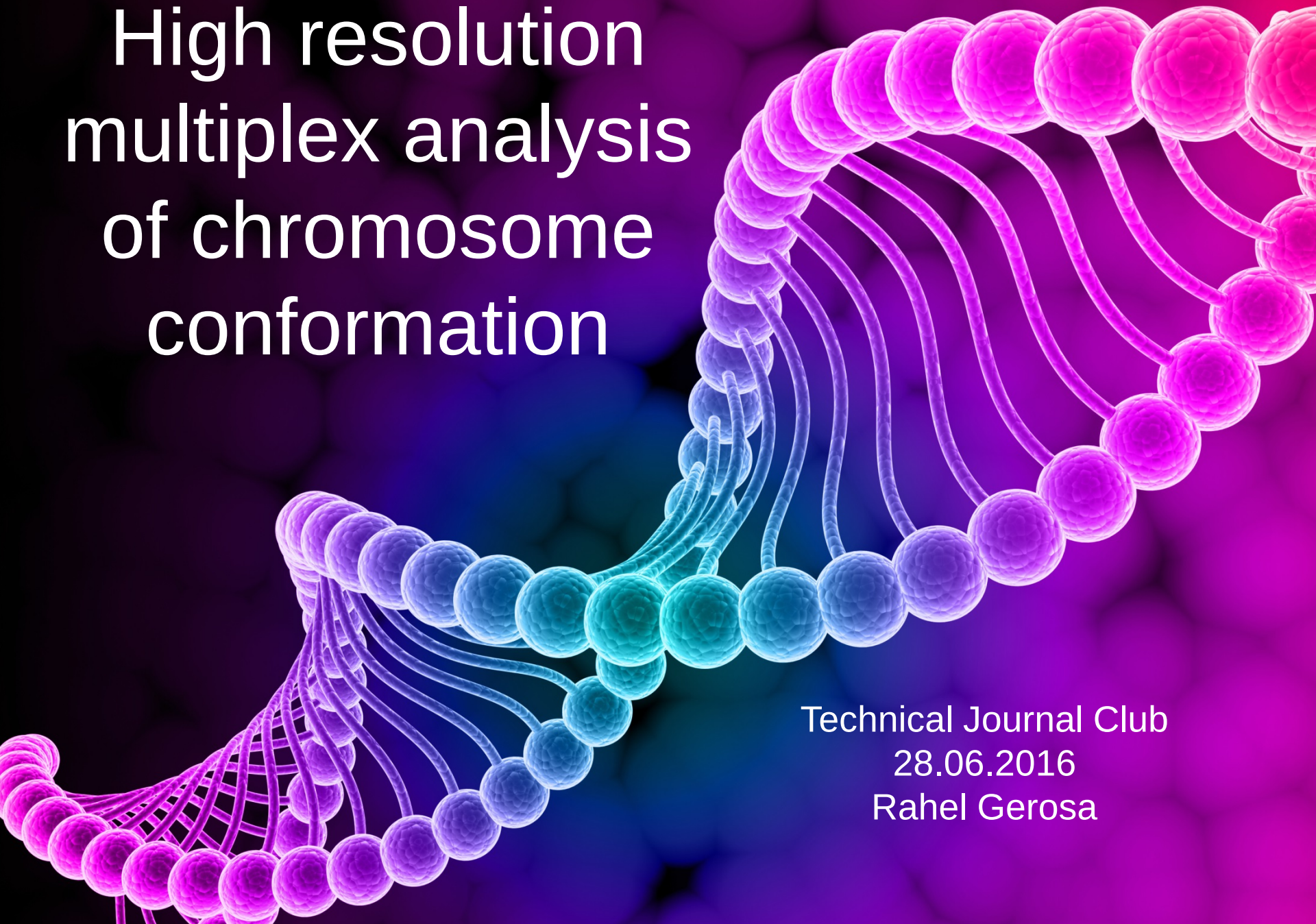
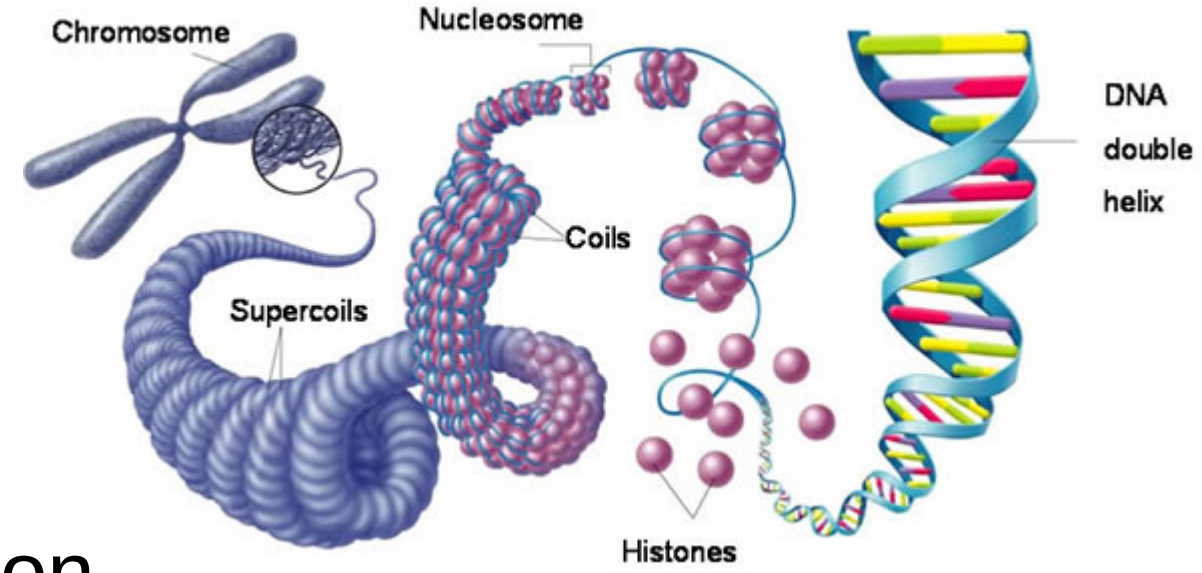
High resolution multiplex analysis of chromosome conformation



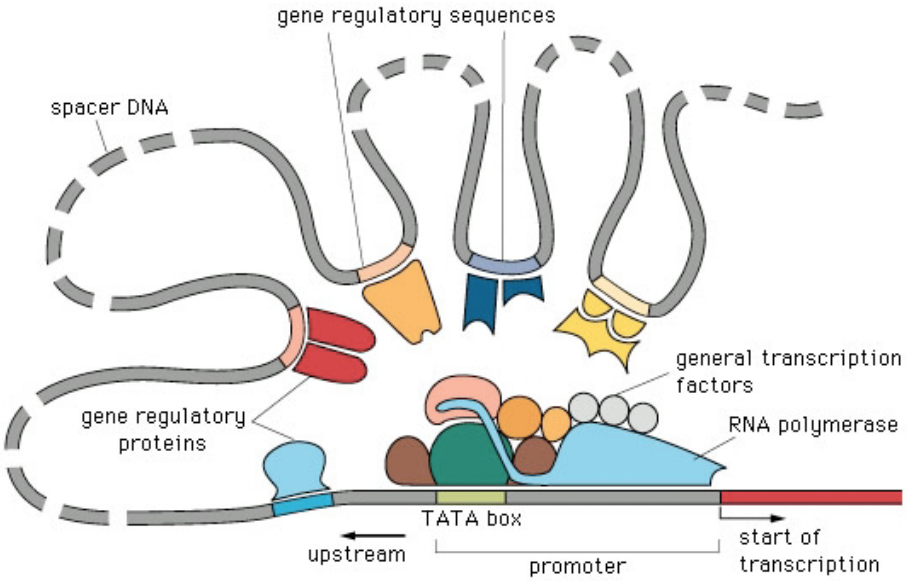
Technical Journal Club
28.06.2016
Rahel Gerosa

Chromosome organization



https://pmgbiology.com/tag/chromosome/

Gene regulation

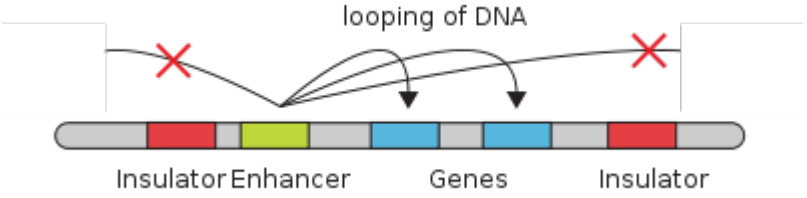


Expression of a single gene is often controlled by multiple regulatory elements that may lie tens to hundreds of kilobases upstream or downstream of their targets

http://pbil.univ-lyon1.fr/baobab/article.php3?id_article=278

Gene regulatory elements

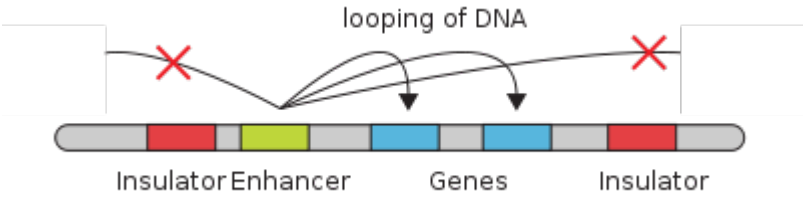
CRE	Cis Regulatory Elements are regions of non-coding DNA which regulate the transcription of nearby genes typically by functioning as binding sites for transcription factors
Intergenic	Non-coding DNA between genes
Intragenic	Introns
Promoter	A region of DNA that initiates transcription of a particular gene. Promoters are located near the transcription start sites of genes, on the same strand and upstream on the DNA
Enhancer	Short (50-1500 bp) region of DNA that can be bound by activators to increase the likelihood that transcription will occur at a gene
Silencer	DNA sequence capable of binding TFs, called repressor what prevents RNA polymerase from transcribing the DNA sequence into RNA



https://en.wikipedia.org/wiki/Insulator_%28genetics%29

Gene regulation

Insulator	Insulator is a genetic boundary element that blocks the interaction between enhancers and promoters
CTCF	Regulates the 3D structure of chromatin by binding together strands of DNA, thus forming chromatin loops, and anchors DNA to cellular structures like the nuclear lamina Since the 3D structure of DNA influences the regulation of genes, CTCF's activity influences the expression of genes CTCF is thought to be a primary part of the activity of insulators, sequences that block the interaction between enhancers and promoters



https://en.wikipedia.org/wiki/Insulator_%28genetics%29

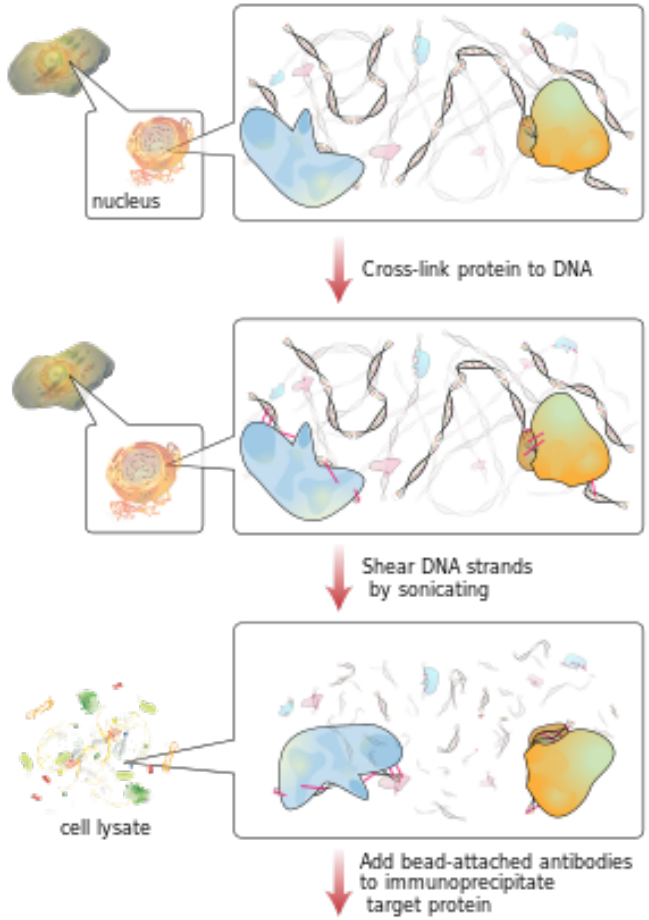
Cis Regulator Elements (CRE)

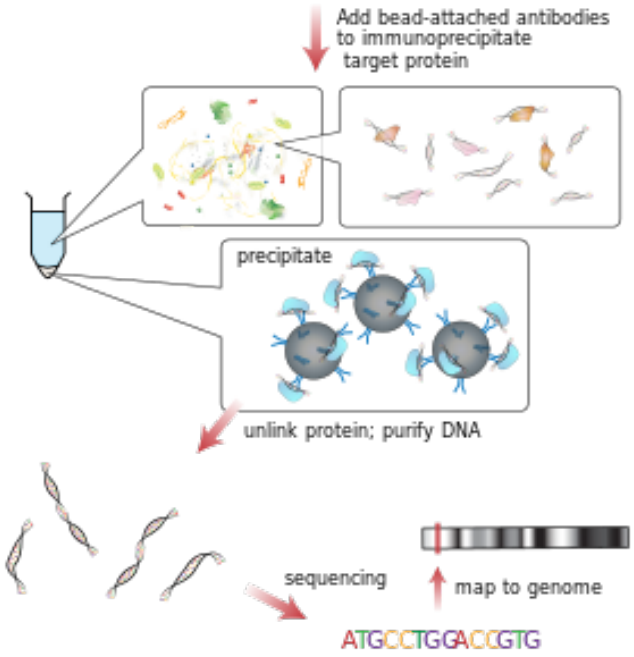
Where are these regulatory elements?

Find gene regulatory elements and investigate their function :

ChIP-Seq	**Ch**romatin **I**mmuno**p**recipitation with sequencing
DNase-Seq	DNase I hypersensitive sites sequencing
ATAC-seq	**A**ssay for **T**ransposase-**A**ccessible **C**hromatin with **seq**uencing

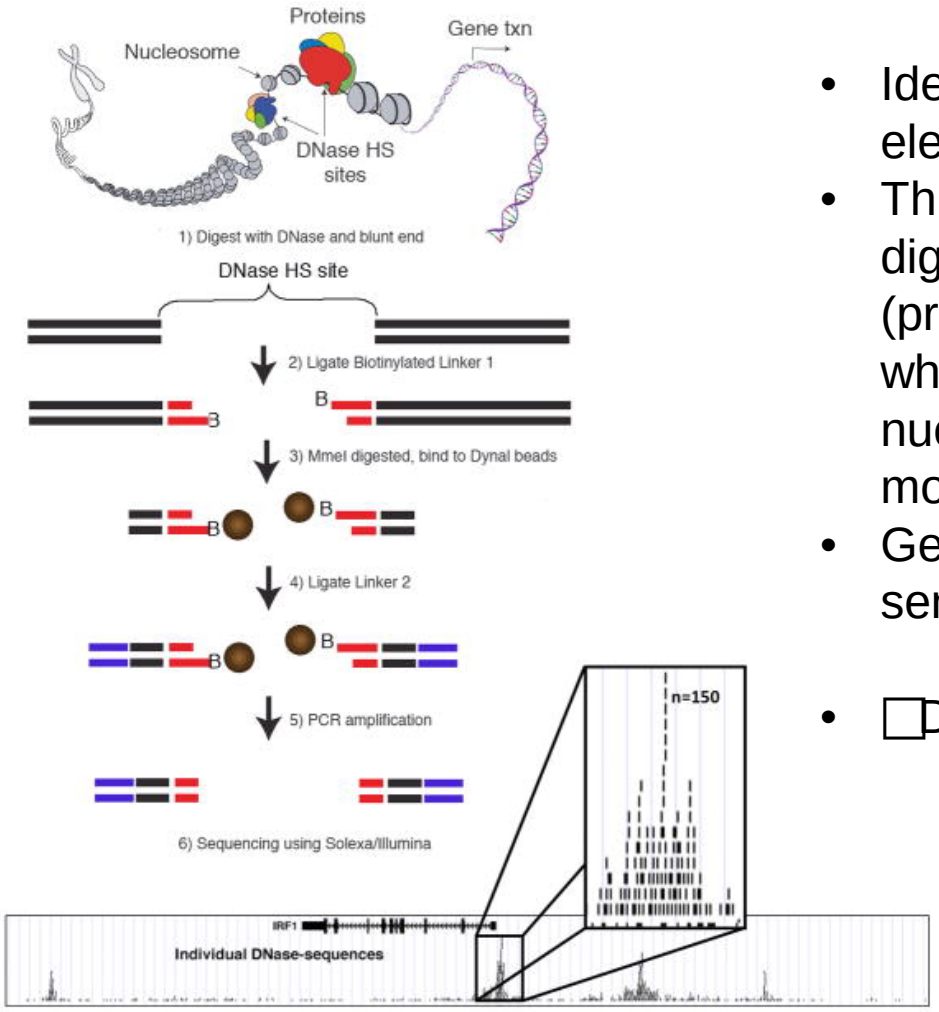
ChIP-Seq
Chromatin Immunoprecipitation with sequencing





https://en.wikipedia.org/wiki/Chromatin_immunoprecipitation

DNase Seq
DNase I hypersensitive sites sequencing

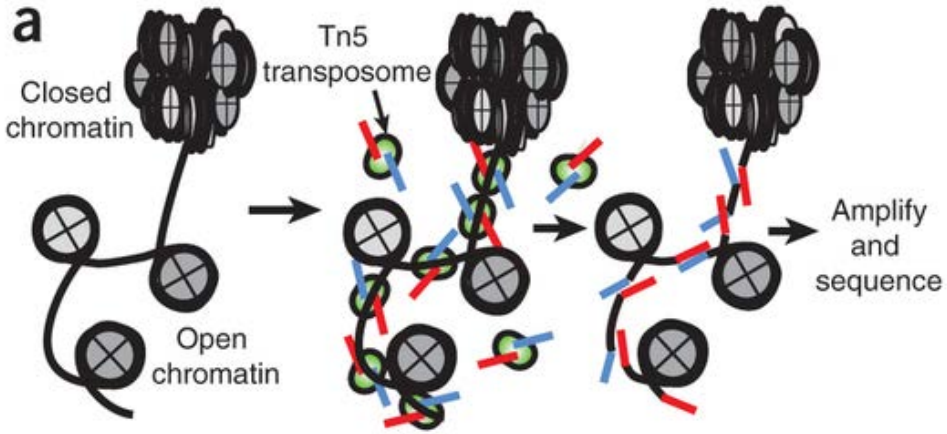


- Identification of active gene regulatory elements
- This method utilizes DNase I to selectively digest nucleosome-depleted DNA (presumably by transcription factors), whereas DNA regions tightly wrapped in nucleosome and higher order structures are more resistant
- Genome-wide sequencing of regions sensitive to cleavage by DNase I

- □DNase I footprint

Lingyun Song and Gregory E. Crawford, PMC, 2013

Assay for Transposase-Accessible Chromatin with sequencing



Buenrostro JD et al., Nature Methods, 2013

Transposase, loaded with sequencing adaptors (red and blue), inserts only in regions of open chromatin and generates sequencing-library fragments that can be PCR-amplified

Cis Regulator Elements (CRE)

Where are these regulatory elements?

Find gene regulatory elements and investigate their function :

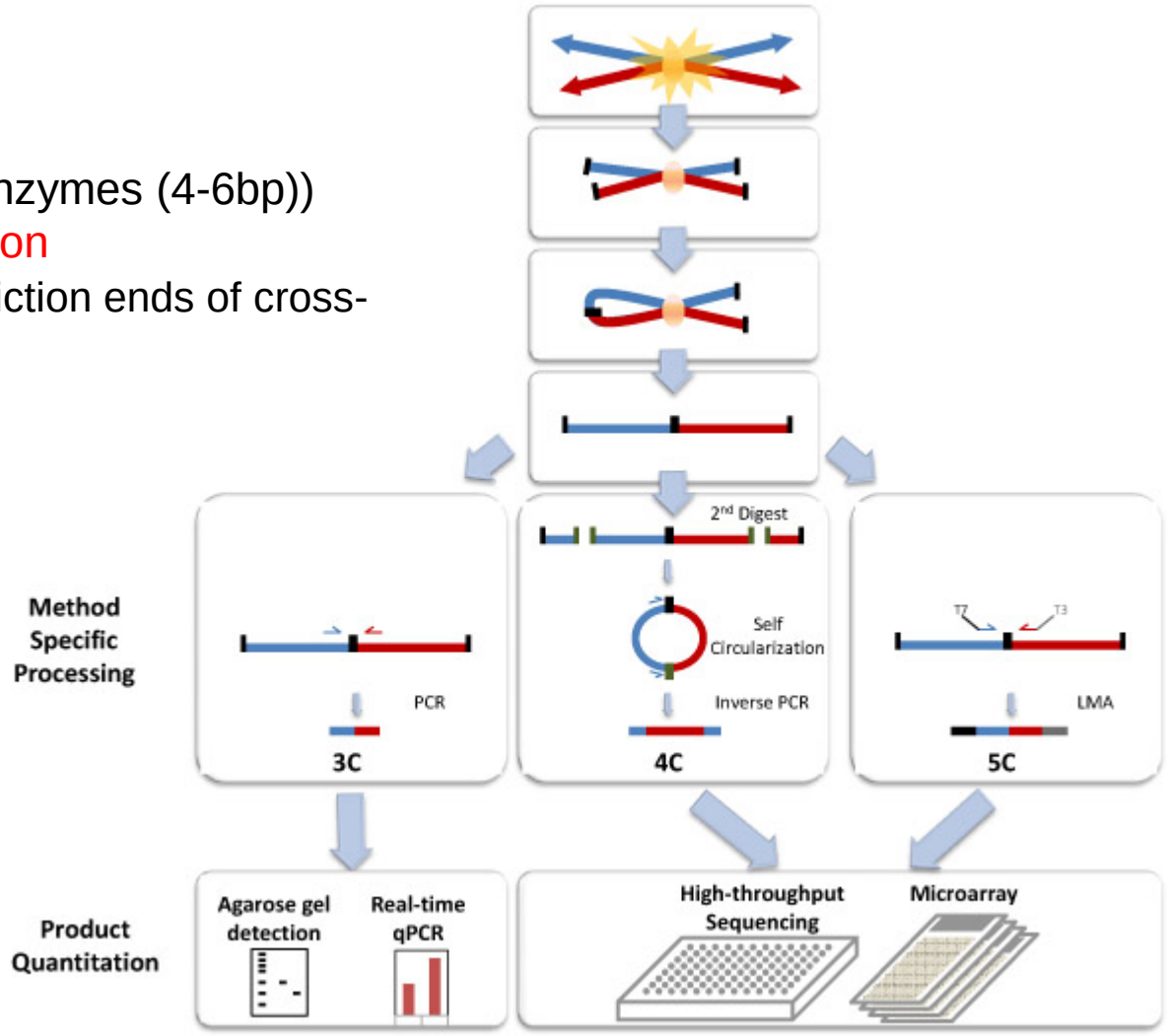
ChIP-Seq	**Ch**romatin **I**mmuno**p**recipitation with sequencing
DNase-Seq	DNase I hypersensitive sites sequencing
ATAC-seq	**A**ssay for **T**ransposase-**A**ccessible **C**hromatin with **seq**uencing

New challenge

Understand the mechanisms by which regularory elements control specific gene promoters at a distance (tens to thousands of kilobase pairs)

Chromosome Conformation Capture Techniques

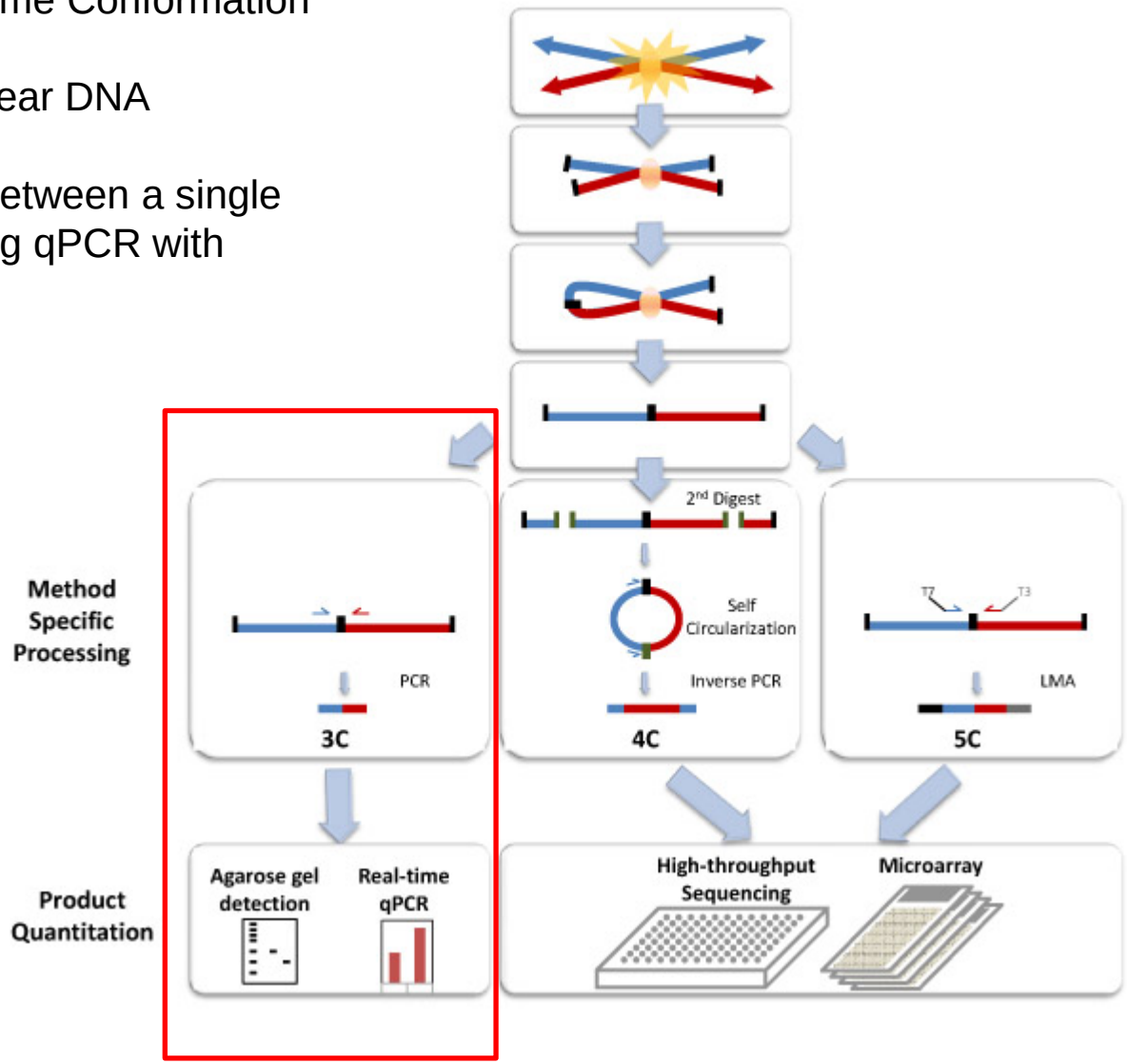
1. Cross-linking
   (formaldehyde)
2. Restriction Digest
   (frequently cutting enzymes (4-6bp))
3. Intramolecular Ligation
   (procedure joins restriction ends of cross-linked segments)
4. Reverse Cross-links
   (proteinase K)



https://en.wikipedia.org/wiki/Chromosome_conformation_capture#/media/File:Chromosome_Conformation_Capture_Technology.jpg

Chromosome Conformation Capture Techniques

- Conventional Chromosome Conformation Capture
- Resulting in a pool of linear DNA fragments (3C library)
- Quantifies interactions between a single pair of genomic loci using qPCR with locus-specific markers

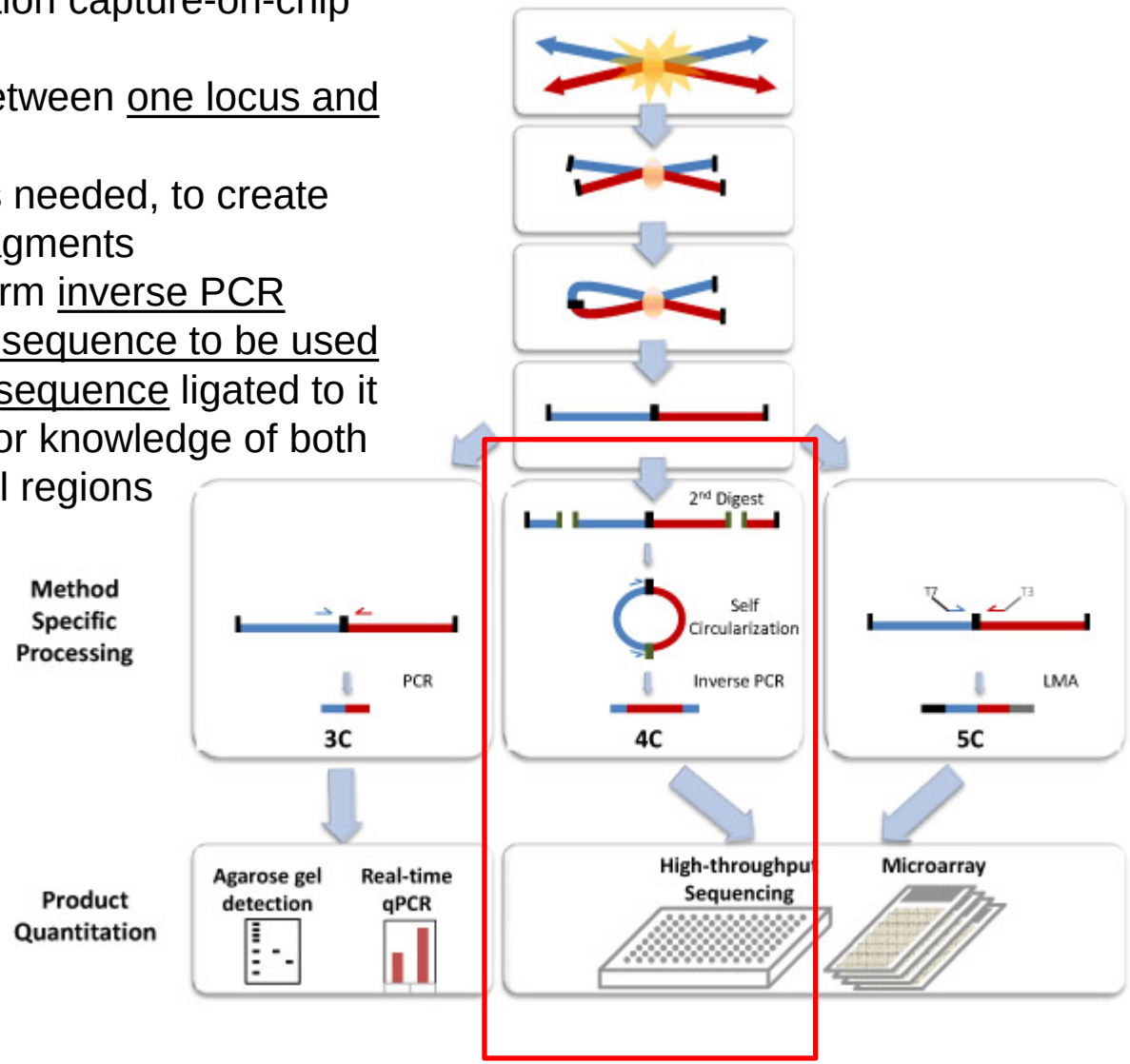


https://en.wikipedia.org/wiki/Chromosome_conformation_capture#/media/File:Chromosome_Conformation_Capture_Technology.jpg

Chromosome Conformation Capture Techniques

- Chromosome conformation capture-on-chip (4C)
- Captures interactions between one locus and all other genomic loci
- A second ligation step is needed, to create self-circularized DNA fragments
- These are used to perform inverse PCR which allows the known sequence to be used to amplify the unknown sequence ligated to it
- Does not require the prior knowledge of both interacting chromosomal regions

Enrichment for SPECIFIC loci

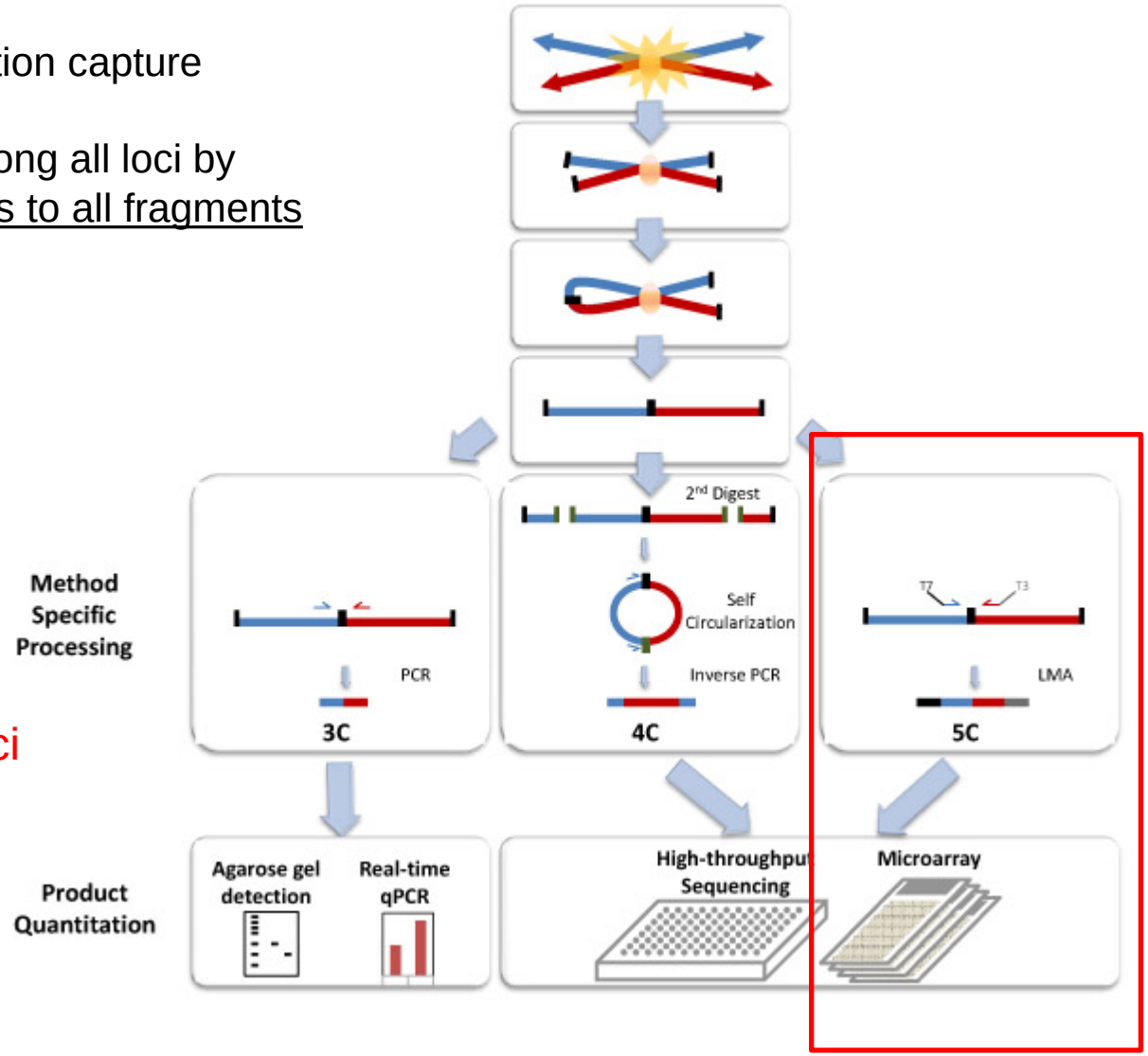


https://en.wikipedia.org/wiki/Chromosome_conformation_capture#/media/File:Chromosome_Conformation_Capture_Technology.jpg

Chromosome Conformation Capture Techniques

- Chromosome conformation capture carbon copy (5C)
- Detects interactions among all loci by <u>ligating universal primers to all fragments</u>

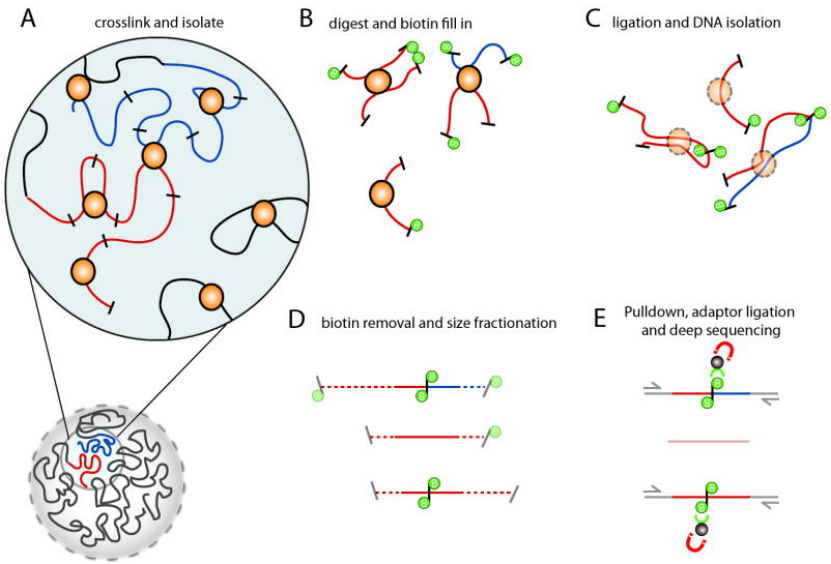
Enrichment for <u>ALL</u> loci



https://en.wikipedia.org/wiki/Chromosome_conformation_capture#/media/File:Chromosome_Conformation_Capture_Technology.jpg

Chromosome Conformation Capture Techniques

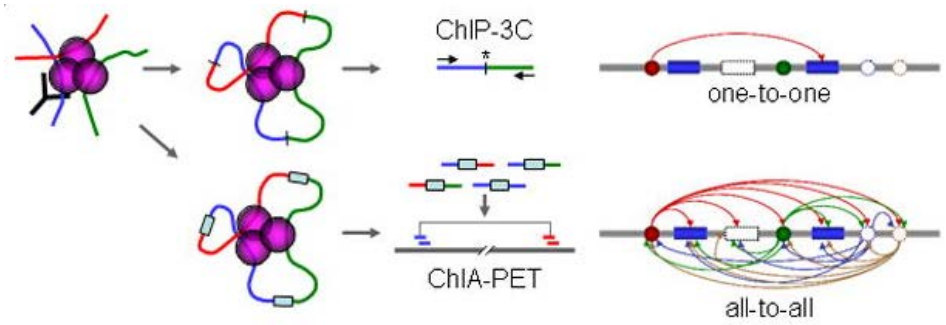
Hi-C



Belton JM et al., Methods, 2012

- All genomic fragments are labeled with a biotinylated nucleotide before ligation, thereby marking ligation junctions
- Junctions are purified by streptavidin-coated magnetic beads

ChIP-loop and ChIA-PET



Fullwood MJ and Ruan Y, PMC, 2009

- The ChIA-PET method □less non-specific interaction noise (like in ChIP-Seq) due to sonication □separates random attachments from specific interaction complexes
- Introduce a linker sequence in the junction of two DNA fragments during ligation
- Analyzed by PET (paired-end tag) sequencing

nature genetics 12 January 2014

Analysis of hundreds of *cis*-regulatory landscapes at high resolution in a single, high-throughput experiment

Jim R Hughes[1], Nigel Roberts[1], Simon McGowan[2], Deborah Hay[1], Eleni Giannoulatou[2], Magnus Lynch[1], Marco De Gobbi[1], Stephen Taylor[2], Richard Gibbons[1] & Douglas R Higgs[1]

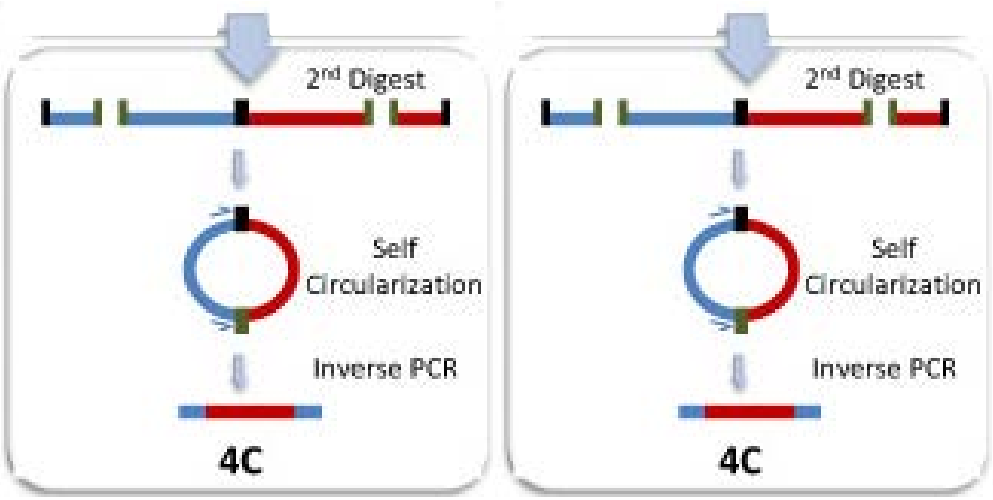
[1]Medical Research Council (MRC) Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. [2]Computational Biology Research Group, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. Correspondence should be addressed to D.R.H. (doug.higgs@imm.ox.ac.uk) or J.R.H. (jim.hughes@imm.ox.ac.uk).

Capture-C

Combines 3C libraries with oligonucleotide capture to enrich for specific loci

Possible Approach

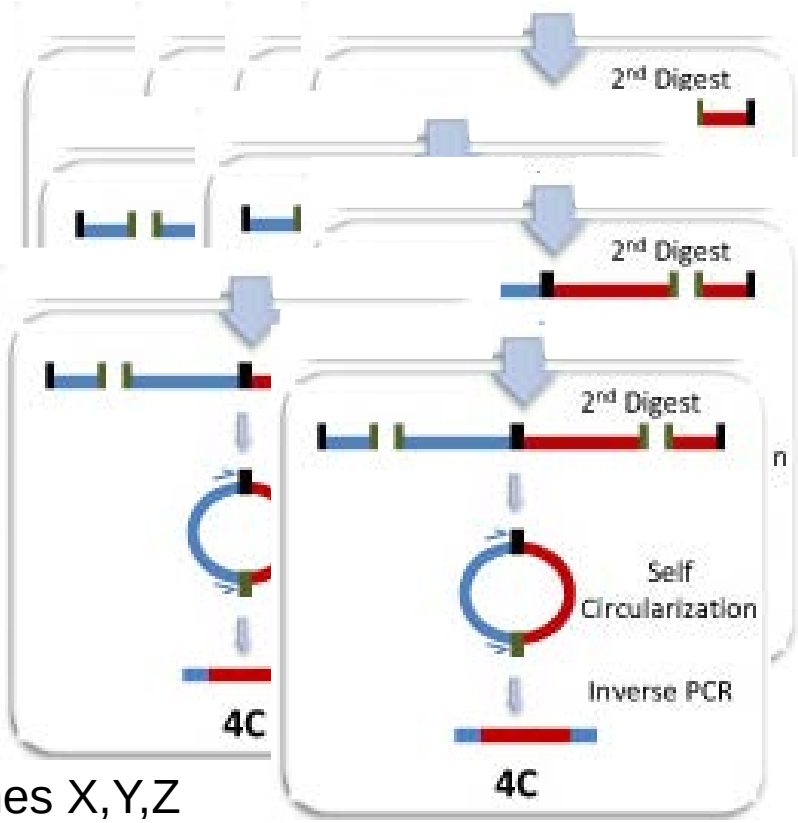
Some versions of the 4C-seq method can map interactions high resolution, but only interrogate a single region of interest at a time



α-globin gene

β-globin gene

Laborious and expensive



Genes X,Y,Z

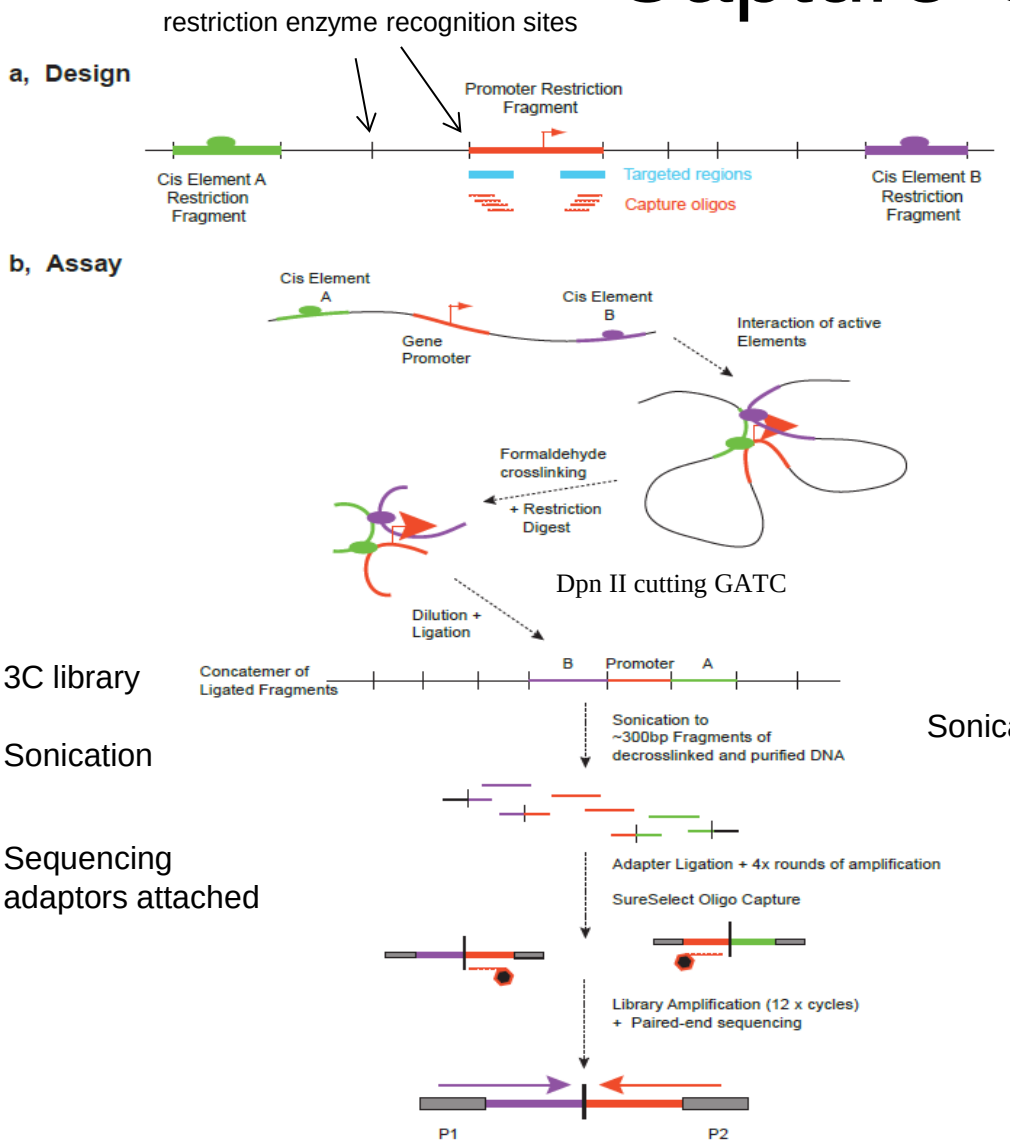
A multiplexed, high-resolution 3C method Capture-C

Capture-C = 3C + OCT + high-throughput sequencing

(targeted DNA capture)

Interrogate cis interactions at hundreds of selected loci at high resolution in a single assay

A multiplexed, high-resolution 3C method Capture-C



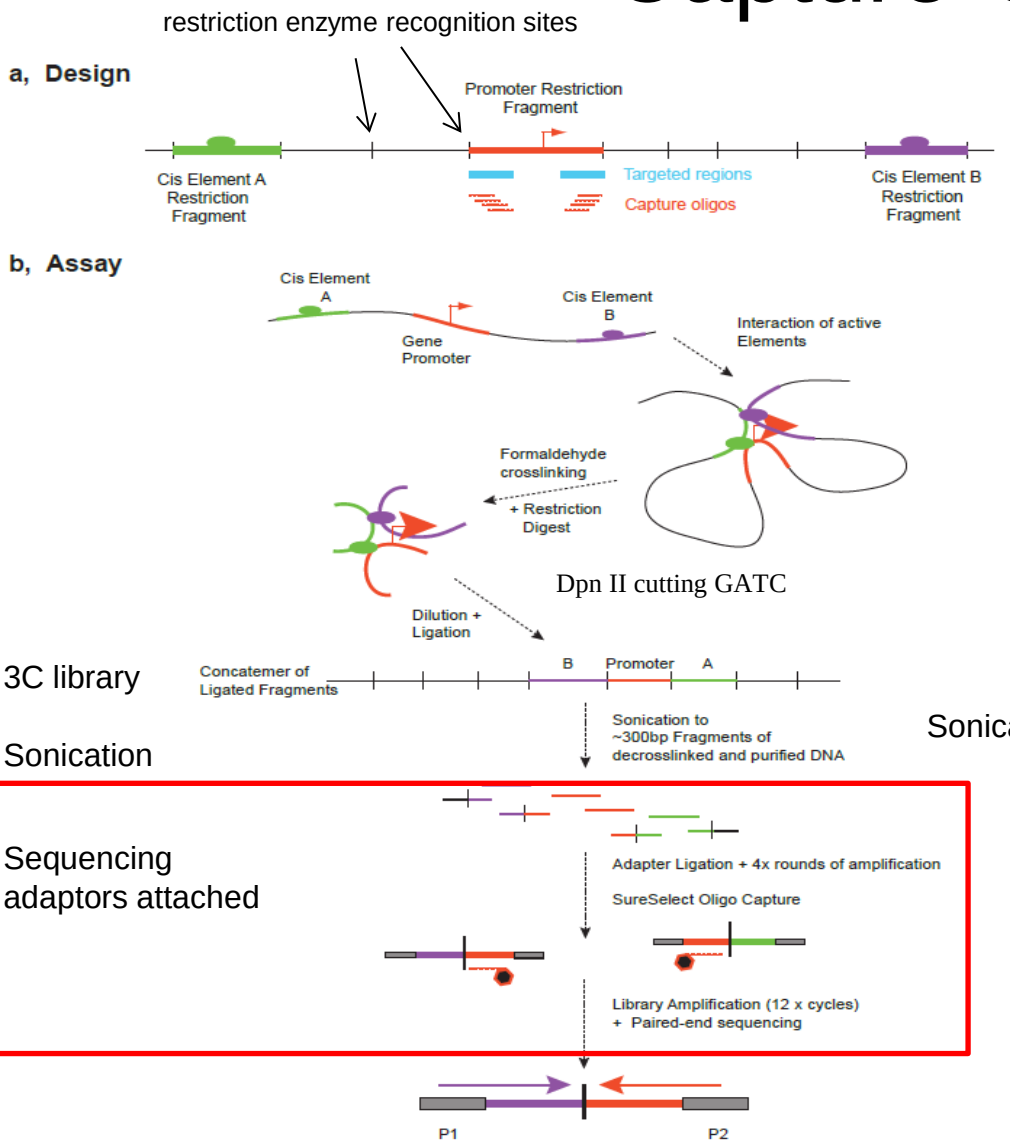
Biotinylated capture probes (flanking DpnII restriction site)

Specific genomic region

Sure Select platform

Sonication reduces non-specific interactions

A multiplexed, high-resolution 3C method Capture-C



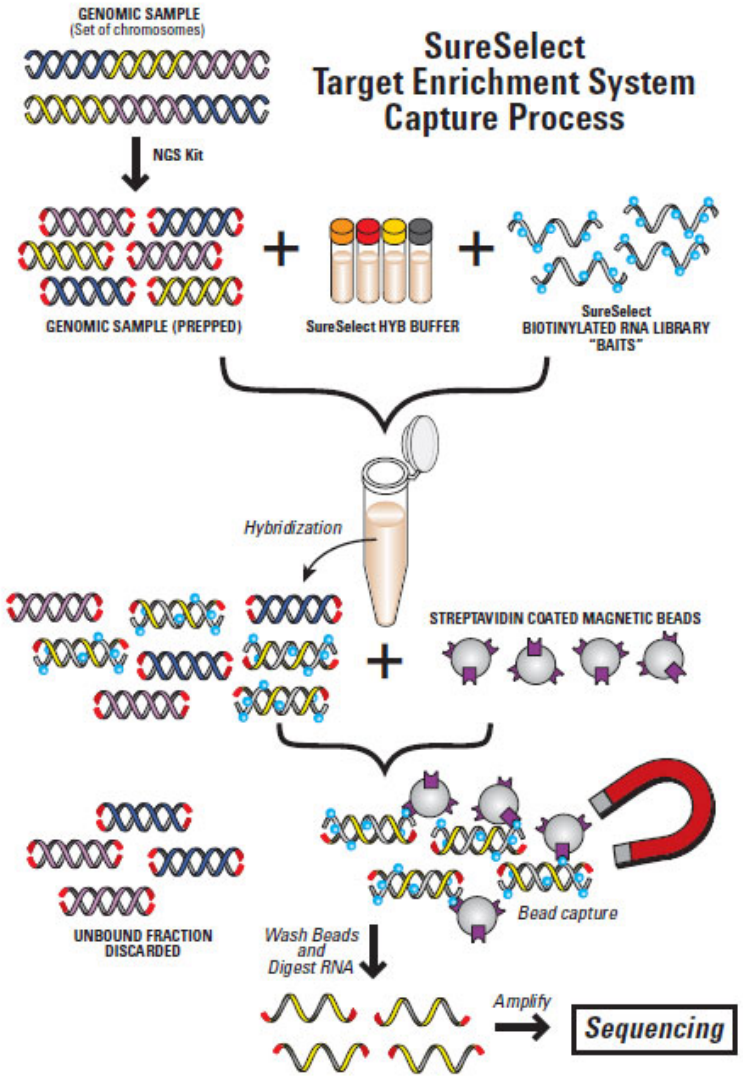
Biotinylated capture probes (flanking DpnII restriction site)

Specific genomic region

Sure Select platform

Sonication reduces non-specific interactions

Oligonucleotide Capture Technology (OCT)



http://www.genomics.agilent.com/article.jsp?pageId=3083

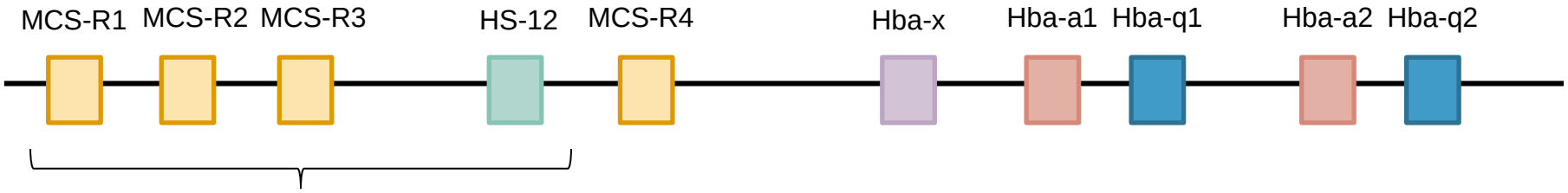
1. Starting from gDNA, a shearing step produces small fragments (Sonication)
2. Prepare library with sequencer specific adaptors and indexes
3. Hybridize sample with biotinylated RNA library baits (ultra long 120mer RNA baits for the highest specificity)
4. Select targeted regions using magnetic streptavidin beads
5. Amplify and load on the sequencer

- Increase sample throughput via multiplexing while achieving the sequencing depth required
- Pooling 8-16 samples prior to enrichment
- 455 preselected promoters

Validation of the Capture-C technique

- Analyzed the *cis* interactions at the mouse α- and β-globin gene loci in erythroid (mouse Ter119+) and nonerythroid (mouse embryonic stem (mES)) cells
- Interactions at the globin loci have been used previously to validate all chromosomal conformation techniques and thus provide the ideal standard by which to assess newly developed techniques

Chromosome 11



Introns of Nprl3

GENES

Hba-x embryonic gene
Hba-a1 fetal and adult tissue
Hba-a2 fetal and adult tissue
Hba-q1 globin-like gene (unknown function)
Hba-q2 globin-like gene (unknown function)

DISTAL REGULATORY ELEMENTS

MCS-R1
MCS-R2
MCS-R3
MCS-R4
HS-12

DNase1 hypersensitive sites

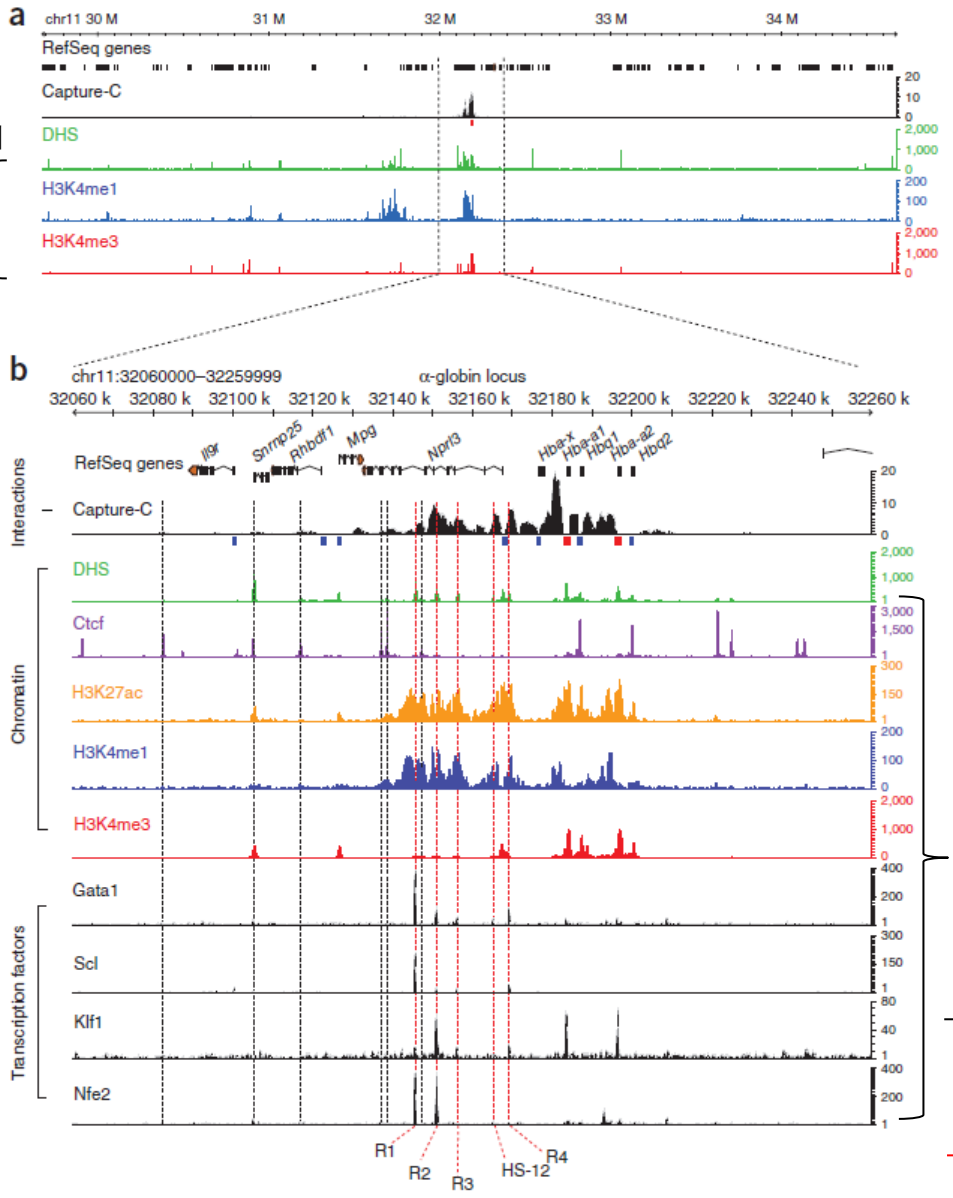
Validation of the Capture-C technique

Chromosome 11

DNase Seq

ChIP Seq

α-globin locus



Interactions per restriction fragment

DNase1 hypersensitive sites

Monomethylation of histone H3 Lys4

Trimethylation of histone H3 Lys4

Capture-C probes designed to the α-globin gene promoters

Capture-C probes designed to surrounding genes (background ctrl)

Chip-seq data

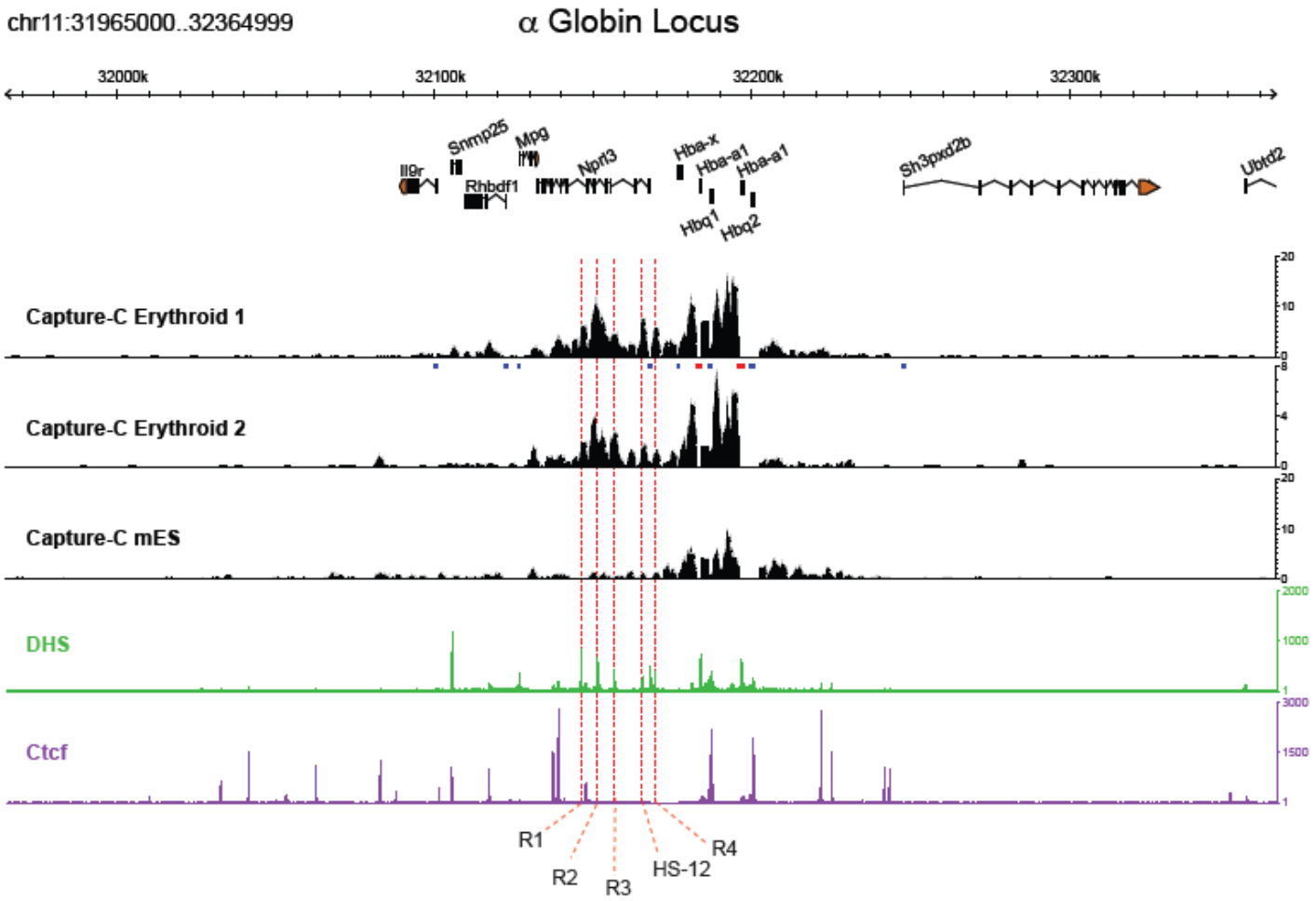
Interactions with regions bound by the Ctcf protein

Positions of known α-globin regulatory elements

Sequences from any capture point will interact with the surrounding genome (in a distance-dependent manner), whether the targeted promoter or regulatory element is active or inactive

The predominant interaction of the α-globin promoters occurs with MCS-R2 which is also known to exert the greatest effct on α -globin expression in vivo

Validation of the Capture-C technique

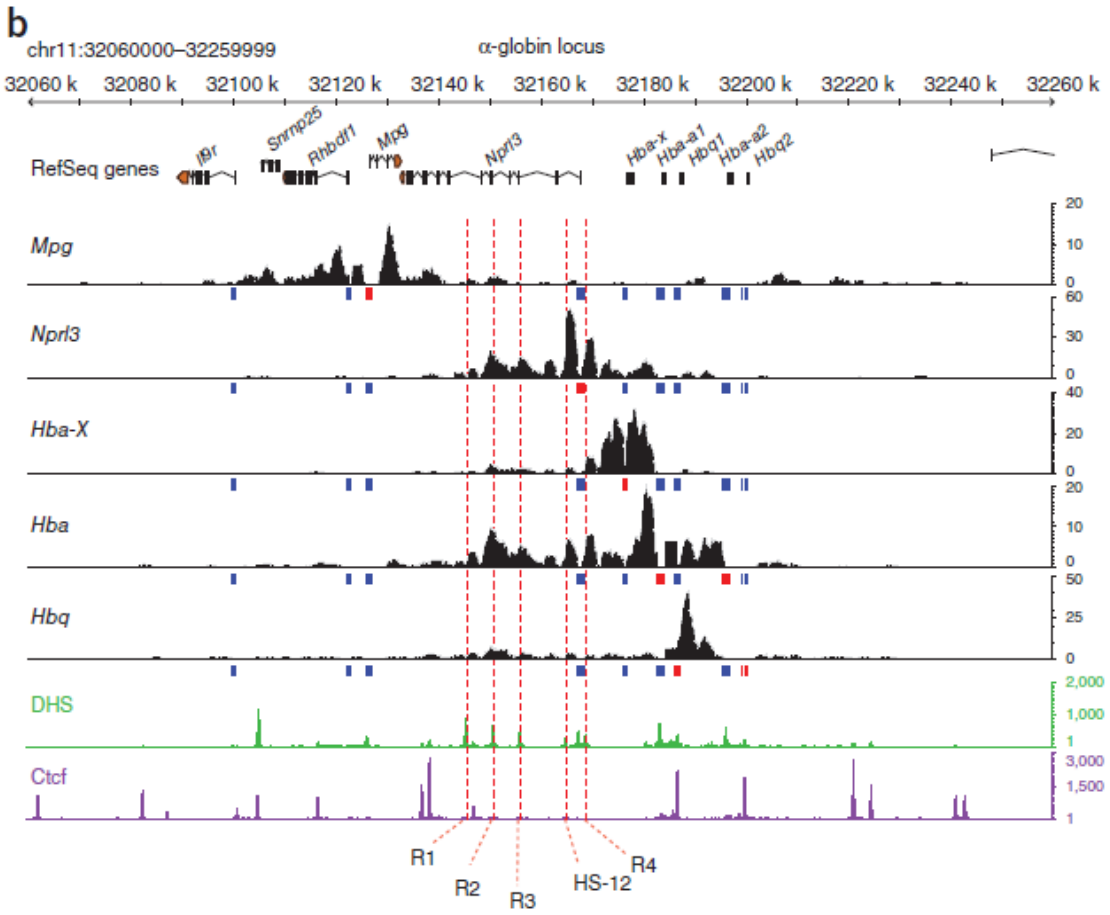


Interactions are consistent in duplicated

Interactions between the α-globin promoters and their distal regulatory elements are seen only when the α -genes are <u>active in erythroid cells</u> and not when they are silent in nonerythroid cells

Validation of the Capture-C technique

- Capture-C is able to simultaneously assay the same locus from a variety of perspectives in a single experiment (different viewponts)
- α-genes interact with their distal regulatory elements □poorly expressed Hbq genes interact only weakly (same for Hba-x)
- The promoter of the adjacent gene, Nprl3, also interacts with the erythroid elements contained within ist introns
- Promoter of the next gene upstream (Mpg) interacts only minimally with these elements



Capture-C agrees with all previous 3C analyses, identifying all known regulatory elements of the globin genes and no additional elements

Multiple loci in one experiment

- 455 preselected promoters
  - α- and β-globin genes as performance controls
  - 71 gene promoters because of either their proximity to the α-globin locus or their known functional importance in erythropoiesis
  - Remainder of 384 promoters □ncreased expression of their associated genes in mouse erythroid Ter119+ cells relative to ES cells

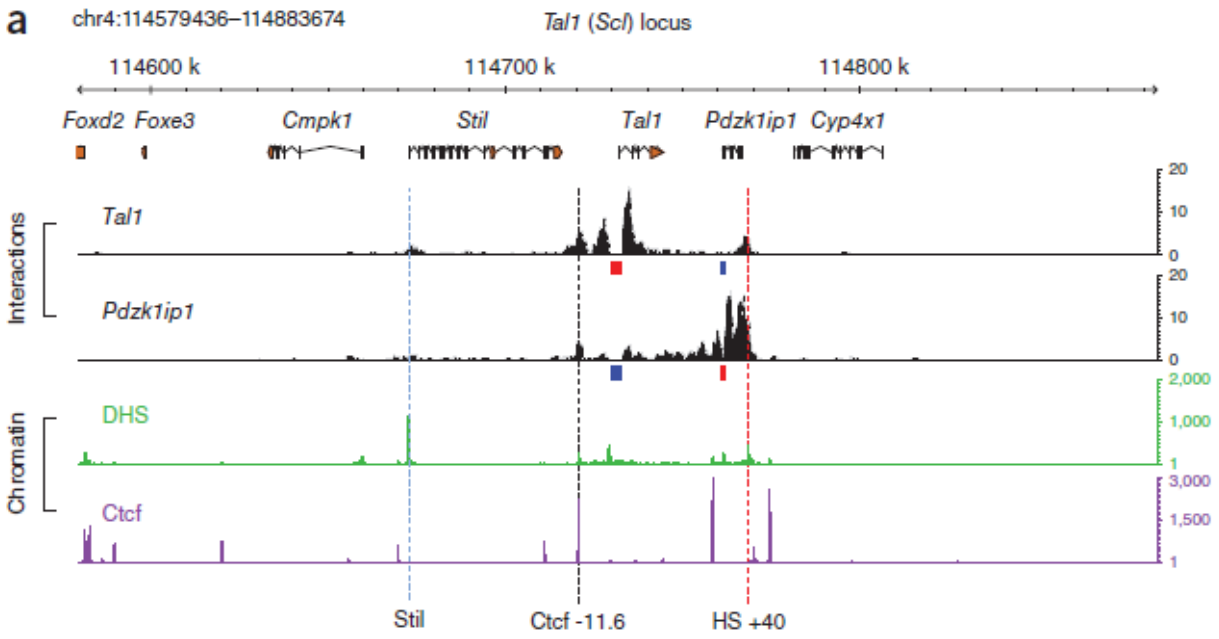
□selected genes whose expression was likely regulated by enhancer-promoter interactions during erythropoiesis

Analysis of the capture efficiency showed that the majority (92%) of the targeted regions were successfully captured

Multiple loci in one experiment

Example

- Regulation well known
- Tal1
- TF essential in erythroid develompent



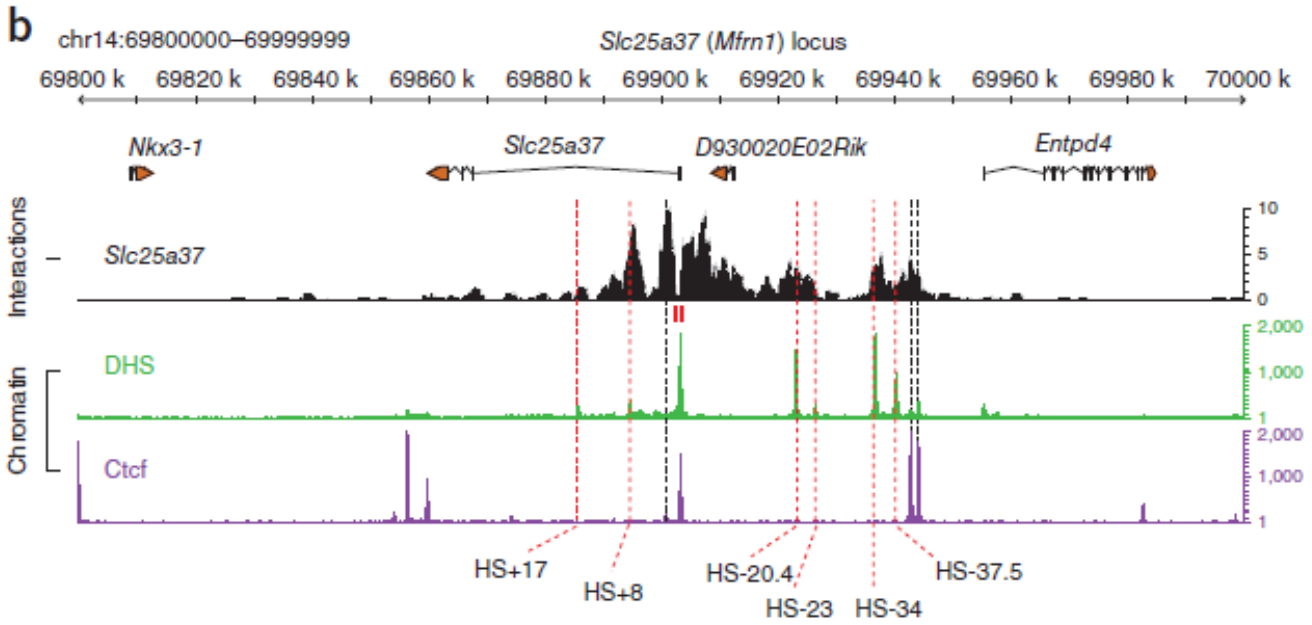
- Chromatin and functional studies have shown that in erythroid cells, a single intergenic erythroid enhancer (+40) lying 40 kb away from Tal1upregulates expression of Tal1 and Pdzk1ip1

- Capture-C from the promoter of either gene clearly identified interactions with the +40 enhancer

Multiple loci in one experiment

Example

- Regulation unknown
- Slc25a37
- Mitochondrial iron transporter



- Also interested in using Capture-C to analyze currently unknown promoter–cis element interactions involving genes of known function

- Capture from the Slc25a37 promoter identified six elements (HS+8, HS+17, HS-20.4, HS-23, HS-34 and HS-37.5), which are variously bound by erythroid transcription factors and bear the chromatin signatures of enhancers

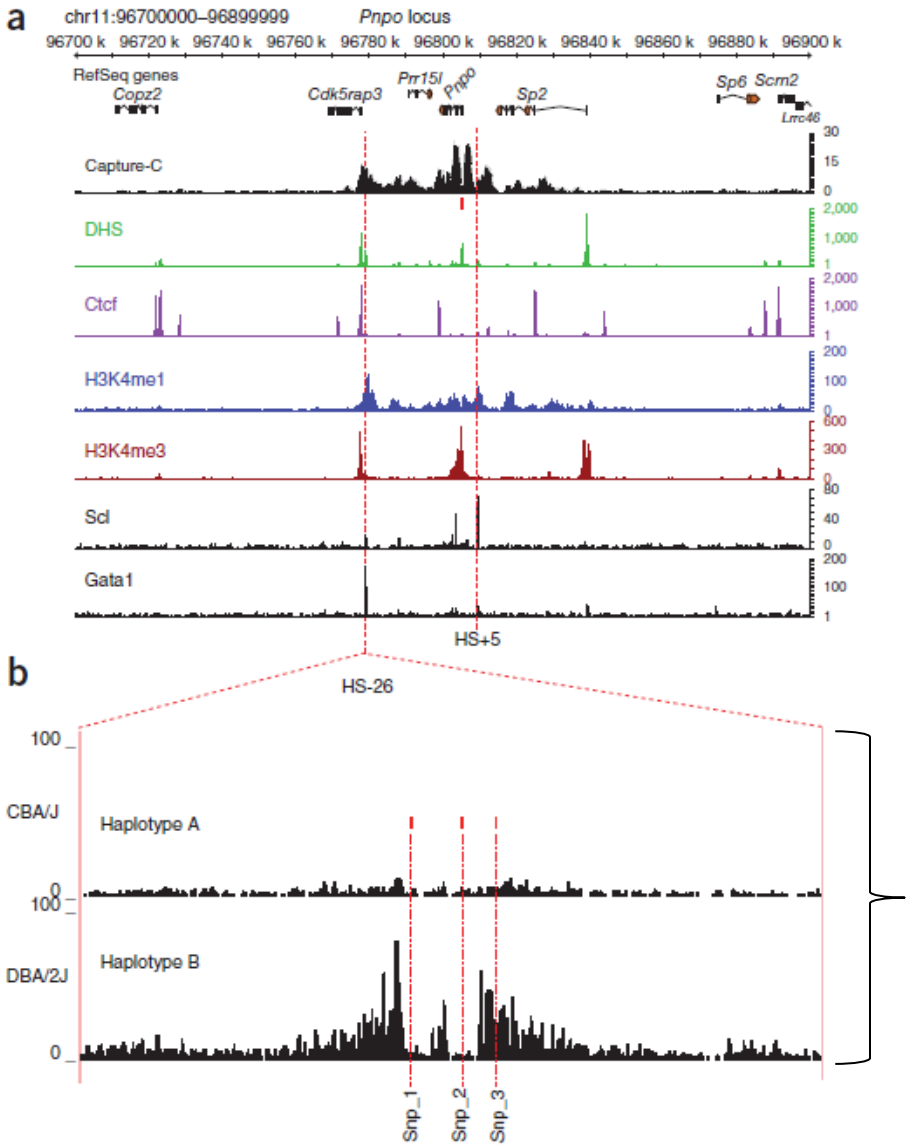
General principals of cis regulation revealed by Capture-C

- Analyzing large numbers of enhancer-promoter interactions:
  - Most interactions occur within ~300 kb of the promoters
  - Frequency of interactions decreases inversely with distance from the transcriptional start sites (TSSs)
  - Interactions between promoters and cis elements are restricted to discrete genomic regions with a clear boundary □What sets the boundaries of such domains is not yet clear

Linking SNPs in regulatory elements to genes they control

- Most SNPs associated with complex diseases lie within a cis-regulatory elements and presumably alter expression of their target genes
- Use Capture-C to link regulatory SNPs to the genes whose expression they affect
- Re-analyzed DNase-seq erythroid data set of eight strains of mice □had a look for SNPs in cis-regulatory elements predicted by Capture-C

Linking SNPs in regulatory elements to genes they control

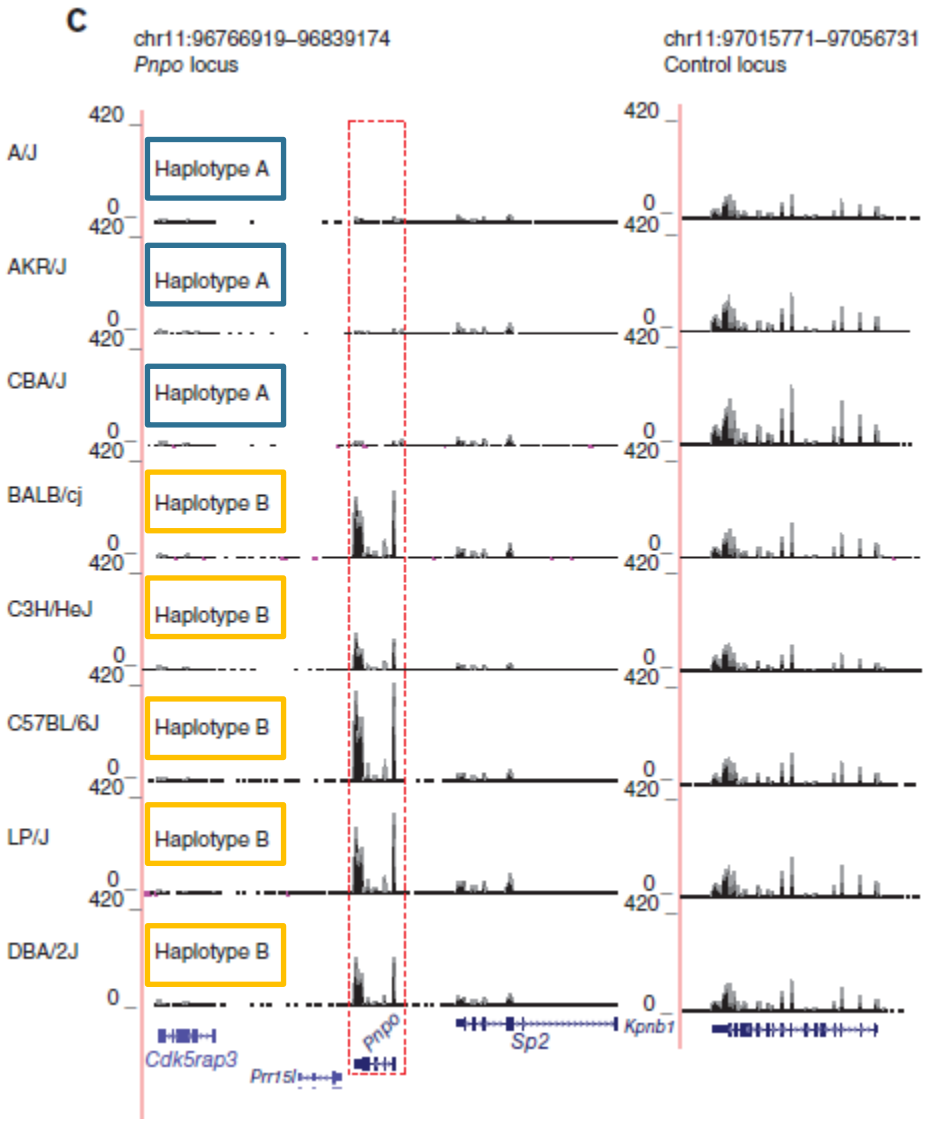


DNase seq

- Pnpo highly upregulated in erythroid cells
- Promoter of Pnpo is linked to a regulatory element with the signature of an enhancer (26kb away from promoter) and binds the erythroid transcription factors Gata1 and Scl
- Found 3 SNPs in this regulatory region

Linking SNPs in regulatory elements to genes they control

Three strains with SNPs in the distal regulatory element have a reduced expression of RNA from the gene Pnpo to which the regulatory element was linked by Capture-C



Conclusion

- Capture-C overcomes many of the limitations of previous methods, providing an unbiased, high-resolution map of cis interactions for hundreds of genes in a single experiment

- Paired-end sequences derived from this method are simply mapped to the genomic sequence to generate a genome-wide map of interaction density with the elements of interest

- Analysis of the mouse α- and β-globin loci as test cases agrees with all previous 3C analyses, identifying all known regulatory elements of the globin genes and no additional elements □was even possible to view all of these interactions from different viewpoints in the same experiment

- Other ‘all versus all’ 3C methods (HiC and ChIA-PET) generate very low numbers (one to tens) of informative interactions from each restriction fragment, Capture-C typically identifies hundreds to thousands of informative interactions from each targeted fragment

- Possible to link SNPs found in regulatory elements to the genes whose expression they influence

| JANUARY 2016 | NATURE METHODS

Multiplexed analysis of chromosome conformation at vastly improved sensitivity

James O J Davies[1], Jelena M Telenius[1], Simon J McGowan[2], Nigel A Roberts[1], Stephen Taylor[2], Douglas R Higgs[1] & Jim R Hughes[1]

[1]Medical Research Council, Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. [2]Computational Biology Research Group, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. Correspondence should be addressed to J.R.H. (jim.hughes@imm.ox.ac.uk).

Next-generation (NG) Capture-C

A high-throughput version of Capture-C

Reasons for new method

- The original Capture-C protocol uses oligos synthesized on a microarray with a design minimum of 40,000, irrespective of the number of desired viewpoints
- So the cost per sample is very high for small designs
- Experimental designs often require much smaller subsets of regions, but from multiple samples
- Capture-C's sensitivity does not easily allow for the analysis of very weak cis interactions
- Capture-C is not applicable for low cell numbers

Factors influencing the method

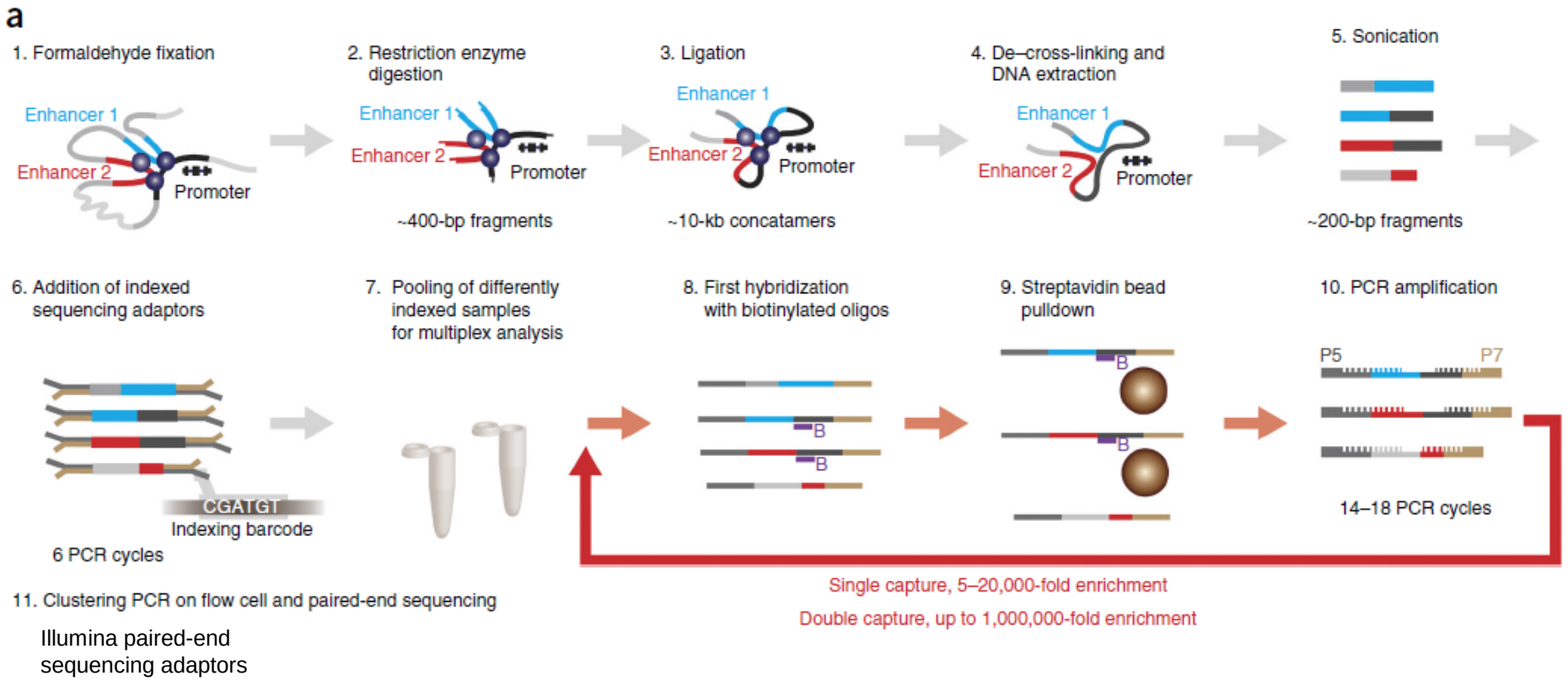
- A maximum of only four interactions can be detected from each region per cell (one from each end of the captured viewpoint fragment on each allele)
- the number of available cells and the complexity of the 3C library determine the maximum number of interactions that can be detected

- The hybridization efficiency of the capture probe is important, and is largely dictated by the underlying sequence

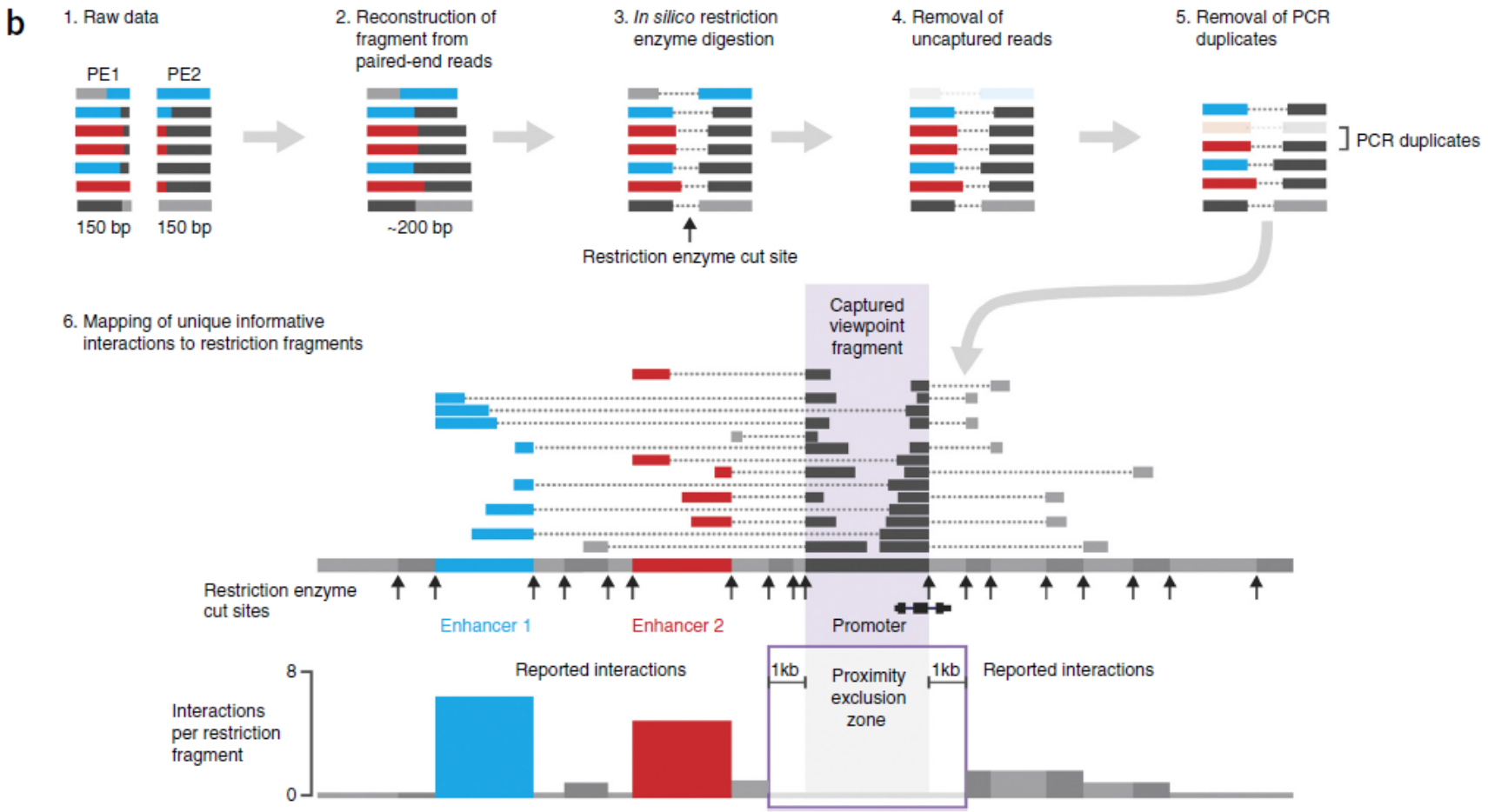
- The efficiency of the assay and depth of sequencing required are determined by the proportion of background fragments from noncaptured DNA contaminating the library

Experimental Workflow

- Redesigned the Capture-C protocol
- Using DNA rather than RNA biotinylated oligos ☐vastly superior efficiency



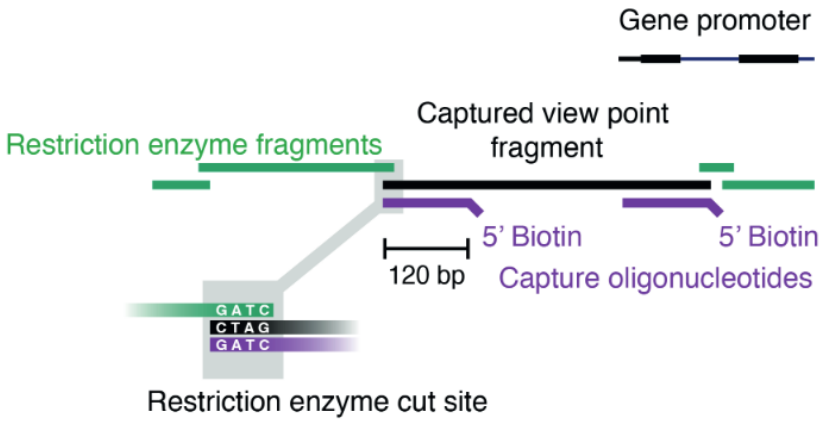
Experimental Workflow



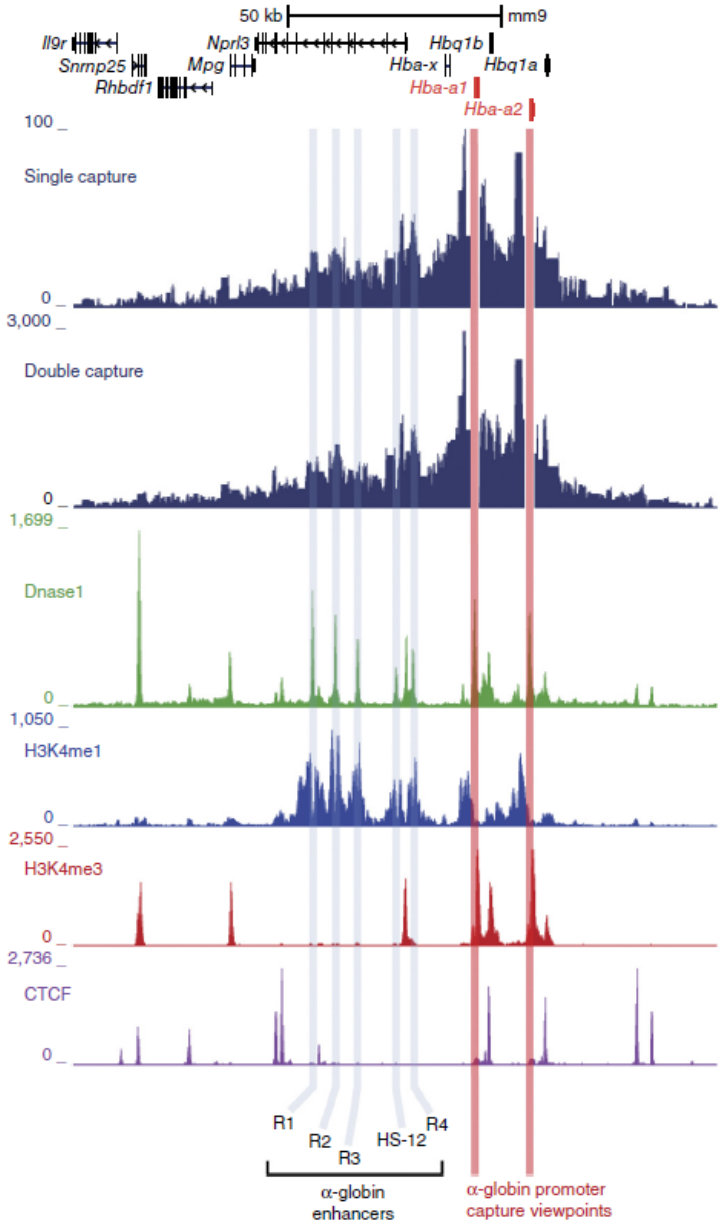
Optimization of the Protocol

- Minimized losses during the addition of sequencing adaptors and by pooling material
- Increased total input to 10 ug of 3C library
- Doubled complexity of the material used for hybridization
- □8 x more ligation junctions than in the previous protocol

- To decrease the amount of background fragments
  - Simplified the library design to use single 120-bp biotinylated DNA oligonucleotides, which included the restriction sites, to capture each end of the target restriction fragment
  - Introduced a second, sequential round of capture □decreased background of uncaptured material
- Sequencing depth was no longer limiting and PCR duplicates could be easily excluded bioinformatically



Testing the new Protocol

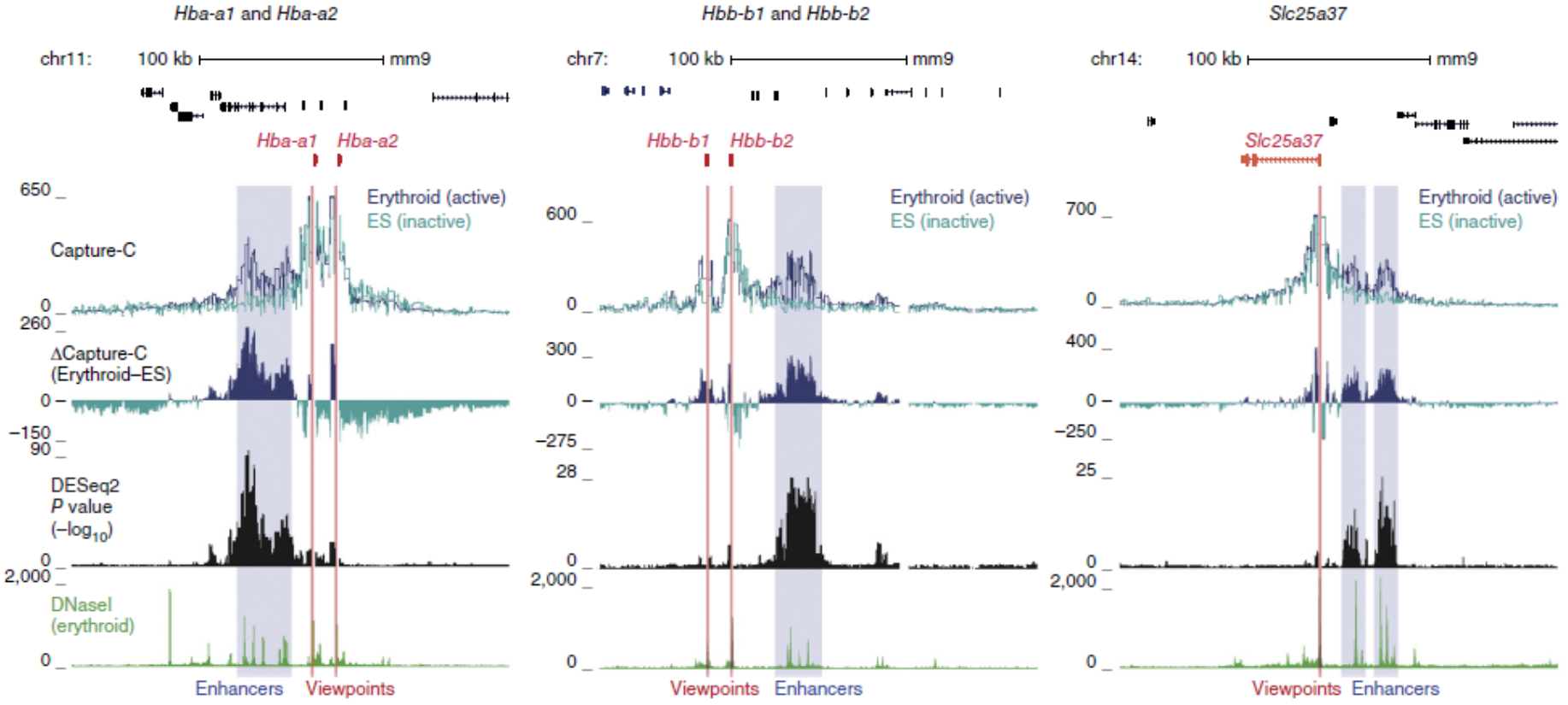


Probes against Hba-a1 and Hba-a2 promoter

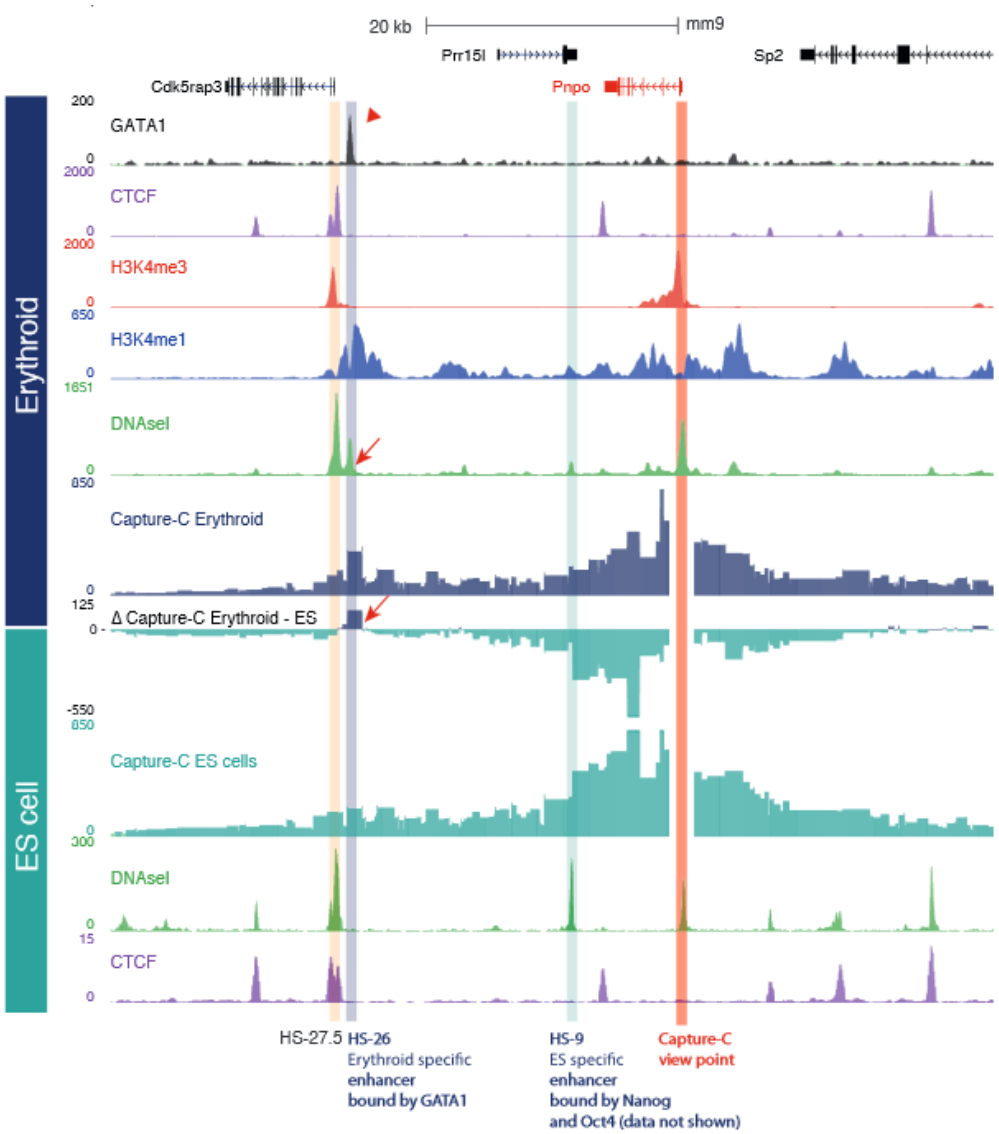
Saw no differences in the local interaction profiles between single and double capture

Testing the new Protocol

- Combined the capture probes for α-globin, β-globin and Slc25a37 (Mitoferrin-1)
- Matched previously determined patterns of interaction
- Interactions with the local erythroid enhancers were clearly and specifically increased in erythroid cells
- For the same depth of sequencing, the double capture increased the sensitivity of the profile 30-fold



Identification of regulatory elements



Subtractive analysis uncovered tissue-specific regulation of genes

For example, Pnpo is specifically upregulated in mouse erythroid cells by an erythroid-specific enhancer (HS-26)

Robust interaction profiles from low numbers of cells

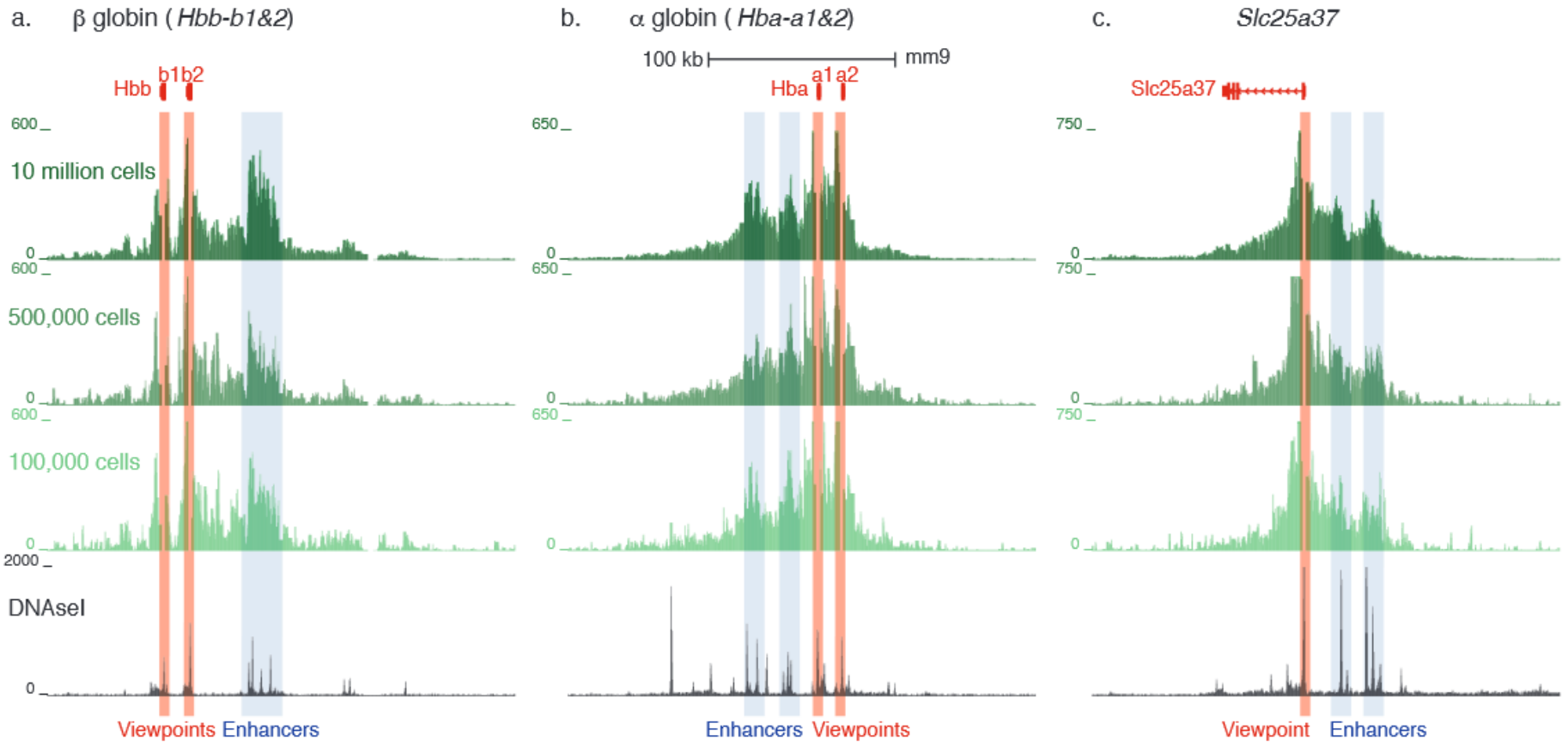
- In human primary tissues, cell numbers are often limited

- Use of low cell numbers did not alter the digestion efficiency, and the amount of DNA extracted per cell was constant □optimized the preparation of hybridization-reaction material for the reduced DNA content of the 3C libraries
  - Smaller volumes
  - More PCR cycles

- Using 100,000 cells □generated ~19,000 interactions, compared to an average of 137,000 when cell number was not limiting

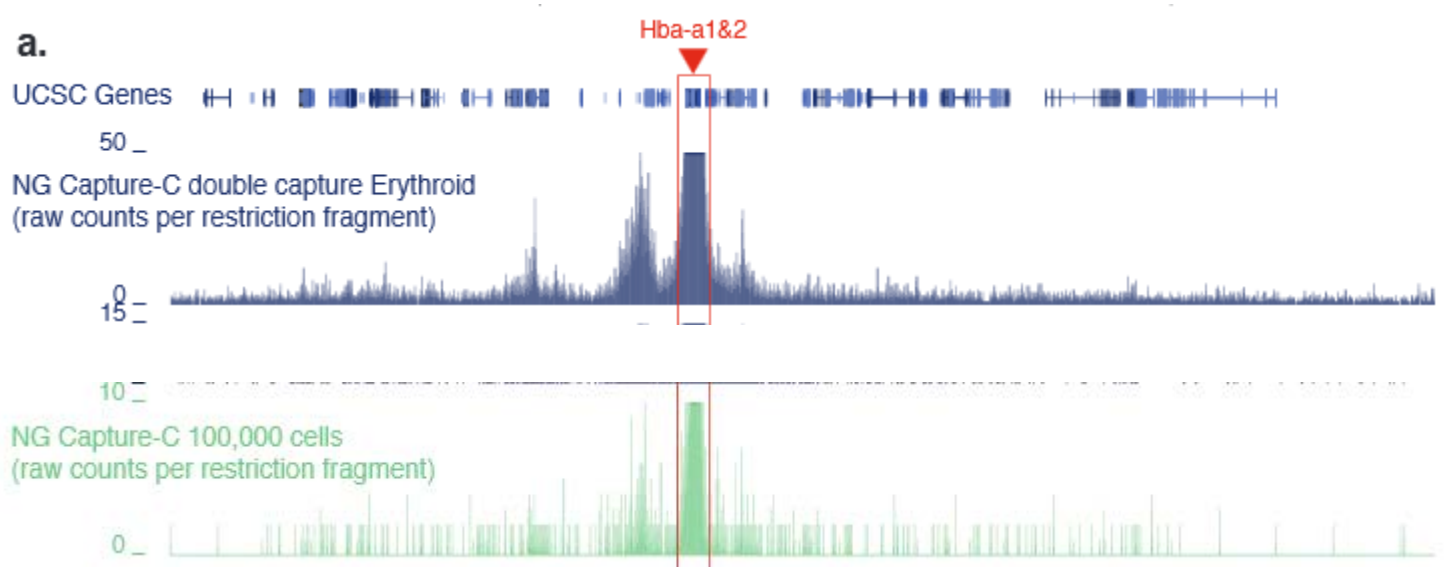
Robust interaction profiles from low numbers of cells

The interaction profiles at the α- and β-globin loci remained virtually unchanged

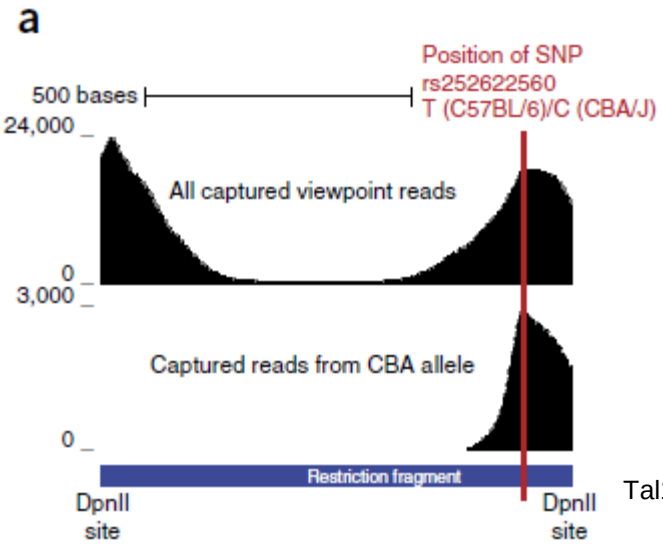


Robust interaction profiles from low numbers of cells

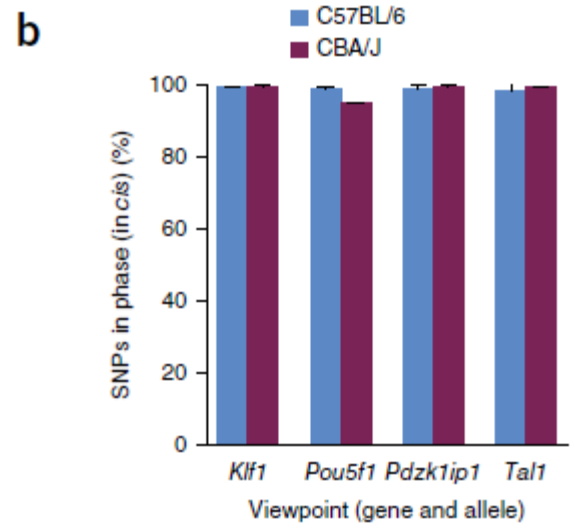
Weak, long-range interactions became difficult to determine reproducibly when only 100,000 cells were used



Generation of SNP-specific interaction profiles



Tal1 promoter fragment

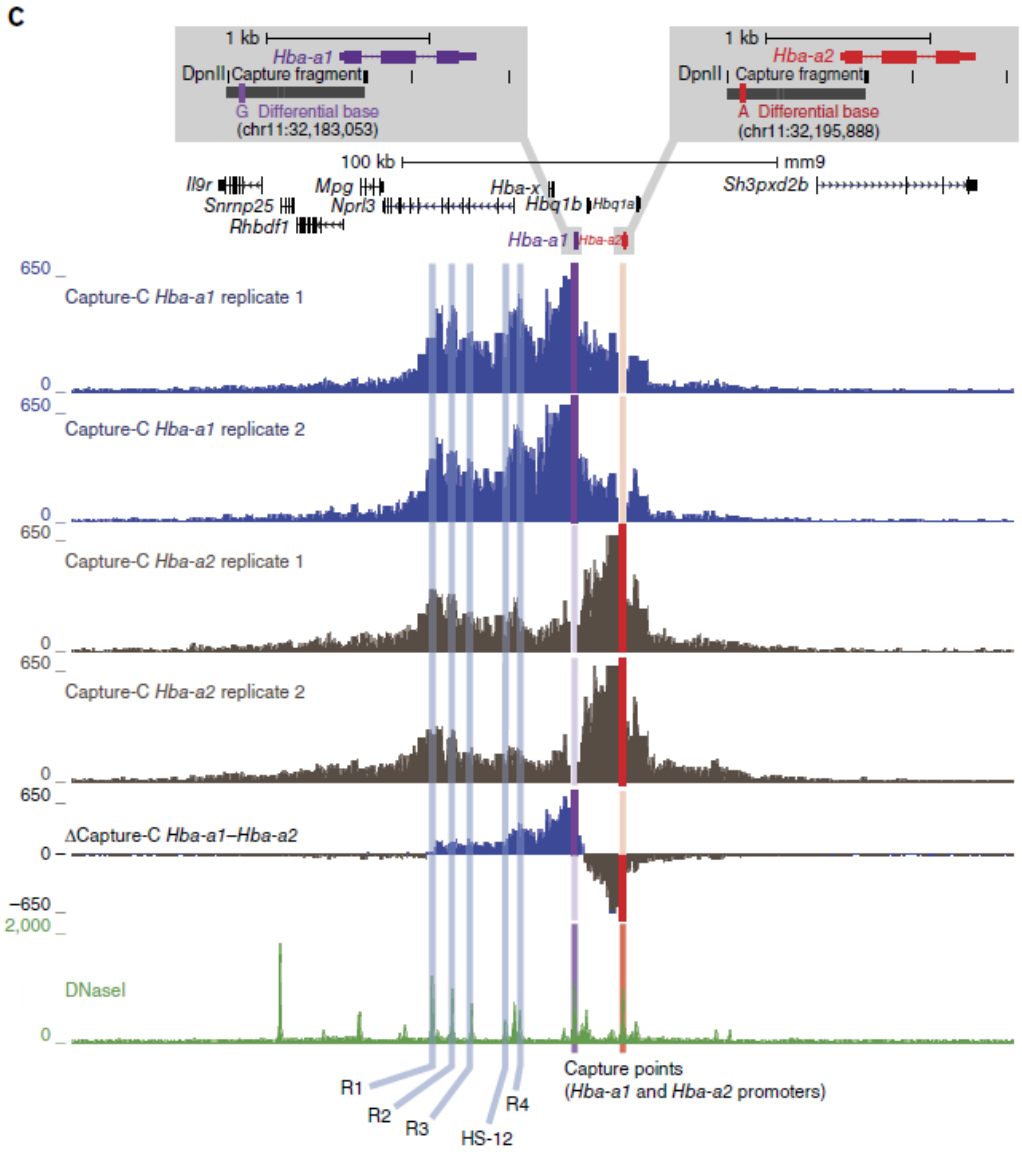


SNPs are frequently heterozygous and may affect the regulatory interactions of the affected allele

Allel-specific interaction profiles

Demonstrates that the chromosome interacts predominately with itself in Cis rather than with its sister chromatid

SNP specific NG-Capture-C



- Can also be applied to non-allelic SNPs Hba-a1 and Hba-a2 differing at only a few positions
- Generated specific interaction profiles for Hba-a1 and Hba-a2
- Hba-a1preferentially interacted with HS-12 and R4
- Hba-a2 and Hba-a1had very similar interactions with the R1 and R2 regulatory elements
- Hba-a2 interacted much more strongly with the chromatin between the two genes

Conclusion

- Exceeds the sensitivity and resolution of detection of current 3C methods
- □ncreased through double capture and the ability to remove PCR duplicates

- NG Capture-C allows the simultaneous capture of multiple samples in a single reaction, greatly increasing the throughput and economy of the assay

- This approach identifies all known regulatory elements at characterized test loci at levels of resolution previously not possible

- Optimized the NG Capture-C method for smaller numbers of cells (~100,000)

- Possible to generate SNP-specific interaction profiles

Thank you for your attention!!