

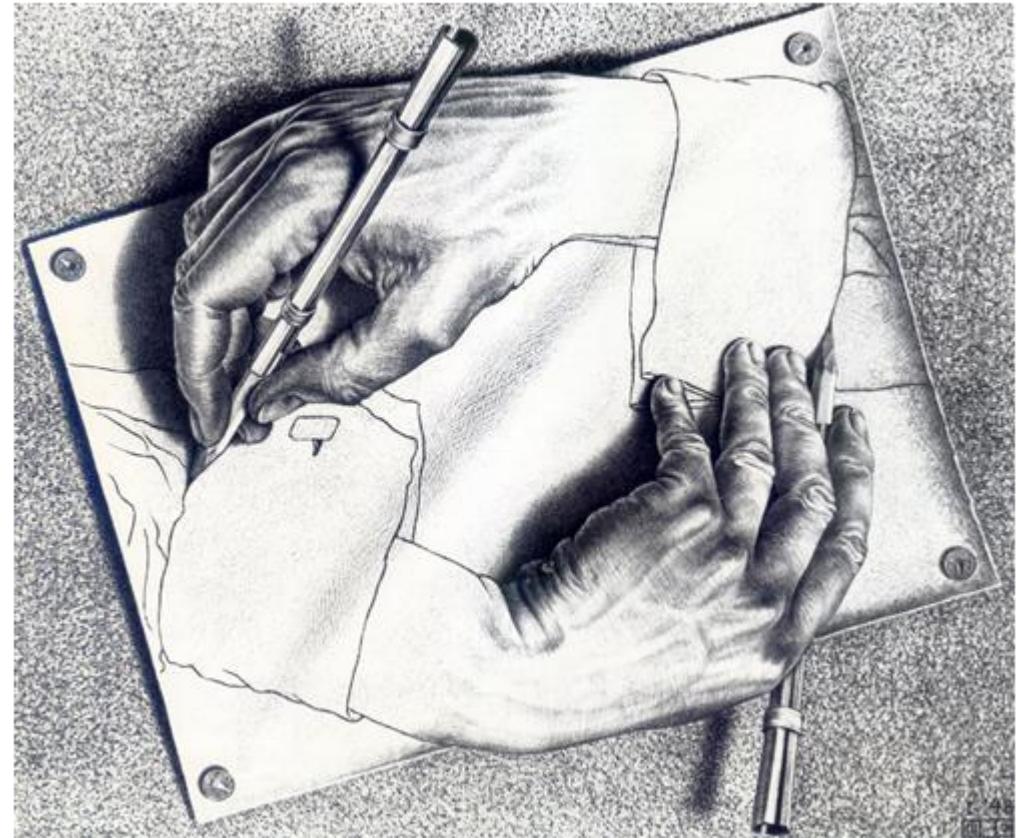
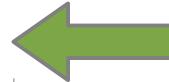
3D genome sequencing

Hi-C data analysis



Outline

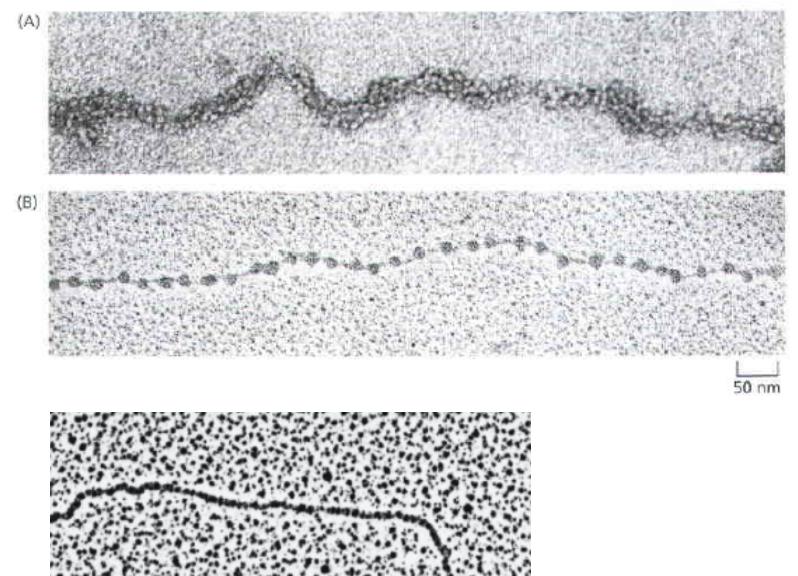
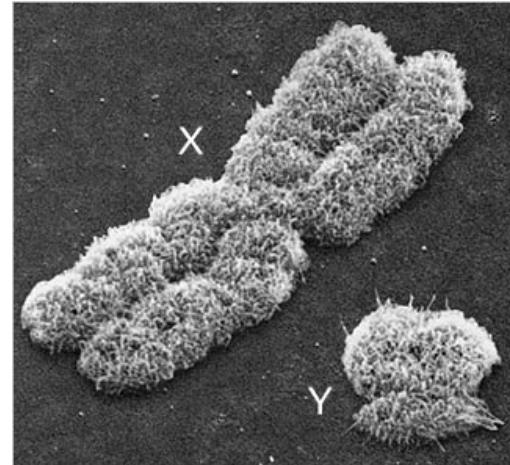
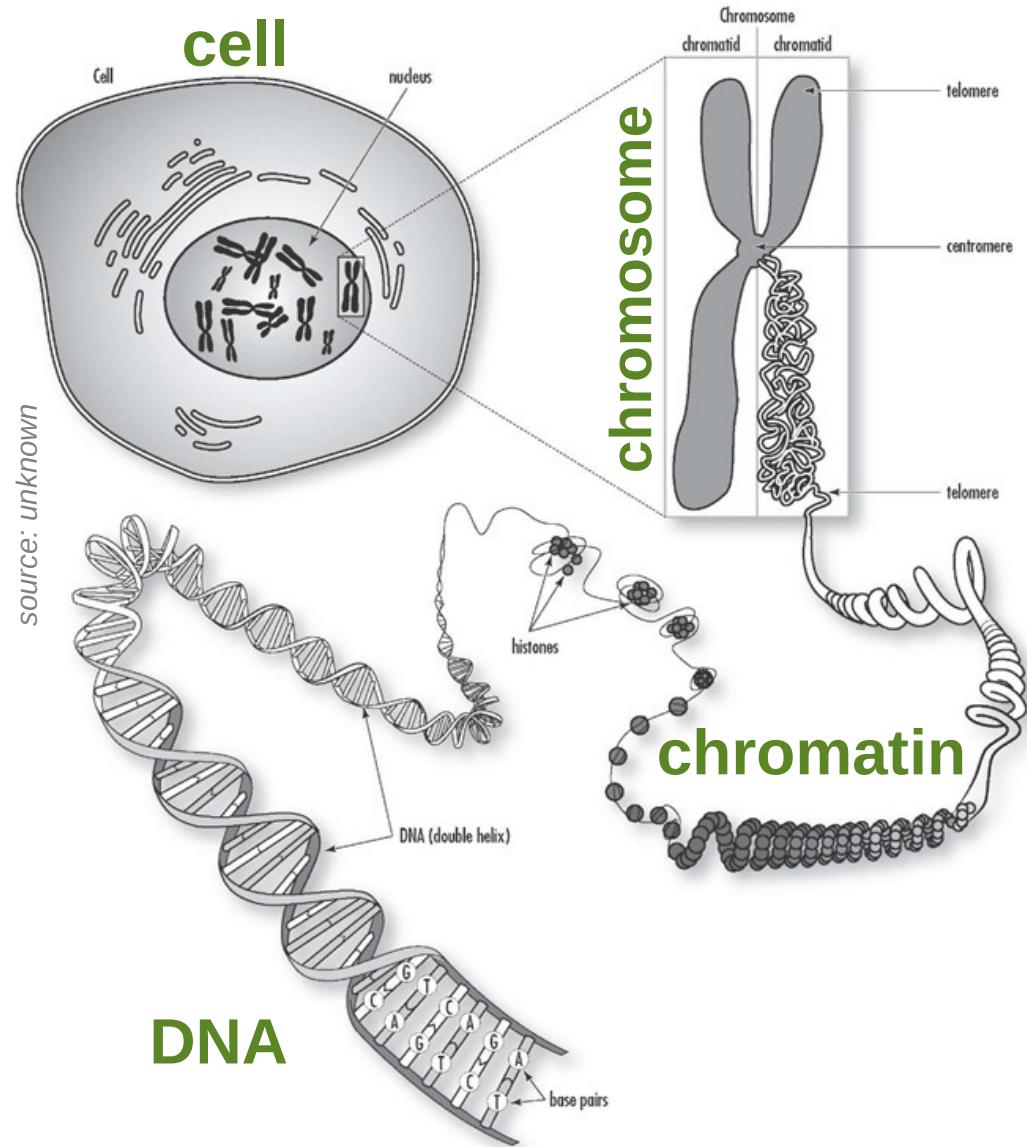
- ◆ Intro
- ◆ Current projects
- ◆ Experiment
- ◆ Data analysis
(part 1)



M.C. Escher, 1948

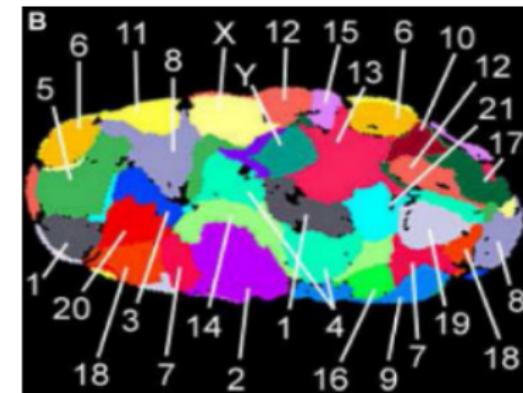
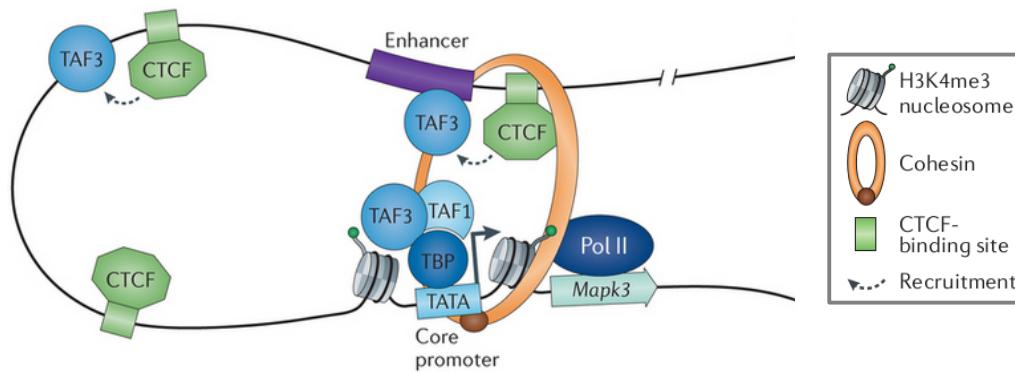
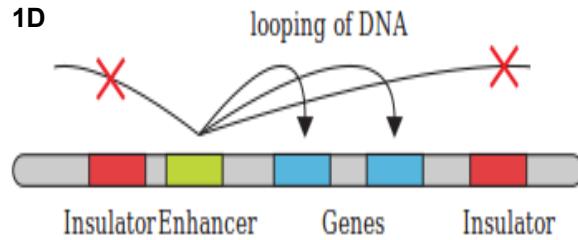
Life, cell, chromosome & DNA

source: unknown

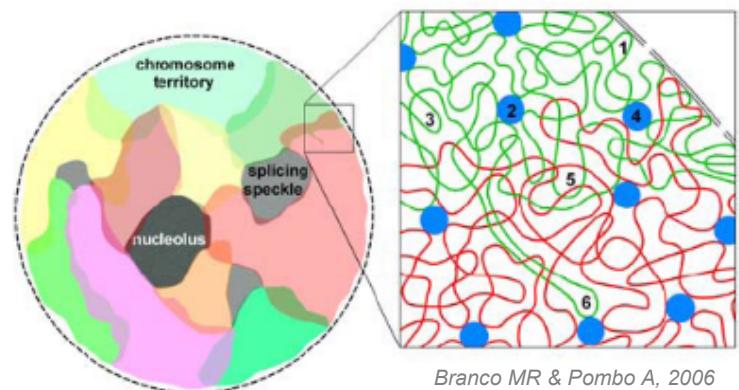


Life, cell, chromosome & DNA

From DNA loops to ... chromosome territories



Bolzer A et al. 2005

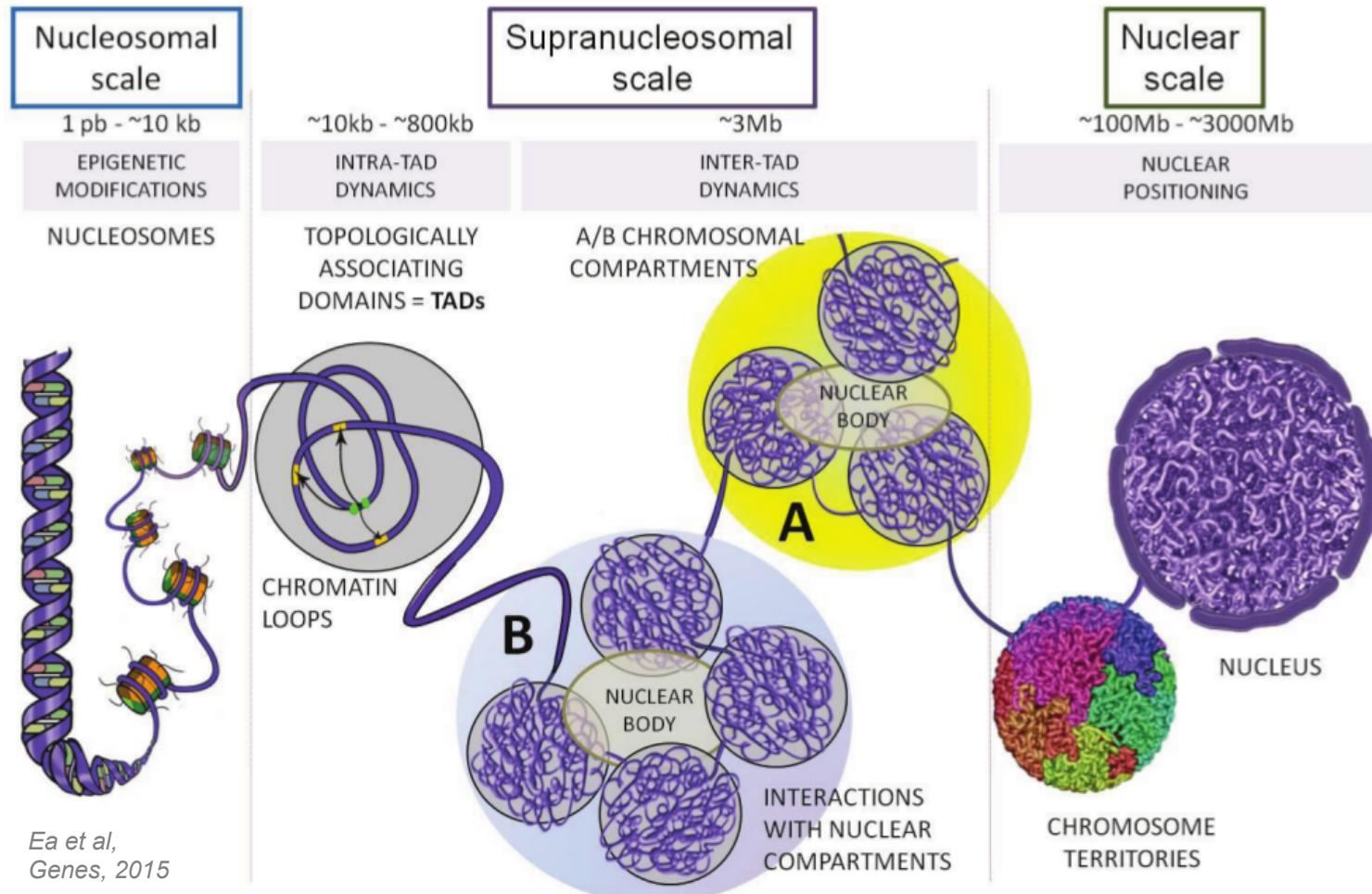


Branco MR & Pombo A, 2006

Genome 3D structure is organized and regulates gene expression

Life, cell, chromosome & DNA

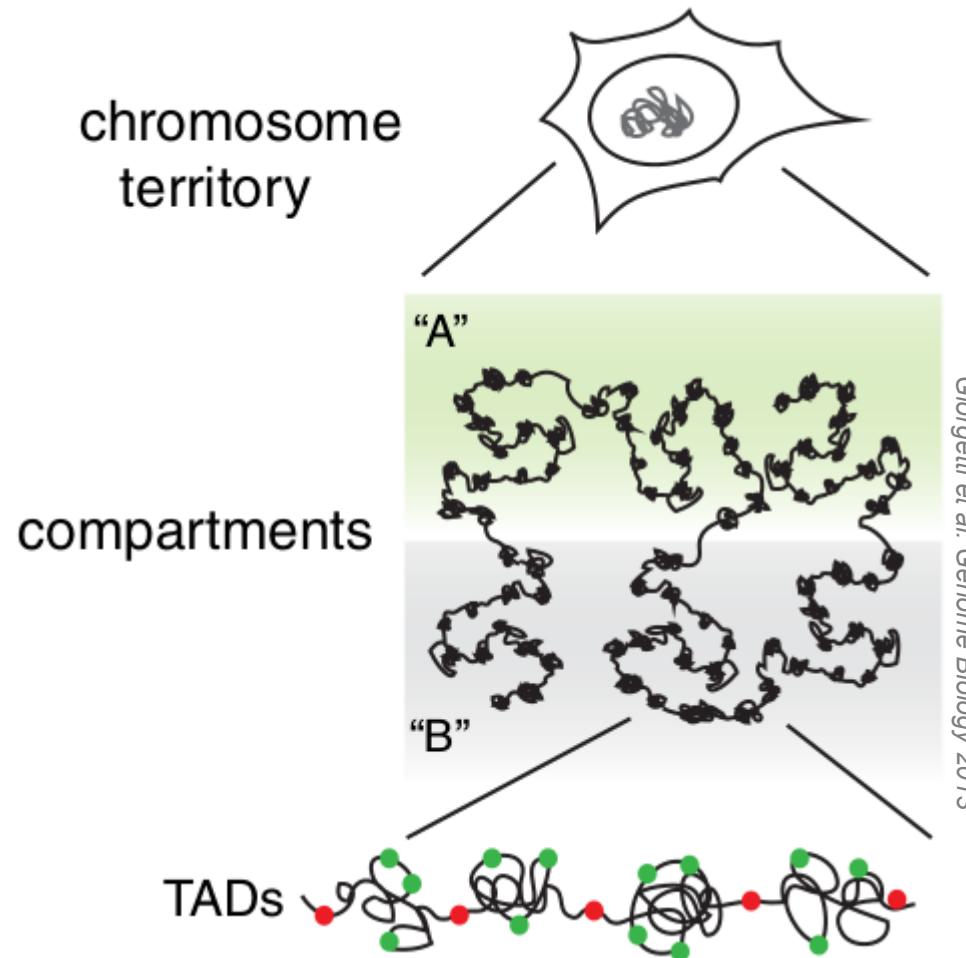
From DNA loops to ... chromosome territories



Genome 3D structure is organized and regulates gene expression

Life, cell, chromosome & DNA

From DNA loops to ... chromosome territories



Genome 3D structure is organized and regulates gene expression

From structure to function

Chromosomal Contact Permits Transcription between Coregulated Genes

Stephanie Fanucchi,¹ Youtaro Shibayama,¹ Shaun Burd,¹ Marc S. Weinberg,^{3,4} and Musa M. Mhlanga^{1,2,*}

¹Gene Expression and Biophysics Group, Synthetic Biology Emerging Research Area, Biosciences Unit, Council for Scientific and Industrial Research, Pretoria, Gauteng 0001, South Africa

²Unidade de Biofísica e Expressão Genética, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, 1649-028 Portugal

³Antiviral Gene Therapy Research Unit, Department of Molecular Medicine and Haematology, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng 2193, South Africa

⁴Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

*Correspondence: yoda@mhlangalab.org

<http://dx.doi.org/10.1016/j.cell.2013.09.051>

SUMMARY

Transcription of coregulated genes occurs in the context of long-range chromosomal contacts that

Nucleic Acids Research Advance Access published February 4, 2015

2012). These highly sensitive assays can nascent mRNA and have revealed the FISH foci in a fraction of the population 2010; Parantonis et al. 2010). This suc-

Nucleic Acids Research 2015 1
doi: 10.1093/nar/gkv046



Spatial re-organization of myogenic regulatory sequences temporally controls gene expression

Akihito Harada¹, Chandrashekara Mallappa², Seiji Okada¹, John T. Butler², P. Baker^{2,3}, Jeanne B. Lawrence², Yasuyuki Ohkawa^{1,2,*} and Anthony N. Im

¹Department of Advanced Medical Initiatives, JST-CREST, Faculty of Medicine, Kyushu University, 812-8582, Japan. ²Department of Cell and Developmental Biology, University of Massachusetts

22. G. A. Way, *Nat. Rev. Genet.* 8, 206–225 (2007).
22. S. B. Carroll, *Cell* 134, 25–36 (2008).

and K. Paszkelewicz and the Exeter Sequencing Service facility for genome sequencing services. This work was also support

TRANSCRIPTION

CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation

Varun Narendra,^{1,2}
Esteban O. Mazzoni

Polycomb and Trithorax identity by establishing

the *Hox* clusters. Here we demonstrate that the CTCF-binding factor (CTCF) functions



HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CON

Institution: SWETS SUBSCRIPTIONSERVICE Sign In via User Nam

Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin

Elizabeth Ing-Simmons^{1,2,7}, Vlad C. Seitan^{1,7}, Andre J. Faure^{3,8}, Paul Flicek^{3,4},

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

Dario G. Lupiáñez,^{1,2} Katerina Kraft,^{1,2} Verena Heinrich,² Peter Krawitz,^{1,2} Francesco Brancati,³ Eva Klar,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 Borsig,¹¹ Stefan A. Haas,¹¹ Marco Osterwalder,¹² Bernd Timmermann,¹³ Jochen Hecht,^{1,14} Malte Spielmann,^{1,2,14} Axel Visel,^{12,15,16} and Stefan Mundlos^{1,2}

¹Max Planck Institute for Molecular Genetics, RG Development & Disease, 14195 Berlin, Germany

²Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany

³Medical Genetics Unit, Policlinico Tor Vergata University Hospital, 00133 Roma, Italy

Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression

E. Josephine Clowney,¹ Mark A. LeGros,^{2,4} Colleen P. Eirene C. Markenskoff-Papadimitriou,³ Markko Myllys,¹ and Stavros Lomvardas^{1,2,3,*}

Leading Edge
Previews

A CRISPR Connection between Chromatin Topology and Genetic Disorders

Bing Ren^{1,*} and Jesse R. Dixon¹

¹Ludwig Institute for Cancer Research, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0653

*Correspondence: bren@ucsd.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.047>

Genome 3D structure is organized and regulates gene expression

Next Genera
USA Congre

27 - 28 October 2015,

Cell

ns to human diseases variants can interrupt altered spatiotemporal

in cases (Figure 1). Remarkably, mice carrying these structures accurately reproduce

From structure to function

Chromosomal Contact Permits Transcription between Coregulated Genes

Stephanie Fanucchi,¹ Youtaro Shibayama,¹ Shaun Burd,¹ Marc S. Weinberg,^{3,4} and Musa M. Mhlanga^{1,2,*}

¹Gene Expression and Regulation Group, Synthetic Biology Emerging Research Area, Biosciences Unit, Council for Scientific and Industrial Research, South Africa

²Unit of Synthetic Biology, Synthetic Biology Emerging Research Area, Biosciences Unit, Council for Scientific and Industrial Research, South Africa

³Antiviral Health Research Institute, University of Cape Town, South Africa

⁴Department of Cell Biology, University of Cape Town, South Africa

*Correspondence: mmhlanga@csir.co.za

<http://dx.doi.org/10.1126/science.1253313>

SUMMARY

Transcriptional control of gene expression is often mediated by long-range interactions between regulatory elements. These interactions are thought to occur through the formation of nucleoprotein complexes that bring distant genomic loci into close proximity. However, the mechanism by which these complexes are formed and how they regulate transcription remains elusive. Here we show that the CTCF-binding factor (CTCF) functions as a nucleoprotein complex that permits transcription between coregulated genes by bringing them into close proximity. Our results suggest that CTCF-mediated gene regulation is achieved through the formation of nucleoprotein complexes that permit transcription between coregulated genes.

**Spatial
sequencing**

Akihito Hara,¹ P. Baker^{2,3},

¹Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

²22. G. A. Way, Nat. Rev. Mol. Cell Biol. 12, S. B. Carroll, Cell

TRANSCRIPTION

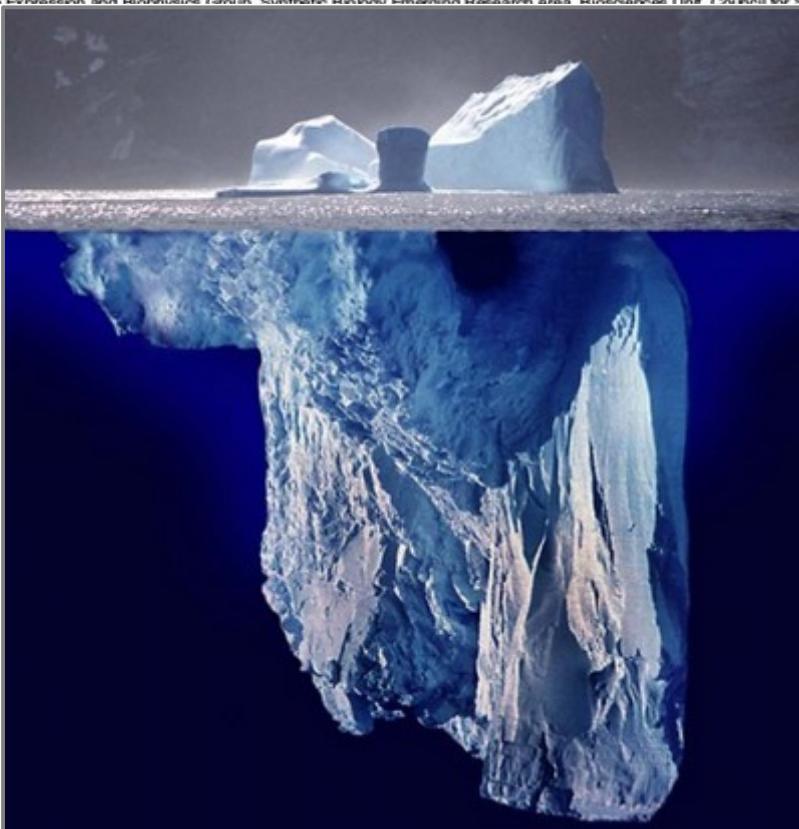
CTCF

chromatin domains at the Hox clusters during differentiation

Varun Narendra,^{1,2} Esteban O. Mazzoni,¹

Polycomb and Trithorax identity by establishing

the Hox clusters. Here we demonstrate that the CTCF-binding factor (CTCF) functions



E. Josephine Clowney,¹ Mark A. LeGros,^{2,4} Colleen P. Eirene C. Markenskoff-Papadimitriou,³ Markko Myllys,^{1,2,*} and Stavros Lomvardas^{1,2,4,*}

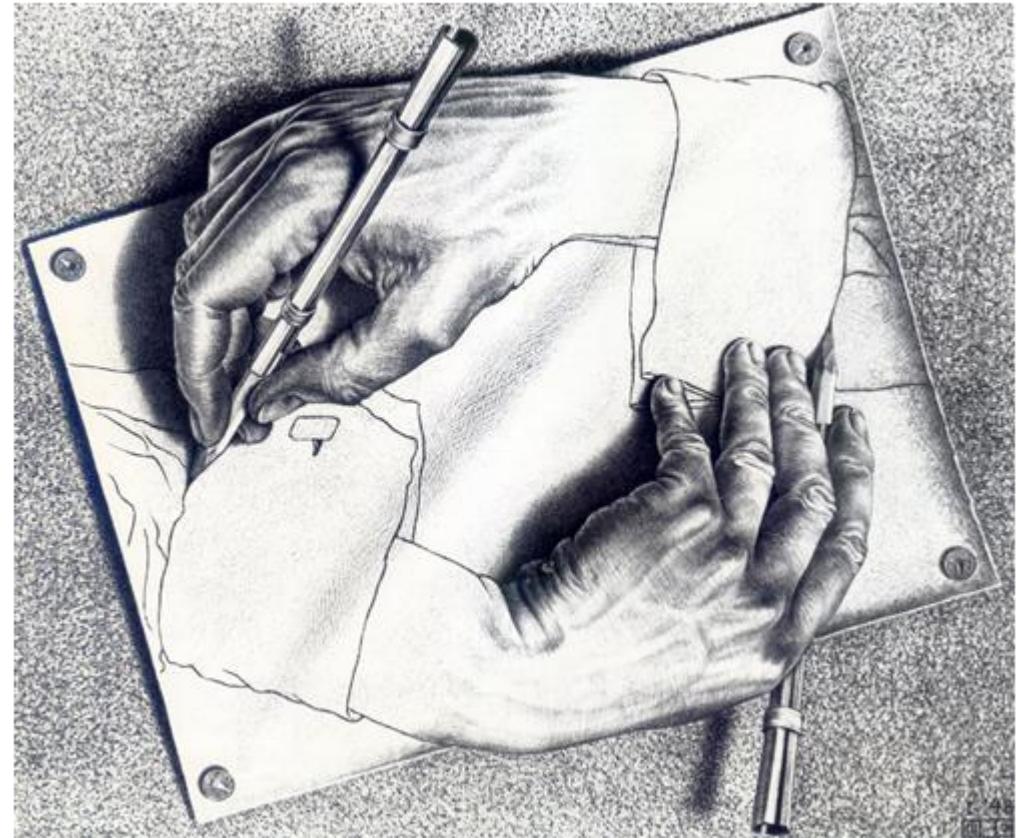
Bing Ren^{1,*} and Jesse R. Dixon¹

¹Ludwig Institute for Cancer Research, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0653
²Correspondence: bren@ucsd.edu
<http://dx.doi.org/10.1126/science.1253313>

Genome 3D structure is organized and regulates gene expression

Outline

- ◆ Intro
- ◆ Current projects
- ◆ Experiment
- ◆ Data analysis
(part 1)



M.C. Escher, 1948

The FR-AgENCODE project

- ◆ a French pilot project of the FAANG consortium
- ◆ goal: improve functional annotation of livestock genomes

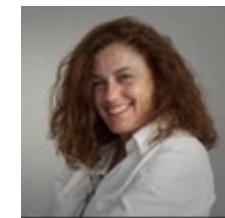


The FR-AgENCODE project

- ◆ a French pilot project of the FAANG consortium
- ◆ goal: improve functional annotation of livestock genomes
- ◆ founding: INRA, France (300KE)
- ◆ 4 INRA sites, 9 labs, 58 scientists
- ◆ 4 species: pig, chicken, cattle, goat
- ◆ primary targets: liver & blood cells (CD4+ & CD8+)
- ◆ molecular assays: RNA-seq, Hi-C & ATAC-seq
- ◆ duration: 2015-2017



FAANG
Functional Annotation of Animal Genomes



E. Giuffra,
INRA GABI



S. Foissac,
INRA GenPhySE

The FR-AgENCODE project

Sampling: 40+ tissues

(liver, CD4+, CD8+, sperm, plasma, heart, lung, skin, fat, duodenum, ileum, jejunum, cerebellum, frontal lobe, olfactory bulb, trigeminal ganglia, hypothalamus, pancreas, adrenals, kidney, muscle, bone, joints, spleen, lymphatic nodes, peyer's patches, ovary, oocytes, oviduct, uterus, mammary gland, acini, testis, seminal vesicle, etc)

2x ♂
2x ♀



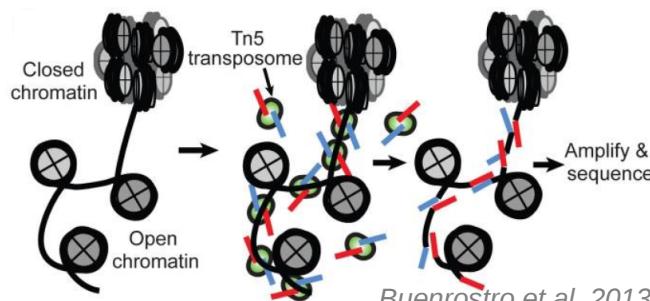
Sus scrofa
(Large White)

Gallus gallus
(White Leghorn)

Bos Taurus
(Holstein)

Capra hircus
(Alpine)

=> INRA CRB-Anim biorepository

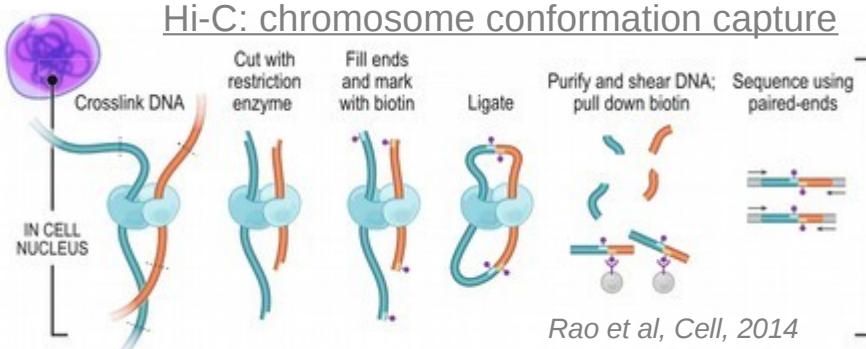


ATAC-seq: chromatin accessibility

RNA-seq: long & short RNAs

Molecular assays: 3 target tissues

transcriptome & chromatin structure profiling
polyA+ RNA-seq (mRNAs & lncRNAs, 130M RP/lib)
small RNA-seq (miRNAs & <200nt RNAs, 40MR/lib)
Hi-C (130M RP/lib) & ATAC-seq (40M RP/lib)

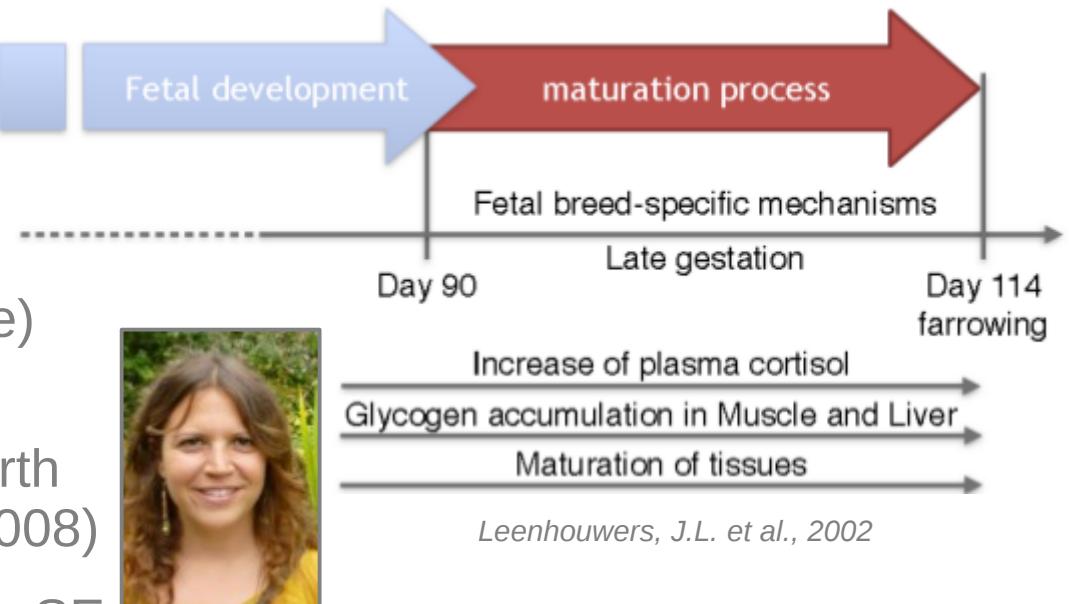


Bioinformatics data analysis

genome annotation, gene expression, lncRNAs & sRNAs
annotation/prediction, chromosome interaction matrices &
contact heatmaps, allele-specific expression, chimeric
transcripts detection, comparative genomics, etc

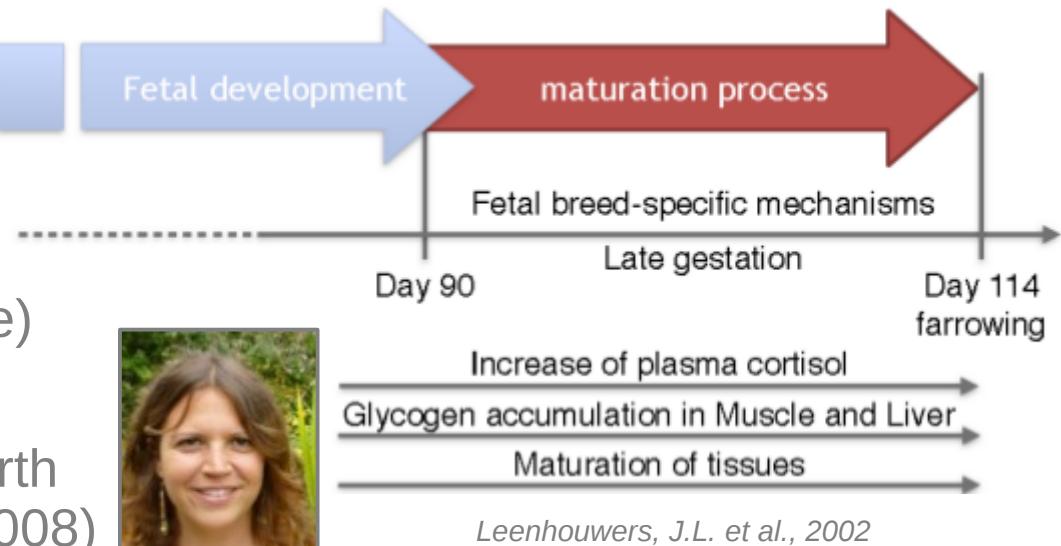
The Porcinet project

- ◆ Context: pig development and maturation process (muscle)
- ◆ Agronomic interest: maturity => piglet mortality at birth (Canario L., 2006, Baxter E.M 2008)
- ◆ PhD thesis: Maria Marti, GenPhySE



The Porcinet project

- ◆ Context: pig development and maturation process (muscle)
- ◆ Agronomic interest: maturity => piglet mortality at birth (Canario L., 2006, Baxter E.M 2008)
- ◆ PhD thesis: Maria Marti, GenPhySE
- ◆ Experimental design: 90 days vs. 110 days of development, 3 replicates
- ◆ Assays: Hi-C, ChiP-Seq (CTCF), RNA-seq



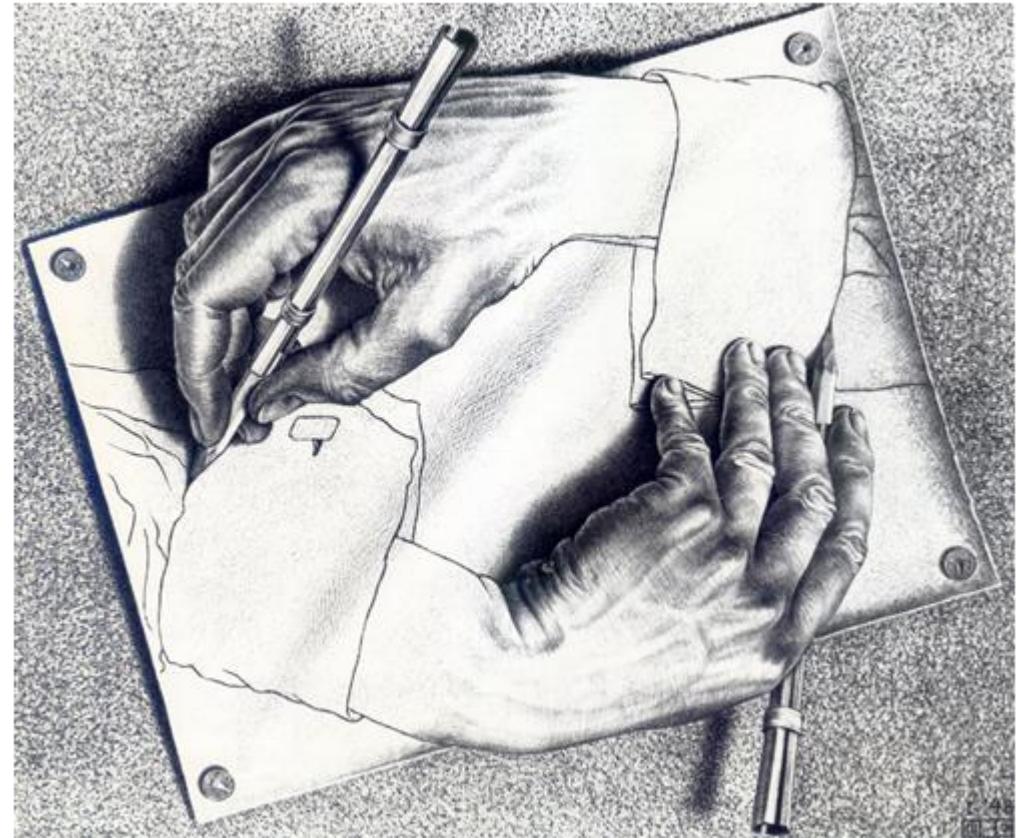
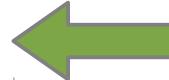
90 days (x3)



110 days (x3)

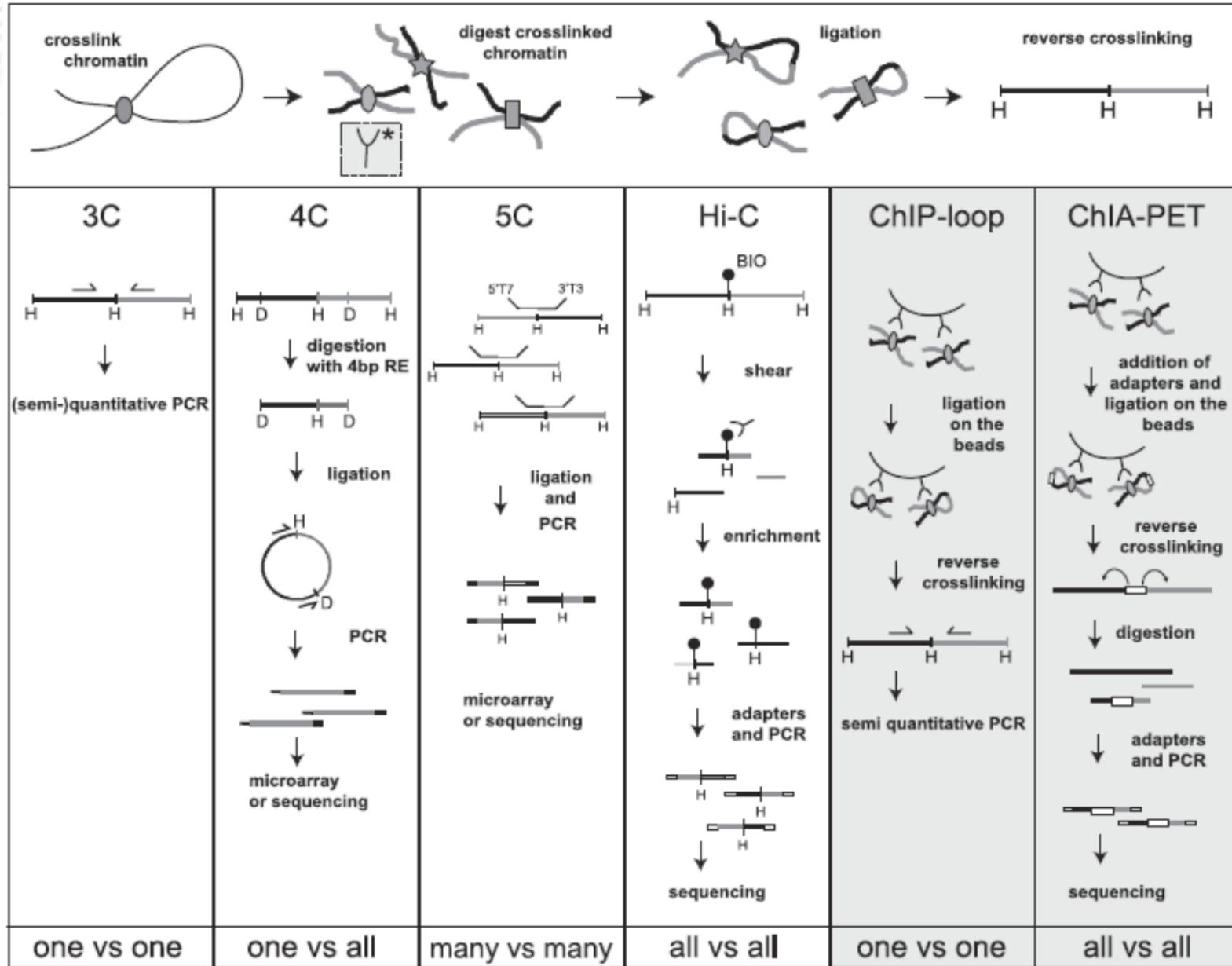
Outline

- ◆ Intro
- ◆ Current projects
- ◆ Experiment
- ◆ Data analysis
(part 1)

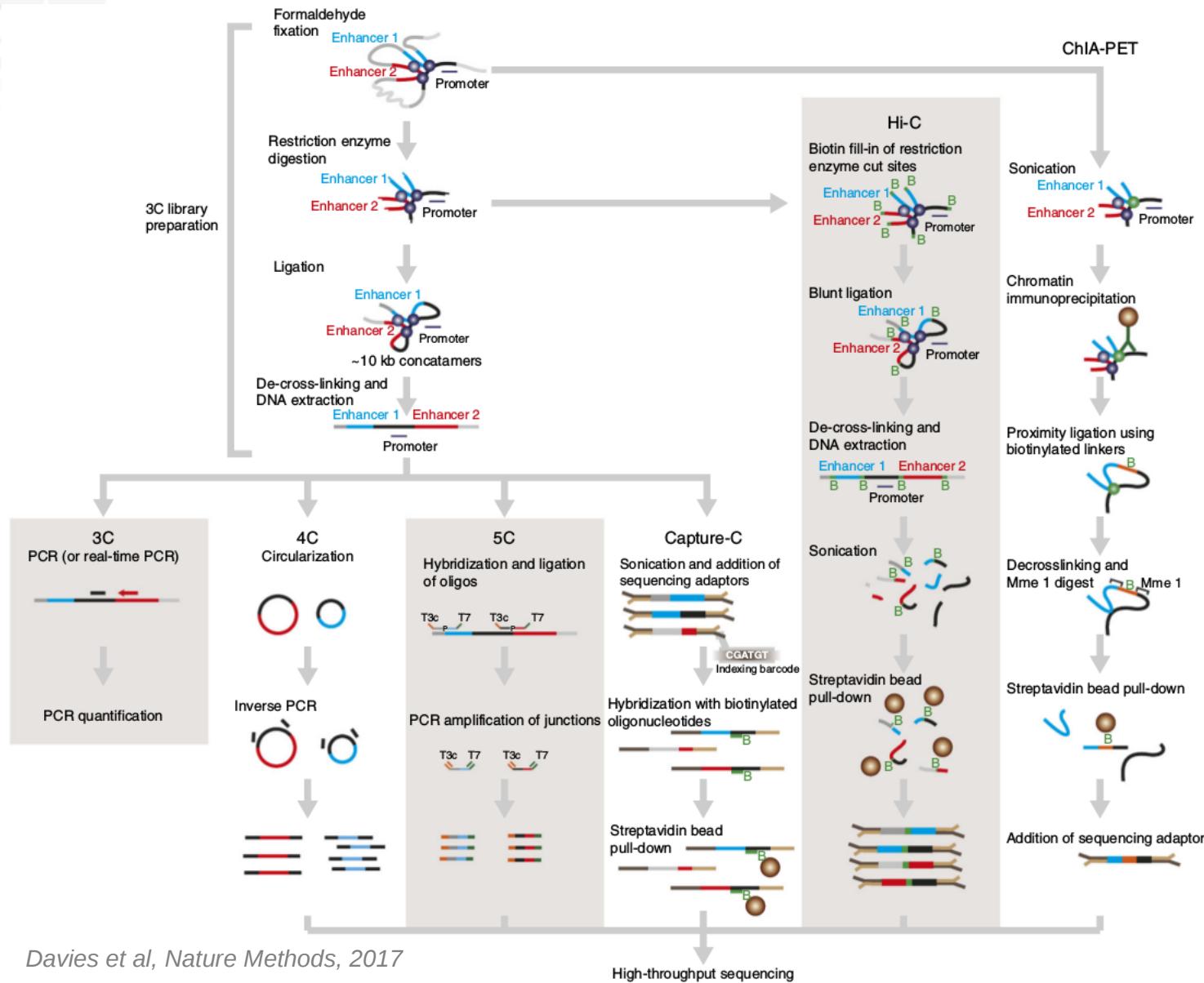


M.C. Escher, 1948

Chromosome Conformation Capture assays



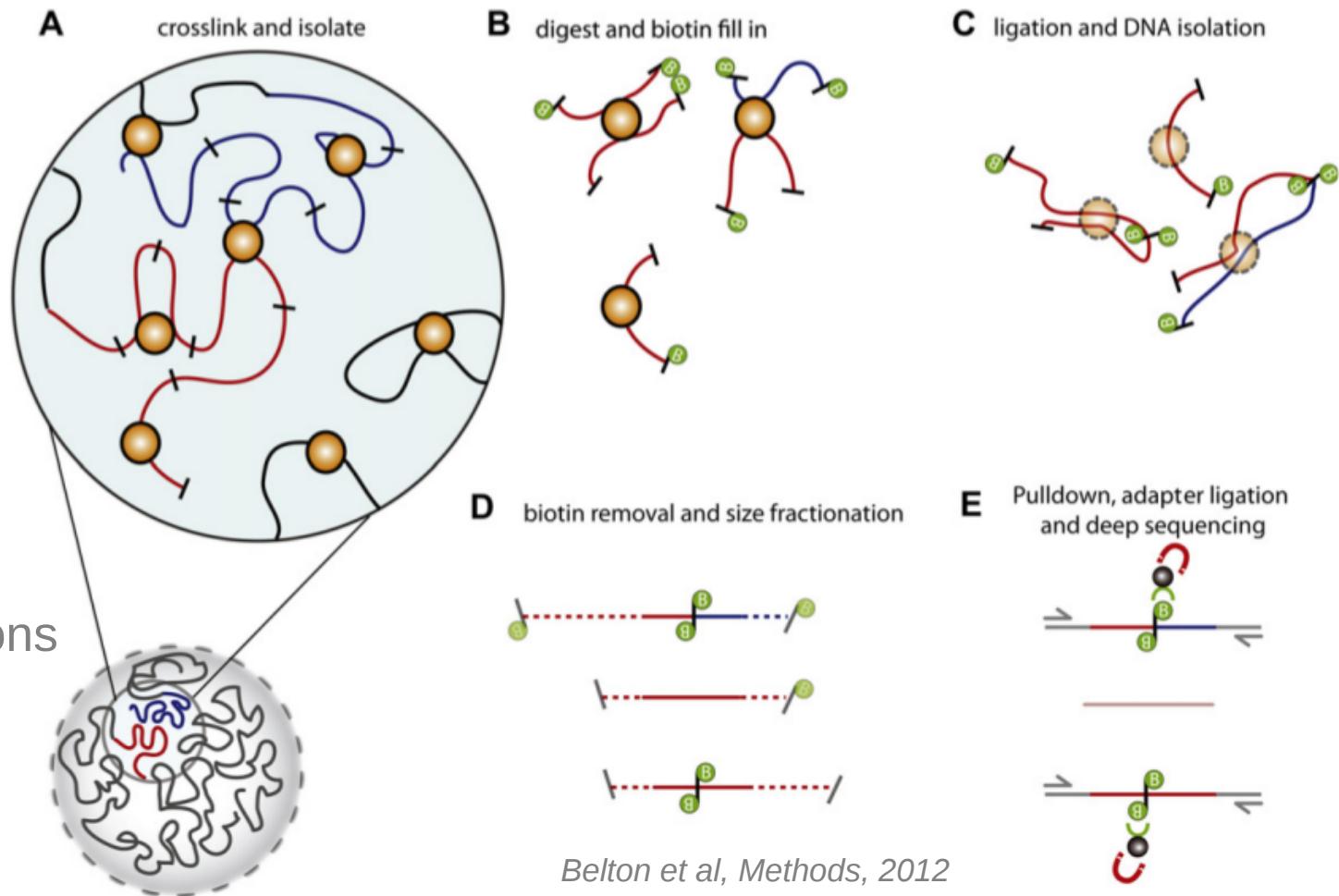
Chromosome Conformation Capture assays



Hi-C: the experiment

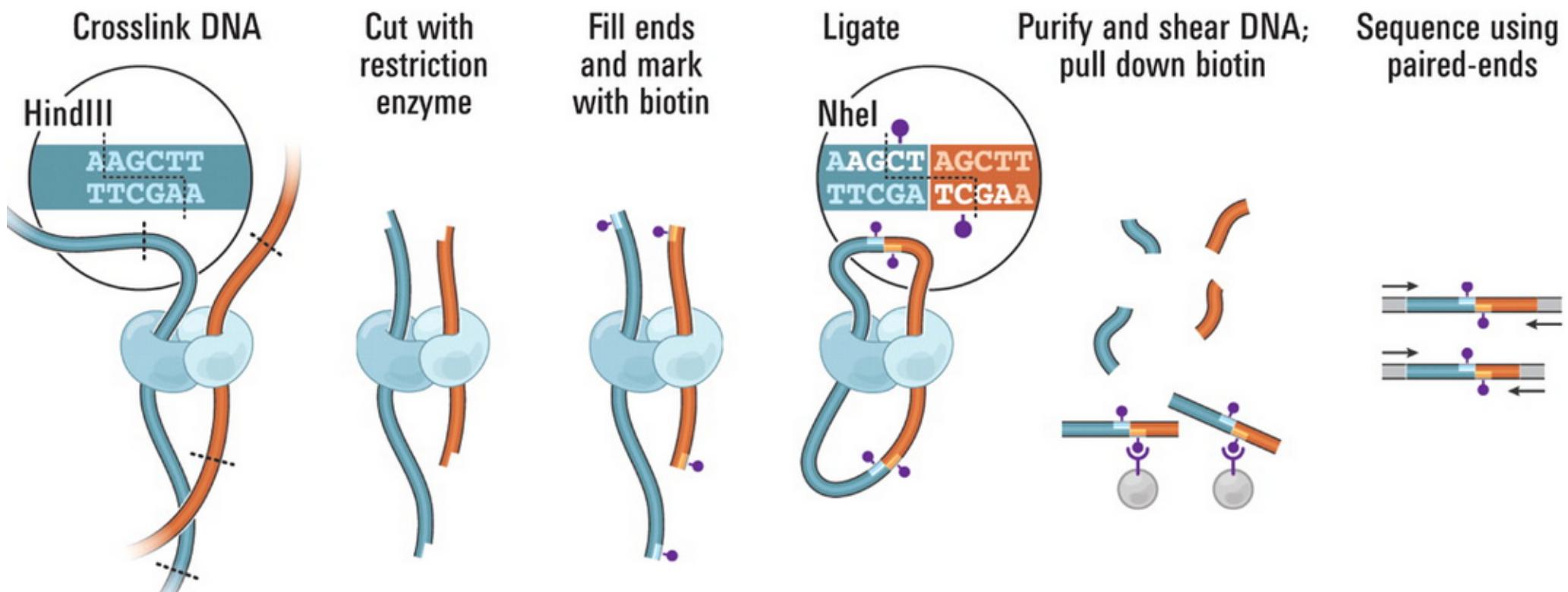
Hi-C: high-throughput chromosome conformation capture
(Lieberman-Aiden et al, Science, 2009, Rao et al, Cell, 2014)

- ◆ crosslink DNA (“fixation”)
- ◆ cleave genome with restriction enzyme
- ◆ biotin-mark and ligate extremities
- ◆ fragment, select biotin-marked junctions
- ◆ sequence fragments (paired-ends)



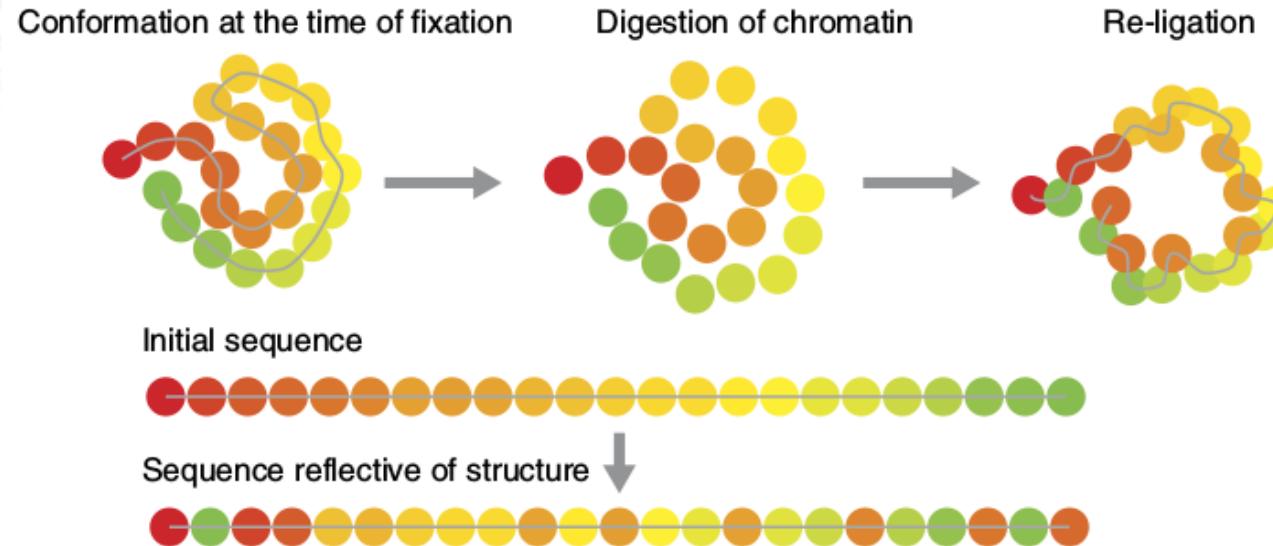
Hi-C: the experiment

Hi-C: high-throughput chromosome conformation capture
(Lieberman-Aiden et al, Science, 2009, Rao et al, Cell, 2014)

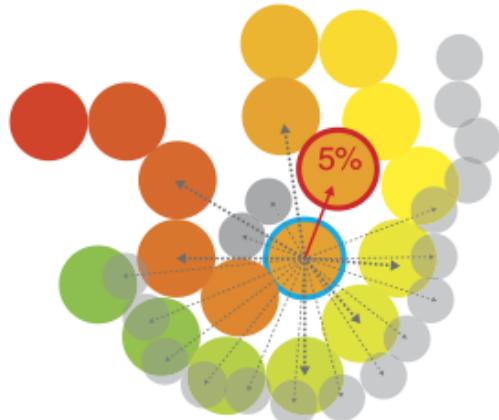


Rao et al, Cell, 2014

Hi-C: the experiment

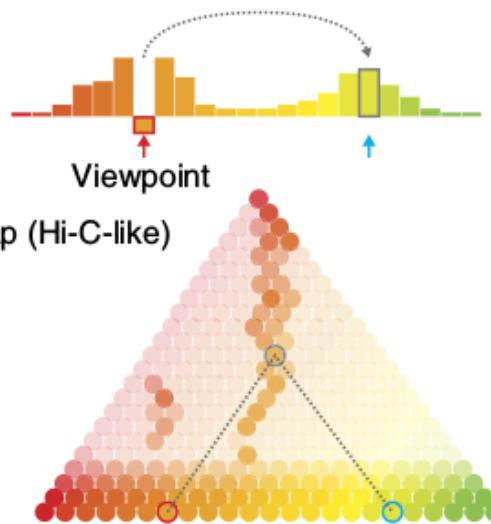


B



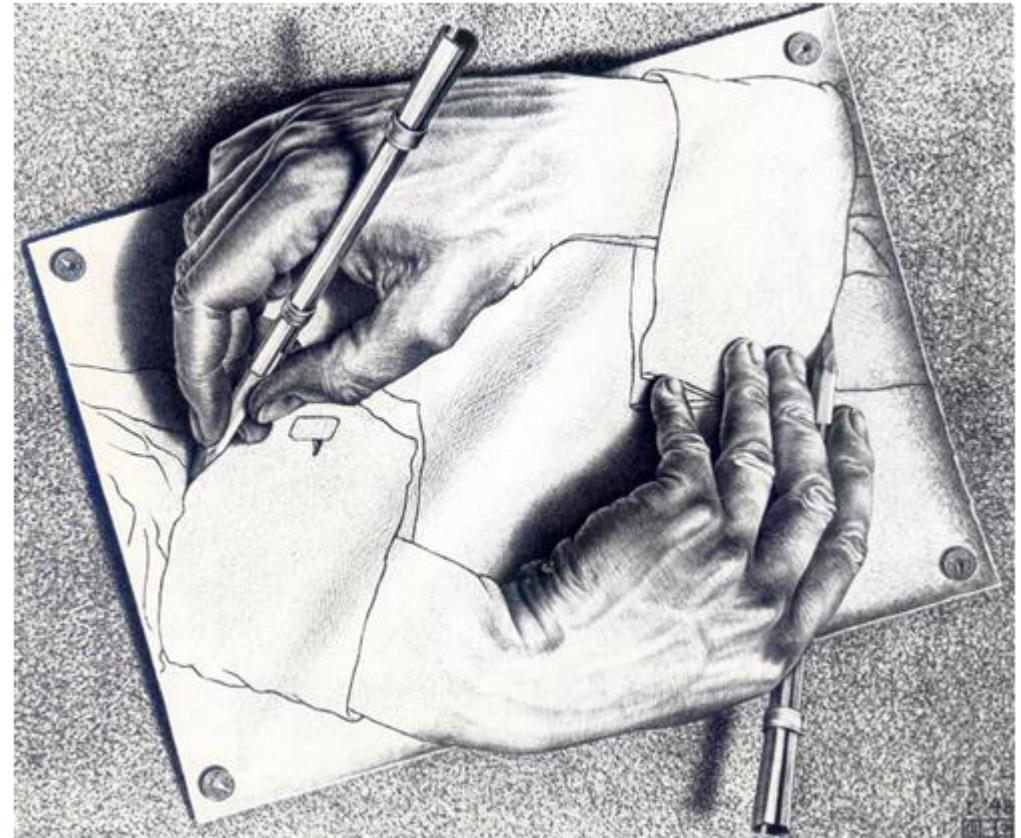
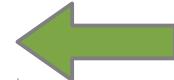
Davies et al, *Nature Methods*, 2017

C Probability distribution from a single viewpoint (4C-like)



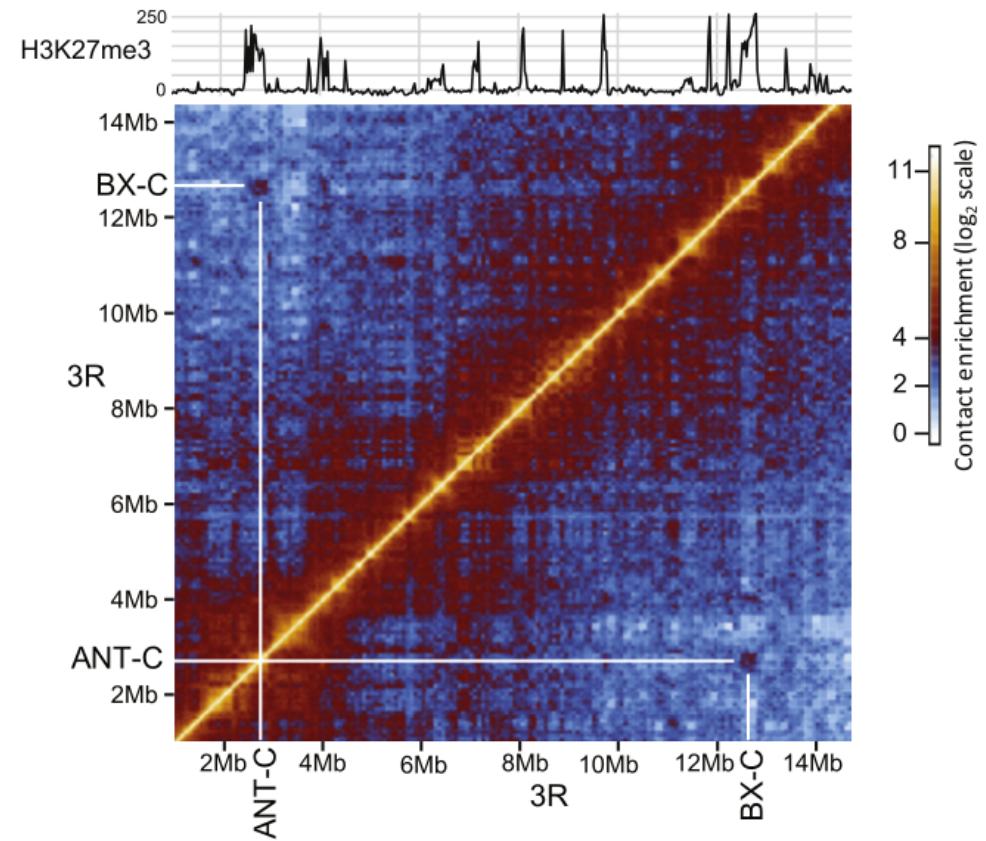
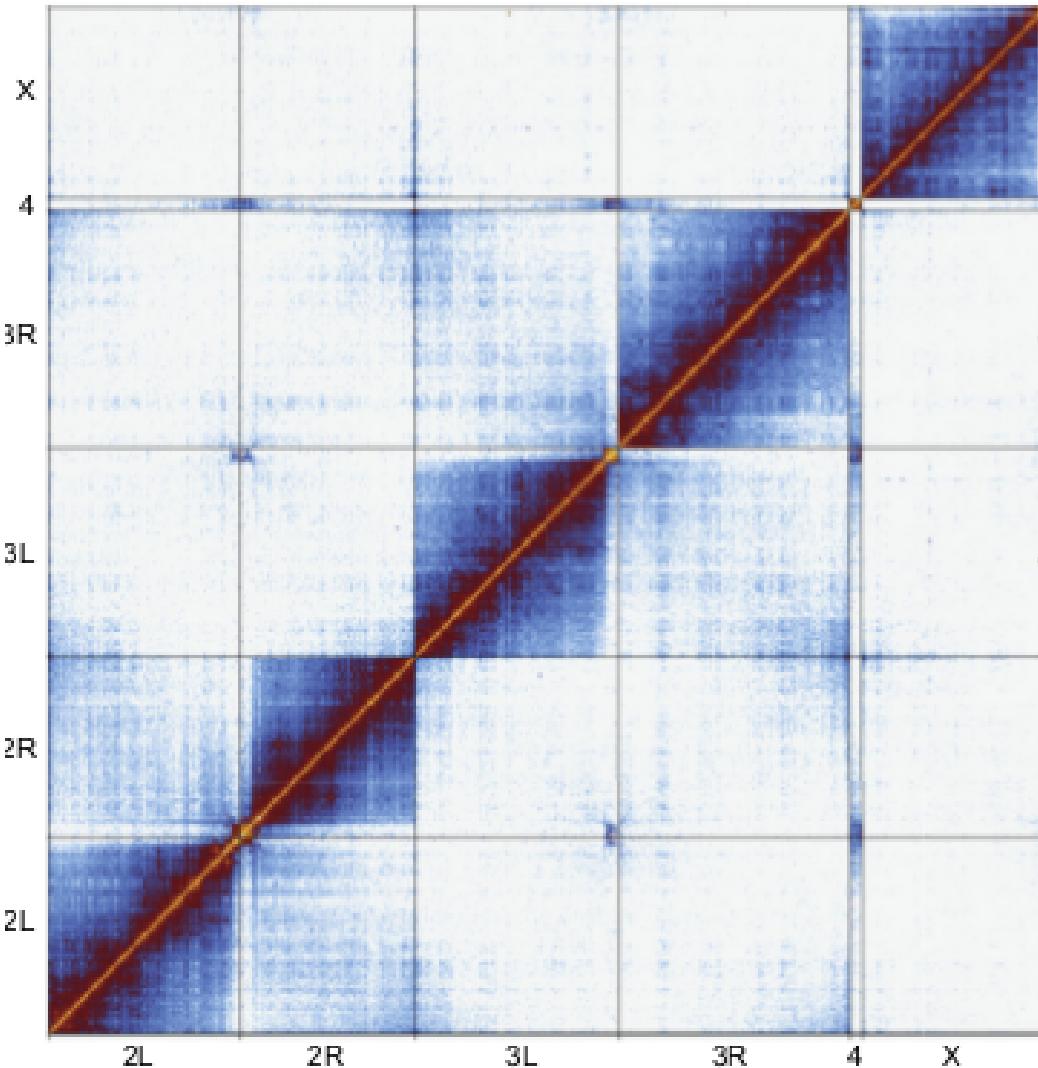
Outline

- ◆ Intro
- ◆ Current projects
- ◆ Experiment
- ◆ Data analysis
(part 1)



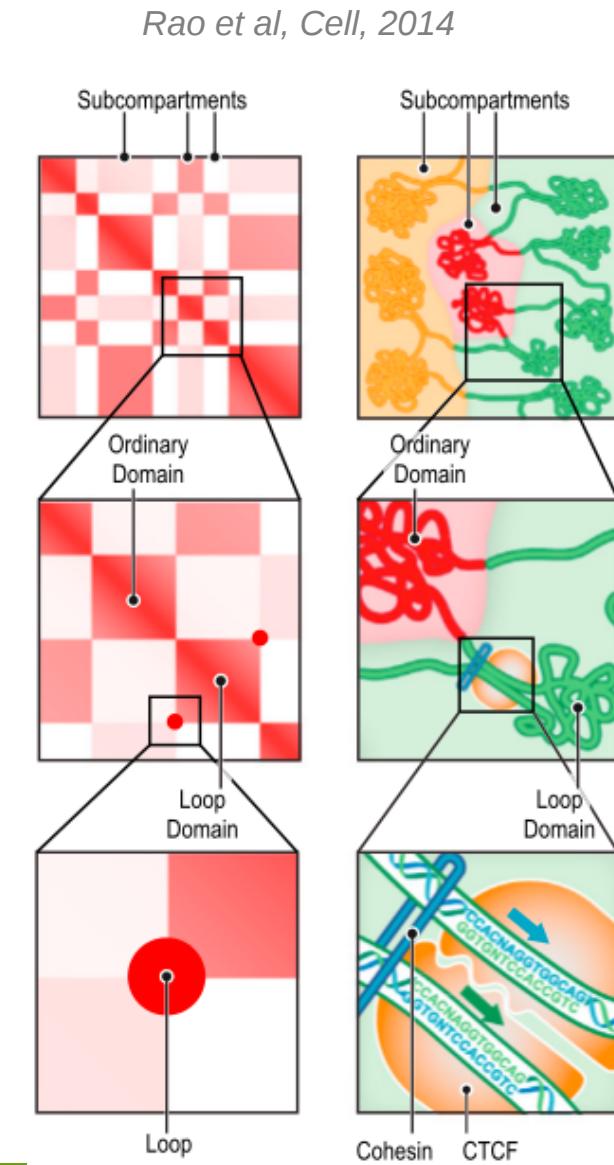
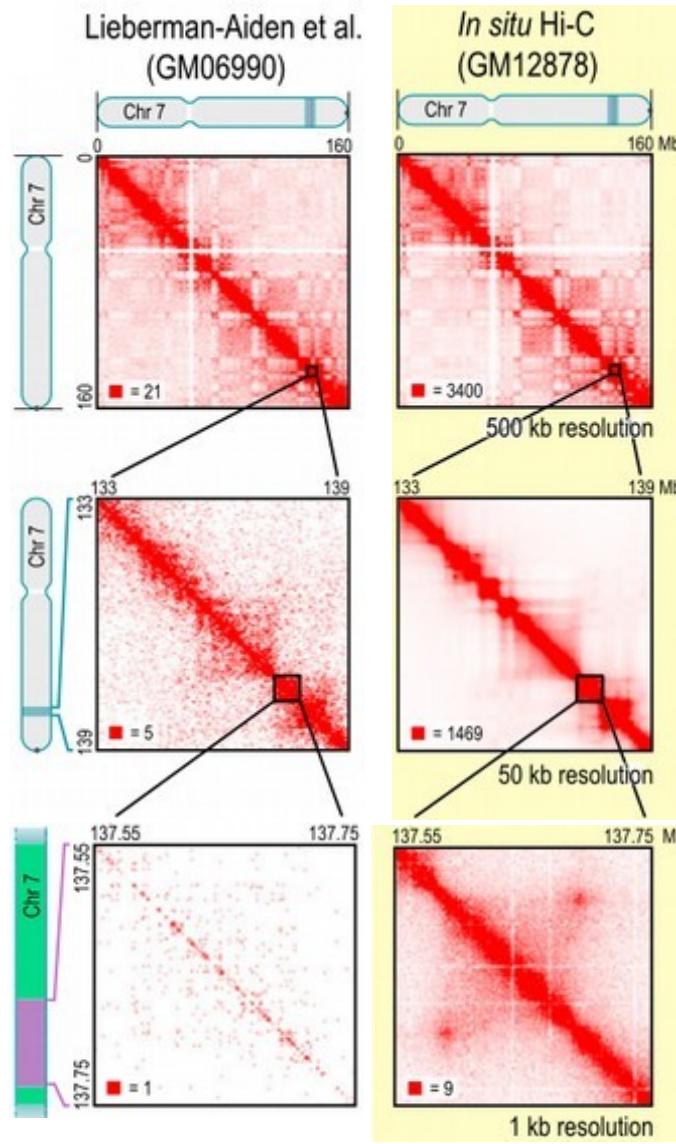
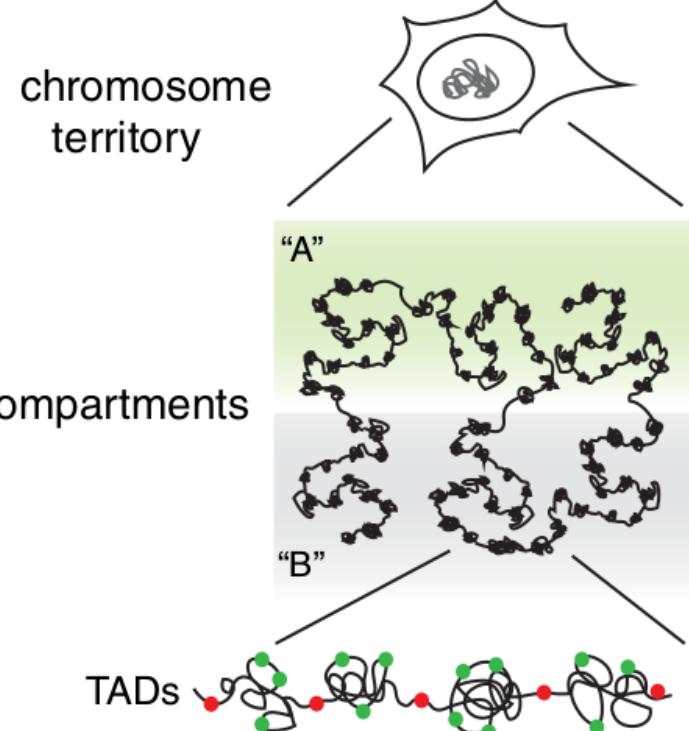
M.C. Escher, 1948

Hi-C data analysis: the contact matrix



Sexton et al, Cell, 2012

Hi-C data analysis: the contact matrix



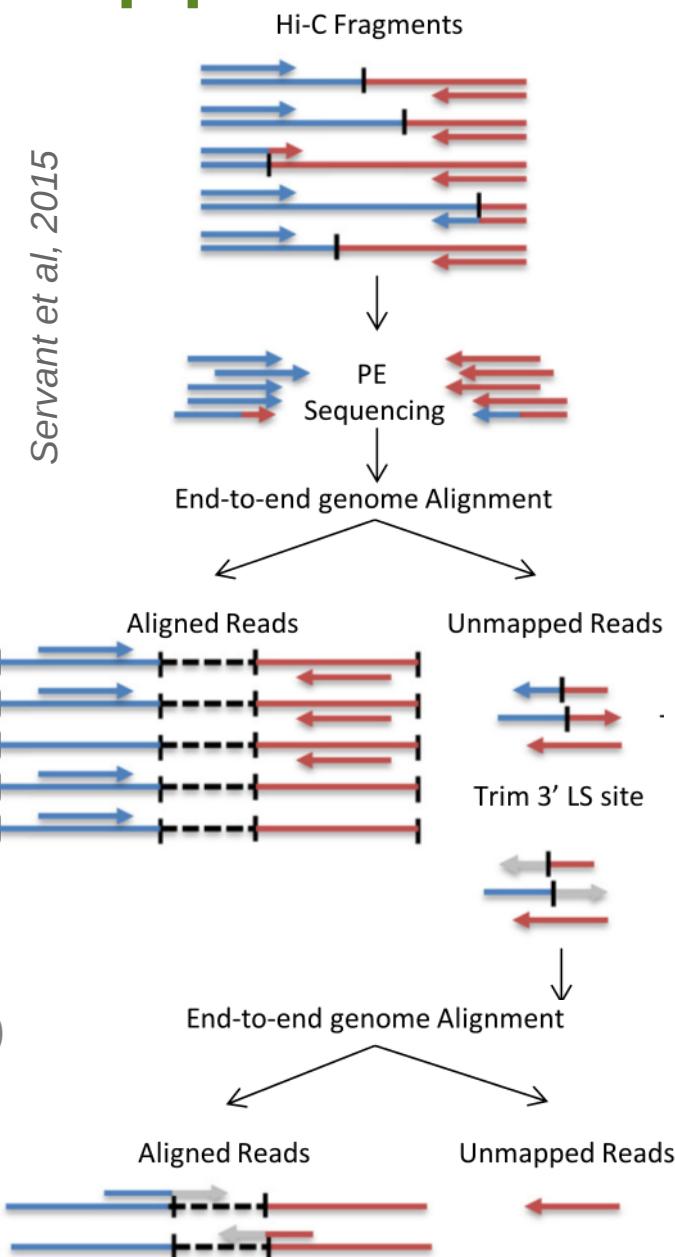
Hi-C data analysis: bioinformatics pipeline

FR-AgENCODE pipeline

- Trim reads (ligation site)
- Map on reference genome
- Discard inconsistent pairs
- Count reads in pairs of genomic bins & generate contact matrix
- Normalize contact matrix (non parametric, matrix balancing)
- Generate html report
- Identify Topologically Associated Domains, *cis* and *trans* interactions
- Identify A and B compartments

Software

- HiC-Pro pipeline (Servant et al 2015)
- Bowtie2 mapping (Langmead et al, 2009)
- ICE normalization (Imakaev et al, 2012)
- HiTC display and A/B comp. (Servant et al, 2012)
- HiFive pipeline (Sauria et al, 2015)
- Armatus TAD finding (Filippova et al, 2014)
- Juicebox browser (Durand et al, 2016)



Hi-C data analysis: FR-AgENCODE QC

Read pairs status after mapping

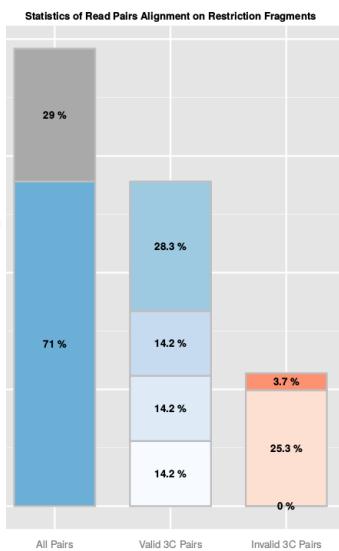
Read Pair Filtering (IMR90)



Dixon et al data

(human, from Servant et al 2015)

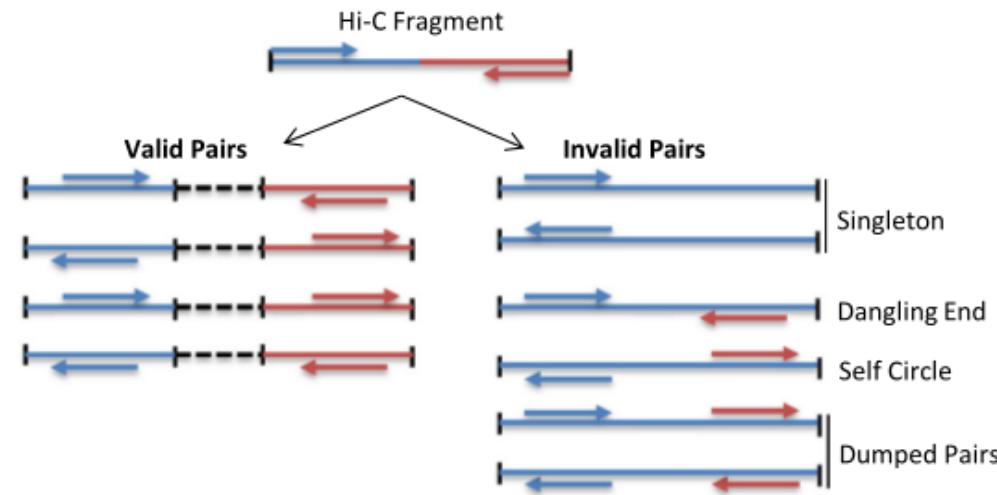
62%
valid pairs



Rao et al data

(mouse, CH12 cells)

71%
valid pairs

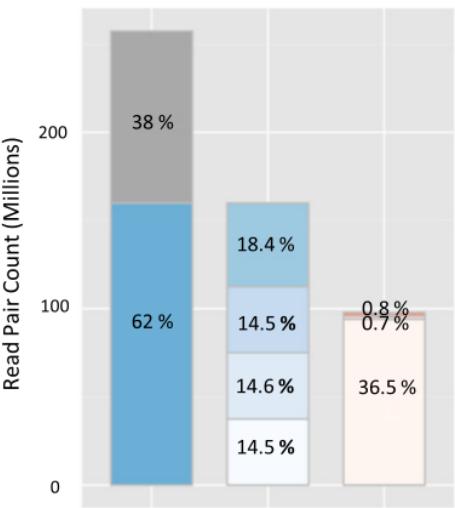


Servant et al 2015

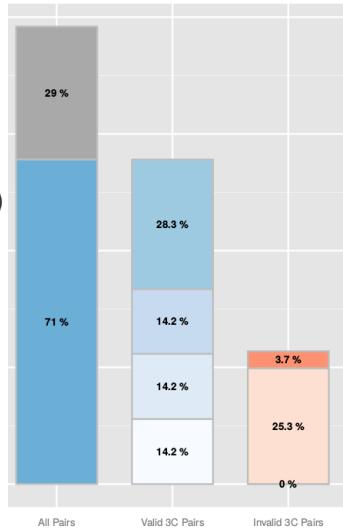
Hi-C data analysis: FR-AgENCODE QC

Read pairs status after mapping

Read Pair Filtering (IMR90)



Statistics of Read Pairs Alignment on Restriction Fragments



Dixon et al data

(human, from Servant et al 2015)

62%
valid pairs

Rao et al data

(mouse, CH12 cells)

71%
valid pairs

Statistics of Read Pairs Alignment on Restriction Fragments



FR-AgENCODE data

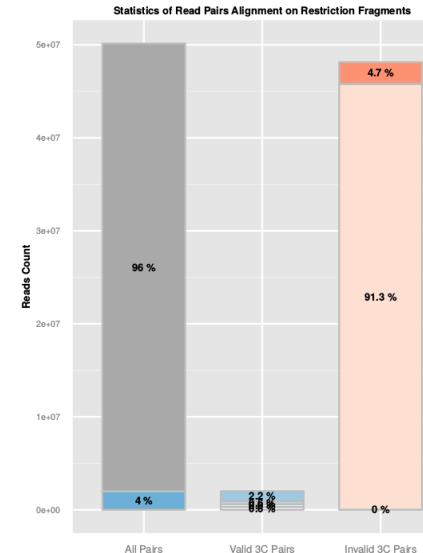
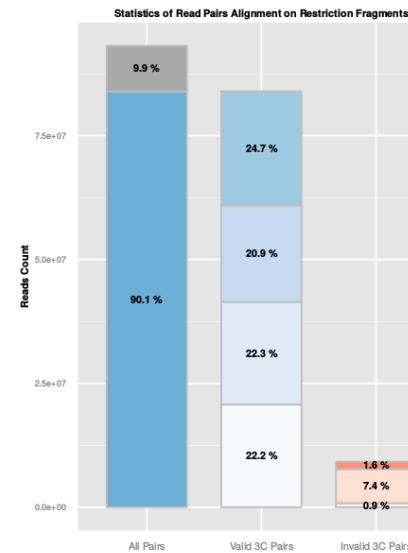
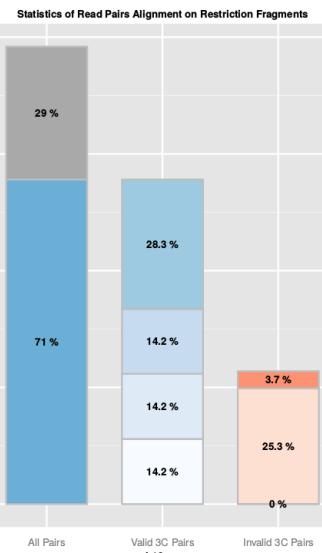
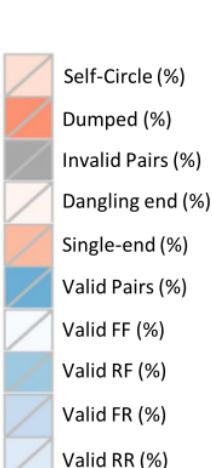
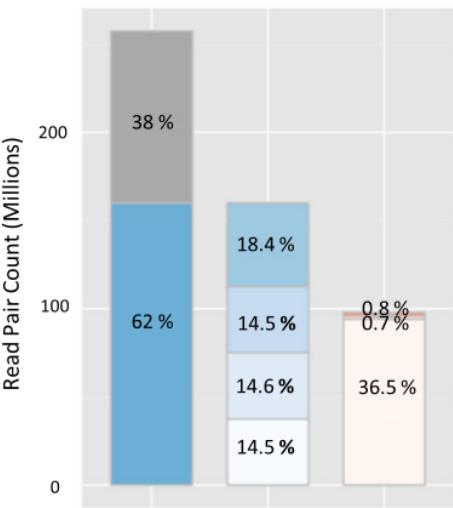
(mouse, STO cells)

90%
valid pairs

Hi-C data analysis: FR-AgENCODE QC

Read pairs status after mapping

Read Pair Filtering (IMR90)



Dixon et al data

(human, from Servant et al 2015)

62%
valid pairs

Rao et al data

(mouse, CH12 cells)

71%
valid pairs

FR-AgENCODE data

(mouse, STO cells)

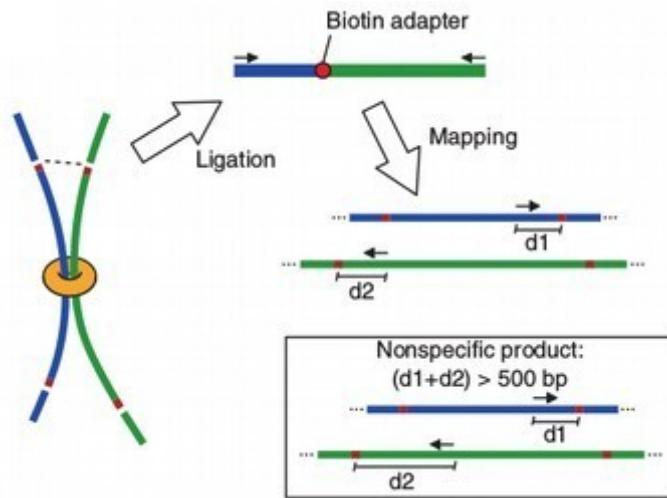
90%
valid pairs

FR-AgENCODE data

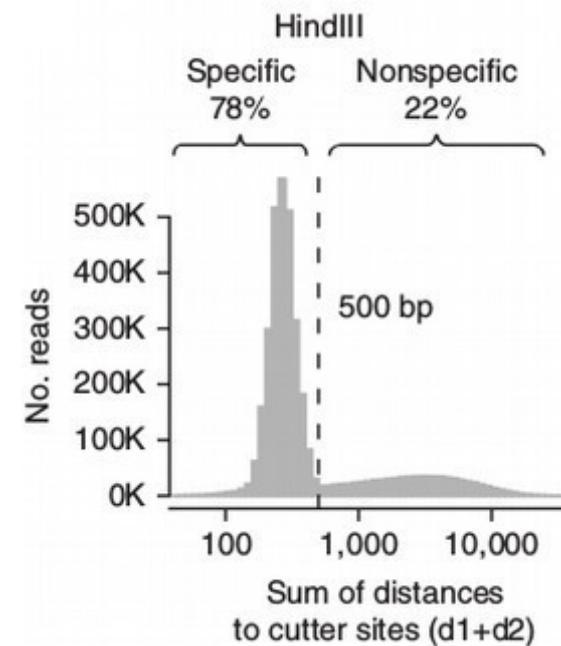
(pig, hepatocytes)

4%
valid pairs

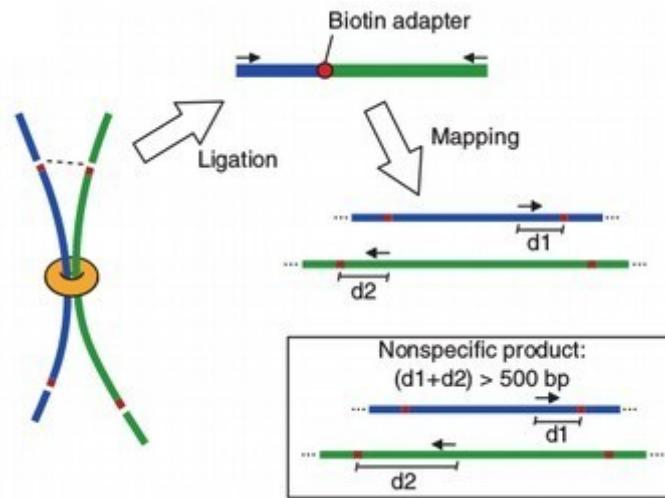
Hi-C data analysis: FR-AgENCODE QC



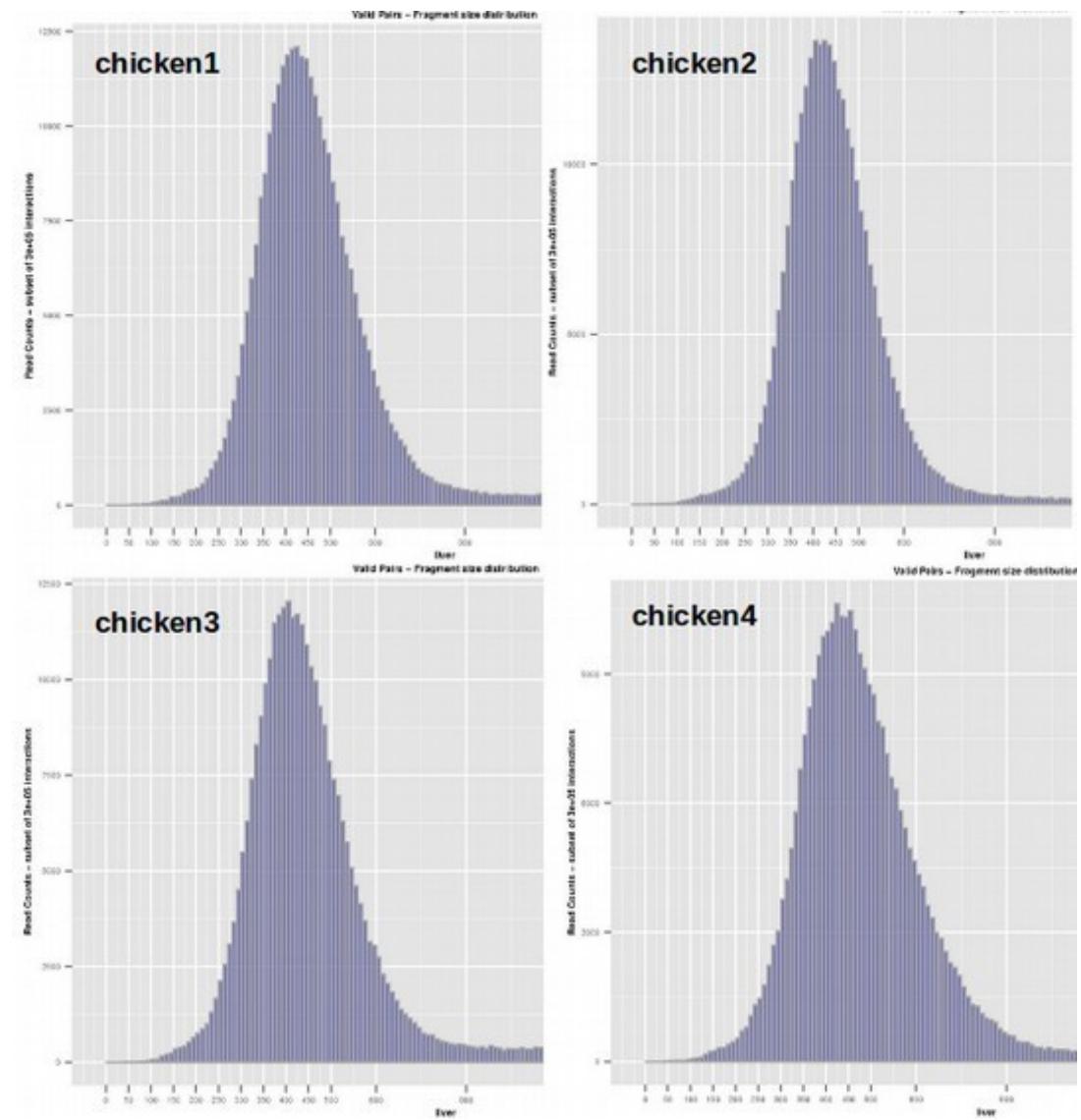
Yaffe & Noble, 2011



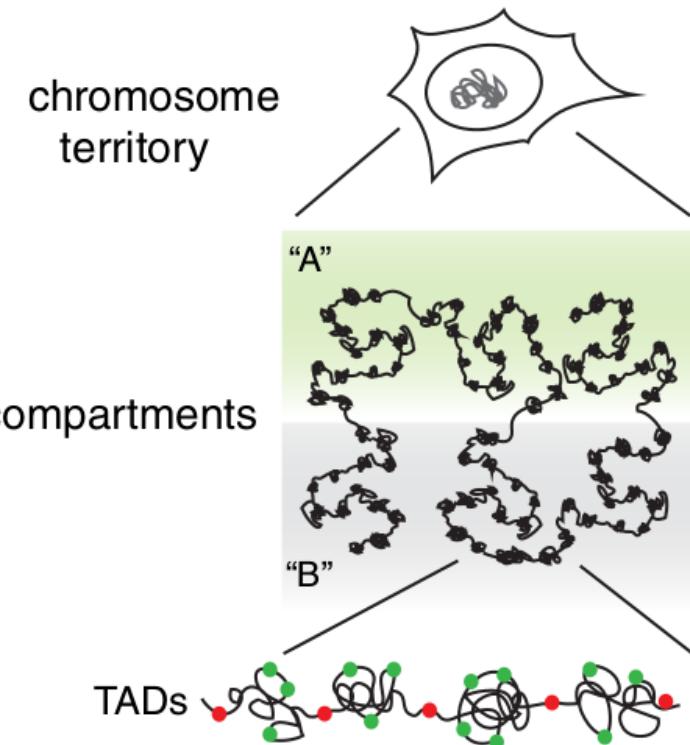
Hi-C data analysis: FR-AgENCODE QC



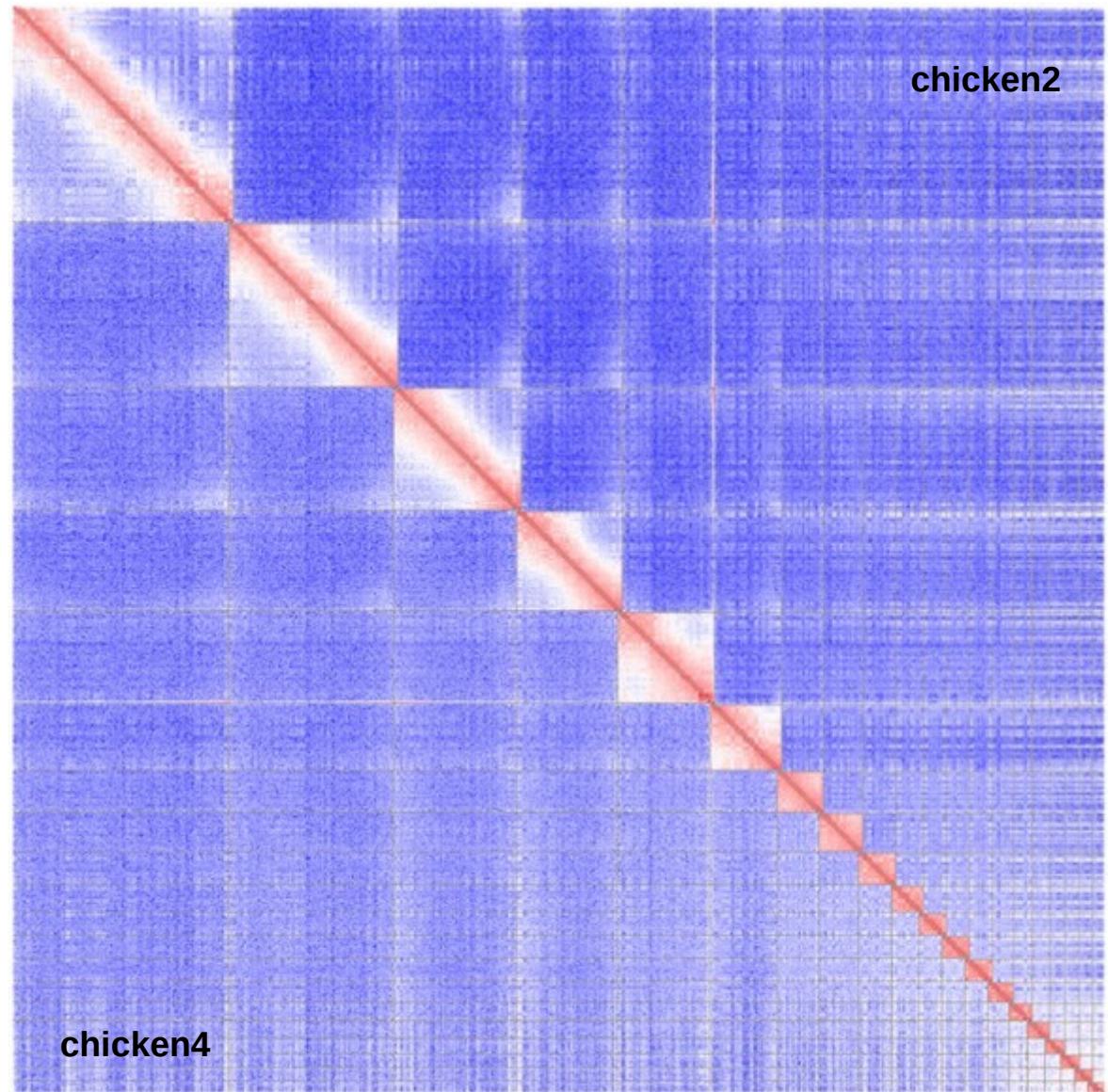
Yaffe & Noble, 2011



Hi-C data analysis: FR-AgENCODE results



Giorgetti et al. *Genome Biology* 2013



Hi-C data analysis: PORCINET results

Sequencing (GeT PlaGE Platform):

Paired-end sequencing: Illumina Hi-Seq3000

1) Low-depth sequencing:

6 pig lib + 7 lib

13 libraries sequenced in 1 line
~ 20 M read pairs / library

2) Full-depth sequencing:

6 pig lib → pool

6 libraries sequenced in 3 lines → **2 libraries / line**
~ 180 M read pairs / library

Other published data:

3) Human data set (Dixon 2015) →

SRR1030718_1.fastq

SRR1030718_2.fastq

~ 170 M read pairs

Mouse data set (Rao 2014) →

SRR1658716_R1.fastq

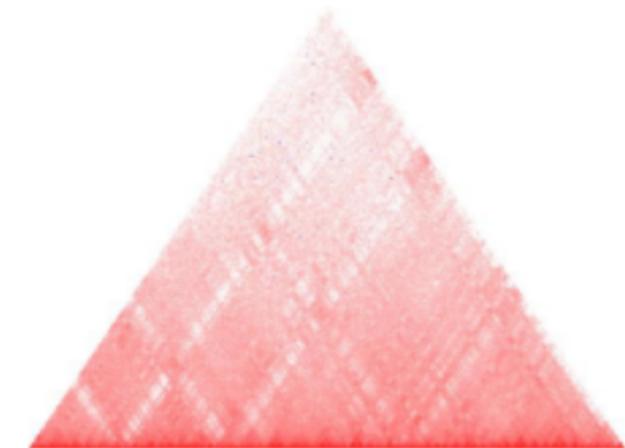
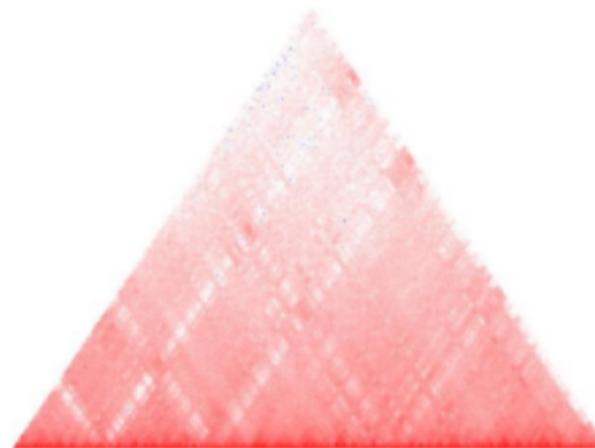
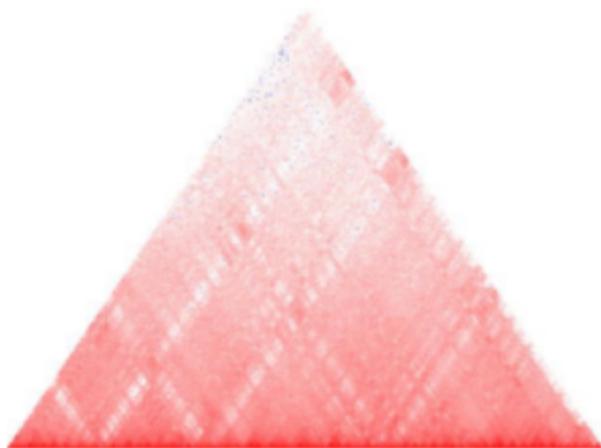
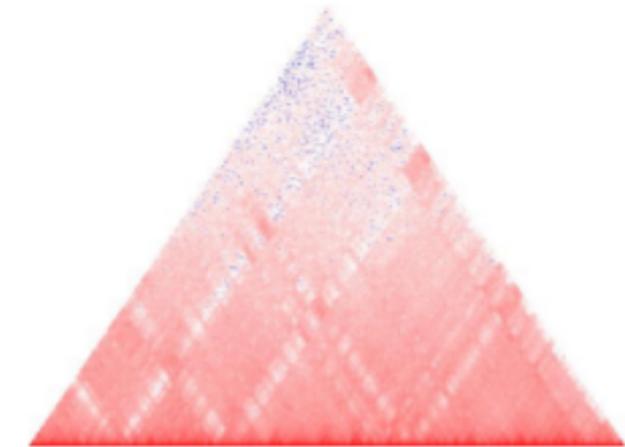
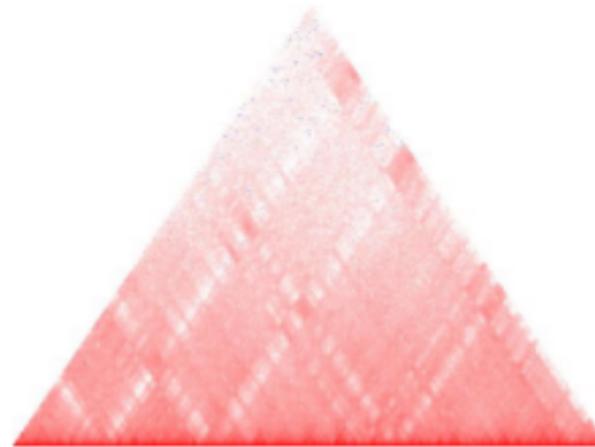
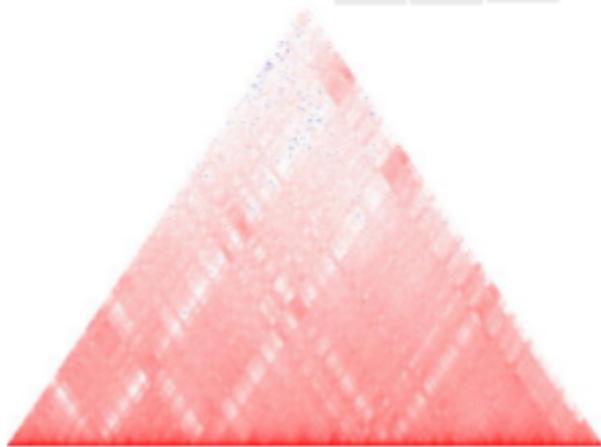
SRR1658716_R2.fastq

~ 110 M read pairs



Hi-C data analysis: PORCINET results

Contact matrices for chr13, 500Kb/bin, Sscrofa v11



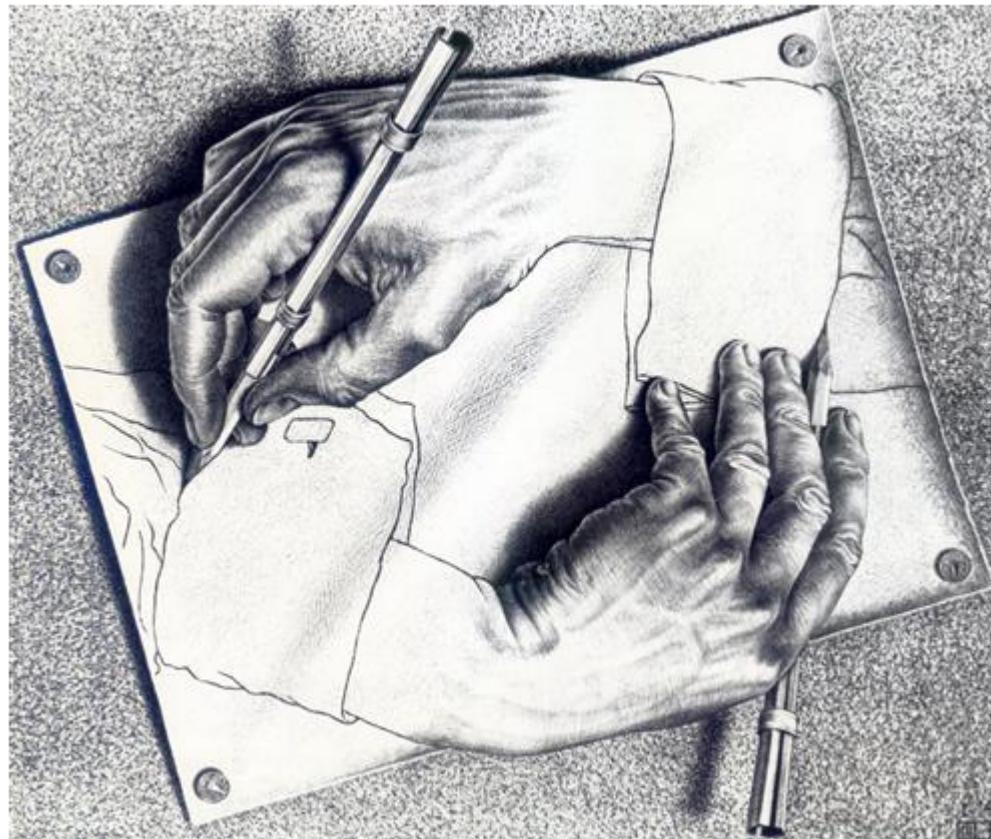
Acknowledgments

FR-AgENCODE members

- ◆ Management: Elisabetta Giuffra (coordination), Sandrine Lagarrigue, Marie Hélène Pinard
- ◆ Sampling: Michèle Tixier-Boichard, Stéphane Fabre et al.
- ◆ Assays: **Diane Esquerré, Hervé Acloque** et al.
- ◆ Analysis: Christophe Klopp, Cédric Cabau, **Sarah Djebali, Matthias Zytnicki, Nathalie Villa-Vialaneix, David Robelin, Magali San Cristobal, Ignacio Gonzalez**, Christine Gaspin, Kylie Munyard, Thomas Derrien, Kévin Muret, Céline Noirot et al.

PORCINET @ GenPhySE

- ◆ **Maria Marti, Martine Yerle, Yvette Lahbib, Florence Mompart, Laurence Liaubet** et al.

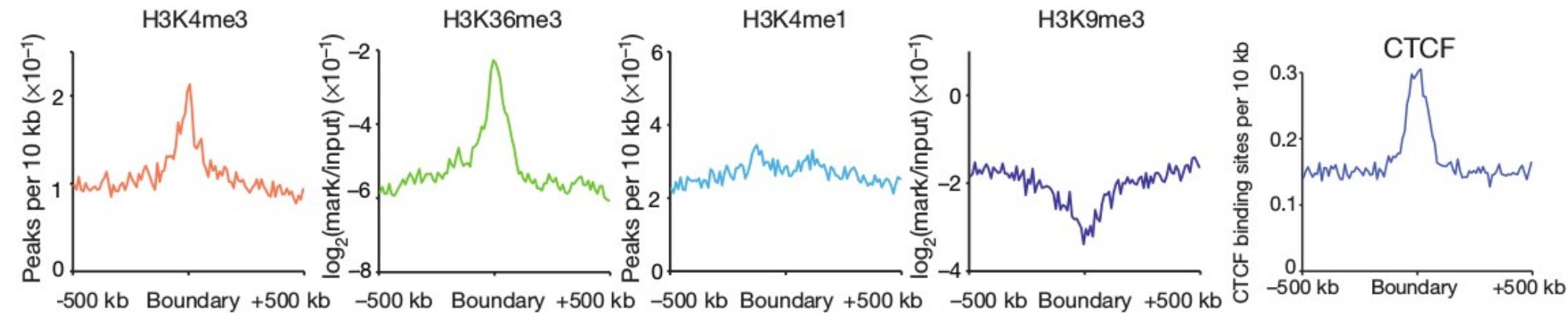


M.C. Escher, 1948

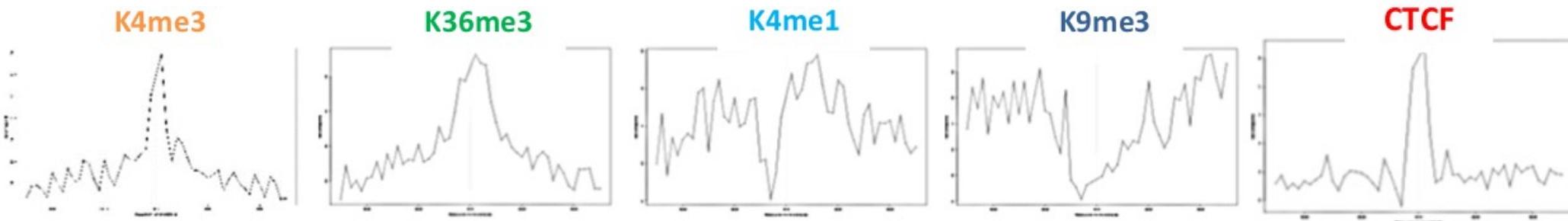
Hi-C data analysis: PORCINET results

Enrichment of genomic features around TAD boundaries

Homo sapiens (Dixon et al, Nature, 2012)

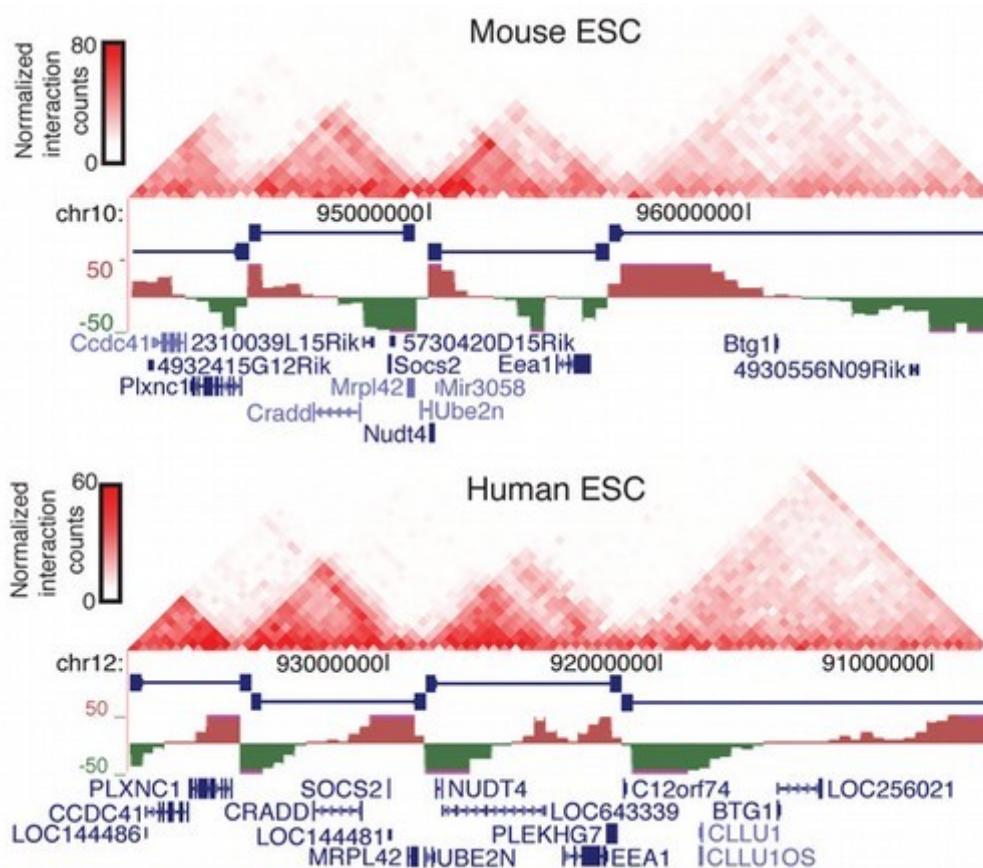


Sus scrofa (PORCINET project results)

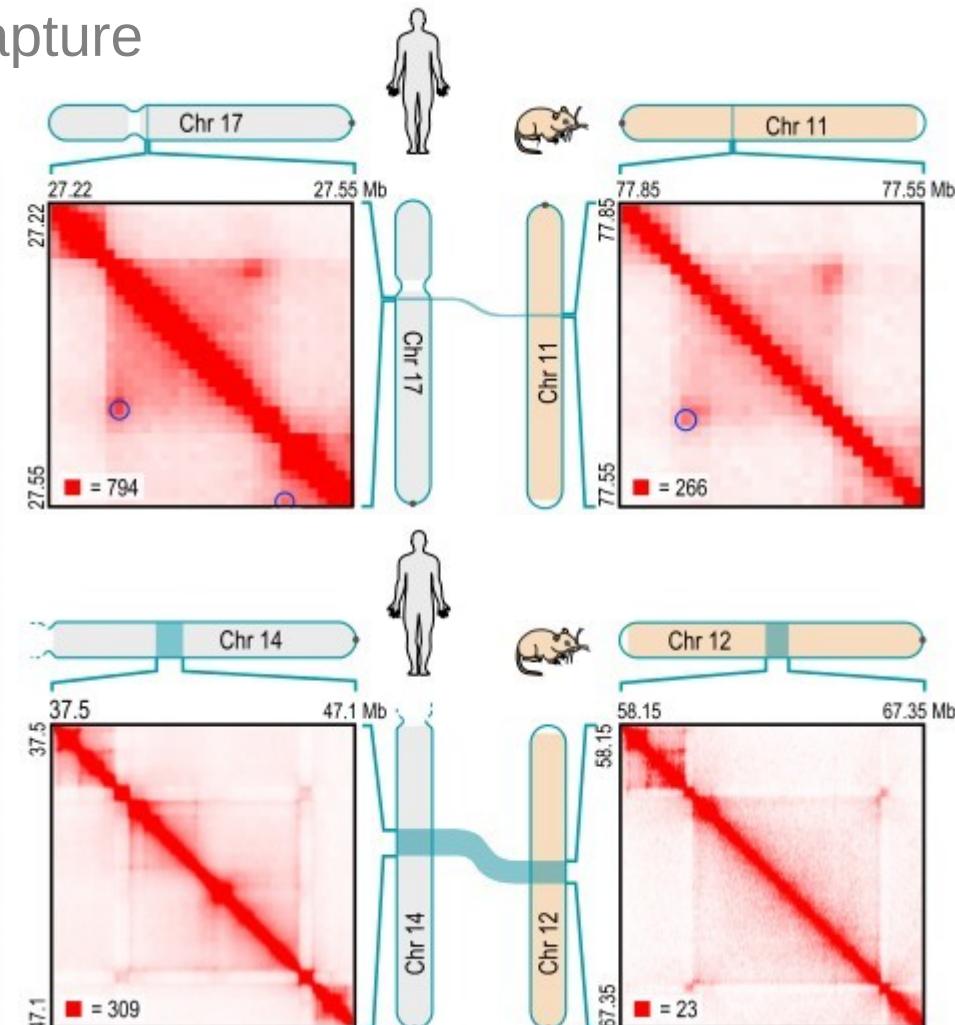


FR-AgENCODE: molecular assays

- ◆ RNA-seq
- ◆ Hi-C: chromosome conformation capture



Dixon et al 2012



Rao et al, 2014