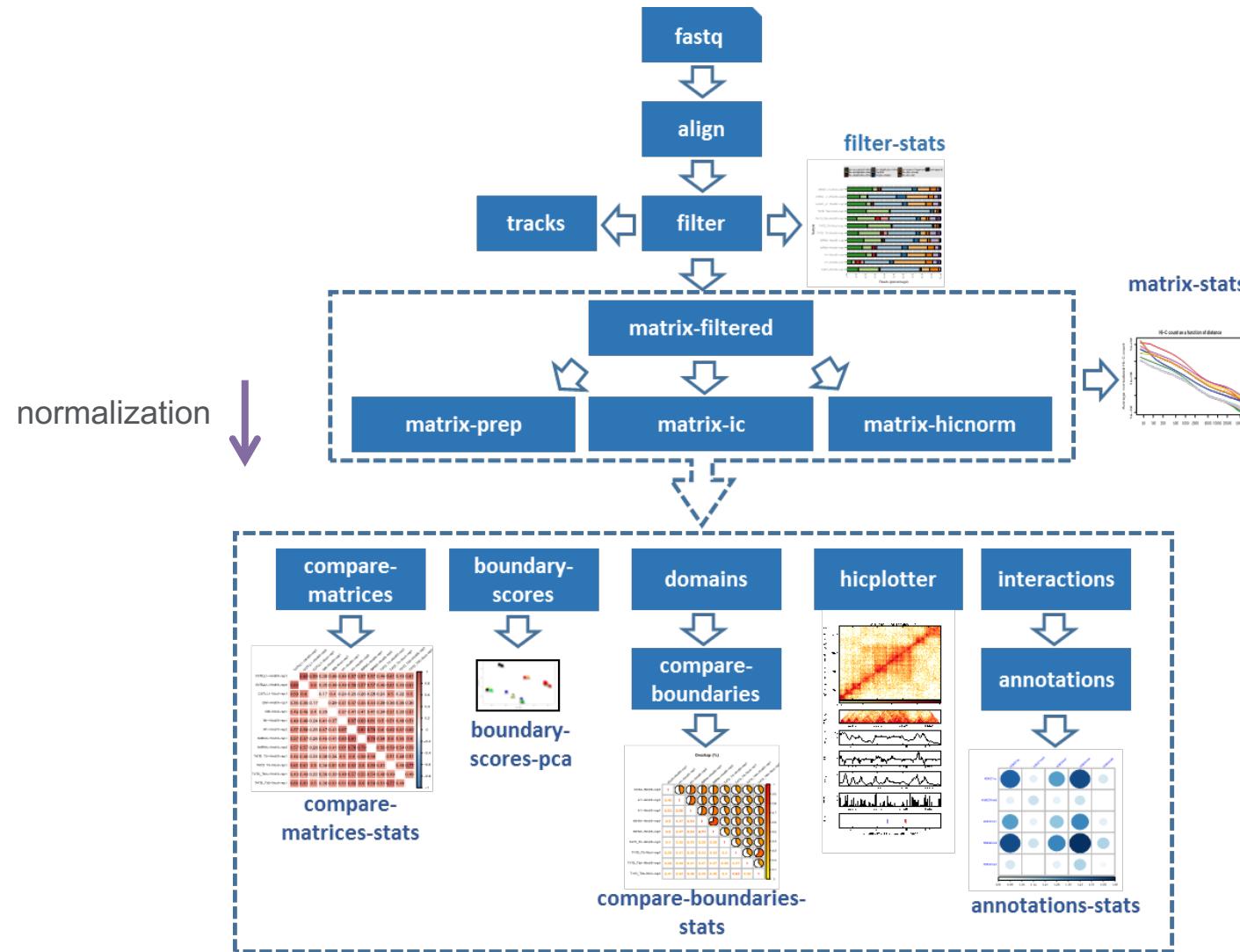


Data quality assessment



- Only 30-45% are “usable” reads available for downstream processing
- Experiments with such percentages of usable reads are considered successful

Normalization of Hi-C matrix



HiC-bench (Lazaris et al., BMC Genomics 2017)

Hi-C biases

nature genetics

Analysis | Published: 16 October 2011

Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture

Eitan Yaffe & Amos Tanay 

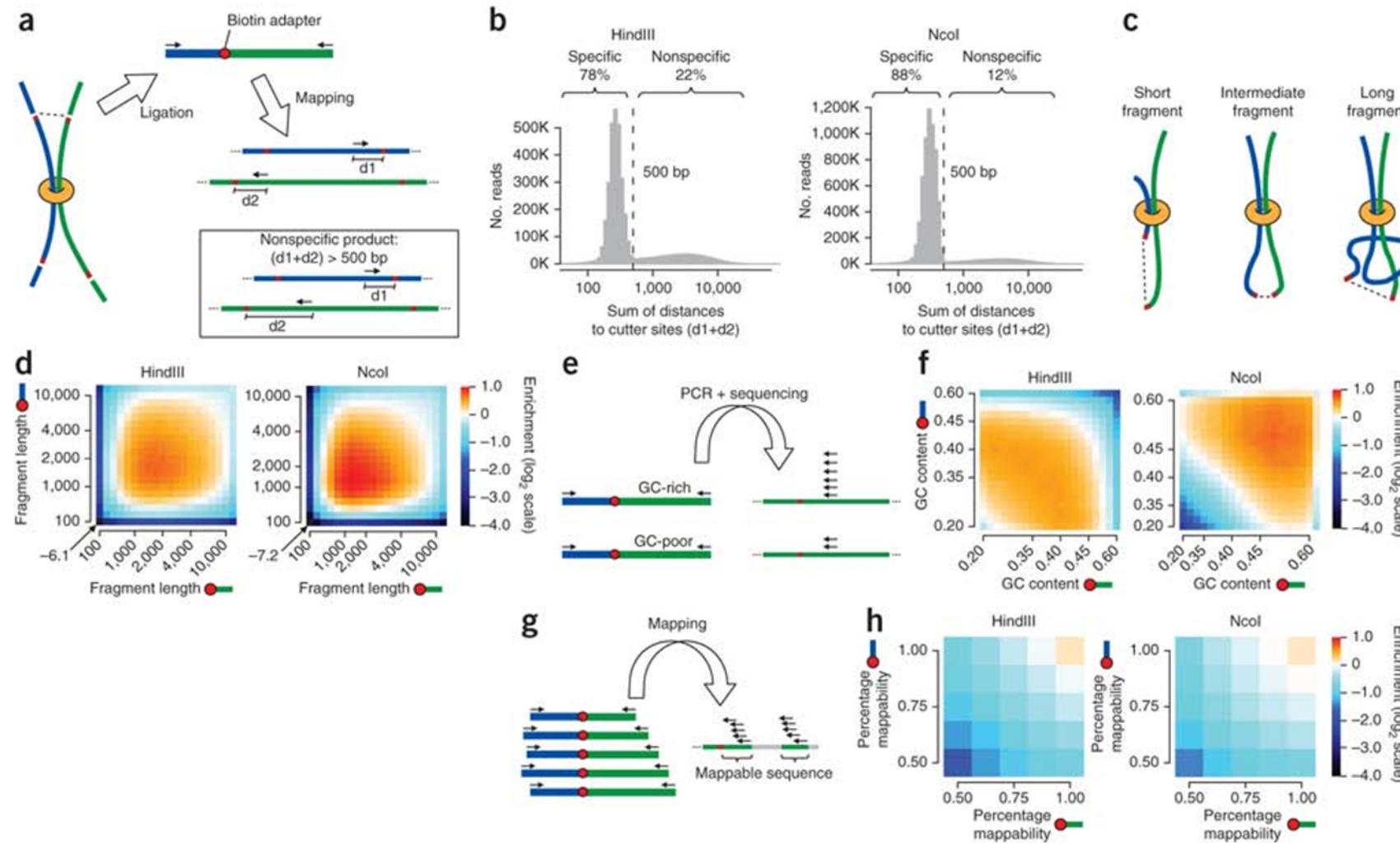
nature methods

Article | Published: 02 September 2012

Iterative correction of Hi-C data reveals hallmarks of chromosome organization

Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker  & Leonid A Mirny 

Sources of Hi-C biases: Fragment length and GC content

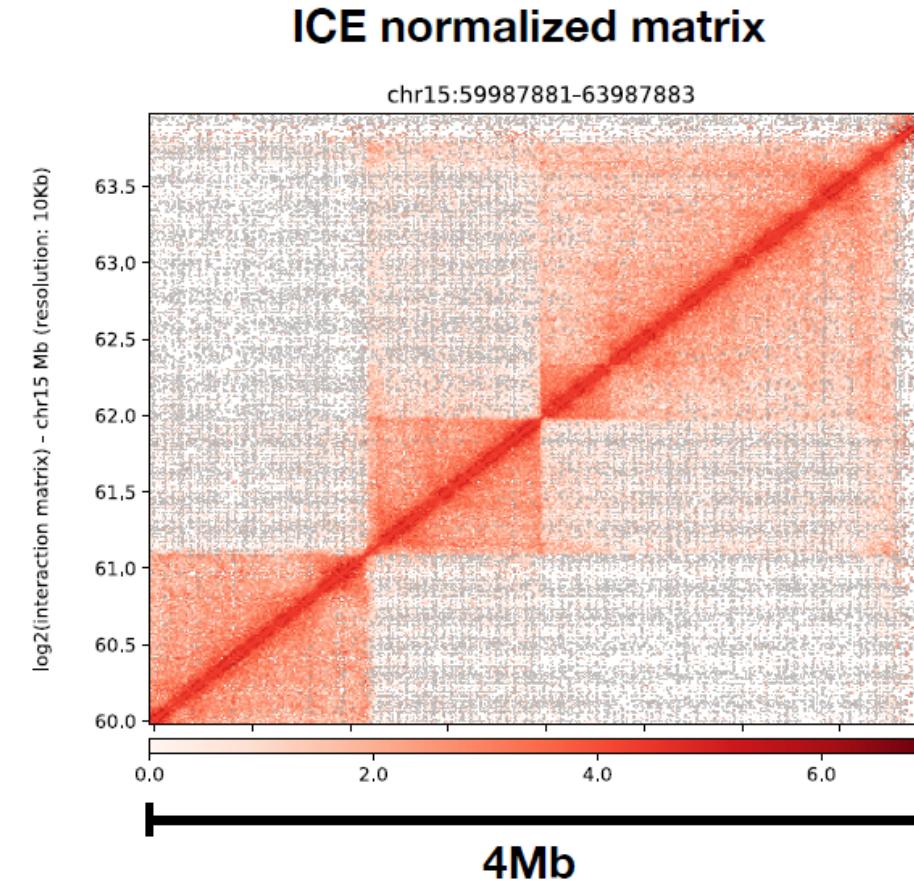
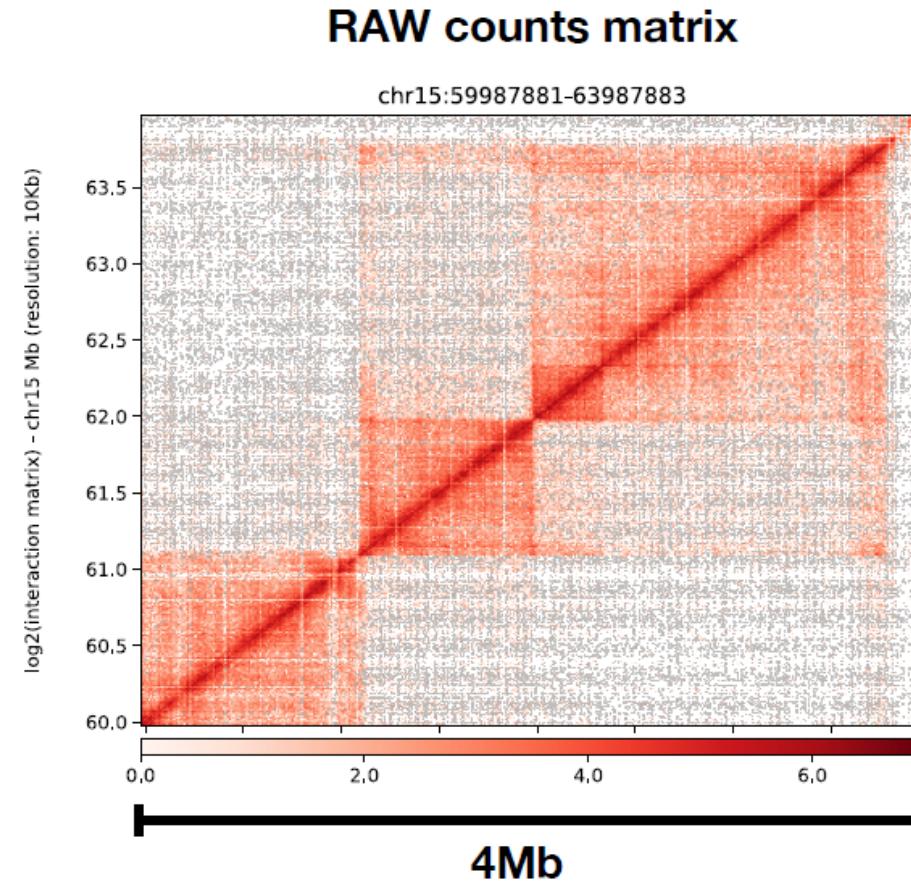


Different biases affect Hi-C contact matrices

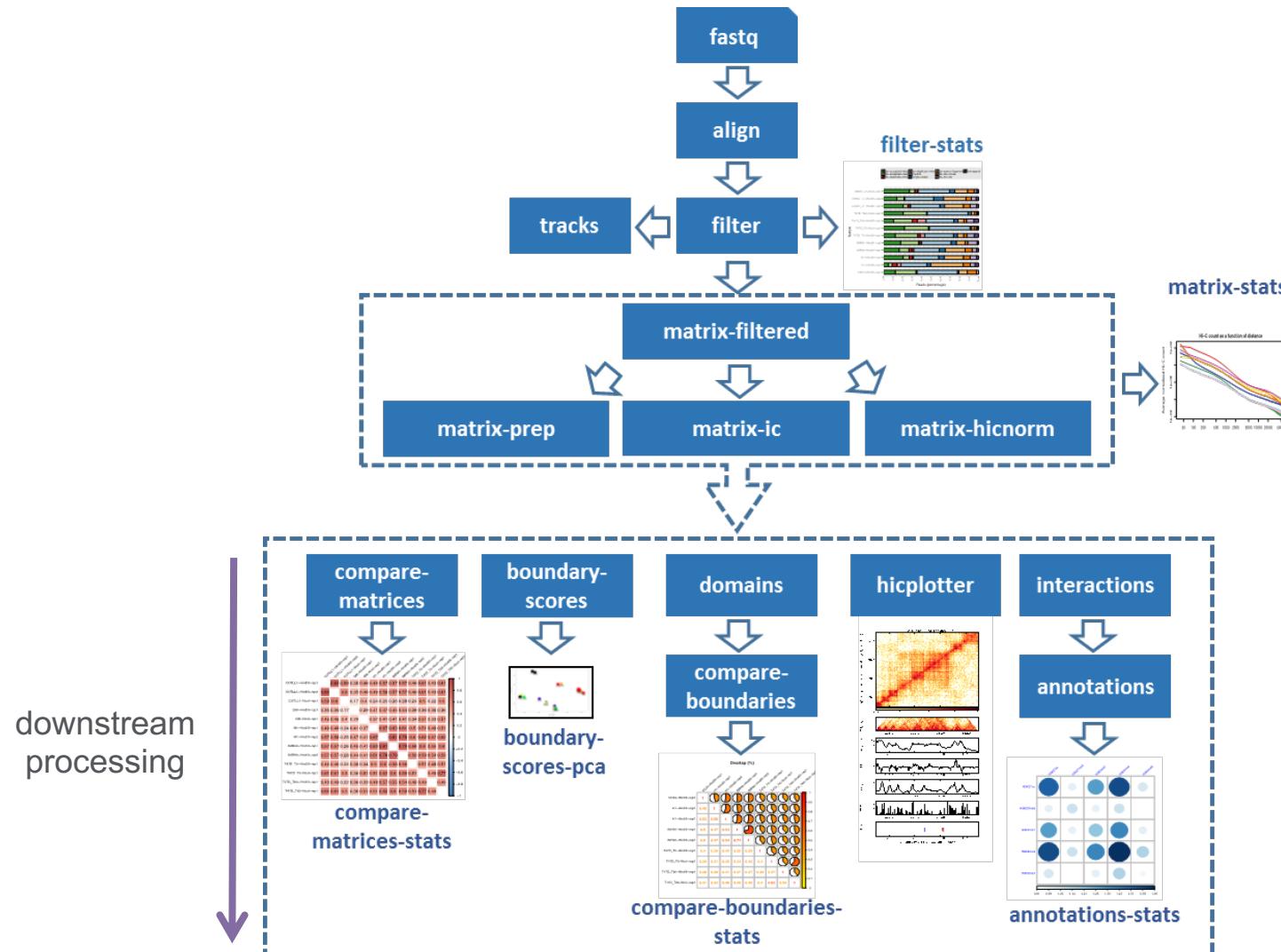
- Linear distance plays a role too
 - the closer two loci are in the linear genome, the more likely they are close in 3D space by chance
- Can be illustrated as an average loss of signal over distance
- Influenced by various factors, such as cell cycle, nuclear volume, technical biases etc.



Example of ICE-normalization



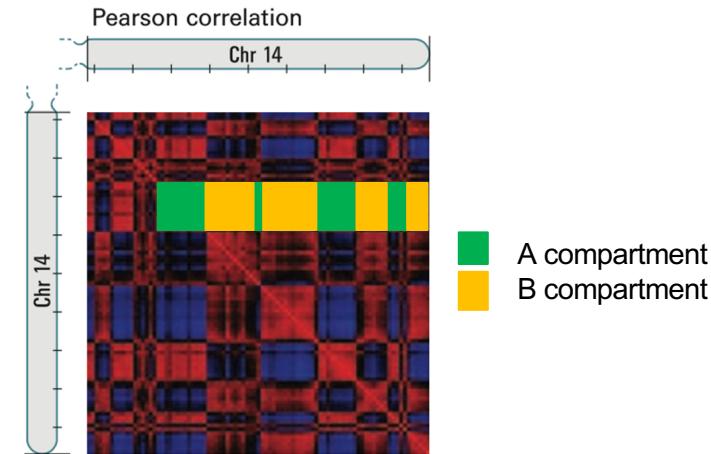
HiC-Bench & Identification of Chromatin Structures



HiC-bench (Lazaris et al., BMC Genomics 2017)

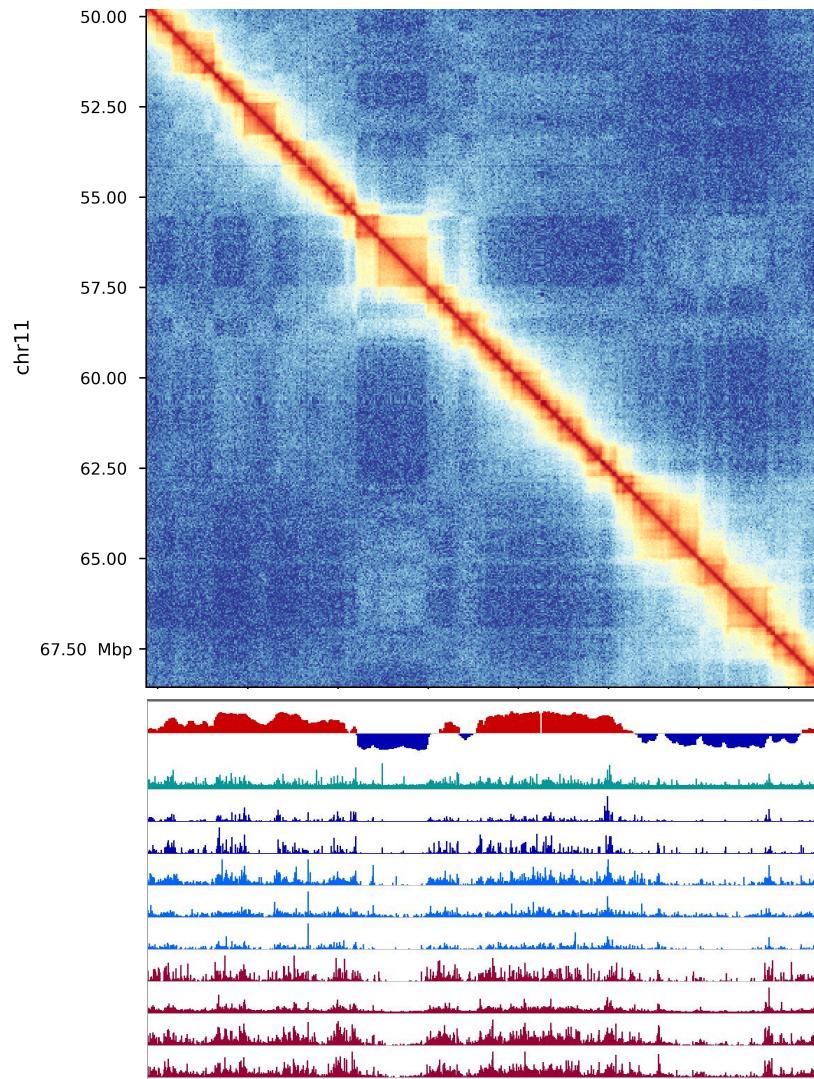
Compartment analysis

- Tools available for compartment analysis
 - Homer (<http://homer.ucsd.edu/homer/>)
 - Cscore (Zheng X, Zhen Y, Bioinformatics, 2018)
- Idea:
 - Identifying chromosome-wide regions of mutually exclusive interactions
 - Performing an eigenvector-analysis of the contact-matrix
 - Calculating Principal Component (PC) 1
- Score does not define A vs. B, which can be achieved by e.g. histone mark enrichments
 - A compartment is enriched with active histone marks
 - B compartment is enriched with repressive histone marks



Example of PC1 track -> compartment assignment via histone marks

Example for compartment analysis and assignment



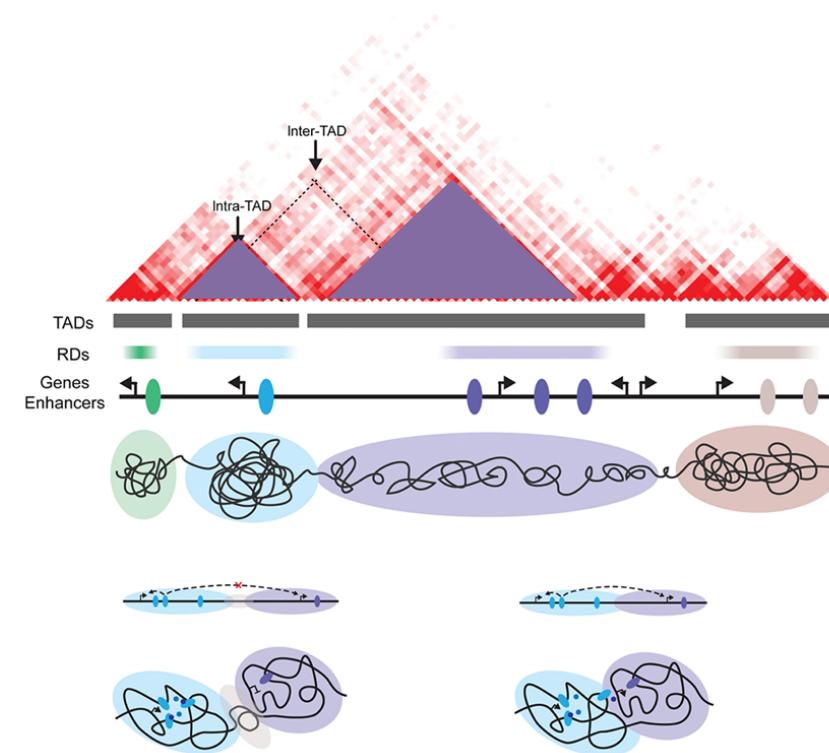
Kindly provided by Alex Polyzos

TAD calling



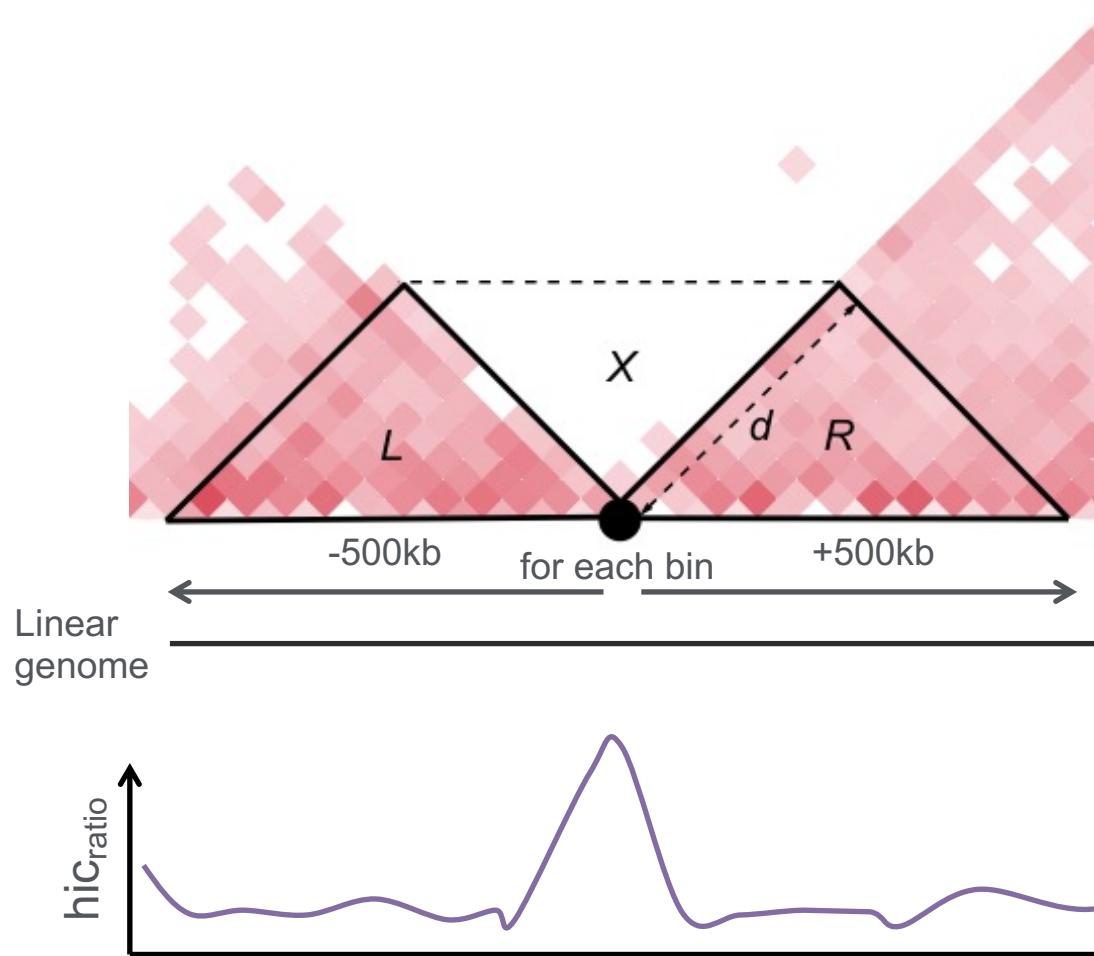
Identified TAD

- Various tools exist to identify TADs from Hi-C data
 - hic-ratio (Lazaris C et al., BMC Genomics, 2017)
 - Developed by Crane et al. (Crane E et al., Nature, 2015)
 - TopDom (Shin H et al., NAR, 2015)
 - TADbit (Serra F et al., Plos Comp Bio, 2017)
 - ...
- Benchmarked in multiple studies:
 - Lazaris C et al., BMC Genomics, 2017
 - Zufferey M et al., Genome Biology, 2018



Remeseiro et al. *WIREs Dev Biol* 2015.

Insulation score to determine TAD boundaries



$$hic_{inter} = \text{sum}(X)$$

$$hic_{intra-max} = \max(\text{sum}(L), \text{sum}(R))$$

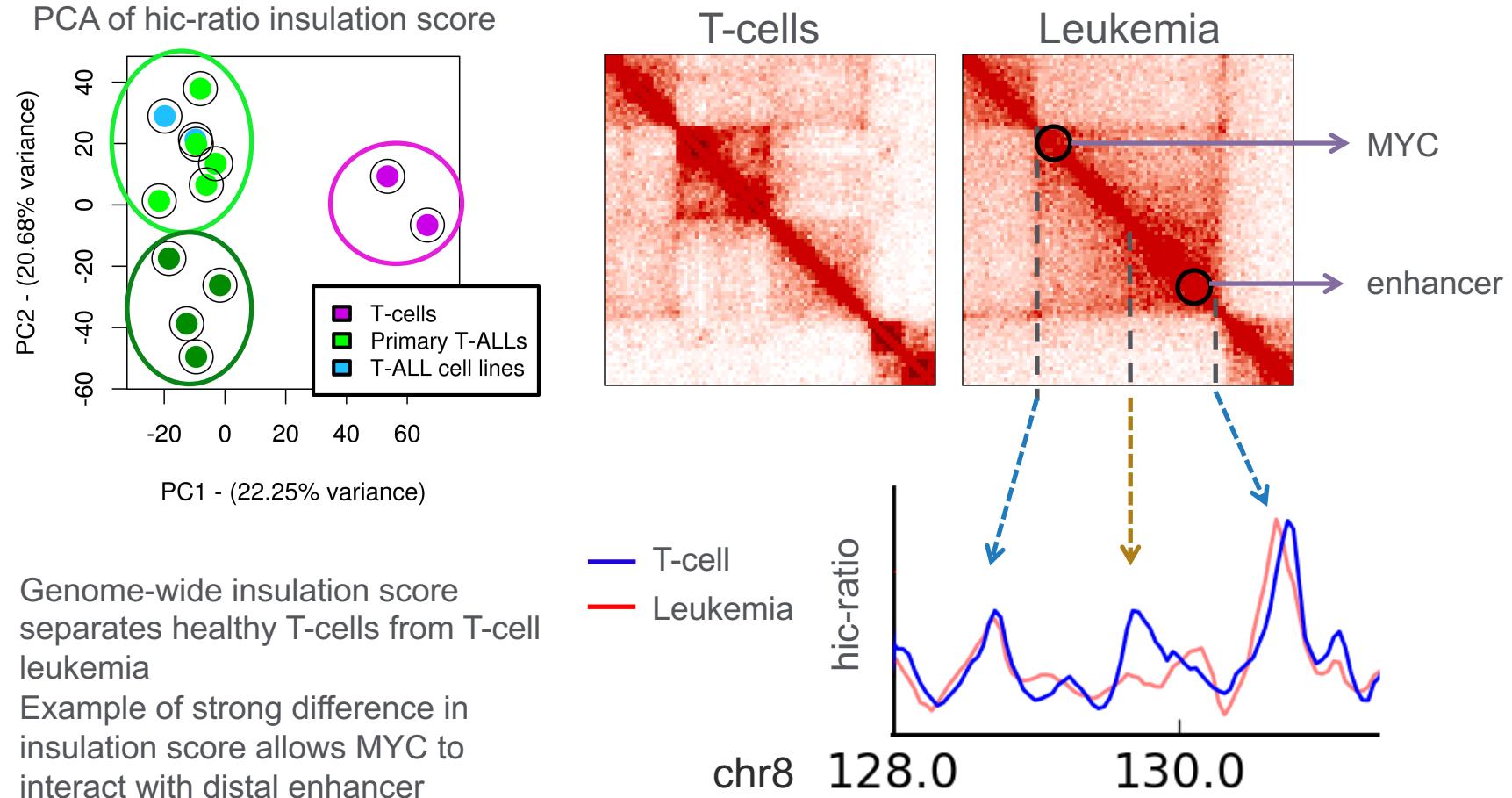
$$hic_{ratio} = hic_{intra-max} / hic_{inter}$$

$hic_{inter} \ll hic_{intra-max} \rightarrow hic_{ratio}$ increases

$hic_{inter} \gg hic_{intra-max} \rightarrow hic_{ratio}$ decreases

→ TAD boundaries are represented by local maxima in the hic-ratio score

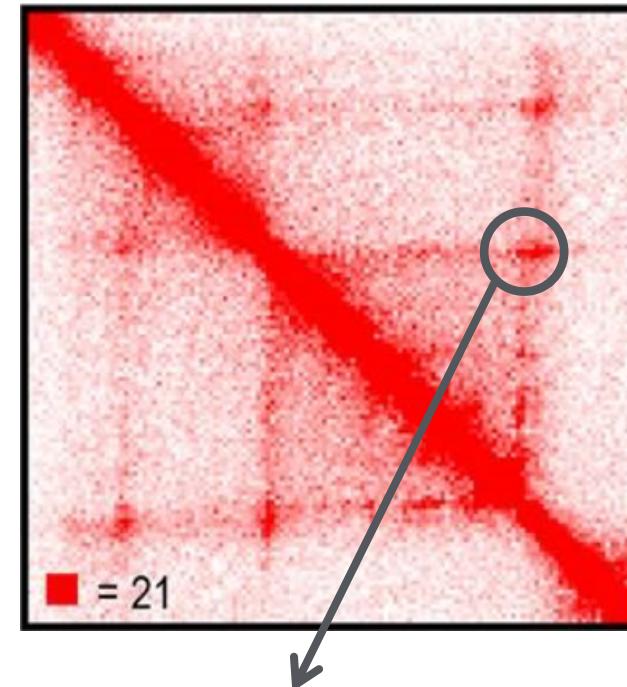
HiC-ratio identifies insulation loss in T-cell leukemia allowing overexpression of the proto-oncogene MYC



- Genome-wide insulation score separates healthy T-cells from T-cell leukemia
- Example of strong difference in insulation score allows MYC to interact with distal enhancer

Loop calling

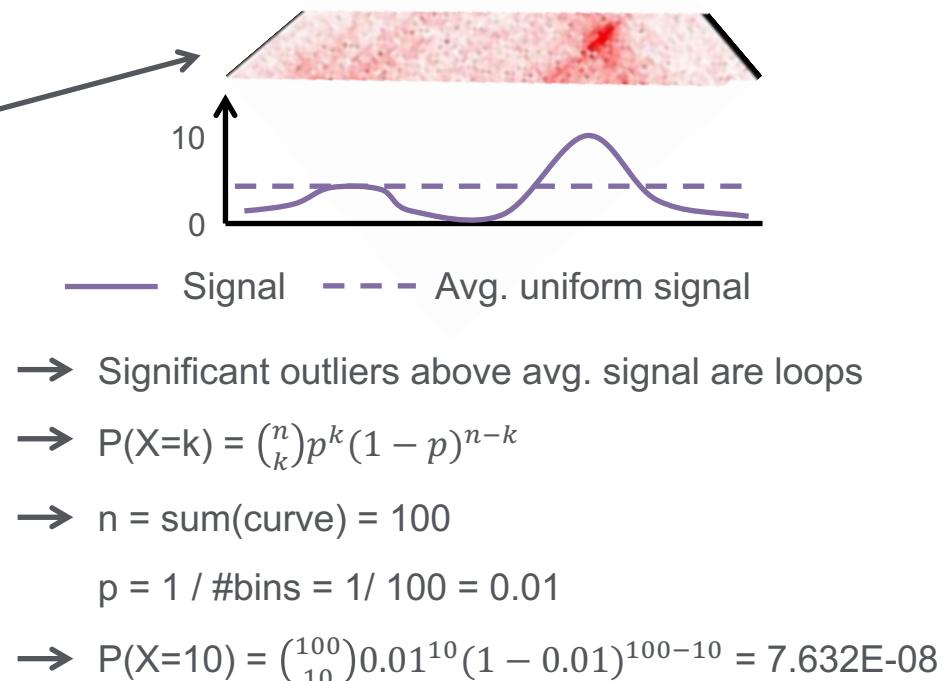
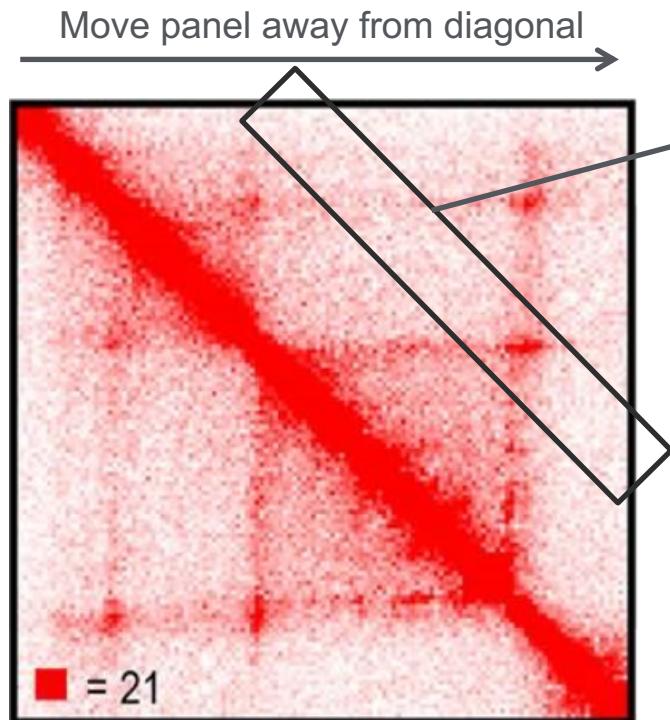
- Tools are being developed, sometimes specifically for other techniques such as HiChIP/ChIA-PET
 - FitHiC / mango (Phanstiel et al., Bioinformatics, 2015)
 - HiCCUPS (Durand N et al., Cell Systems, 2016)
 - Origami (Weintraub A et al., Cell, 2017)
 - HiChIPPER (Lareau et al., Nature Methods, 2018)



Loop / contact

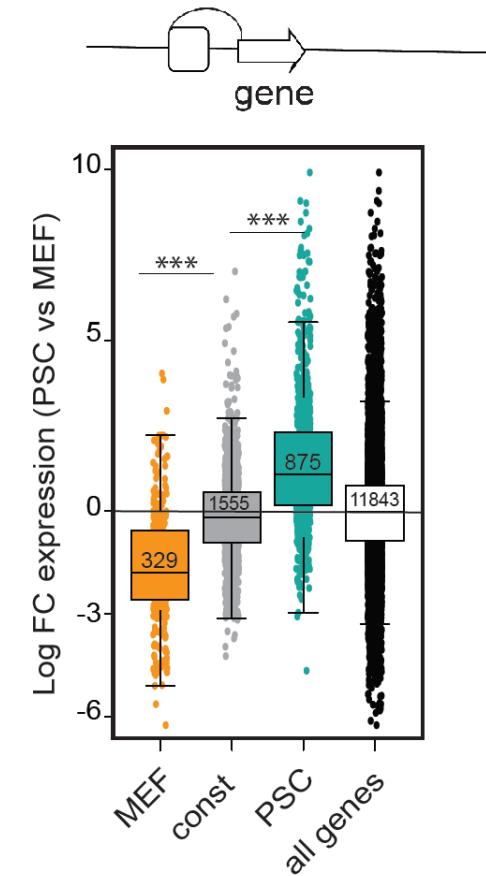
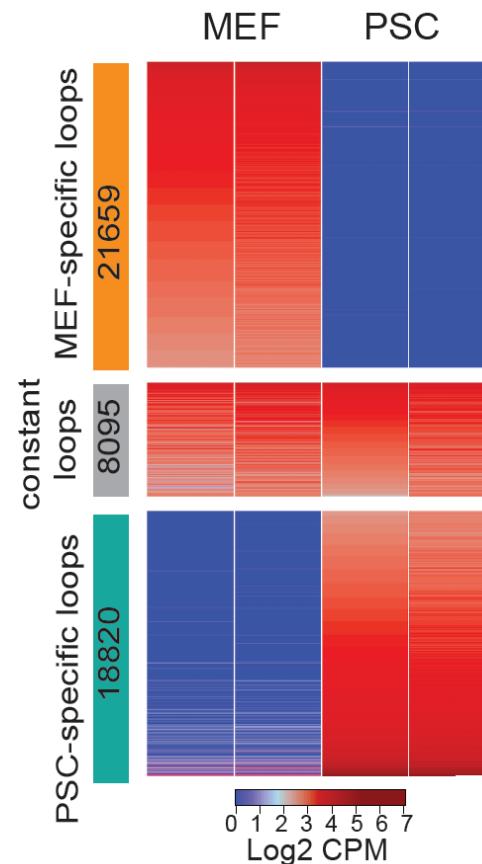
Loop calling - mango

- Each linear distance is treated separately to overcome loss of signal over far distances
- Task: identifying significant outliers per distance
- Signal can be modeled with the binomial distribution



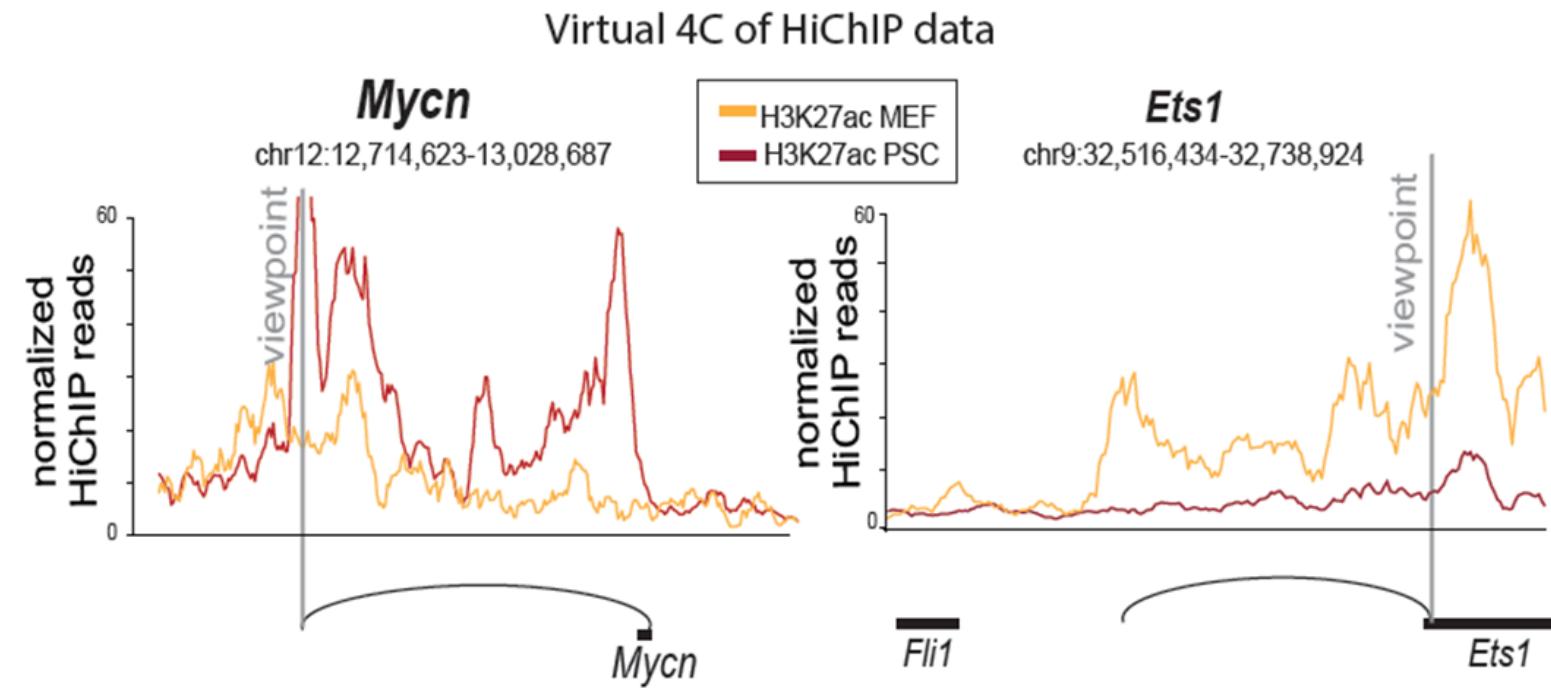
HiChIP loop calling reveals cell type specific interactions of stem cell genes

- Differences in loop formation between stem cells (PSC) and embryonic fibroblasts (MEF)
 - We found differences of cell-type specific loops in both directions, as well as highly stable loops across both conditions
- Gene expression is mediated by cell-type specific enhancers and loops
 - PSC specific loop shows higher gene expression in PSC of connected genes
 - MEF specific loop shows higher gene expression in MEF of connected genes



HiChIP loop calling reveals cell type specific interactions of stem cell genes

- Virtual 4C extracts interactions from the Hi-C matrix from 1 row
→ depicts all nearby 3D interactions with 1 positions
- Cell-type specific loops are accompanied by enhancer activity and expression changes



with Dafne Di Giammartino, Alex Polyzos and Effie Apostolou

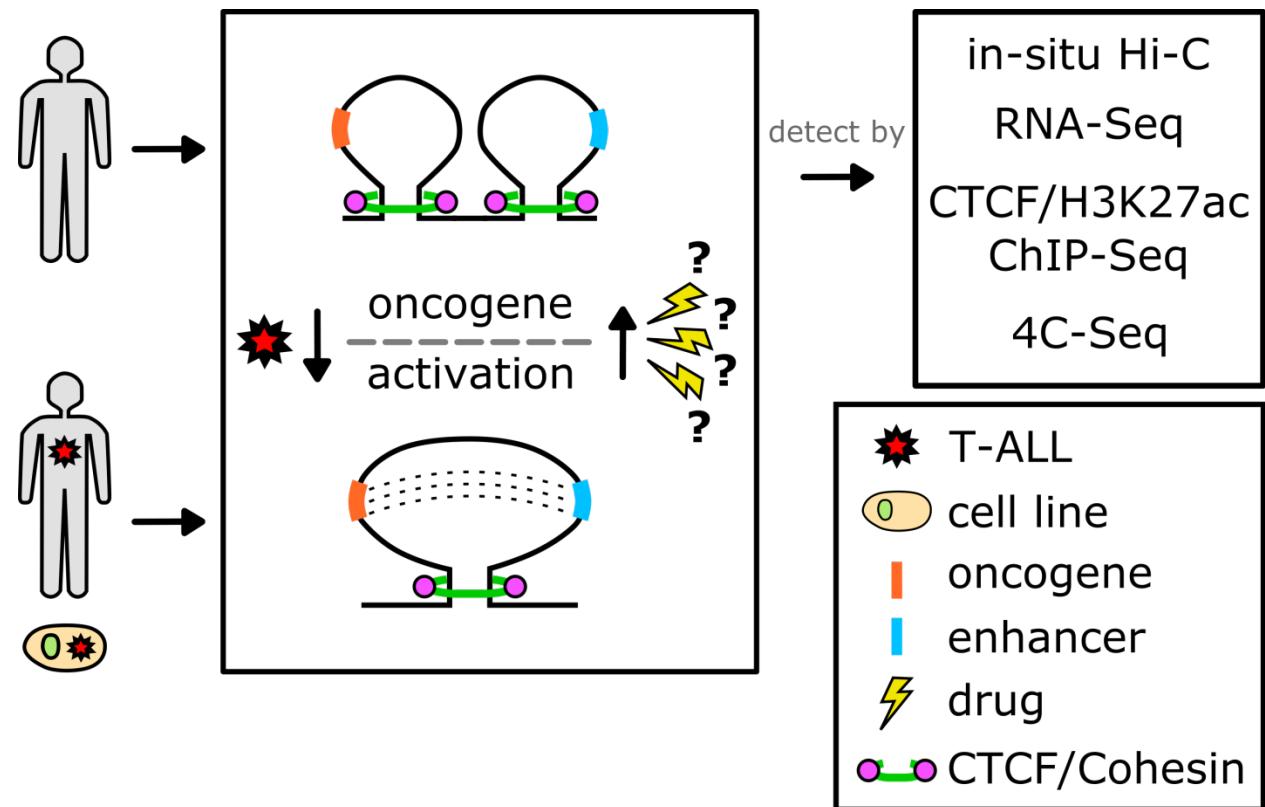
Pipelines & visualization

- Juicer + Juicebox (Lieberman-Aiden lab: <https://github.com/aidenlab/juicer> + <https://aidenlab.org/juicebox/>)
- HiGlass (Gehlenborg lab: <https://higlass.io/>)
- HiC-plotter (<https://github.com/kcakdemir/HiCPlotter>)
- HiC-Explorer (<https://github.com/deeptools/HiCExplorer>)
- HicPro (<https://github.com/nservant/HiC-Pro>)

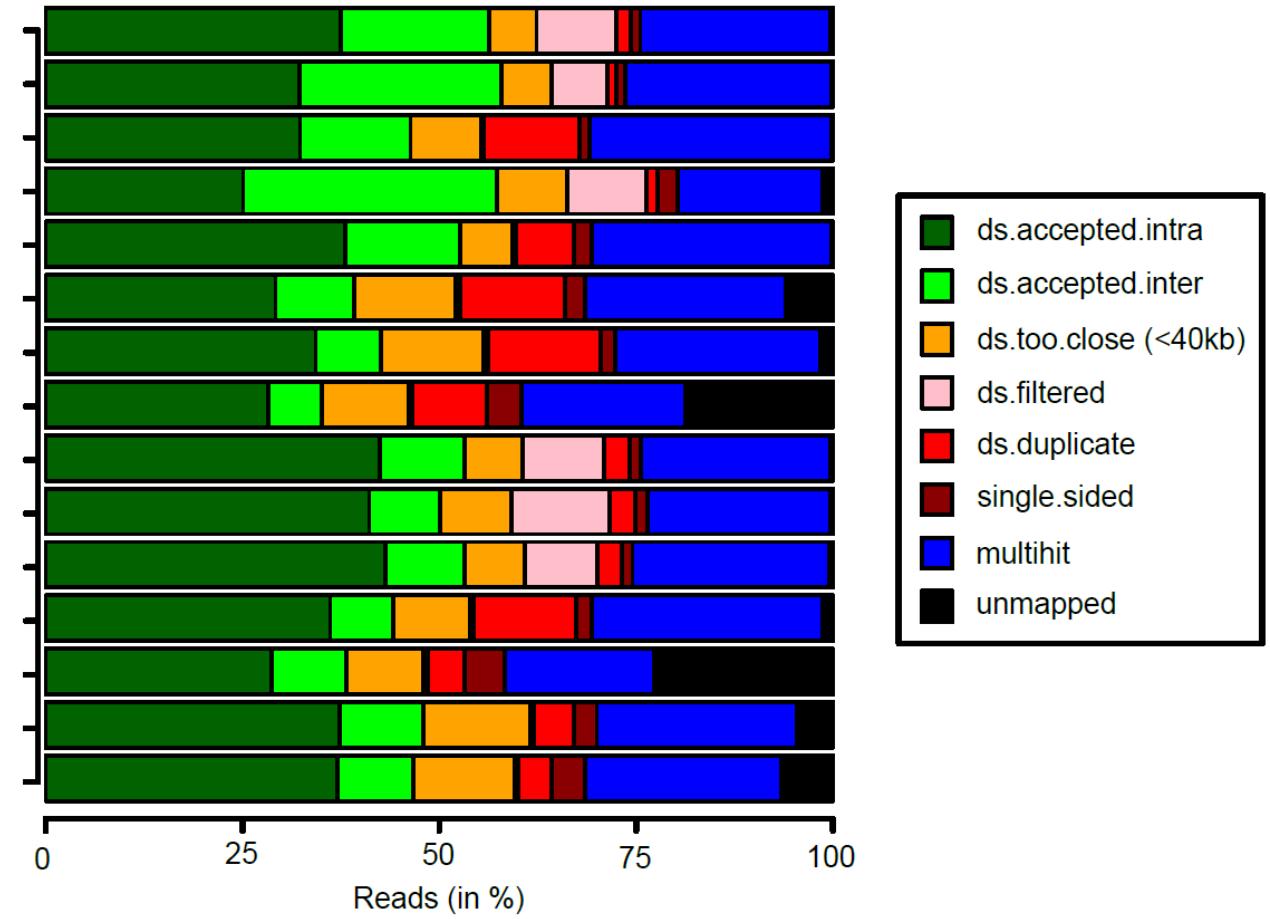
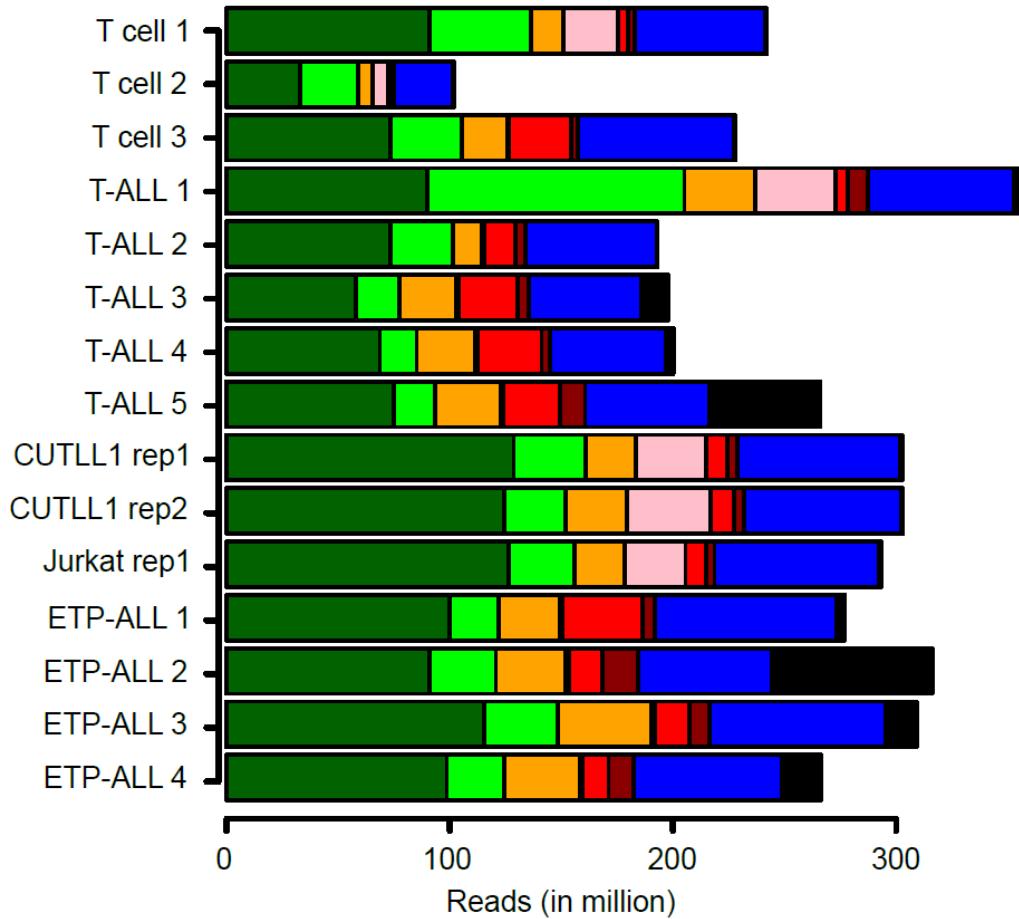
CASE STUDY: THE 3D GENOME OF ACUTE LEUKEMIA PATIENTS

Study design

- Profiled primary patient samples using Hi-C
 - Performed Hi-C in prototypical leukemic cell lines
 - Profiled patient samples and cell lines by RNA-seq and CTCF ChIP-seq
 - Studied changes in intra/inter TAD interactions and their impact on gene expression

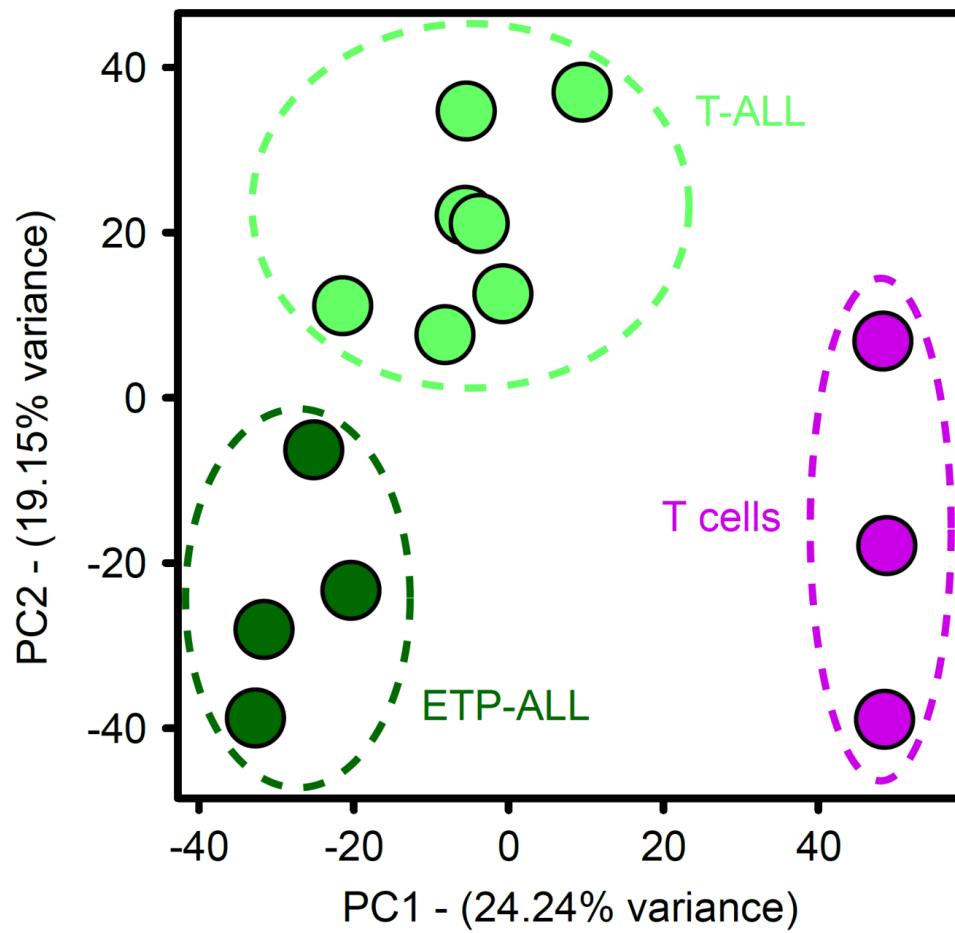


Data quality assessment

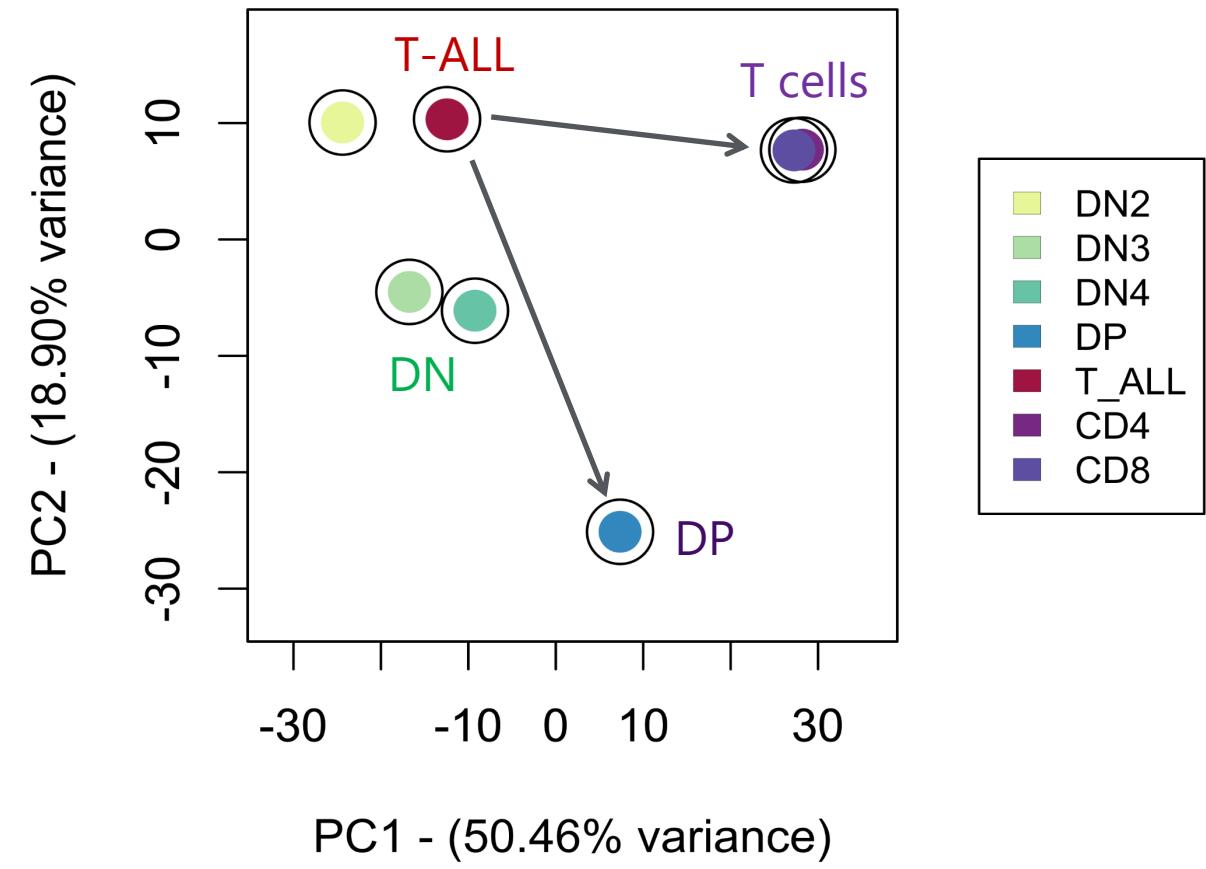


Leukemia samples cluster separately from controls

Human Hi-C data



Mouse Hi-C data



TAD analysis outline

TAD comparison: leukemia vs controls

Case #1:
TADs with conserved boundaries

Intra-TAD changes

Case #2:
TADs with gained/lost boundaries

TAD insulation changes

Changes in intra-TAD topology

TAD comparison: leukemia vs controls

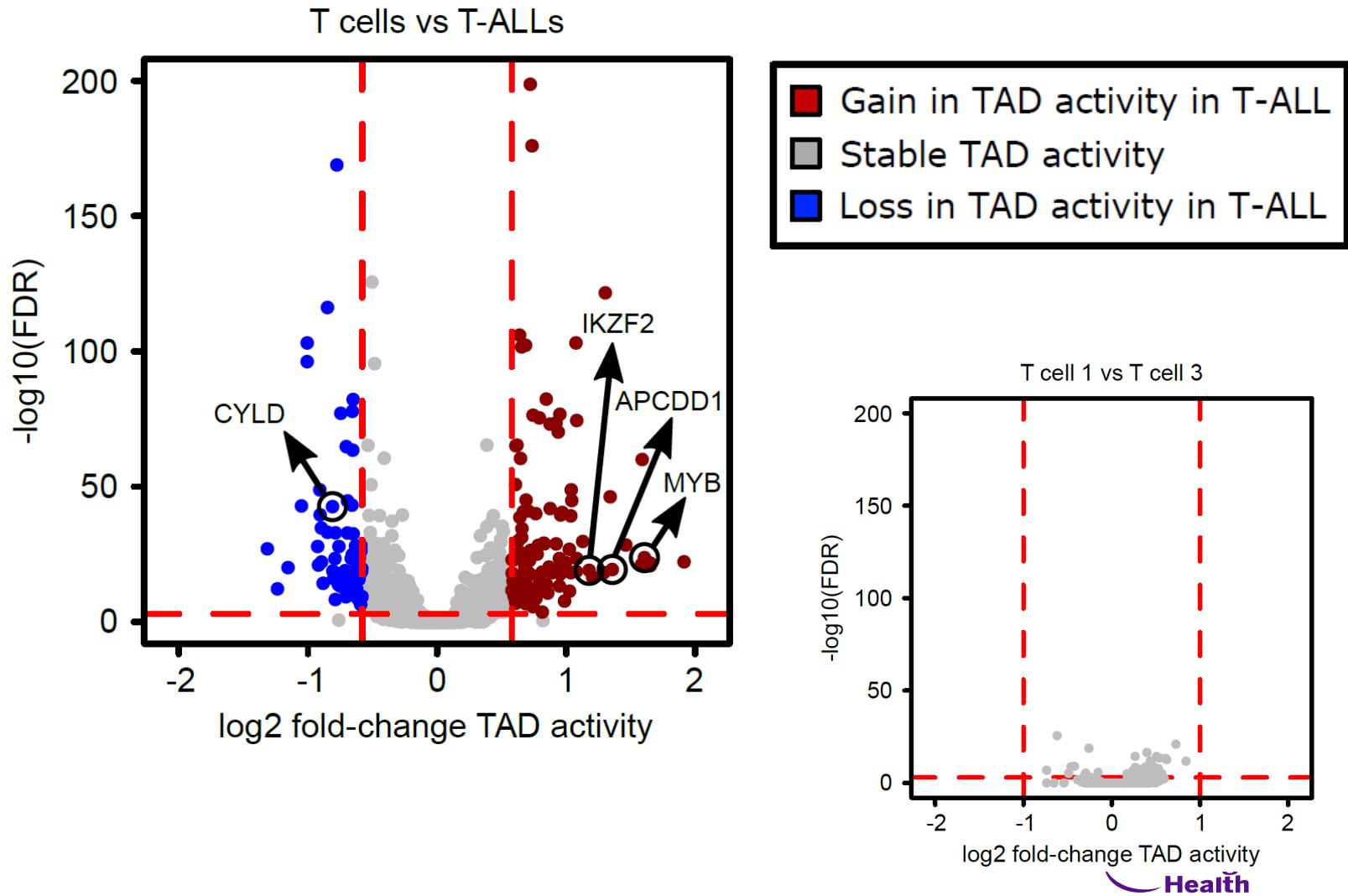
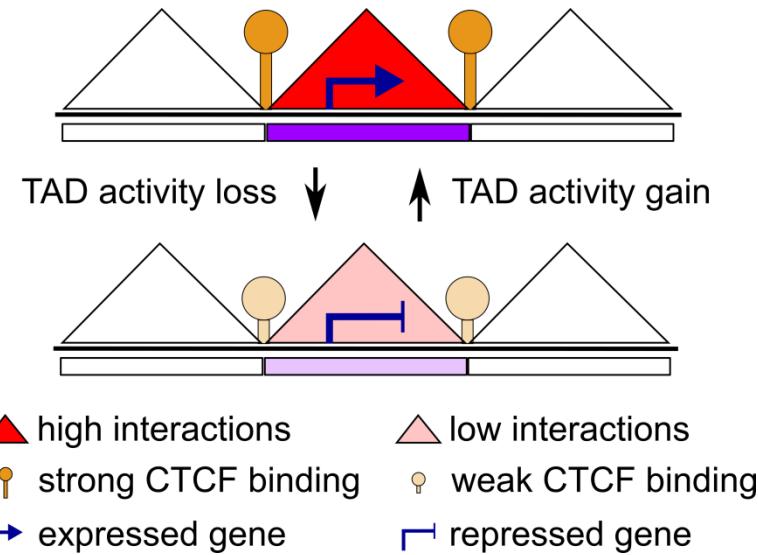
Case #1:
TADs with conserved boundaries

Intra-TAD changes

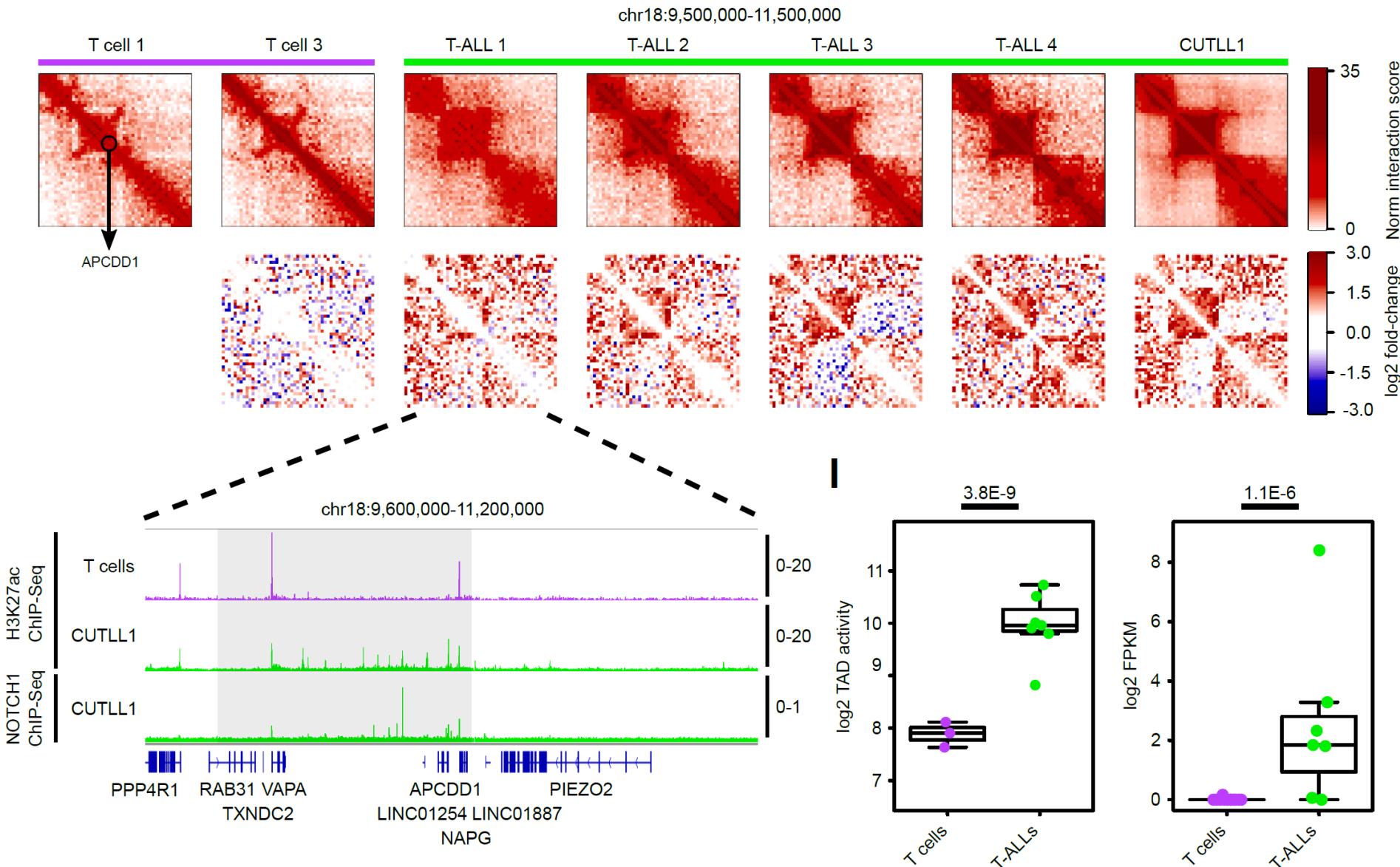
Case #2:
TADs with gained/lost boundaries

TAD insulation changes

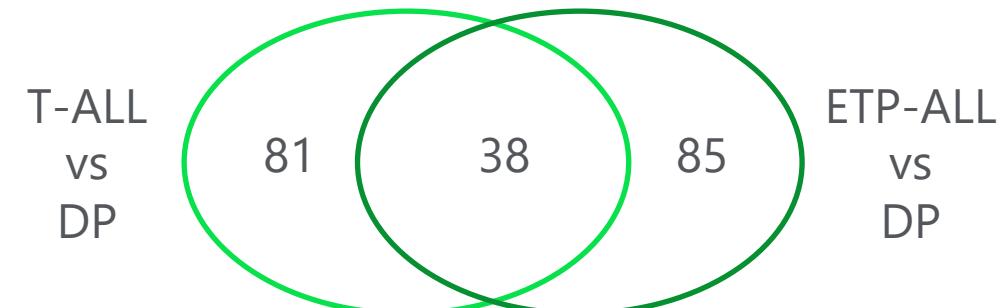
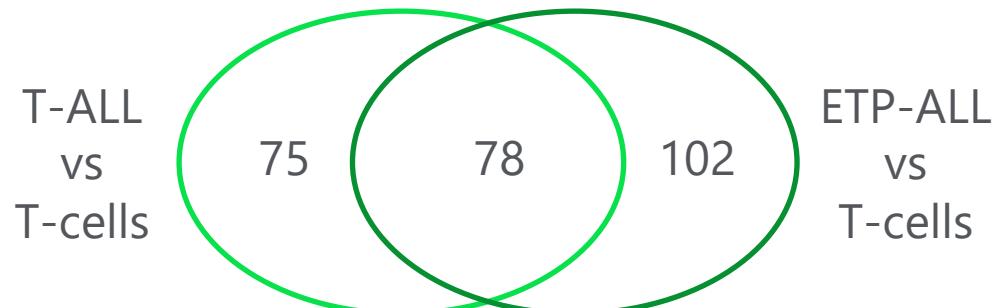
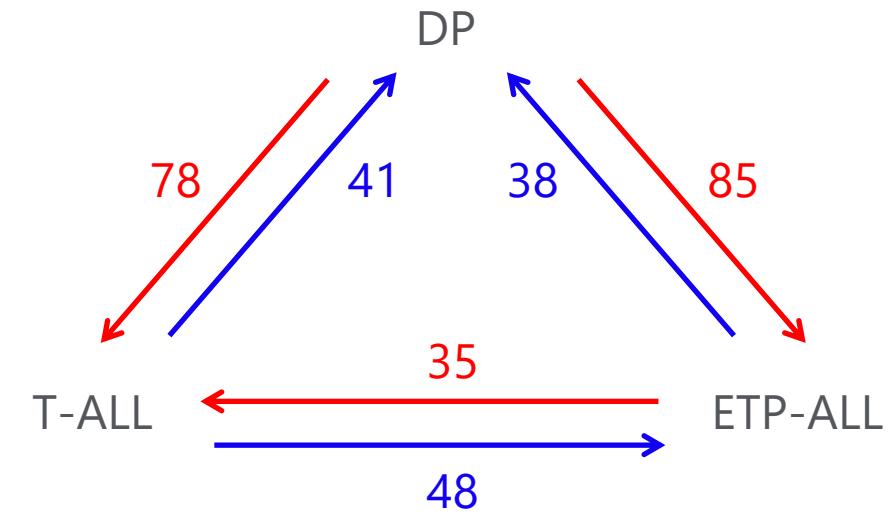
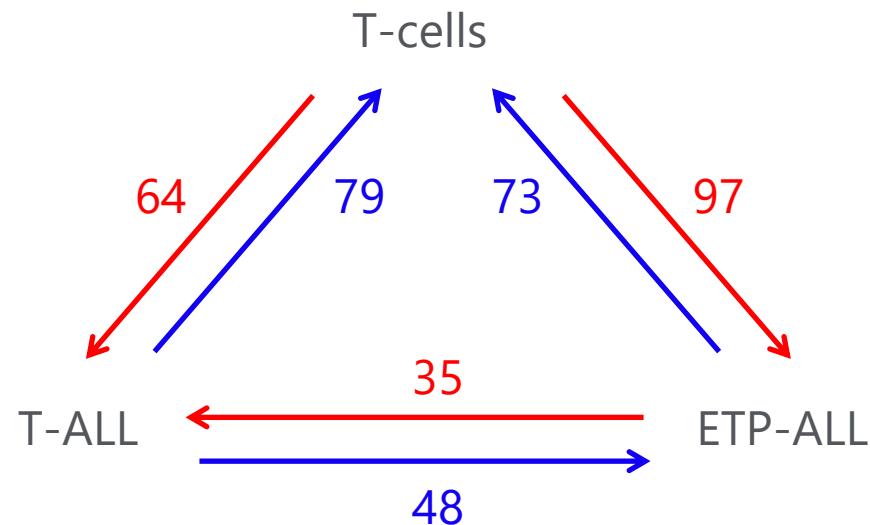
Changes in intra-TAD interactions (TAD “activity”)



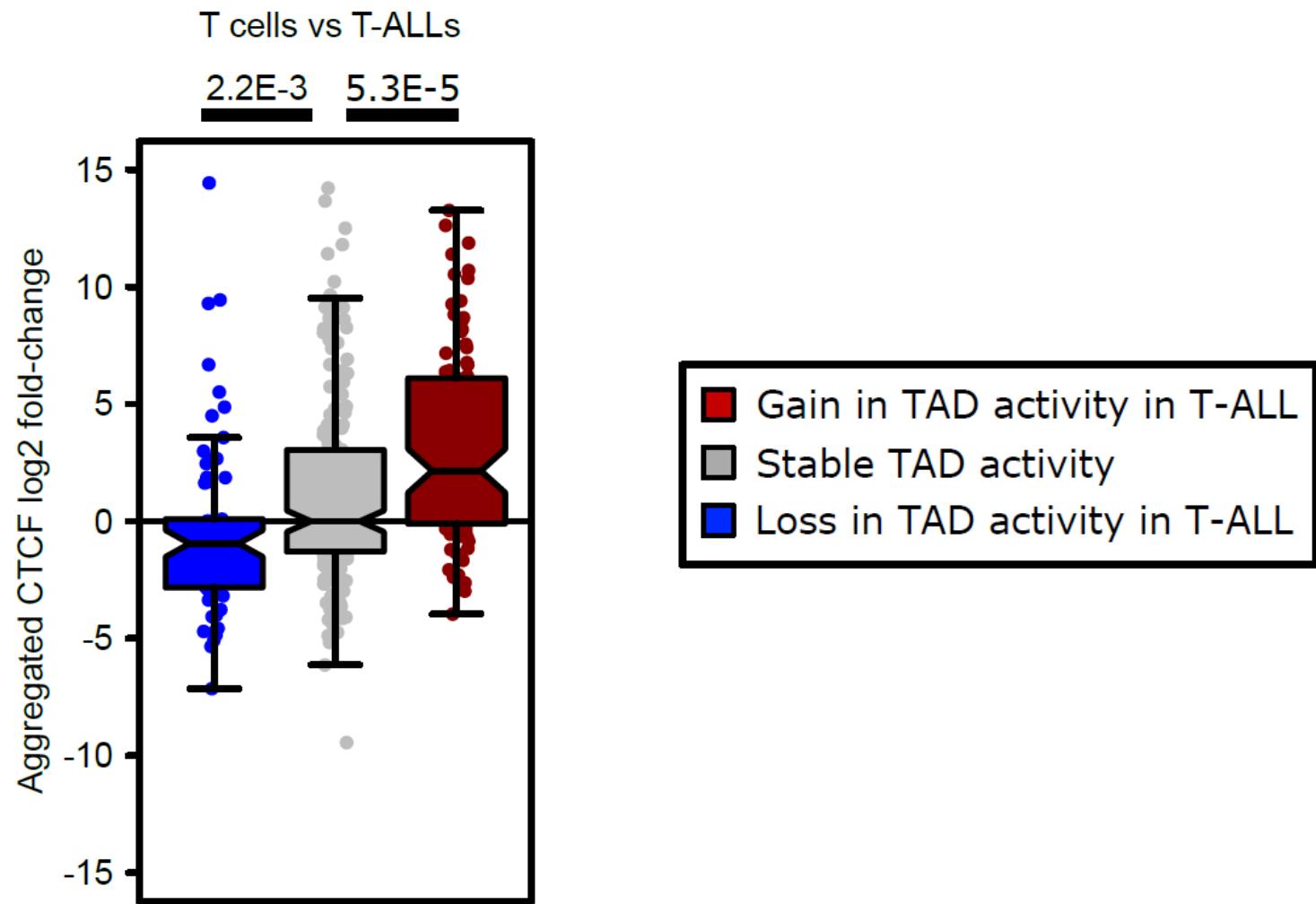
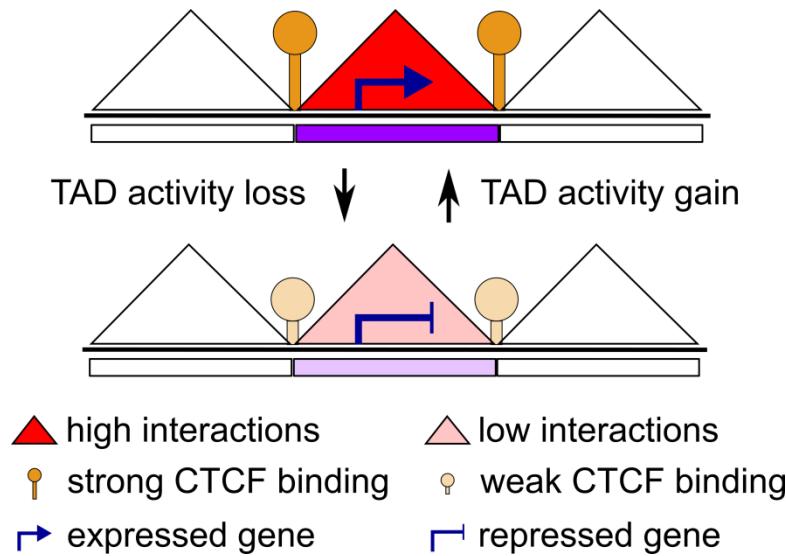
Intra-TAD activity example: APCDD1



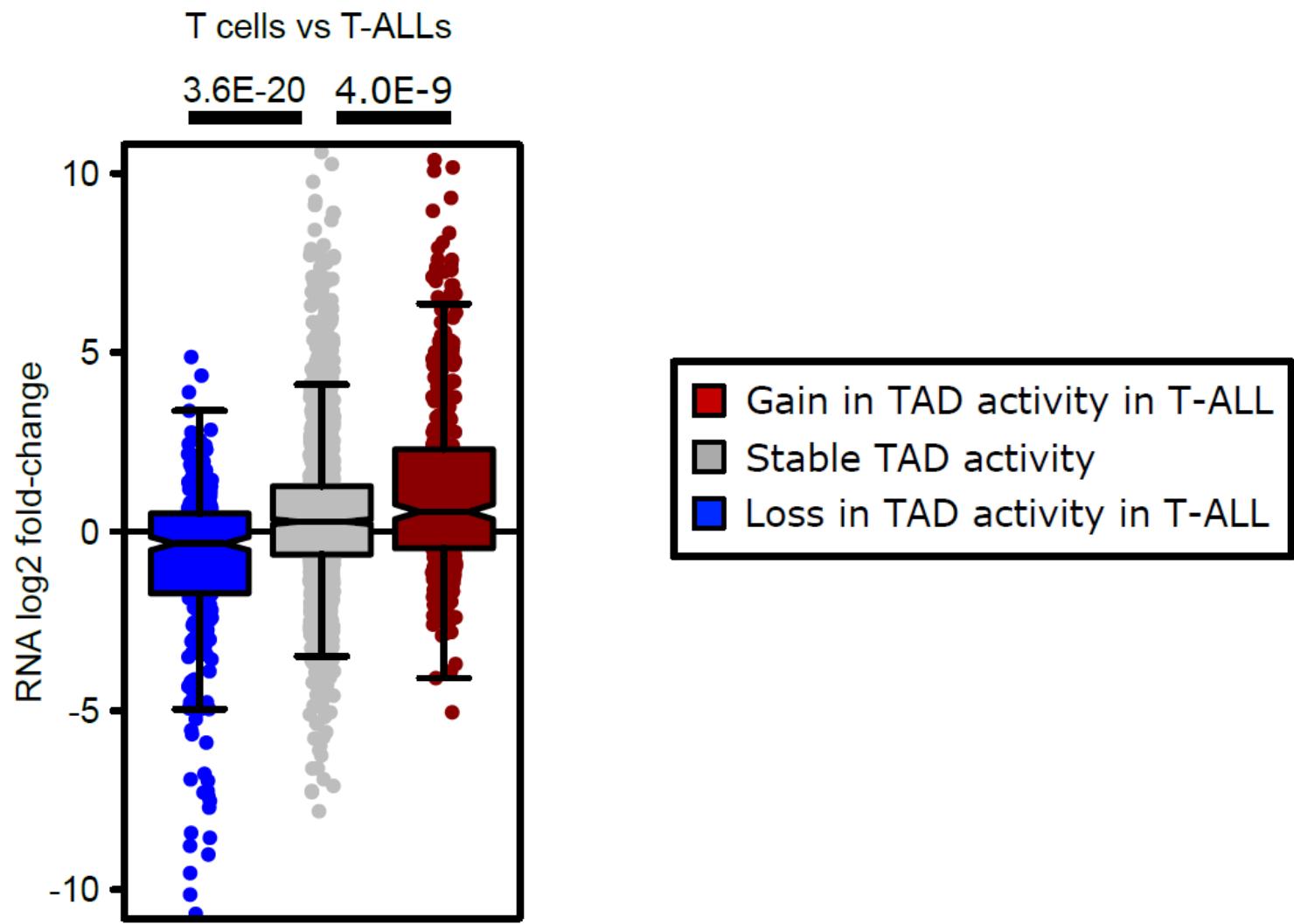
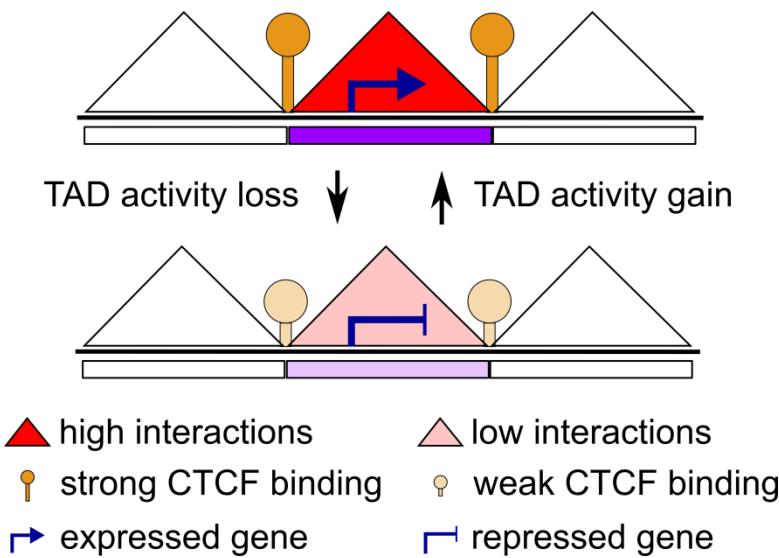
T-ALL, ETP-ALL, T cells and DP cells



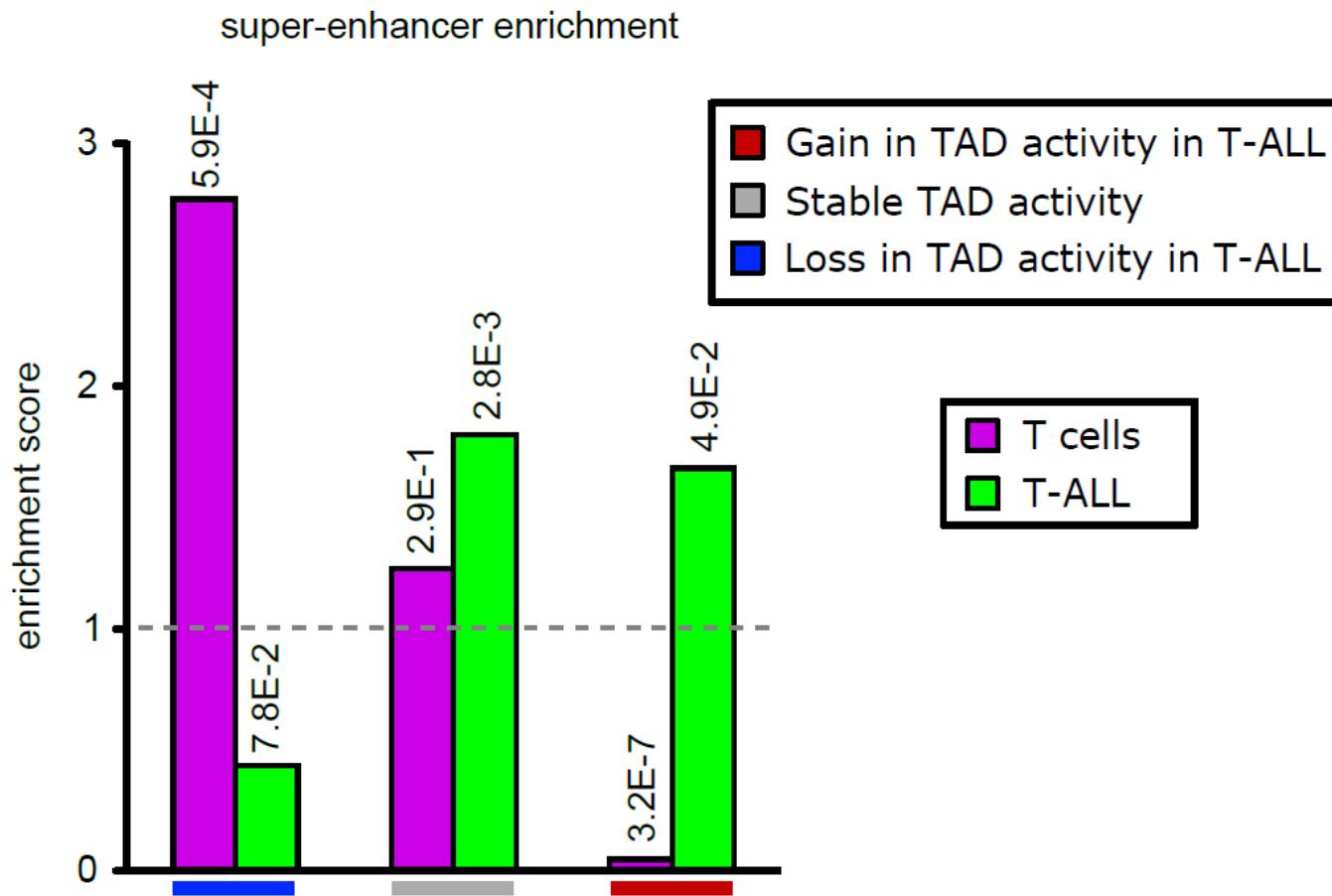
TAD activity associated with stronger CTCF binding



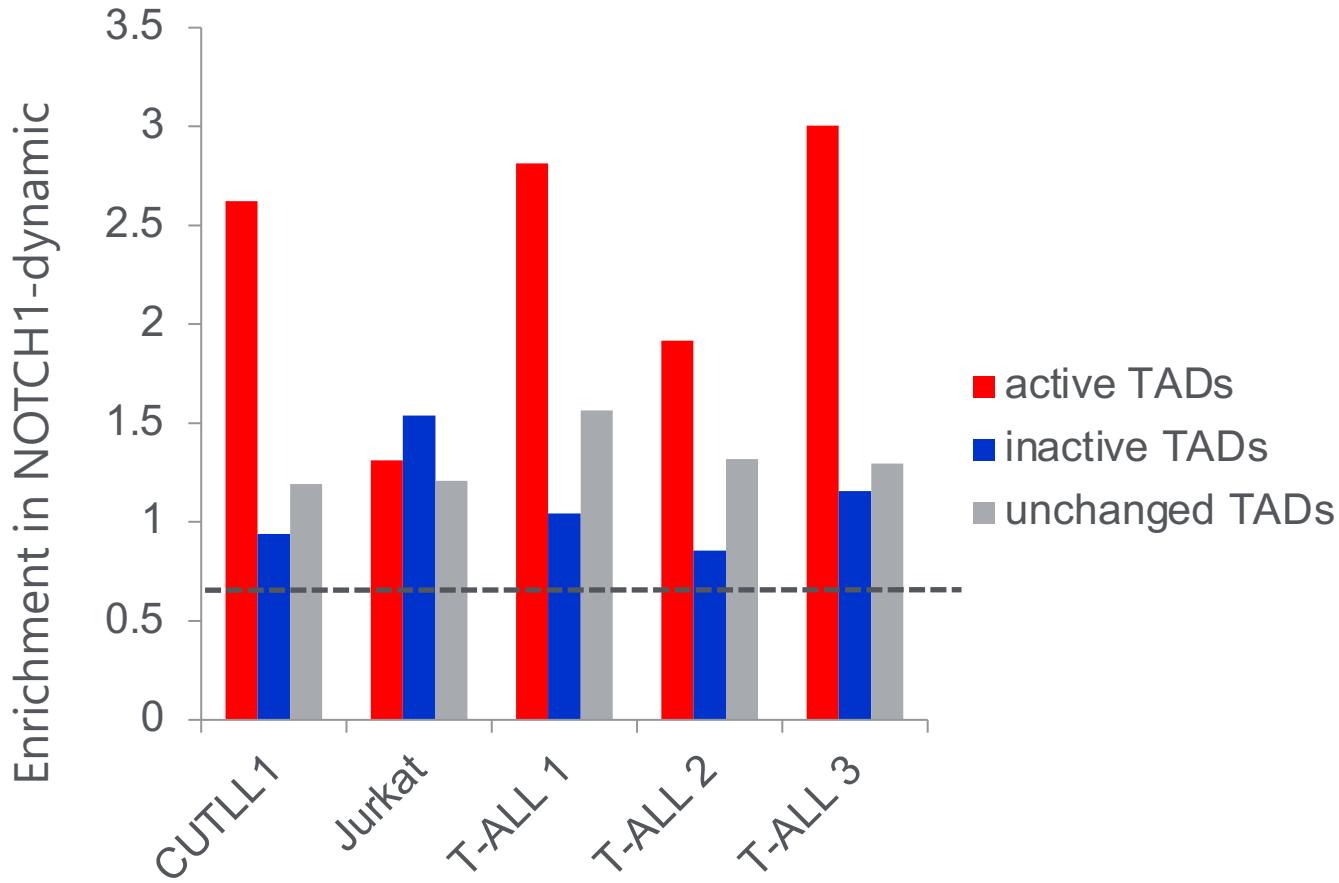
TAD activity correlated with gene expression changes



Intra-TAD activity changes enriched in super-enhancers

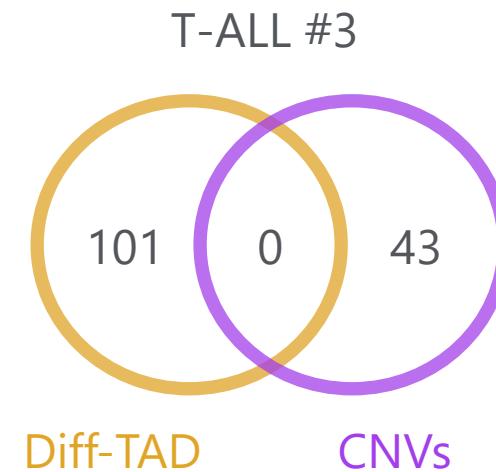
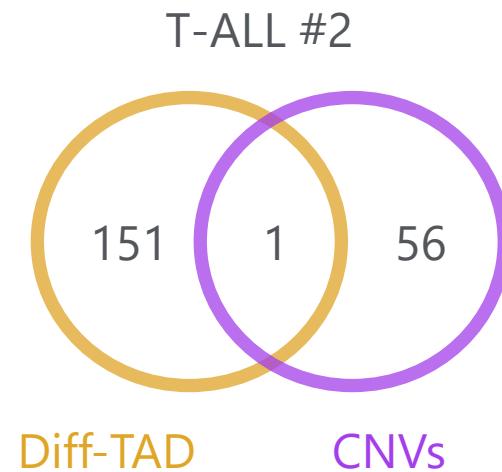
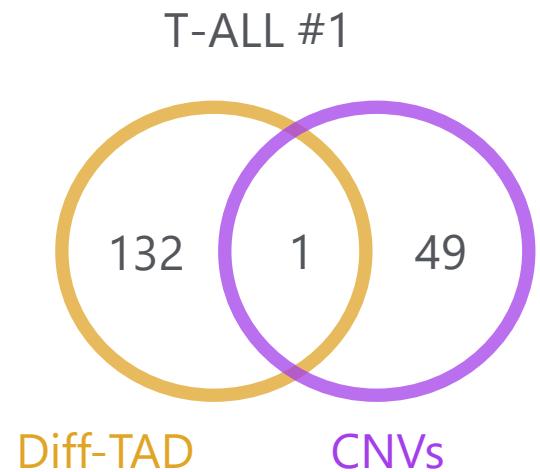


TAD activation overlaps with NOTCH1-dynamic sites



CNVs do not affect intra-TAD activity in T-ALL

- Performed WGS in primary T-ALLs
- Overlapped CNVs (gain/loss) with increased/decreased intra-TAD activity



Changes in inter-TAD topology

TAD comparison: leukemia vs controls

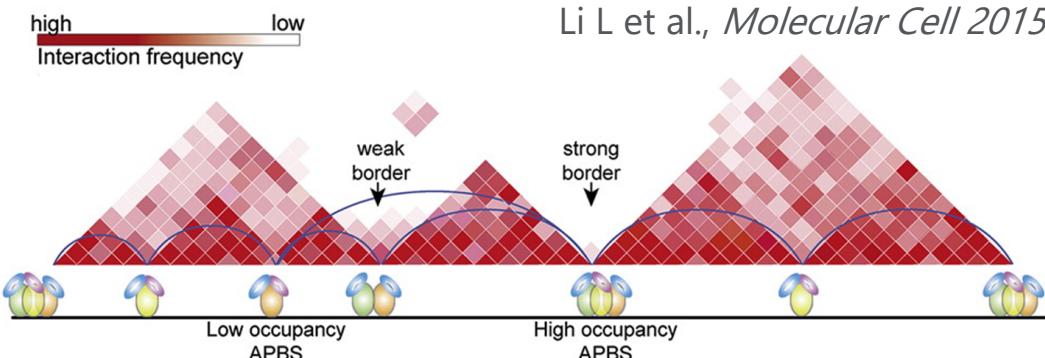
Case #1:
TADs with conserved boundaries

Intra-TAD changes

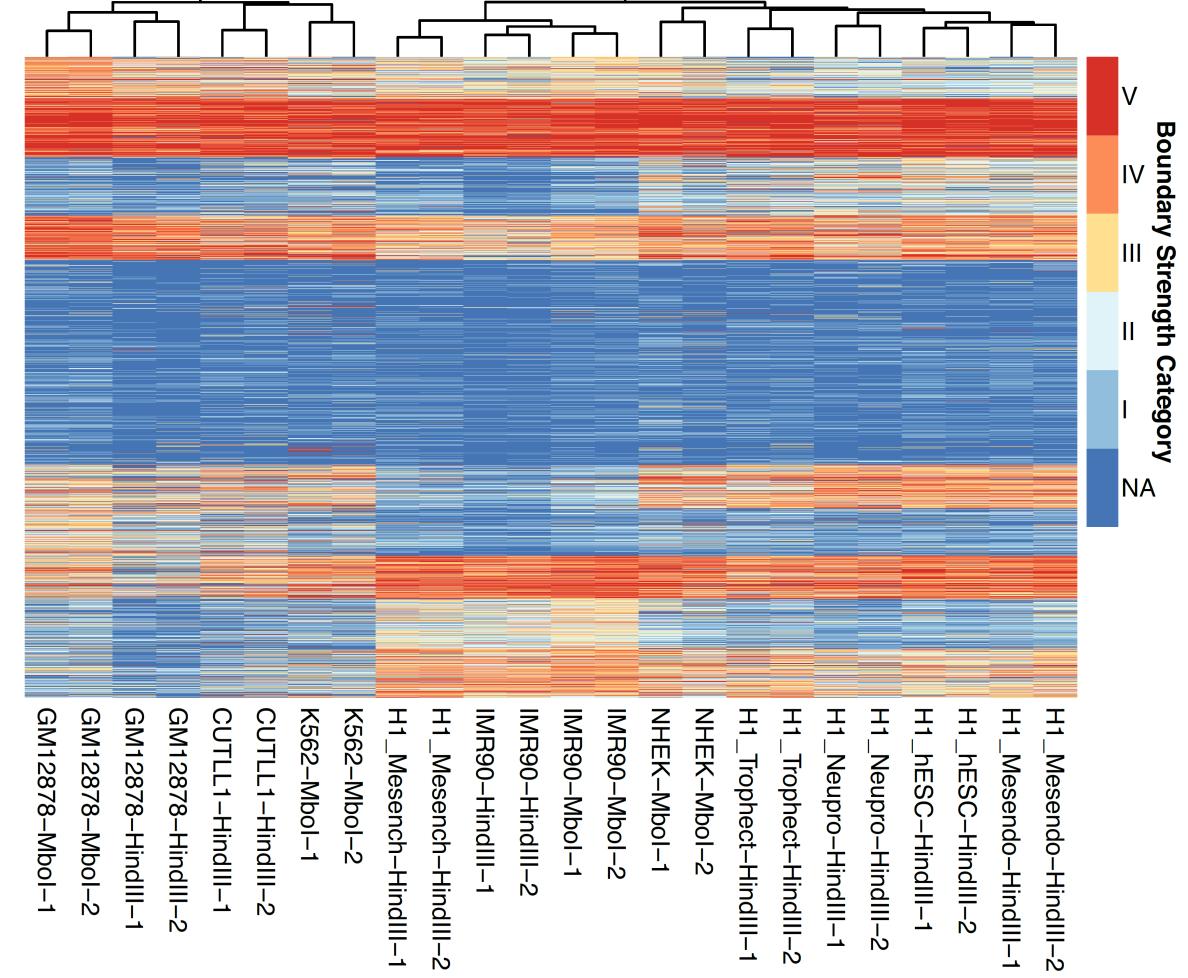
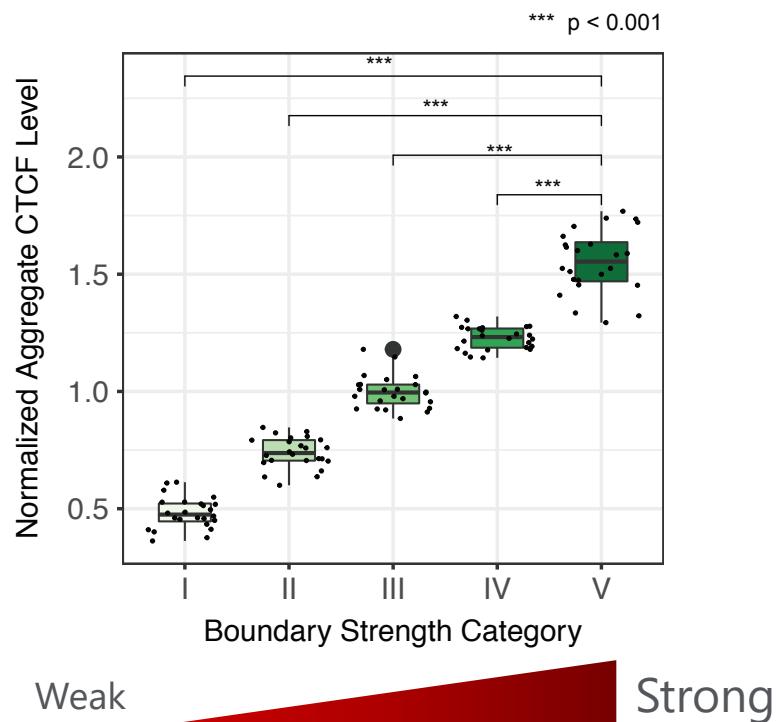
Case #2:
TADs with gained/lost boundaries

TAD insulation changes

TAD boundary strength correlated with CTCF binding

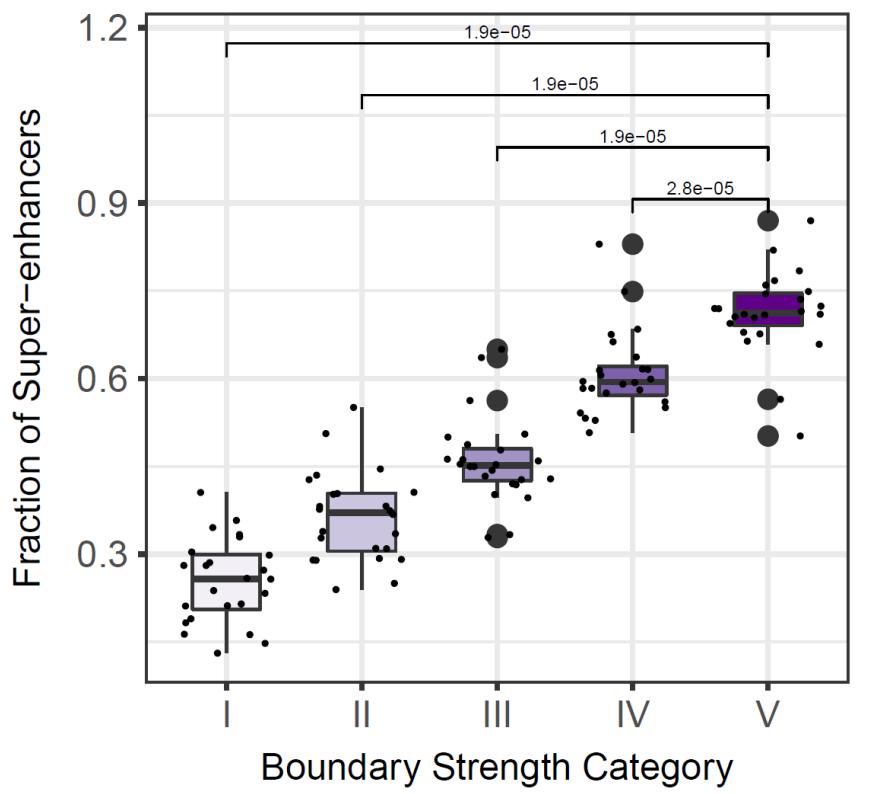


Li L et al., *Molecular Cell* 2015

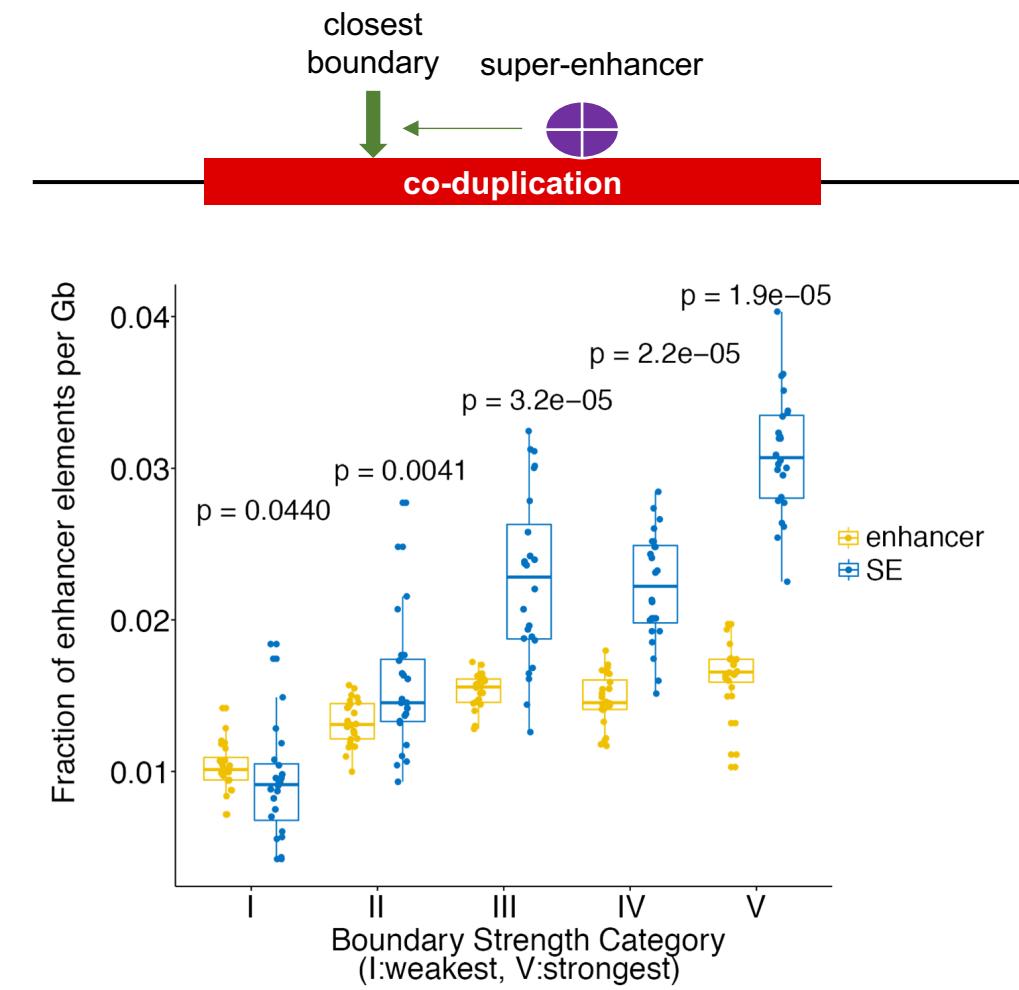


Gong*, Lazaris* et al., *Nature Communications* (2018)

Super-enhancers are insulated by stronger boundaries

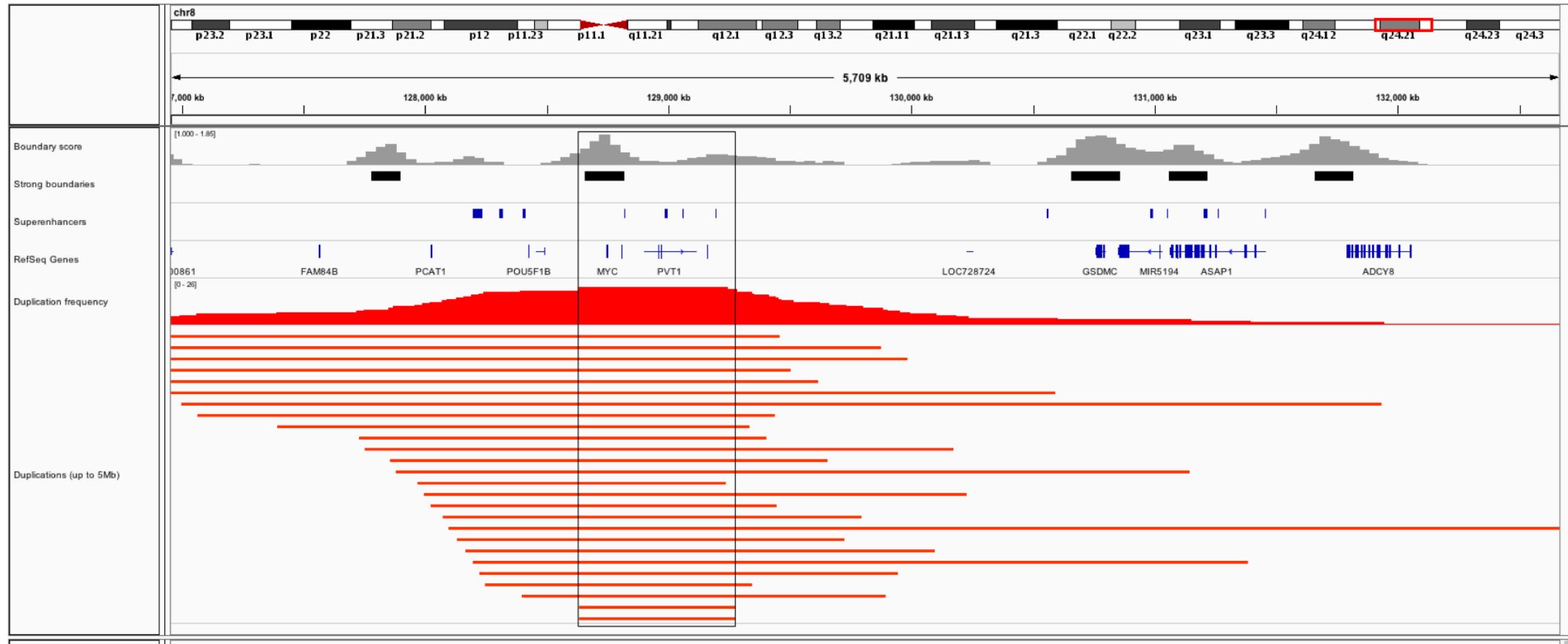


Weak Strong

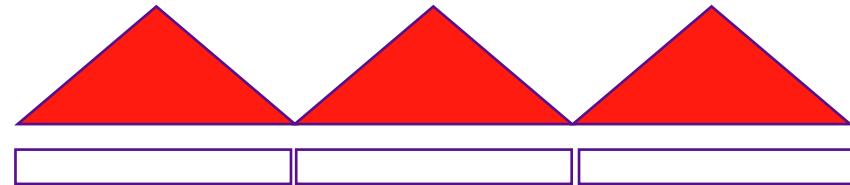


Weak Strong

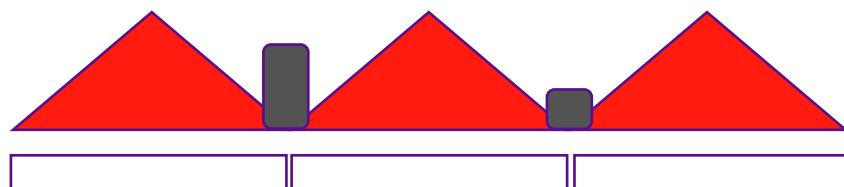
Super-enhancer/boundary co-duplication at MYC



Visual summary



TAD detection (HiC-bench)



Detection of boundaries
of differential strength



Deletion



Duplication



Co-duplication

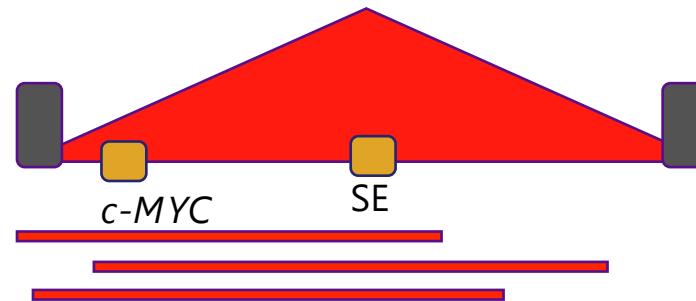
TAD

Strong boundary

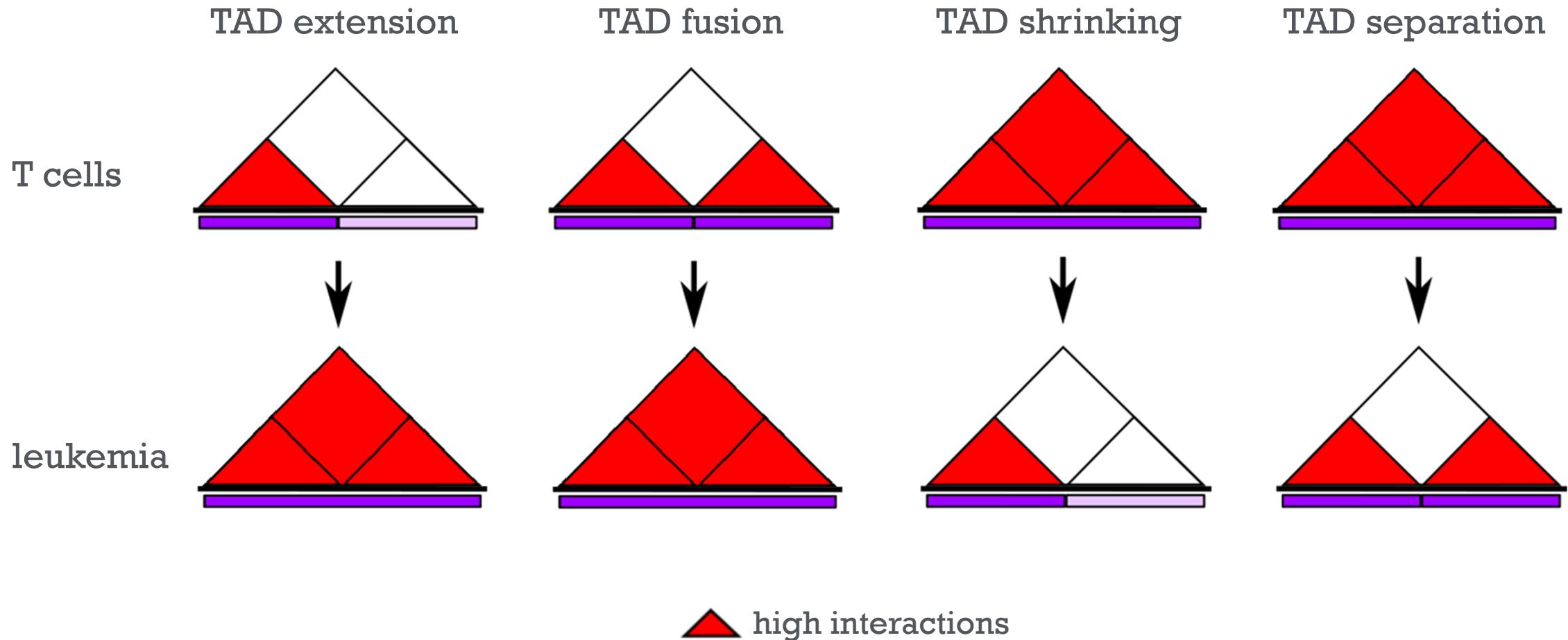
Weak boundary

Pan-cancer
analysis

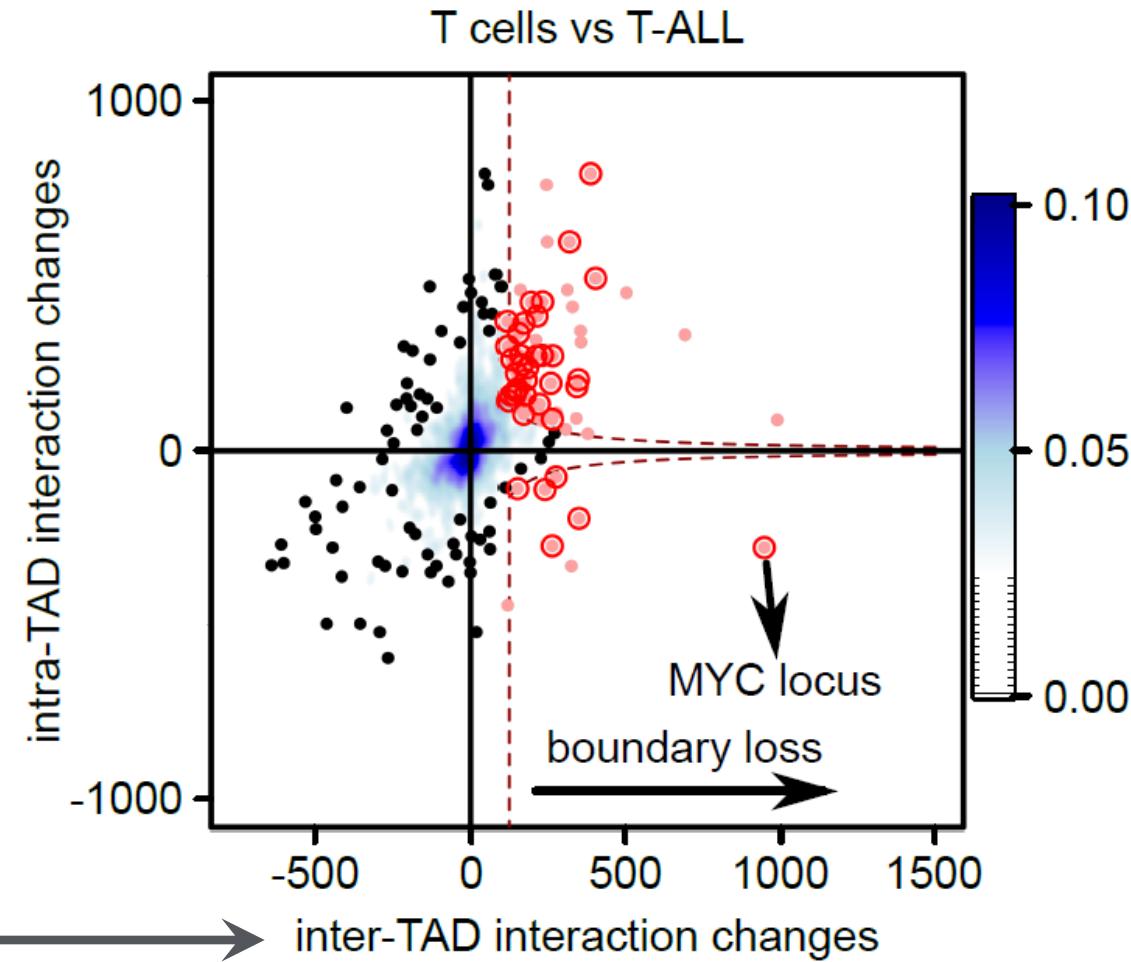
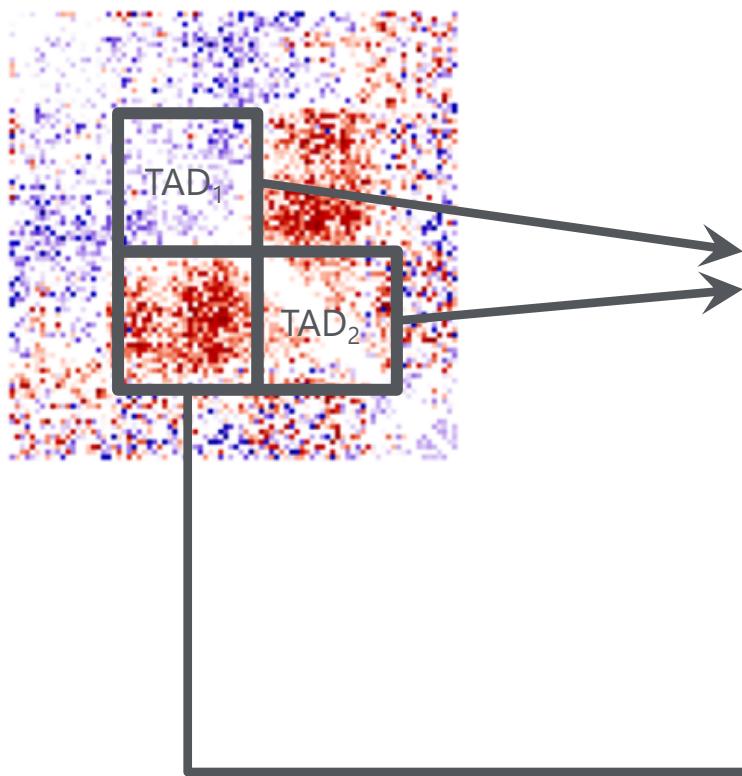
MYC



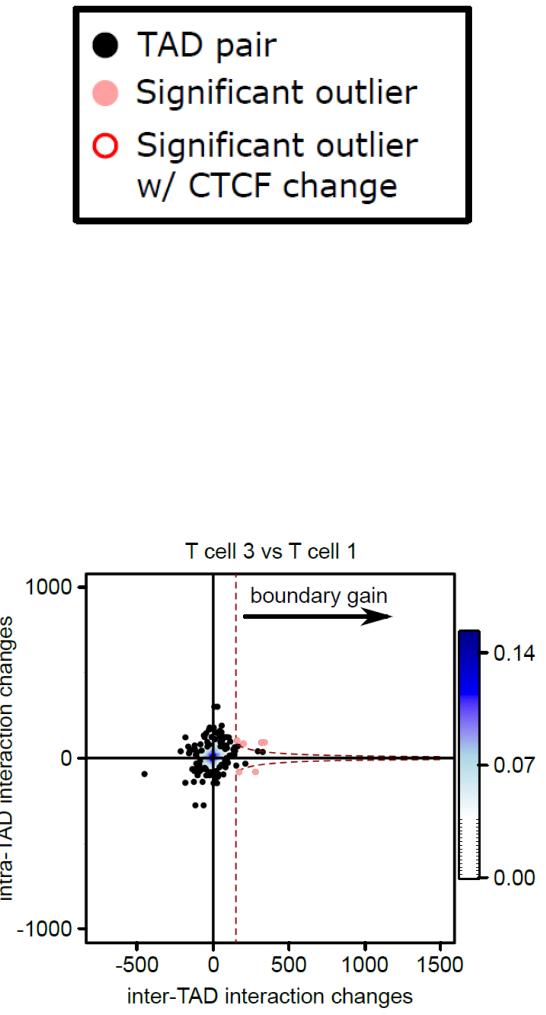
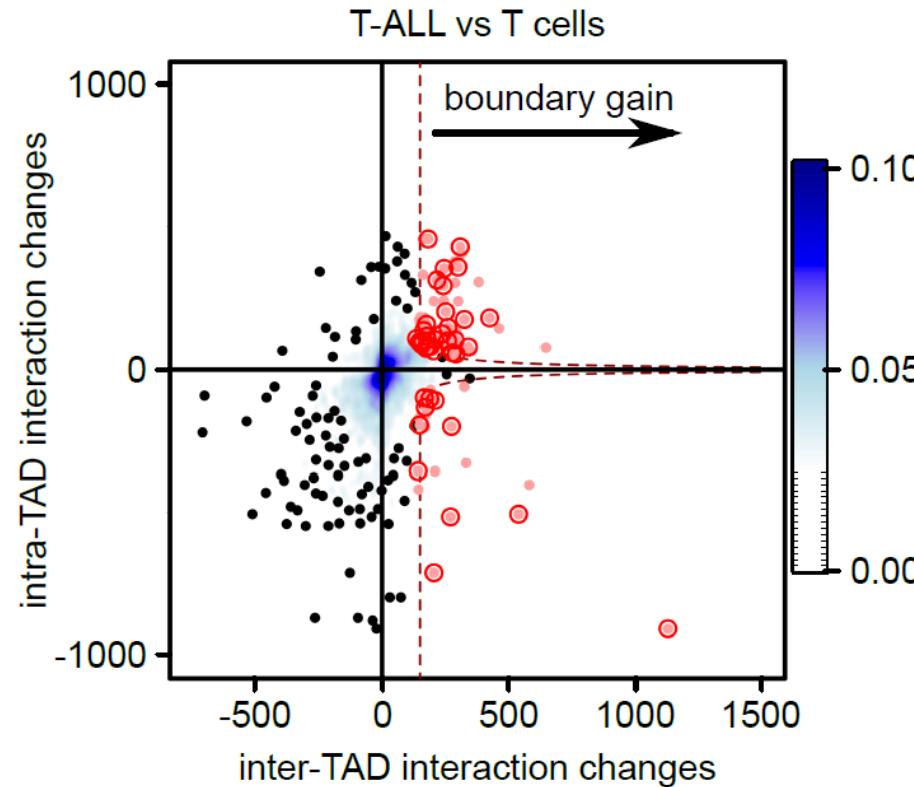
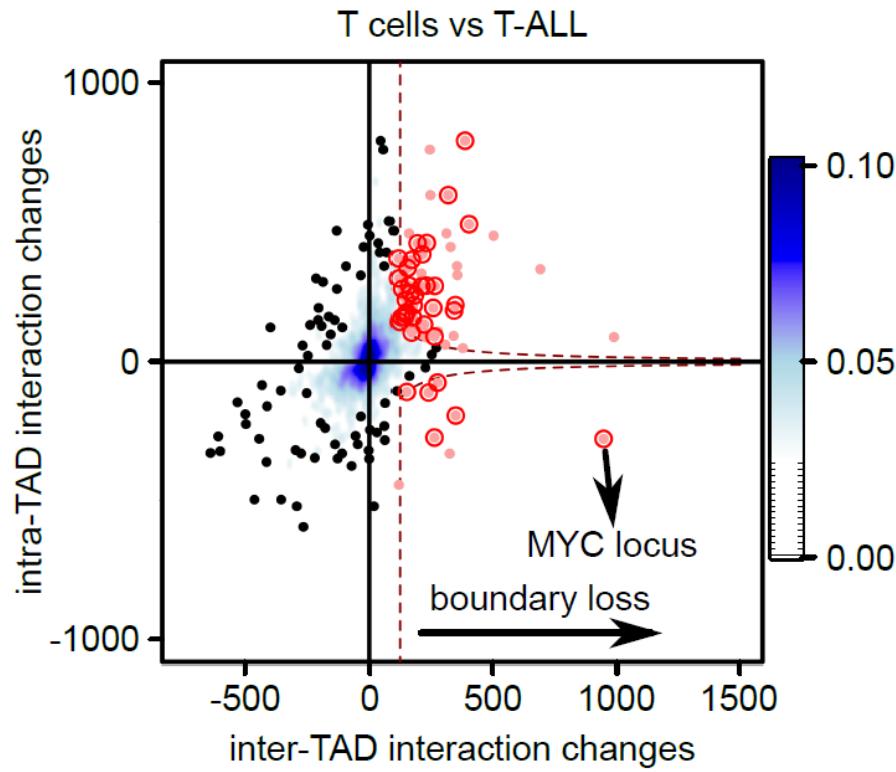
Disrupted TAD boundaries in leukemia



Disrupted TAD boundaries in leukemia



Disrupted TAD boundaries in leukemia



- TAD pair
- Significant outlier
- Significant outlier w/ CTCF change

MYC super-enhancer

- MYC is regulated by a distant super-enhancer cluster during hematopoiesis and in T-ALL
- Conserved in human and mouse

nature
International journal of science

A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies

Carsten Bahr, Lisa von Paleske, Veli V. Uslu, Silvia Remeseiro, Naoya Takayama, Stanley W. Ng, Alex Murison, Katja Langenfeld, Massimo Petretich, Roberta Scognamiglio, Petra Zeisberger, Amelie S. Benk, Ido Amit, Peter W. Zandstra, Mathieu Lupien, John E. Dick, Andreas Trumpp  & François Spitz 

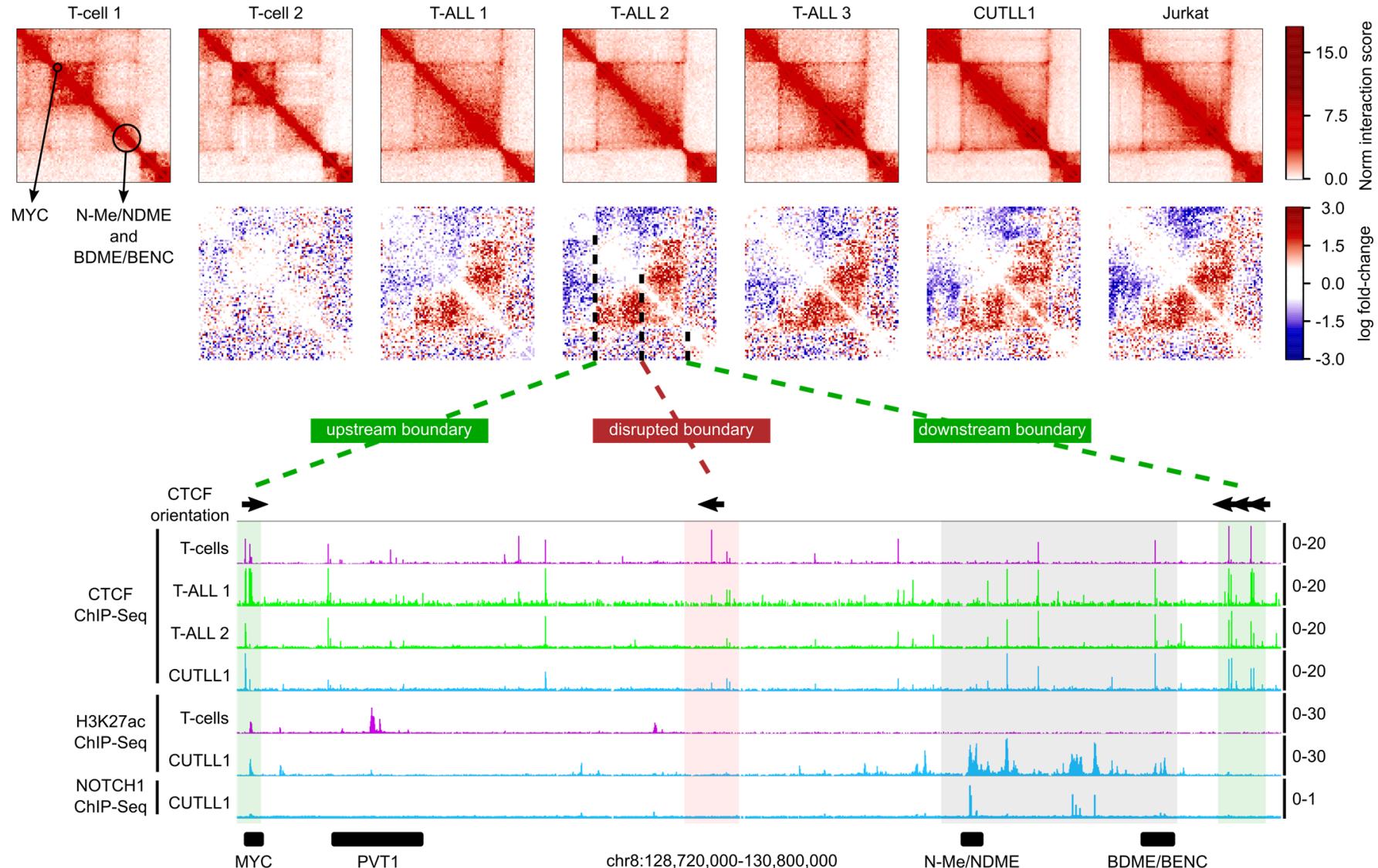
nature medicine

A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia

Daniel Herranz, Alberto Ambesi-Impiombato, Teresa Palomero, Stephanie A Schnell, Laura Belver, Agnieszka A Wendorff, Luyao Xu, Mireia Castillo-Martin, David Llobet-Navás, Carlos Cordon-Cardo, Emmanuelle Clappier, Jean Soulier & Adolfo A Ferrando 

MYC TAD "fusion" in leukemia patients

chr8:128,000,000-131,500,000



SUMMARY

- Chromatin architecture in the nucleus can be organized in different hierarchies
 - Chromosome territories, compartments, TADs, loops/contacts
 - Organization is cell-type specific
- Hi-C reveals most of these structures
- Analysis of Hi-C data is a multi-step process and depends on the aim of the study
 - Alignment, filtering, matrix generation
 - Compartment calling, TAD calling
 - Visualization
- Integrating additional epigenetics data is useful to determine the role of the nuclear organization



FUTURE DIRECTIONS

- Single-cell Hi-C
- Phase-separation around strong “super-enhancers”
- Inter-chromosomal interactions
- Dozens of examples for cell-type / disease specific chromatin alterations

Assignment: Basic analysis of Hi-C data

- Use mouse Hi-C and matched H3K27ac ChIP-seq data available on Big Purple:
 - Hi-C fastq files + ChIP-seq H3K27ac peaks
 - </gpfs/data/courses/bmscga2604>
- Process the Hi-C paired-end sequencing data using HiC-bench or HiCExplorer
 - Align against the respective reference genome and generate filtered valid pairs
 - Generate .hic files for visualization
 - Create and normalize Hi-C contact matrices
- Perform downstream analyses
 - Call A/B compartments (use H3K27ac ChIP-seq data to discriminate A vs B)
 - Compare A vs B compartments per chromosome in terms of gene density, housekeeping gene density and H3K27ac peak density
- Visualize a 10Mb locus and include a track that shows the compartment score



THANK YOU

