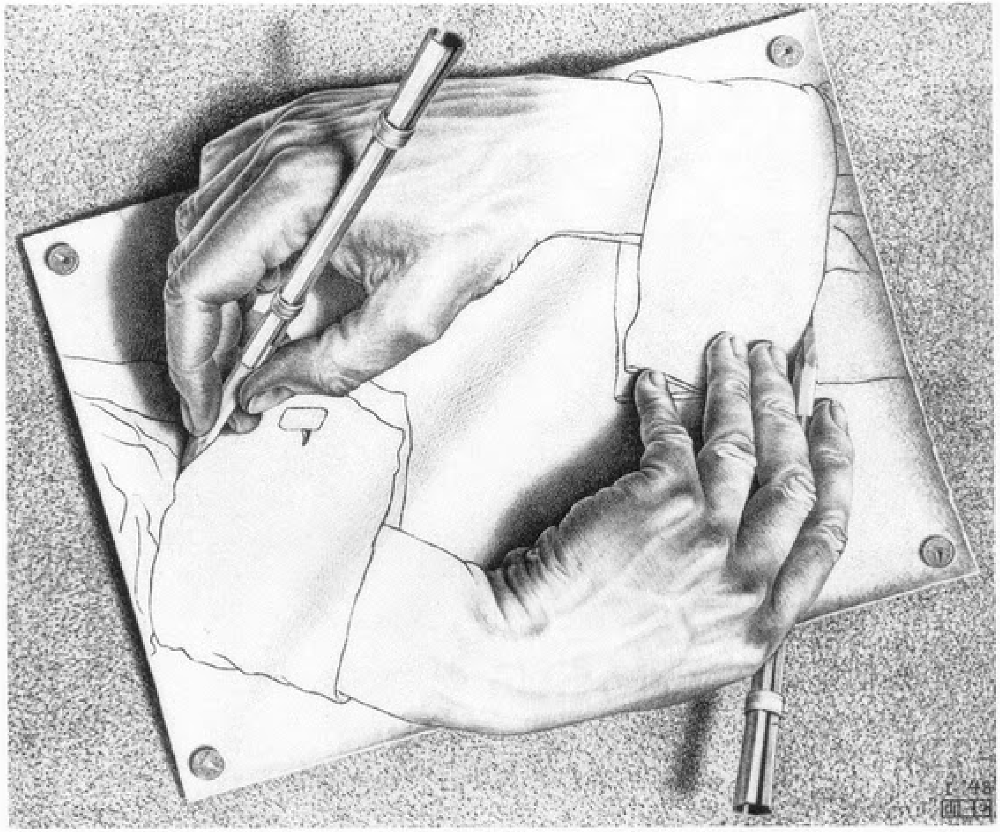
– Not SMRT yet smart –
3D genomics with Hi-C sequencing

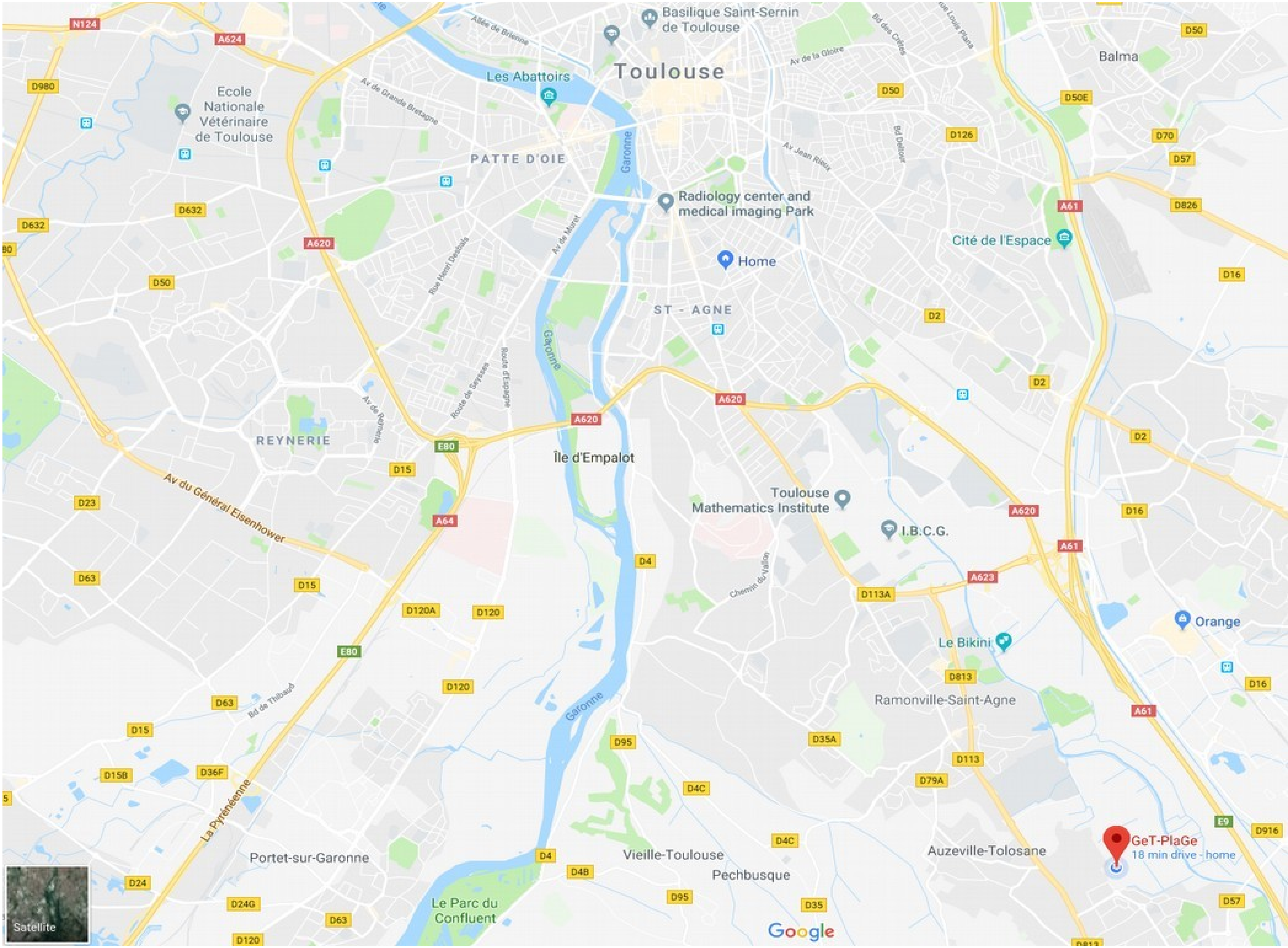
Sylvain Foissac, GenPhySE
November 28, 2017. Get Genotoul, INRA Toulouse, France

From structure ...



https://en.wikipedia.org/wiki/M._C._Escher

... to function

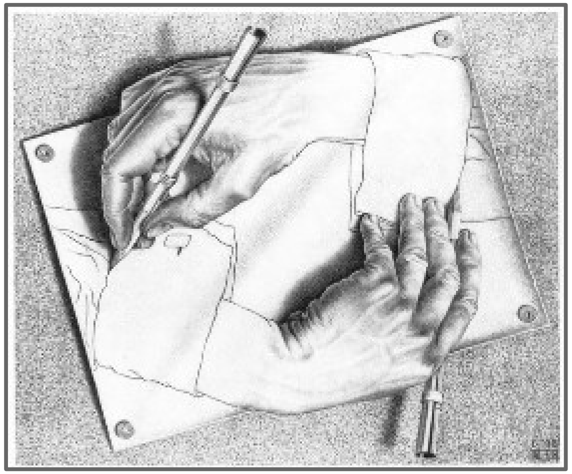

Basilique Saint-Sernin de Toulouse
Toulouse
Les Abattoirs
Ecole Nationale Vétérinaire de Toulouse
PATTE D'OIE
Radiology center and medical imaging Park
Home
Cité de l'Espace
Balma
ST - AGNE
REYNERIE
Île d'Empalot
Toulouse Mathematics Institute
I.B.C.G.
Orange
Le Bikini
Ramonville-Saint-Agne
Auzeville-Tolosane
GeT-PlaGe
18 min drive - home
Vieille-Toulouse
Pechbusque
Portet-sur-Garonne
Le Parc du Confluent
Garonne
Av du Général Eisenhower
La Pyrénéenne
A620
A64
A61
A623
A624
N124
E80
E9
D980
D632
D50
D23
D63
D15
D15B
D36F
D24
D24G
D120
D120A
D4
D4B
D4C
D95
D35
D35A
D113
D113A
D79A
D813
D16
D916
D57
D2
D126
D70
D826
D50E
Satellite
Google


Chapelle des Carmélites
Musée Saint-Raymond
Copy Shop
Corep Lautman
The Geo
Basilique Saint-Sernin de Toulouse
Syndicat CGT
Sept
Lycée Général Saint-Sernin
2D
Map
Google

Intro: from structure to function

Draw me a gene...

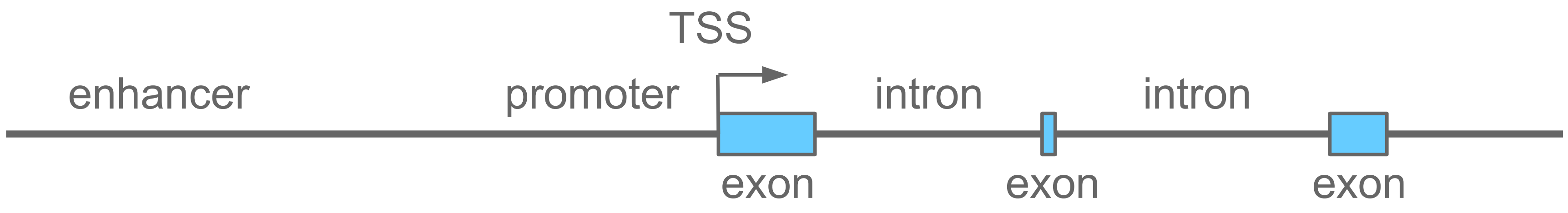
De Saint Exupéry A.,1943



Intro: from structure to function

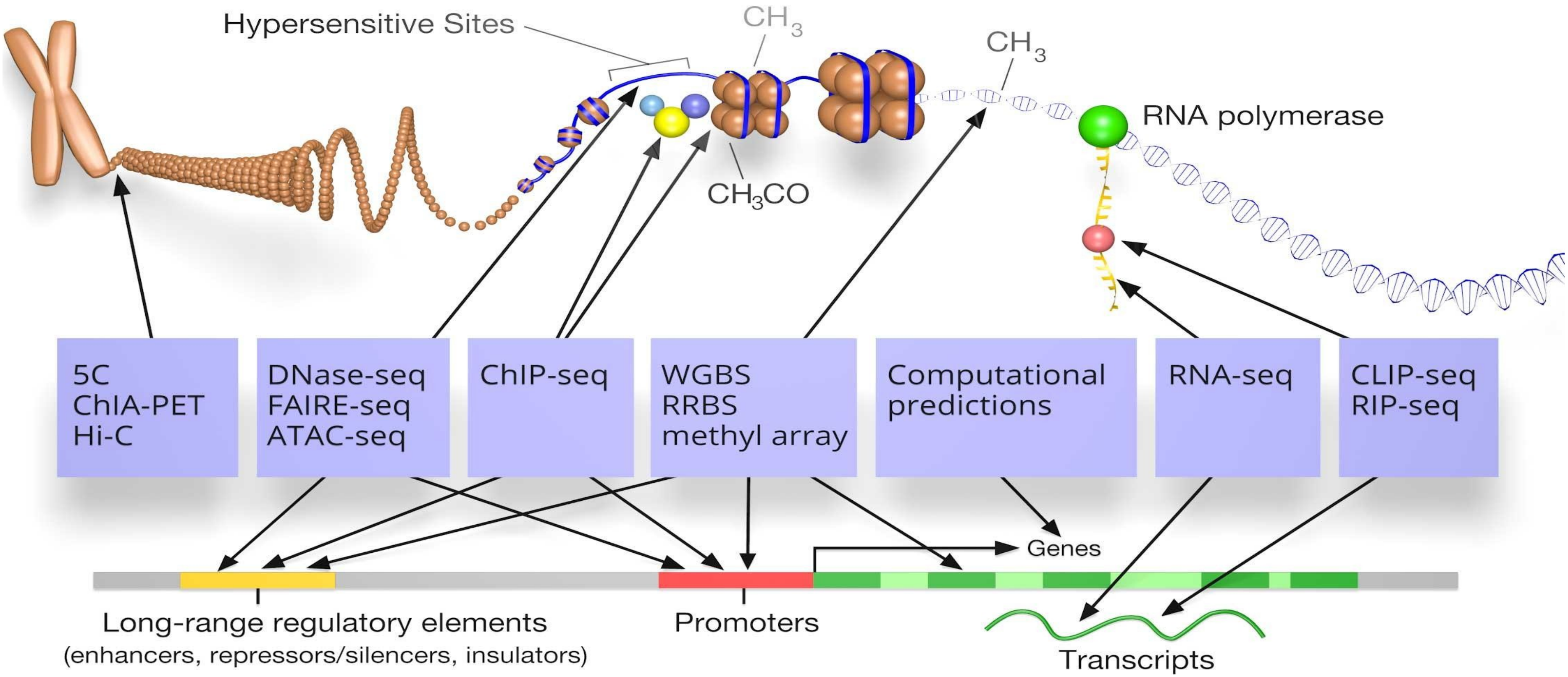
genomic DNA

Intro: from structure to function



genomic DNA

Intro: from structure to function



5C
ChIA-PET
Hi-C

DNase-seq
FAIRE-seq
ATAC-seq

ChIP-seq

WGBS
RRBS
methyl array

Computational predictions

RNA-seq

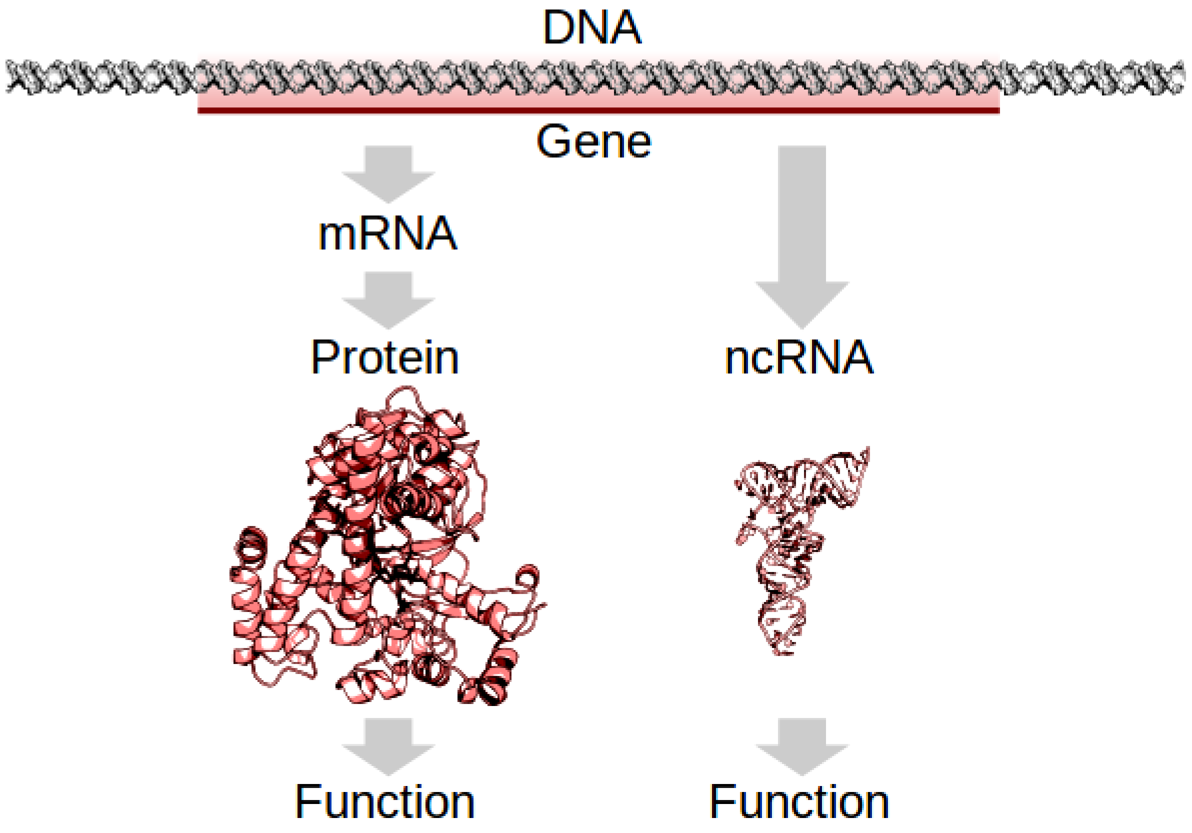
CLIP-seq
RIP-seq

www.encodeproject.org

Human: ENCODE

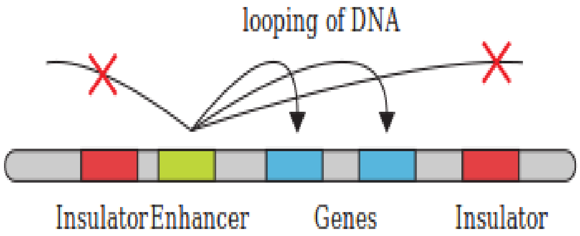
Livestock: FAANG Functional Annotation of Animal Genomes

Intro: from structure to function

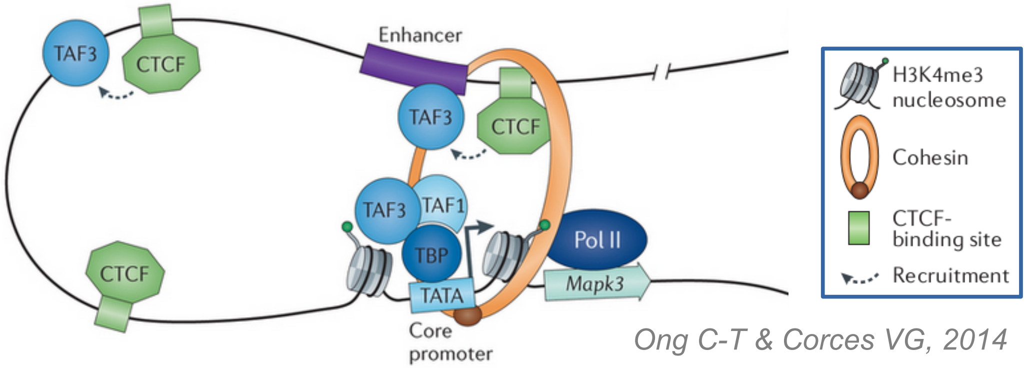


https://en.wikipedia.org/wiki/Gene

Intro: from structure to function

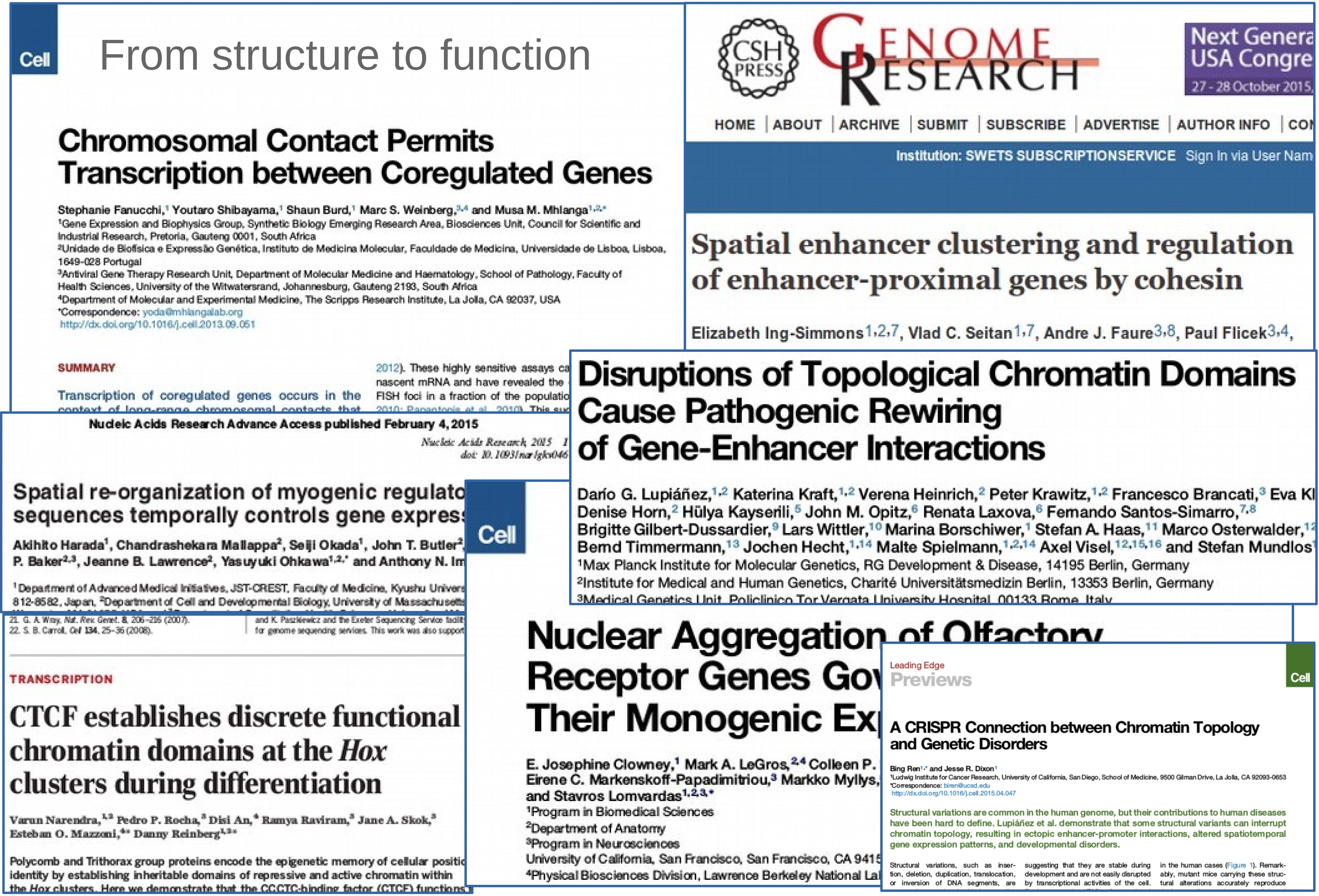


https://en.wikipedia.org/wiki/Enhancer



Ong C-T & Corces VG, 2014

Intro: from structure to function



Cell

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,*]
[1]Gene Expression and Biophysics Group, Synthetic Biology Emerging Research Area, Biosciences Unit, Council for Scientific and Industrial Research, Pretoria, Gauteng 0001, South Africa
[2]Unidade de Biofísica e Expressão Genética, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, 1649-028 Portugal
[3]Antiviral 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
[4]Department 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

2012). These highly sensitive assays ca
nascent mRNA and have revealed the
FISH foci in a fraction of the populatio
2010; Papantonis et al., 2010). This su

CSH PRESS GENOME RESEARCH

Next Genera
USA Congre
27 - 28 October 2015,

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

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],

---

Nucleic Acids Research Advance Access published February 4, 2015
Nucleic Acids Research, 2015 1
doi: 10.1093/nar/gkv046

Spatial re-organization of myogenic regulato sequences temporally controls gene express

Akihito Harada[1], Chandrashekara Mallappa[2], Seiji Okada[1], John T. Butler[2],
P. Baker[2,3], Jeanne B. Lawrence[2], Yasuyuki Ohkawa[1,2,*] and Anthony N. Im
[1]Department of Advanced Medical Initiatives, JST-CREST, Faculty of Medicine, Kyushu Univers
812-8582, Japan, [2]Department of Cell and Developmental Biology, University of Massachusetts

---

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 K
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 Borschiwer,[1] 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]Max Planck Institute for Molecular Genetics, RG Development & Disease, 14195 Berlin, Germany
[2]Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany
[3]Medical Genetics Unit, Policlinico Tor Vergata University Hospital, 00133 Rome, Italy

---

21. G. A. Wray, *Nat. Rev. Genet.* 8, 206–216 (2007).
22. S. B. Carroll, *Cell* 134, 25–36 (2008).

and K. Paszkiewicz and the Exeter Sequencing Service facili
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] Pedro P. Rocha,[3] Disi An,[4] Ramya Raviram,[3] Jane A. Skok,[3]
Esteban O. Mazzoni,[4*] Danny Reinberg[1,2*]

Polycomb and Trithorax group proteins encode the epigenetic memory of cellular positic identity by establishing inheritable domains of repressive and active chromatin within the Hox clusters. Here we demonstrate that the CCCTC-binding factor (CTCF) functions

---

Cell

Nuclear Aggregation of Olfactory Receptor Genes Gov Their Monogenic Ex

E. Josephine Clowney,[1] Mark A. LeGros,[2,4] Colleen P.
Eirene C. Markenskoff-Papadimitriou,[3] Markko Myllys,
and Stavros Lomvardas[1,2,3,*]
[1]Program in Biomedical Sciences
[2]Department of Anatomy
[3]Program in Neurosciences
University of California, San Francisco, San Francisco, CA 9415
[4]Physical Biosciences Division, Lawrence Berkeley National La

---

Leading Edge
Previews

Cell

A CRISPR Connection between Chromatin Topology and Genetic Disorders

Bing Ren[1,*] and Jesse R. Dixon[1]
[1]Ludwig Institute for Cancer Research, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0653
*Correspondence: biren@ucsd.edu
http://dx.doi.org/10.1016/j.cell.2015.04.047

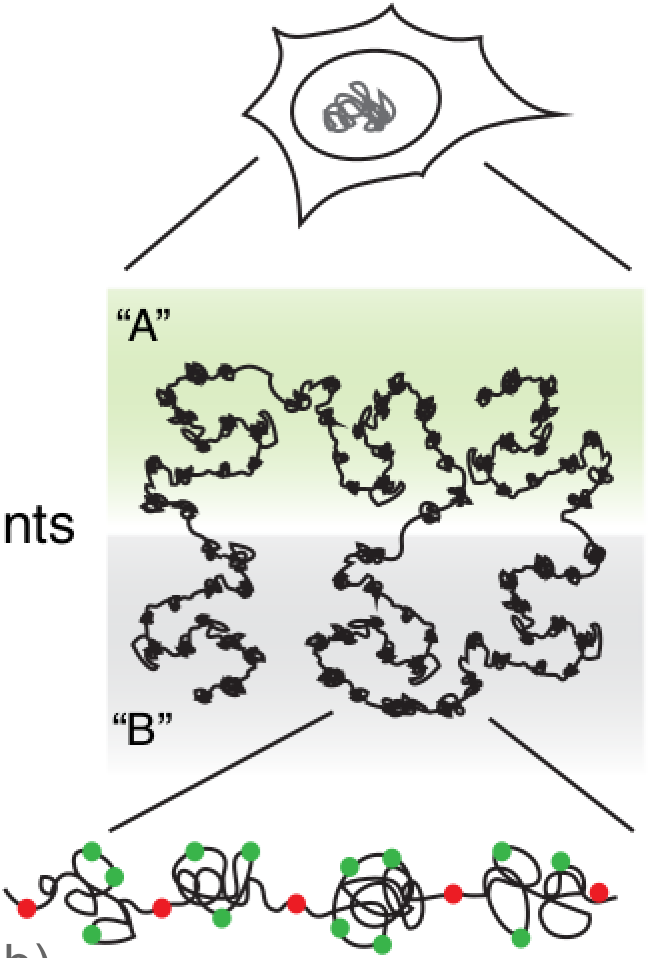
Structural variations are common in the human genome, but their contributions to human diseases have been hard to define. Lupiáñez et al. demonstrate that some structural variants can interrupt chromatin topology, resulting in ectopic enhancer-promoter interactions, altered spatiotemporal gene expression patterns, and developmental disorders.

Structural variations, such as insertion, deletion, duplication, translocation, or inversion of DNA segments, are

suggesting that they are stable during development and are not easily disrupted by transcriptional activities of the cell.

in the human cases (Figure 1). Remarkably, mutant mice carrying these structural alterations accurately reproduce

Intro: from structure to function



compartments
(~n x Mb)

TADs
(~n x 0.1 Mb)

A: open, expressed

B: closed, repressed

Giorgetti et al. Genome Biology 2013

Intro: from structure to function

- **How to profile** the 3D structure of the genome in **livestock species**?
- What are the **main features** of this 3D organization?
- Can we **link structure & function**?

The FR-AgENCODE project

Elisabetta Giuffra

Sylvain Foissac

FR-AgENCODE: a French pilot project of the FAANG action
INRA funding: 300KE, SelGen metaprogramme
4 research units, 60+ participants

<u>Goal: improve the functional annotation of animal genomes</u>

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, hypotalamus, pancreas, andrenals, kidney, muscle, bone, joints, spleen, lymphatic nodes, peyer's patches, ovary, oocytes, oviduct, uterus, mammary gland, acini, testis, seminal vesicle, etc)

♀x2

♂x2

Sus scrofa
(Large White)

Gallus gallus
(White Leghorn)

Bos Taurus
(Holstein)

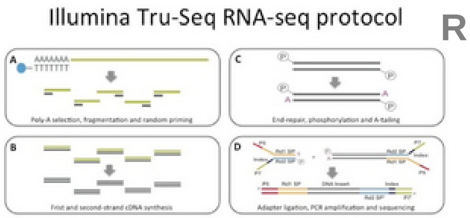
Capra hircus
(Alpine)

EMBL-EBI **4,115 BioSamples entries available at INRA biorepository** CRB ANIM

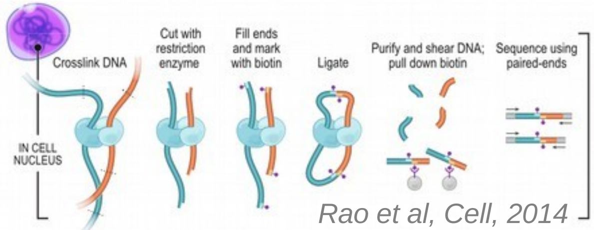
Sequencing assays on 3 target tissues

Transcriptome & chromatin structure profiling: polyA+ RNA-seq (130M RP/lib), small RNA-seq (40MR/lib), Hi-C (130M RP/lib), ATAC-seq (50M RP/lib)

~20B reads (3Tb of seq. data)

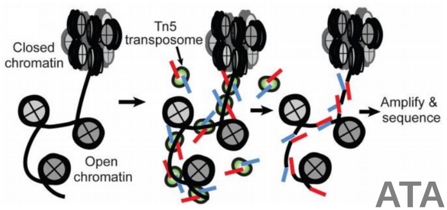


RNA-seq: transcriptome



Rao et al, Cell, 2014

Hi-C: chromosome conformation

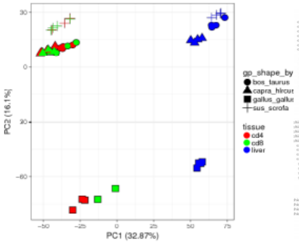


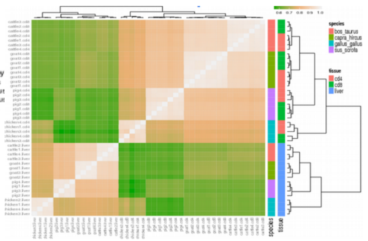
Buenrostro et al, 2013

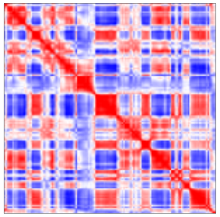
ATAC-seq: chromatin accessibility

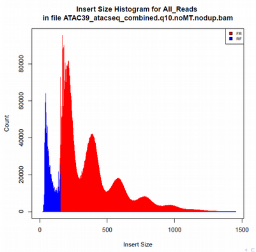
Data analysis

Annotation of genes, transcripts, regulatory regions and topological domains. Comparative and integrative analysis.



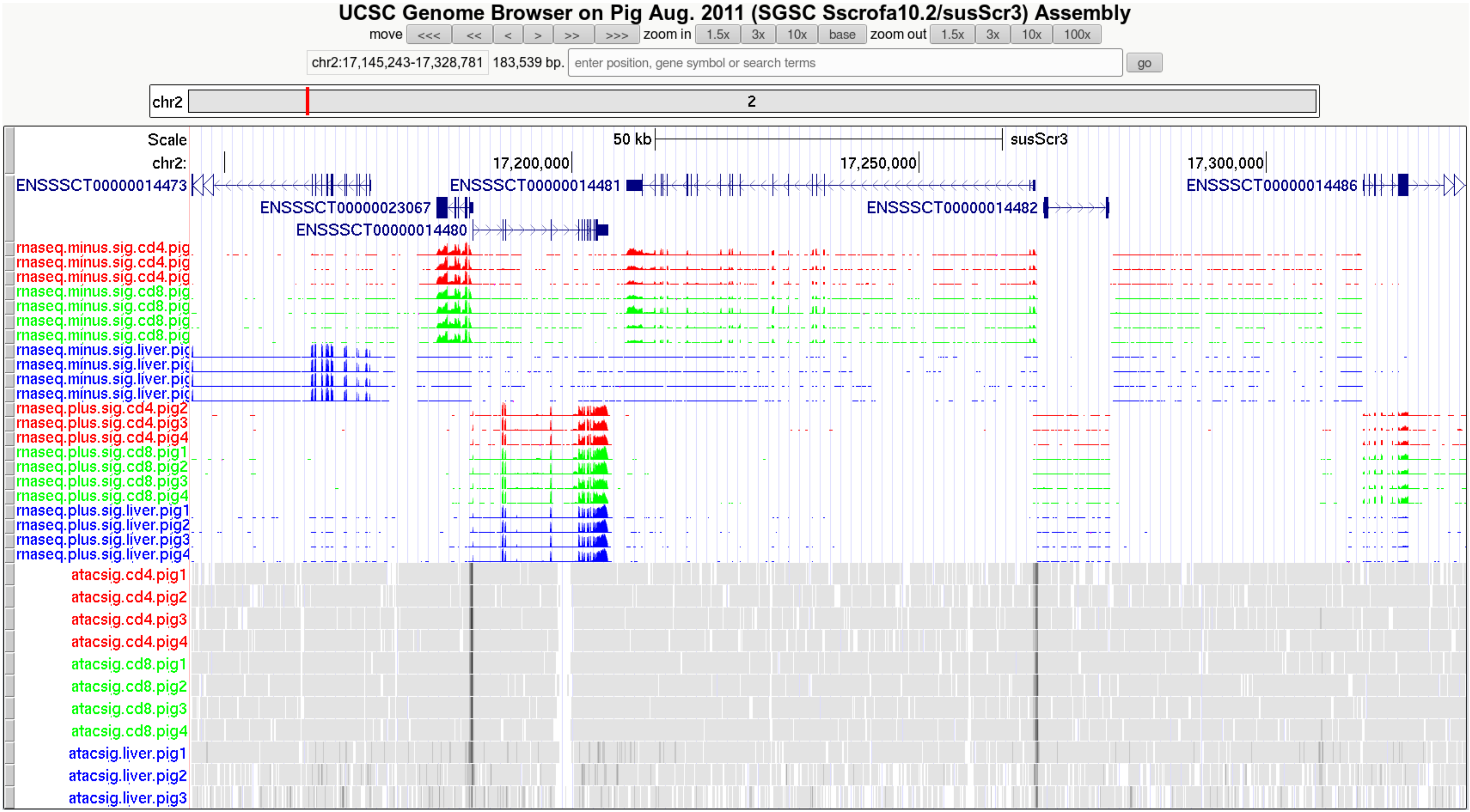




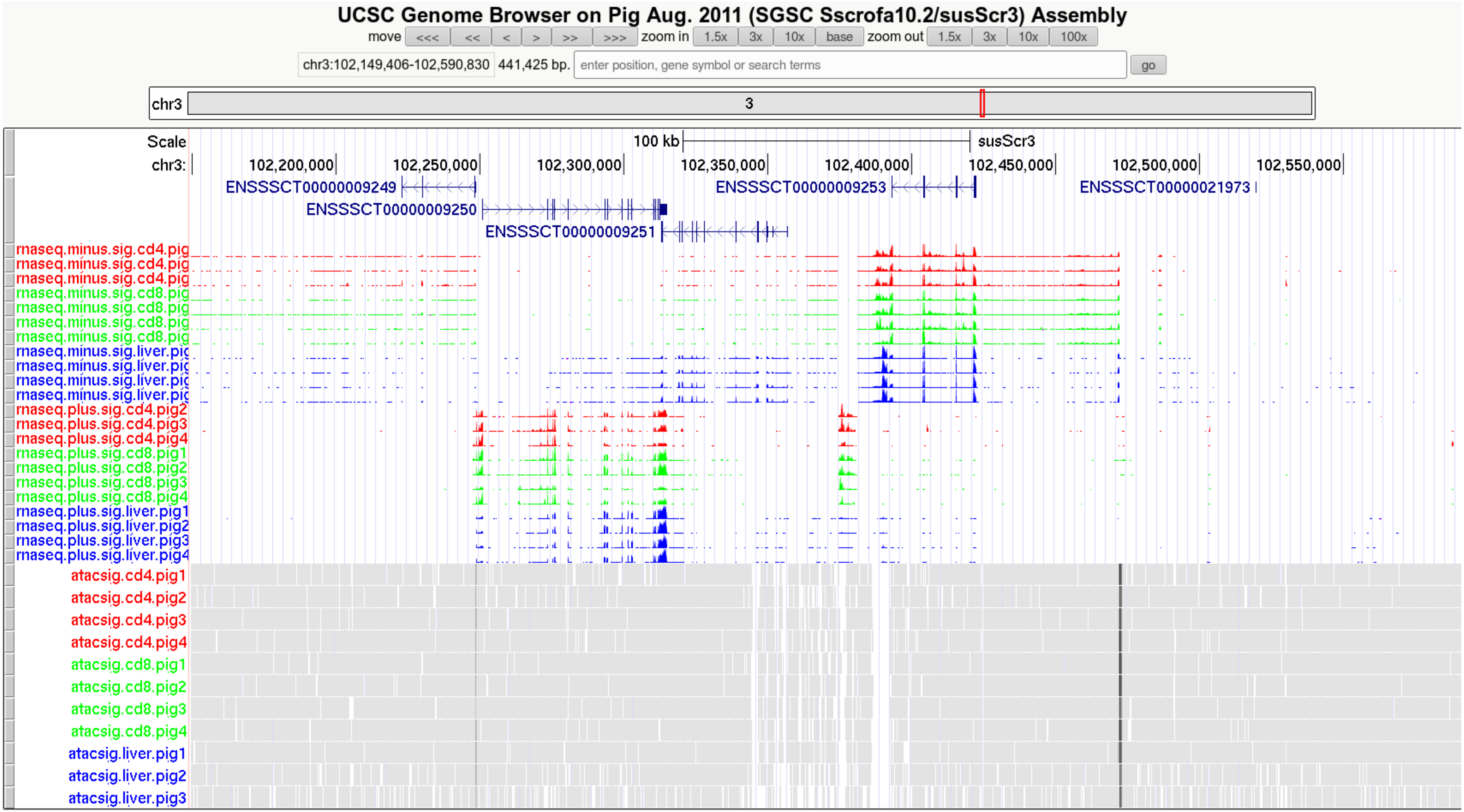


GenoToul GeT **3D genomics with Hi-C // Toulouse, Nov. 28, 2017 // sylvain.foissac@inra.fr** INRA SCIENCE & IMPACT

FR-AgENCODE data: RNA-seq & ATAC-seq



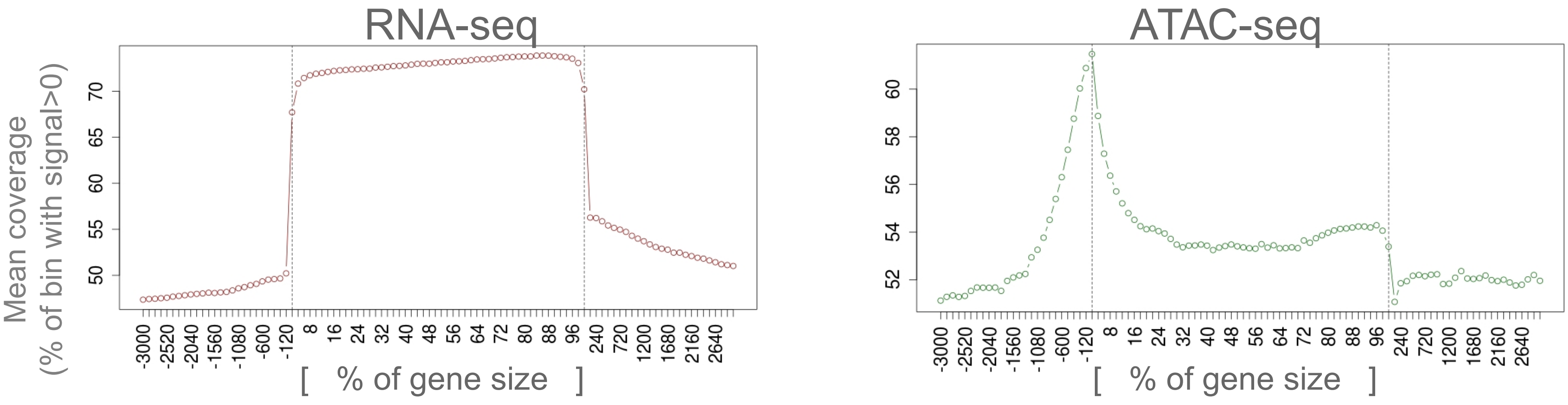
FR-AgENCODE data: RNA-seq & ATAC-seq



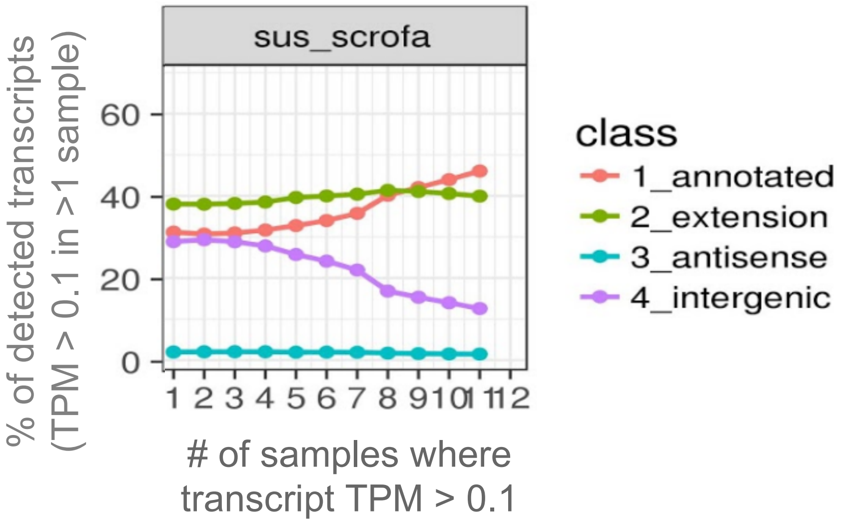
UCSC Genome Browser on Pig Aug. 2011 (SGSC Sscrofa10.2/susScr3) Assembly

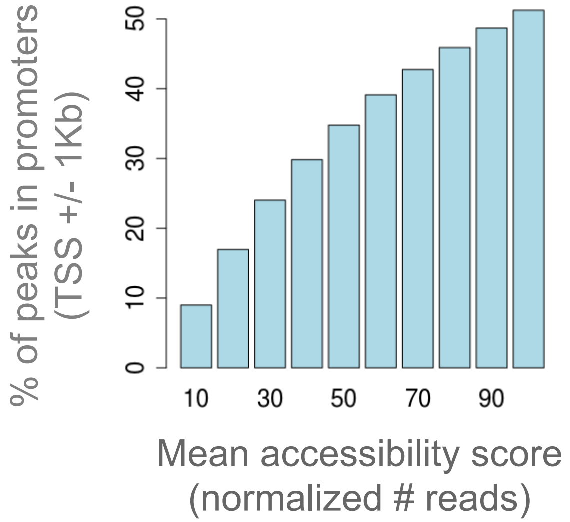
chr3:102,149,406-102,590,830 441,425 bp.

FR-AgENCODE results: RNA-seq & ATAC-seq



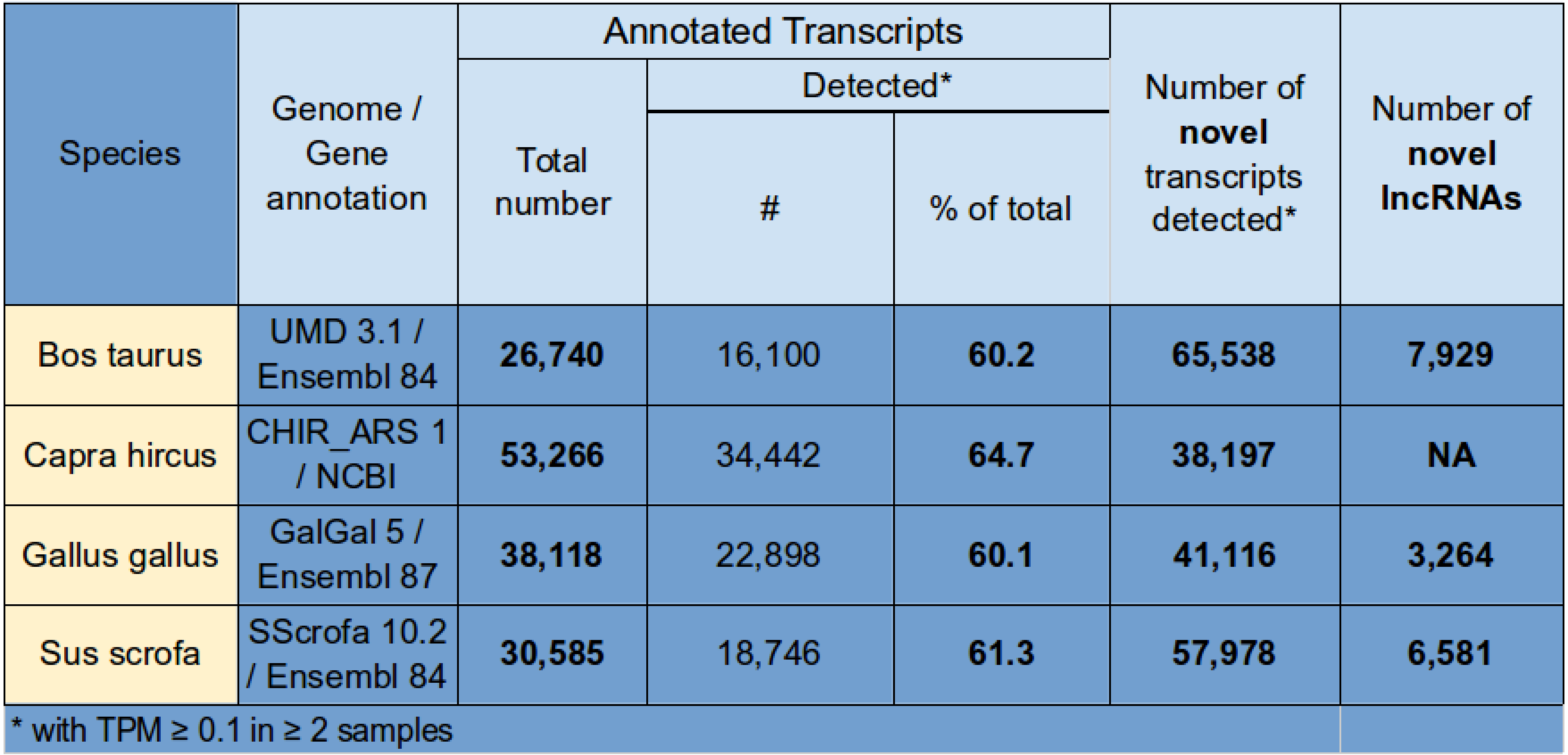
Relative position to annotated pig genes in bp (+/-3Kb)





Annotated transcripts are more expressed than novel ones
ATAC-seq reads and peaks accumulate in promoters

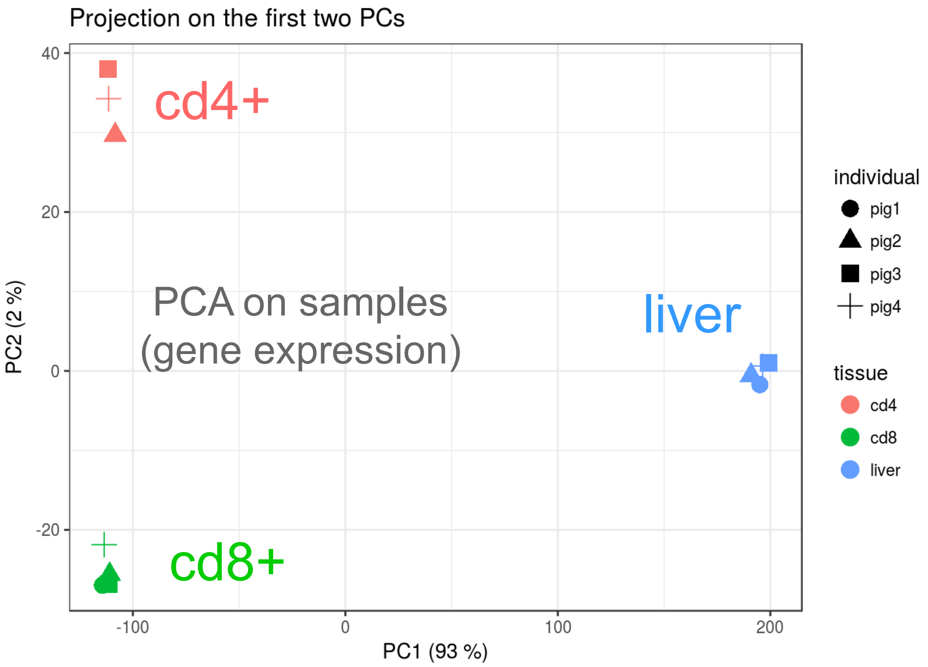
FR-AgENCODE results: RNA-seq & ATAC-seq



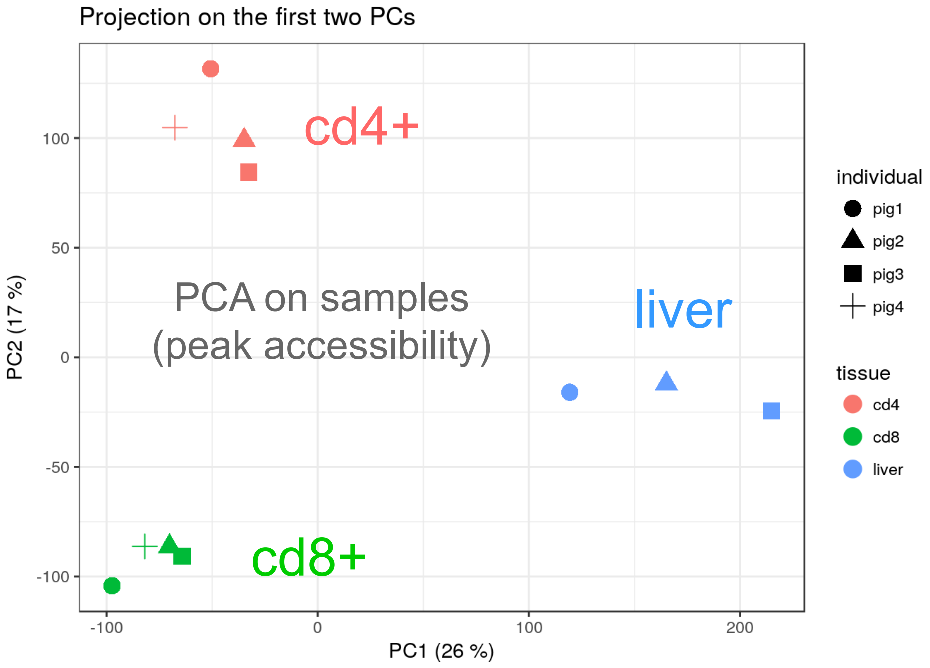
Species	Genome / Gene annotation	Annotated Transcripts			Number of **novel** transcripts detected*	Number of **novel** lncRNAs
		Total number	Detected*			
			#	% of total		
Bos taurus	UMD 3.1 / Ensembl 84	**26,740**	16,100	**60.2**	**65,538**	**7,929**
Capra hircus	CHIR_ARS 1 / NCBI	**53,266**	34,442	**64.7**	**38,197**	**NA**
Gallus gallus	GalGal 5 / Ensembl 87	**38,118**	22,898	**60.1**	**41,116**	**3,264**
Sus scrofa	SScrofa 10.2 / Ensembl 84	**30,585**	18,746	**61.3**	**57,978**	**6,581**

* with TPM ≥ 0.1 in ≥ 2 samples

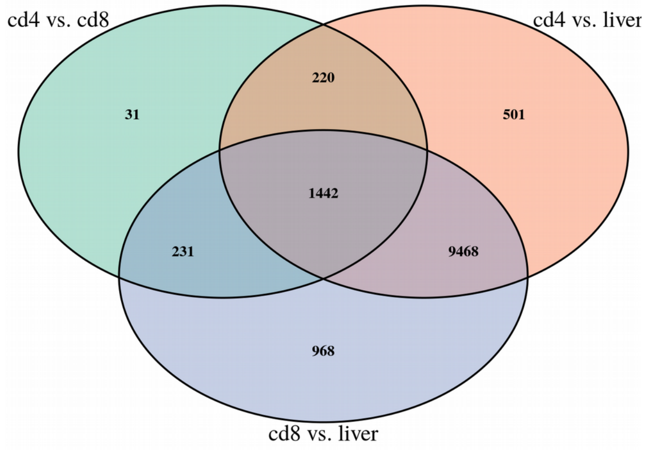
FR-AgENCODE results: RNA-seq & ATAC-seq



Filtering: TPM>0.1 in min 2 samples
Normalization: TMM (edgeR)
N=15,928

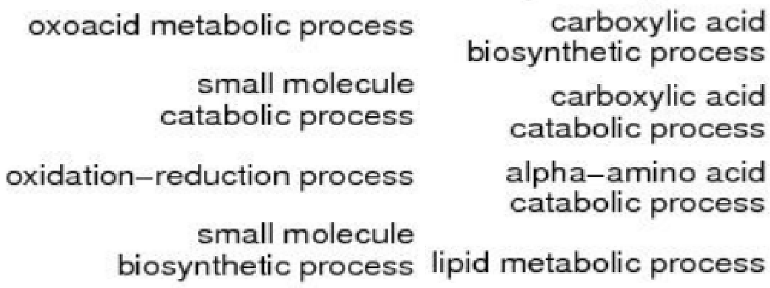


Filtering: none
Normalization: LOESS (csaw)
N=120,914



Differential gene expression *(S. scrofa)*

Top- significant GO terms (Biol. Proc.) for over-expressed genes in liver vs. T cells


oxoacid metabolic process

small molecule catabolic process

oxidation–reduction process

small molecule biosynthetic process

carboxylic acid biosynthetic process

carboxylic acid catabolic process

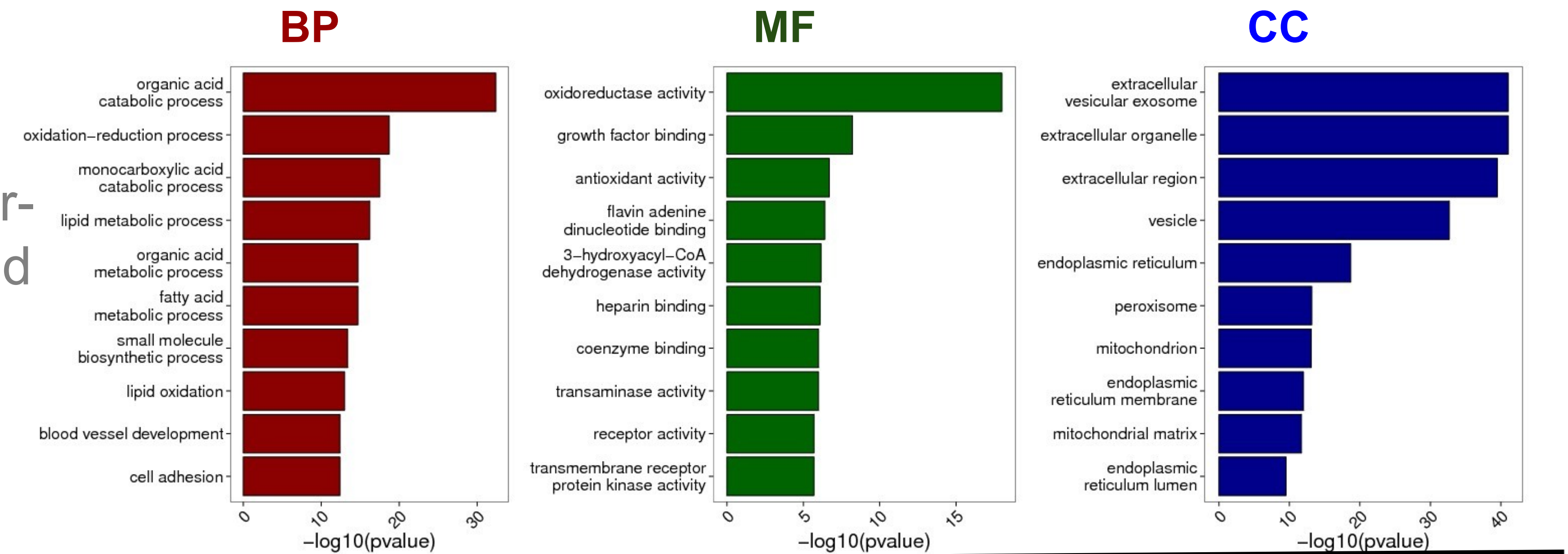
alpha–amino acid catabolic process

lipid metabolic process

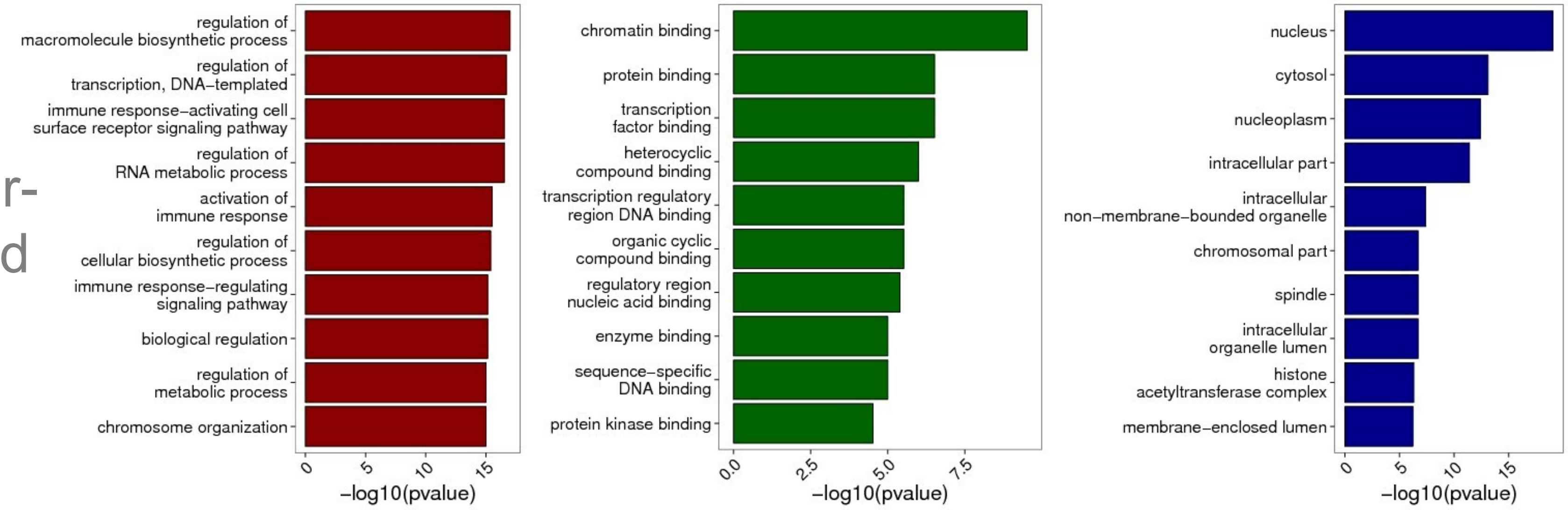
FR-AgENCODE results: RNA-seq & ATAC-seq

Gene Ontology analysis

Liver over-expressed genes

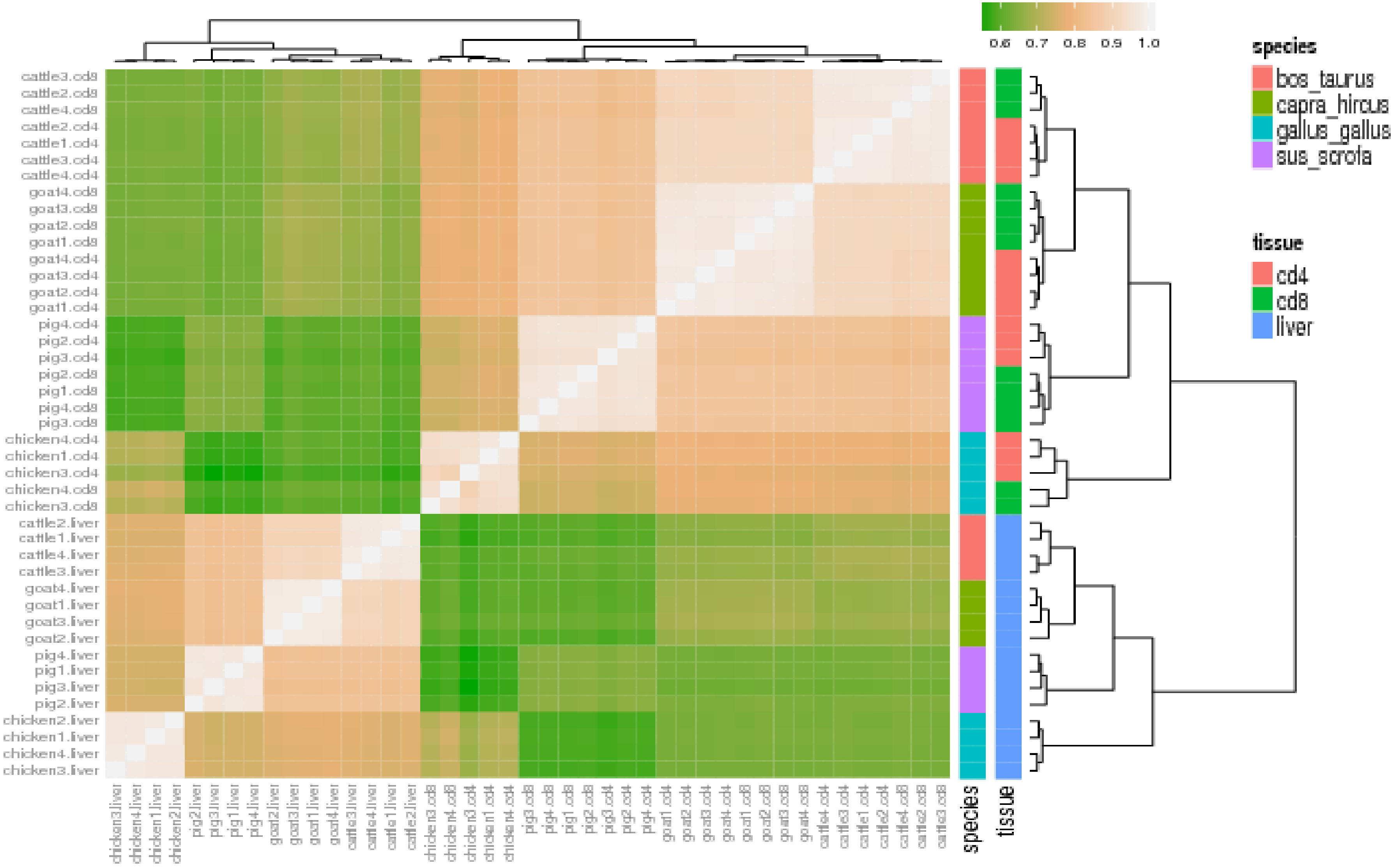


T cell over-expressed genes



FR-AgENCODE results: RNA-seq & ATAC-seq

Expression of orthologous genes: multi-species clustering

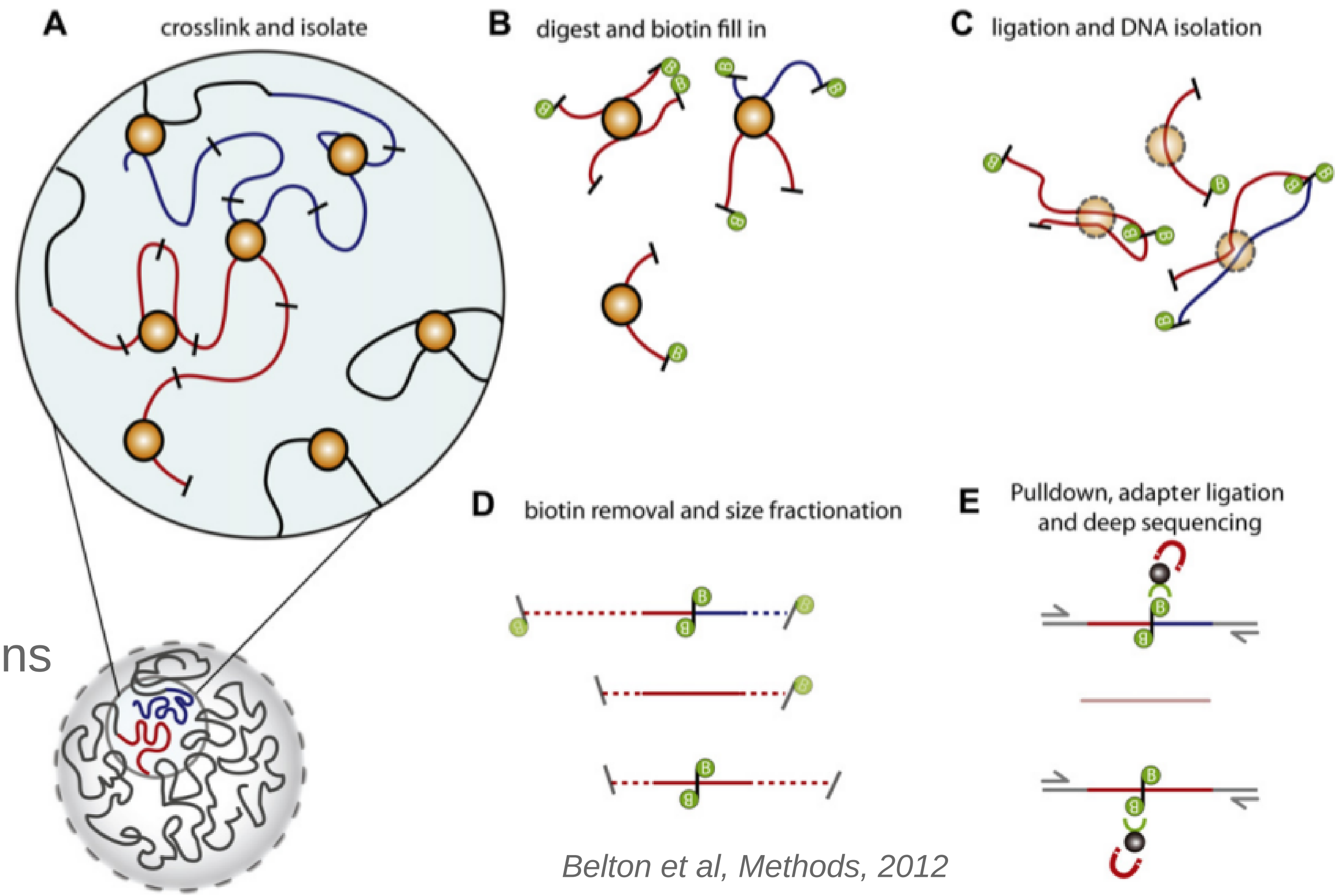


Hi-C : the protocol

Hi-C: chromosome conformation capture by high-throughput sequencing

(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)

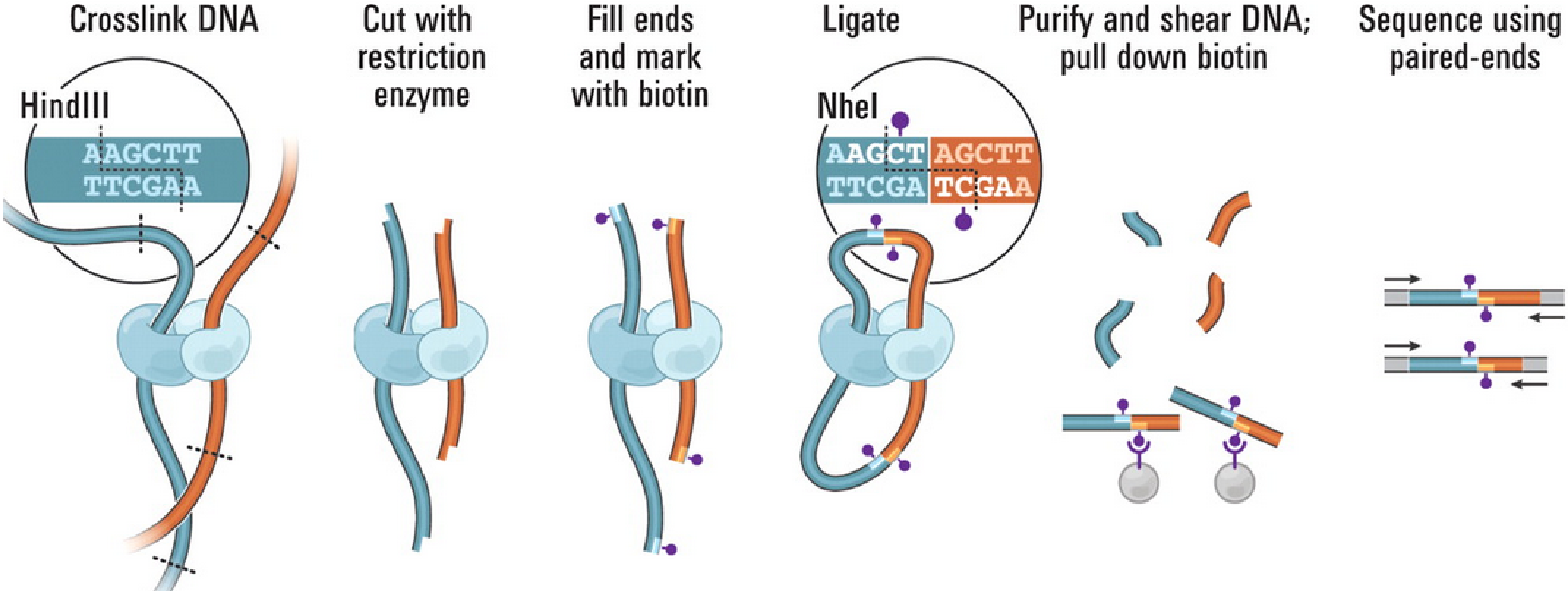


Belton et al, Methods, 2012

Hi-C : the protocol

Hi-C: chromosome conformation capture by high-throughput sequencing

(Lieberman-Aiden et al, Science, 2009, Rao et al, Cell, 2014)



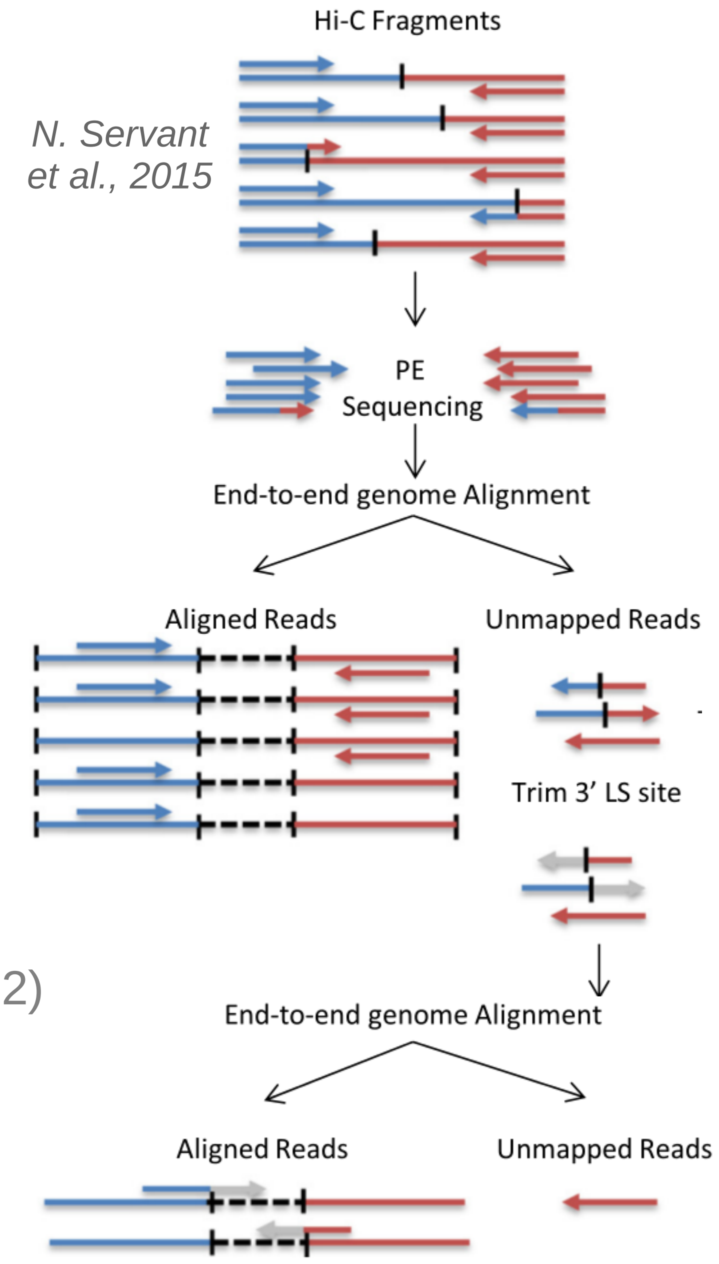
Rao et al, Cell, 2014

Hi-C : the analysis pipeline

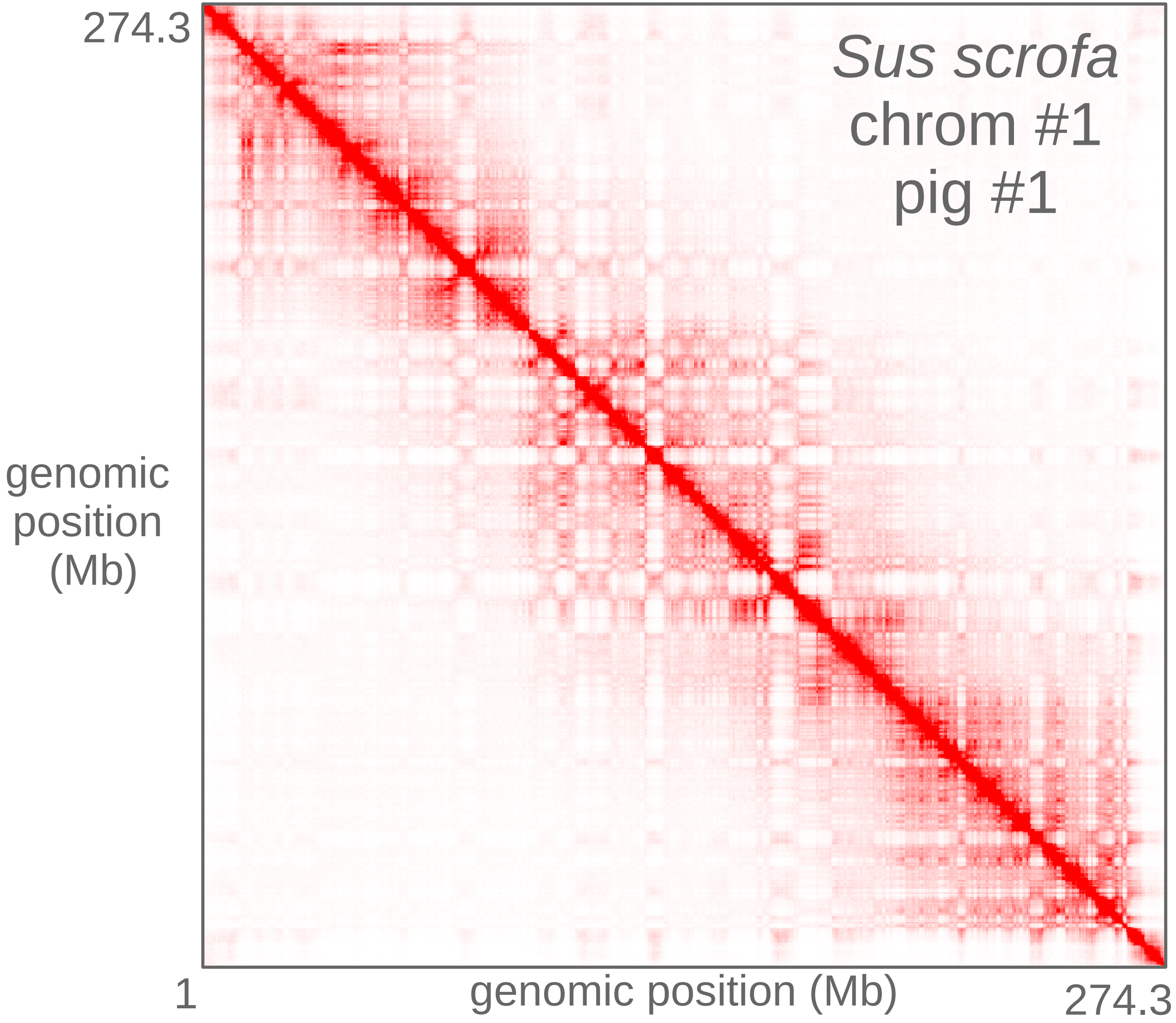
FR-AgENCODE pipeline
- Trim reads (ligation site)
- **Map on reference genome**
- Discard inconsistent pairs
- **Count reads in pairs of genomic bins & generate contact matrix**
- Normalize contact matrix (non parametric, matrix balancing)
- Generate html report
- **Identify Topological Associated Domains**, *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)

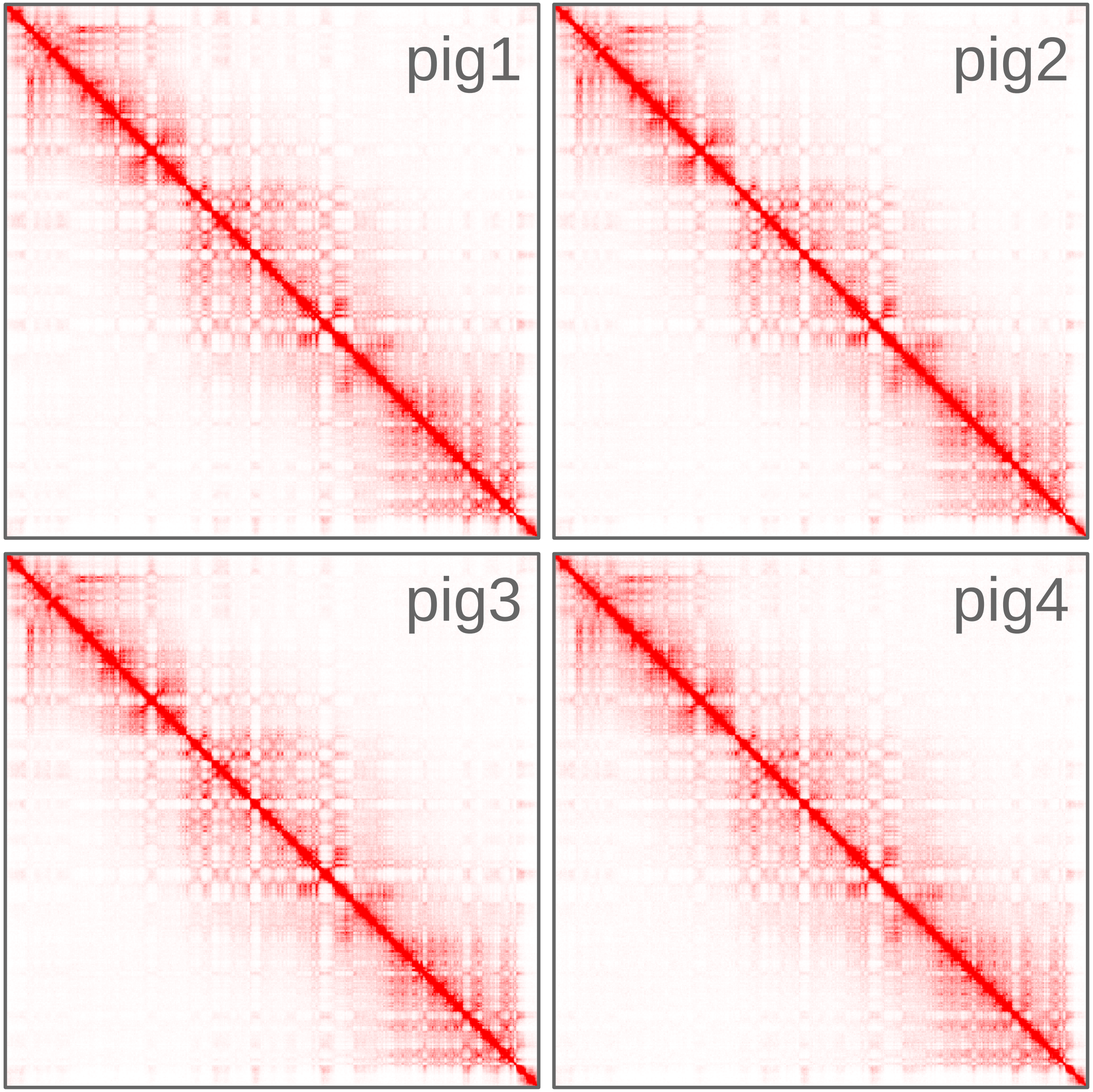
N. Servant et al., 2015



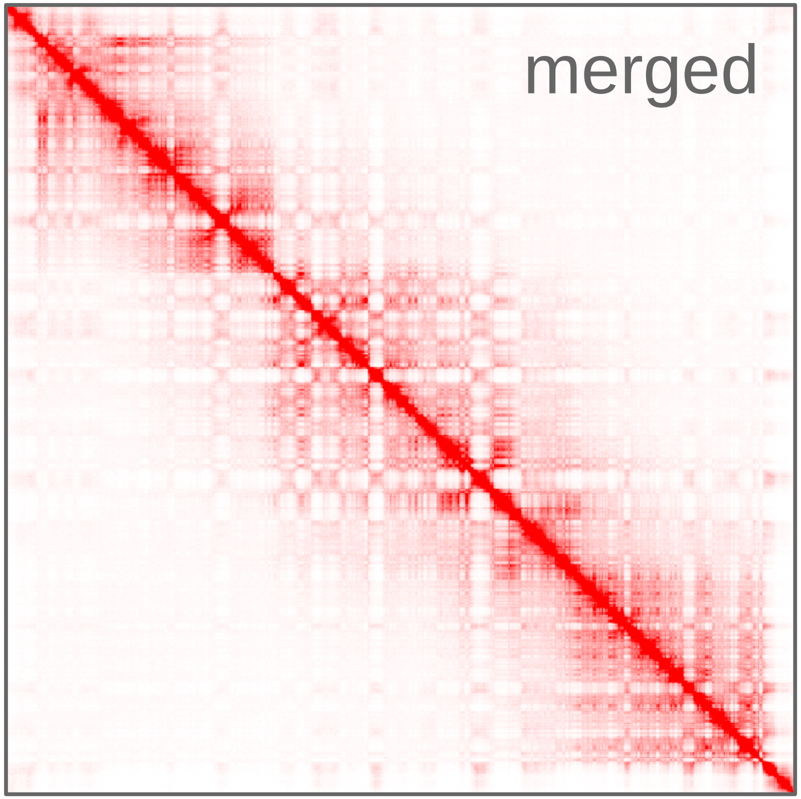
Hi-C results : the interaction matrix



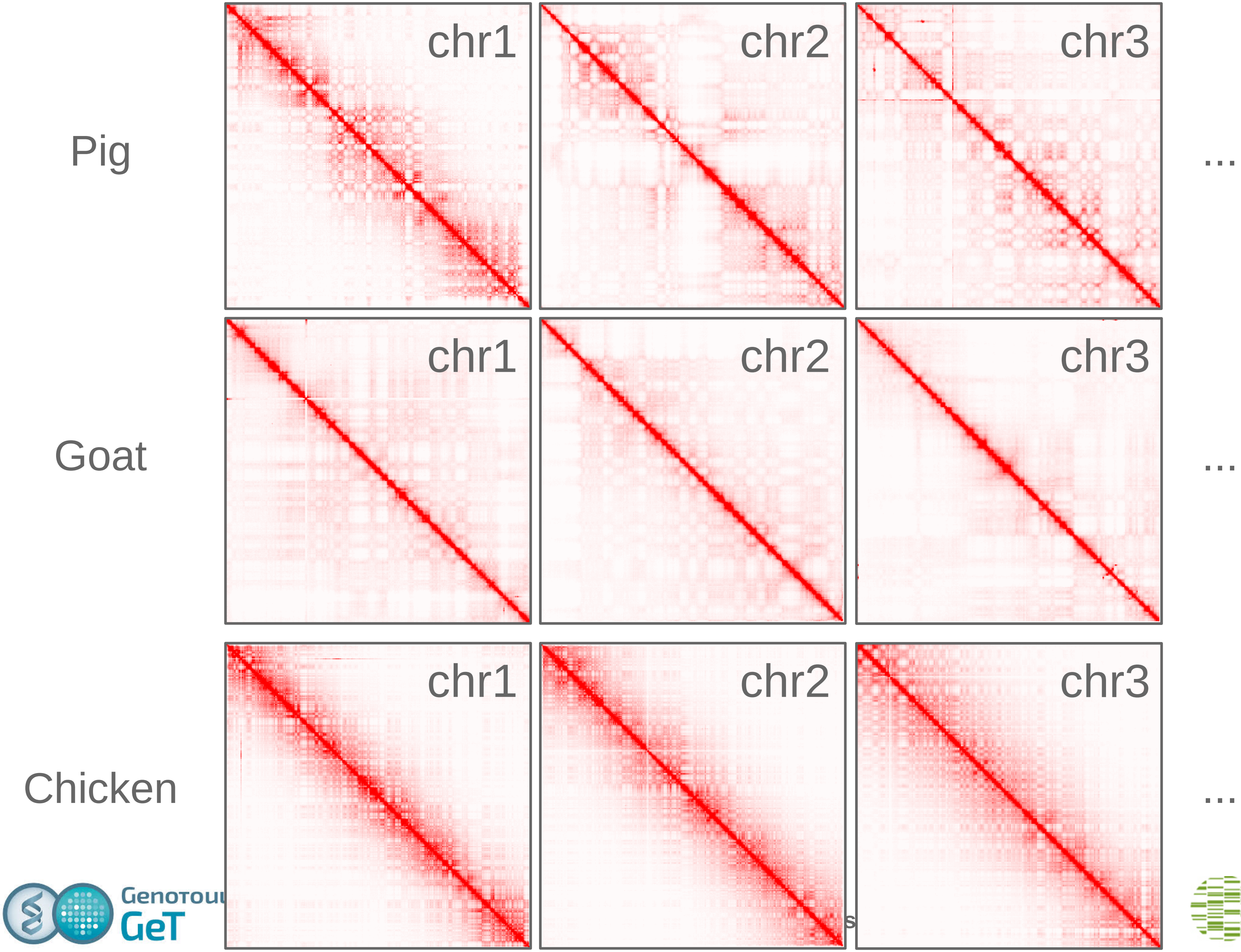
Hi-C results : the interaction matrix



Sus scrofa
chr1



Hi-C results : the interaction matrix



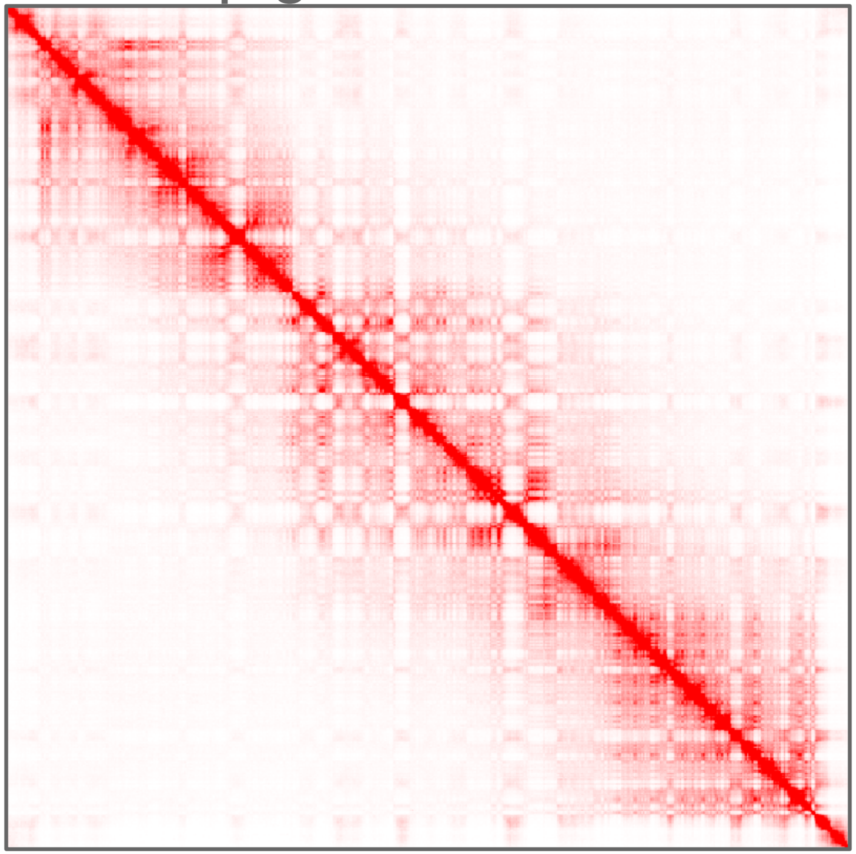
Hi-C results : finding A/B compartments

- Normalization 1: matrix balancing (GC%, repeats, RS density, etc)
- Normalization 2: observed/expected counts (genomic distance)
- Pearson correlation matrix
- Principal Component Analysis => 1st PC sign

Hi-C results : finding A/B compartments

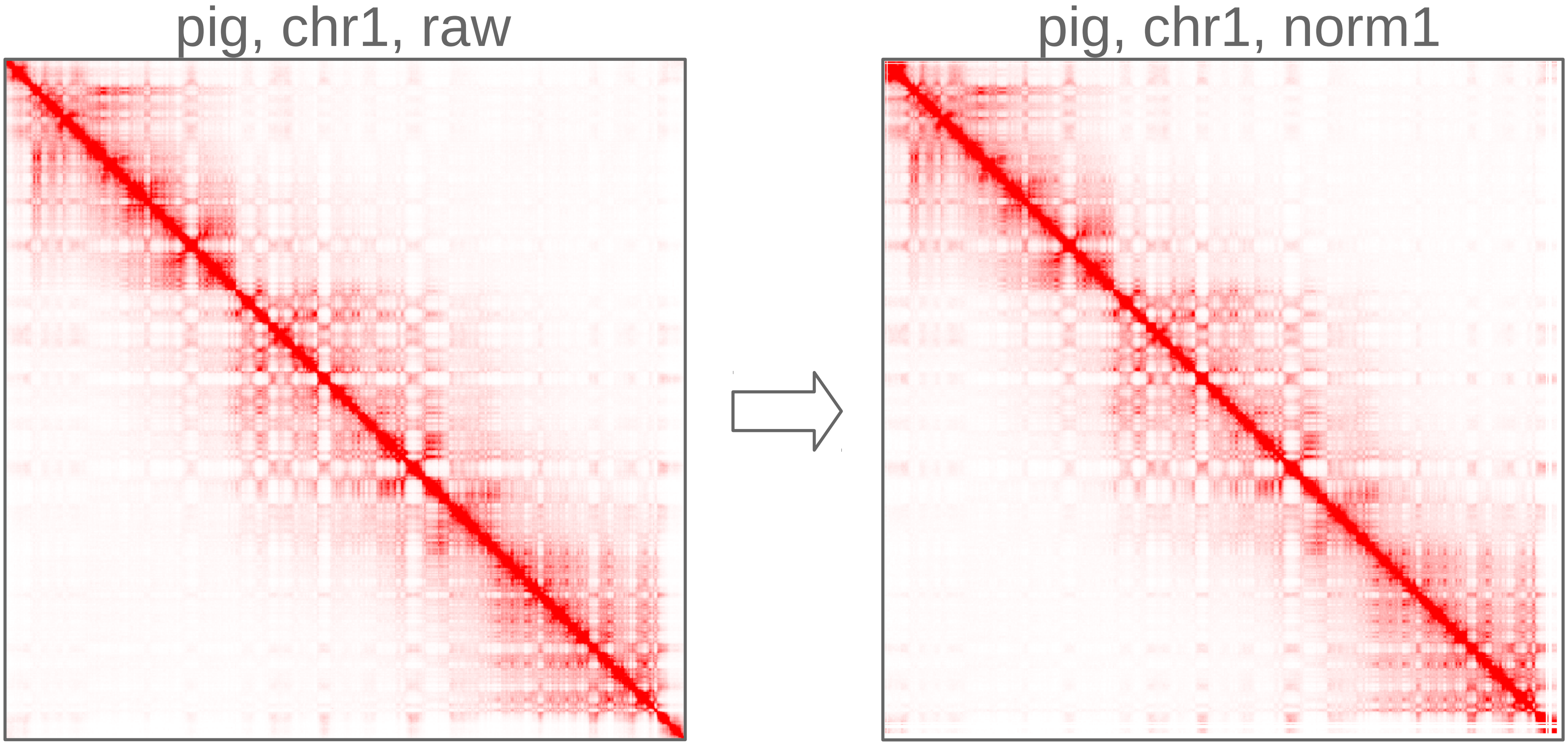
- Normalization 1: “ICE” matrix balancing (GC%, repeats, RS density, etc)

pig, chr1, raw



Hi-C results : finding A/B compartments

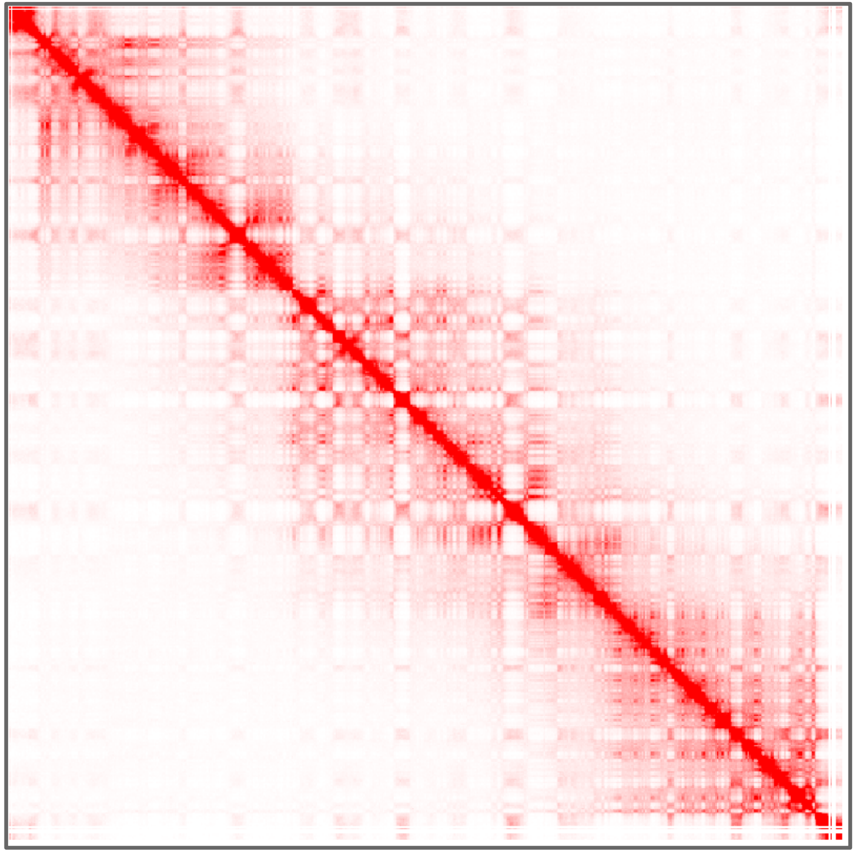
- Normalization 1: “ICE” matrix balancing (GC%, repeats, RE density, etc)



Hi-C results : finding A/B compartments

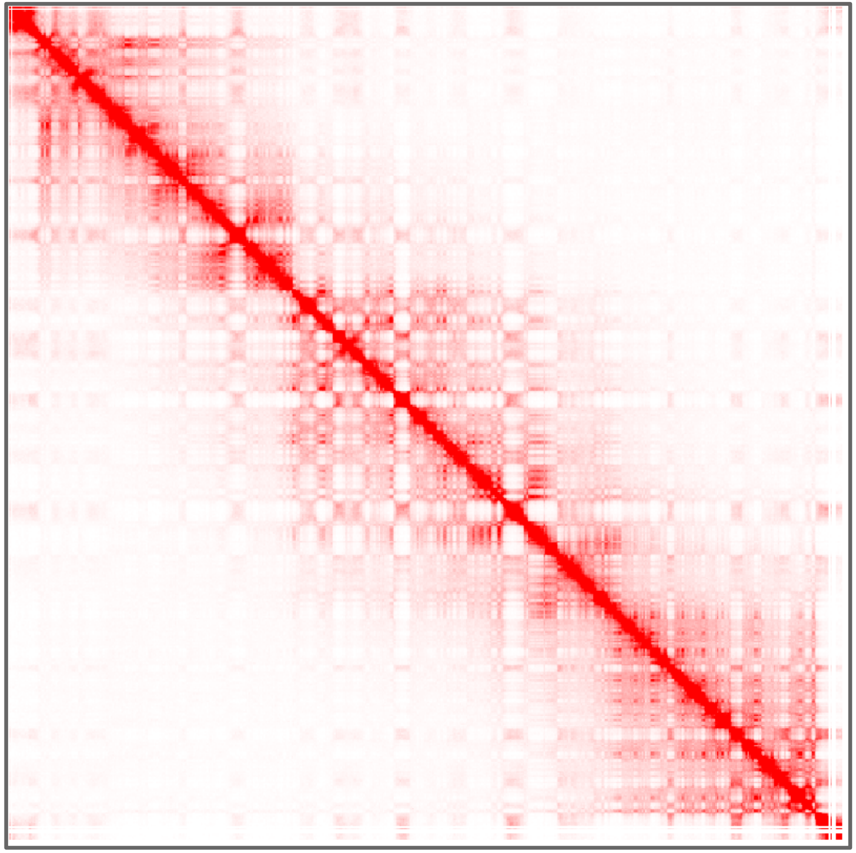
- Normalization 2: observed/expected counts (genomic distance)

pig, chr1, norm1

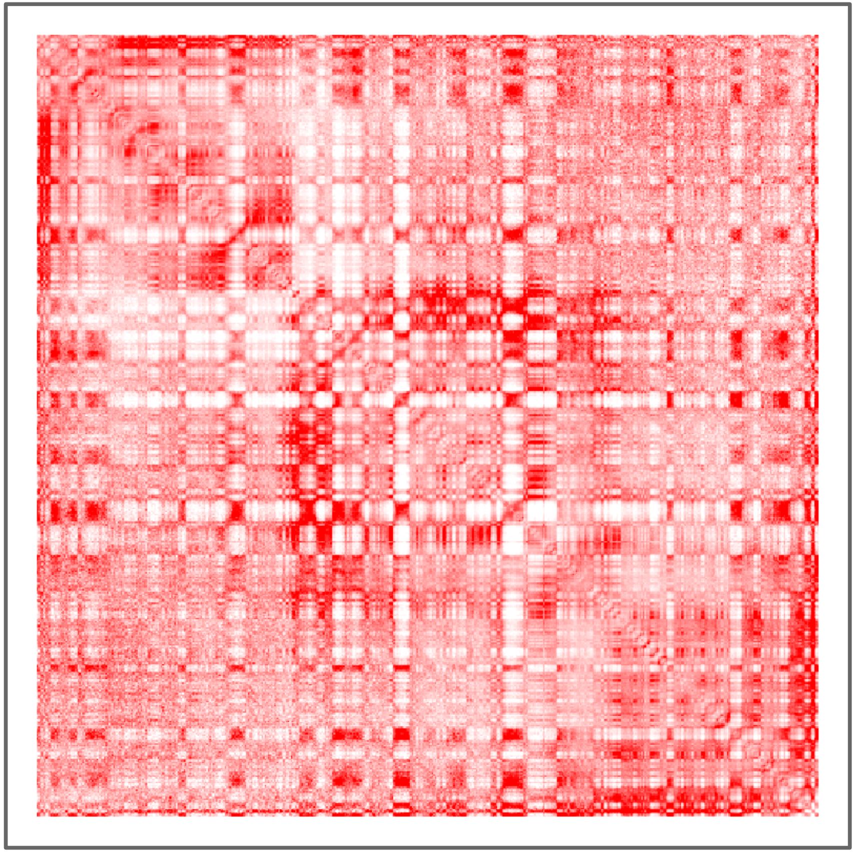


Hi-C results : finding A/B compartments

- Normalization 2: observed/expected counts (genomic distance)



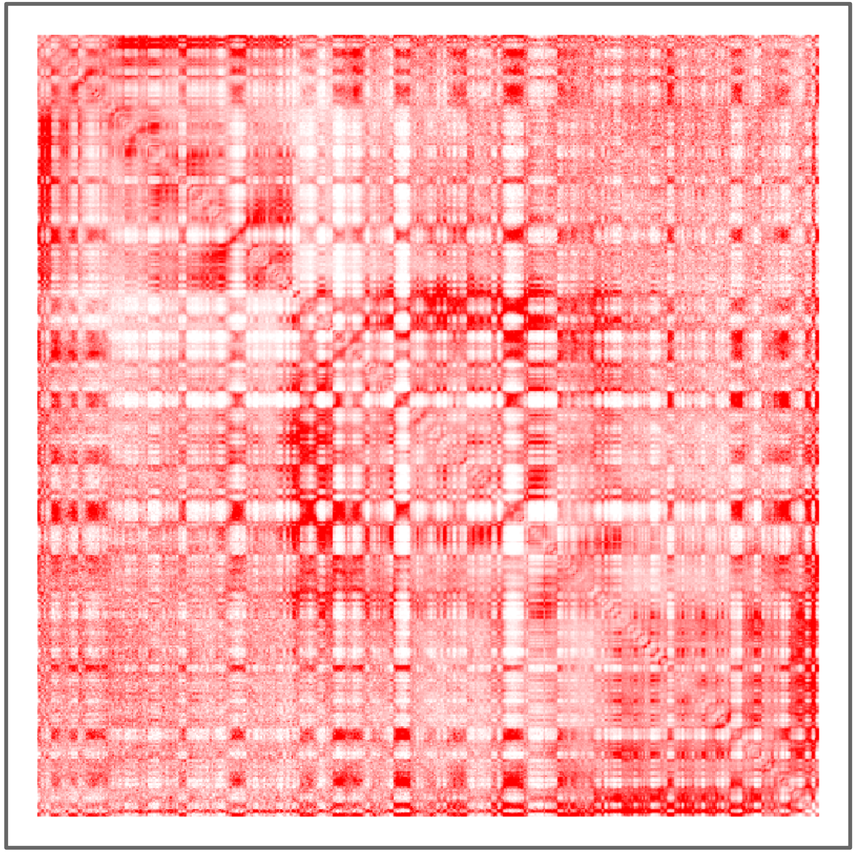
pig, chr1, norm1

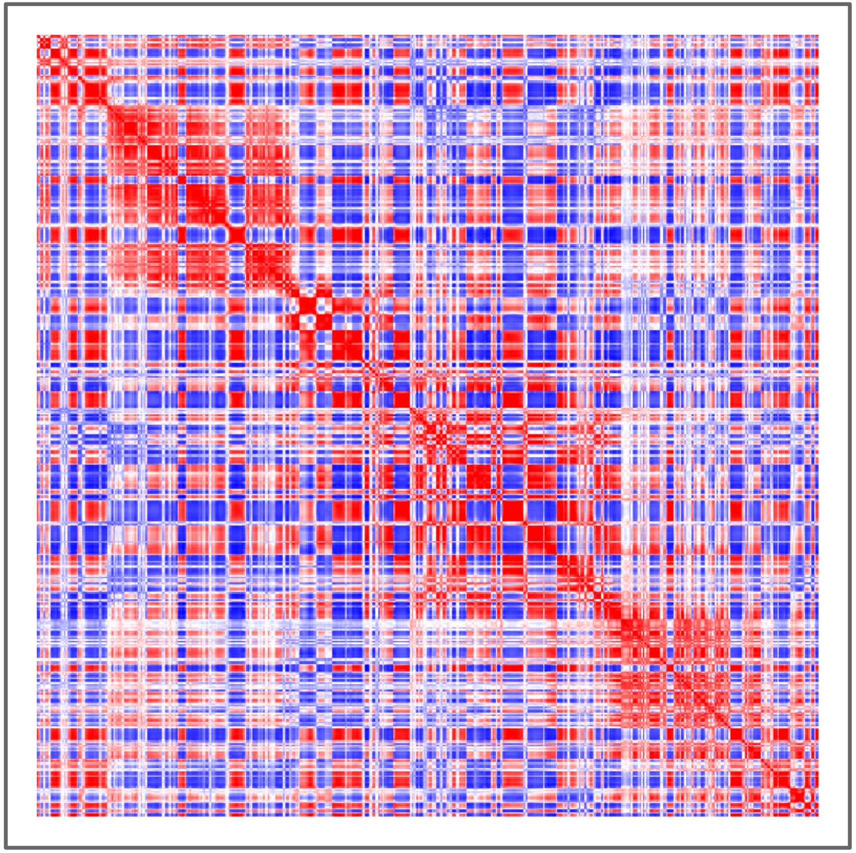


pig, chr1, norm2

Hi-C results : finding A/B compartments

- Pearson correlation matrix

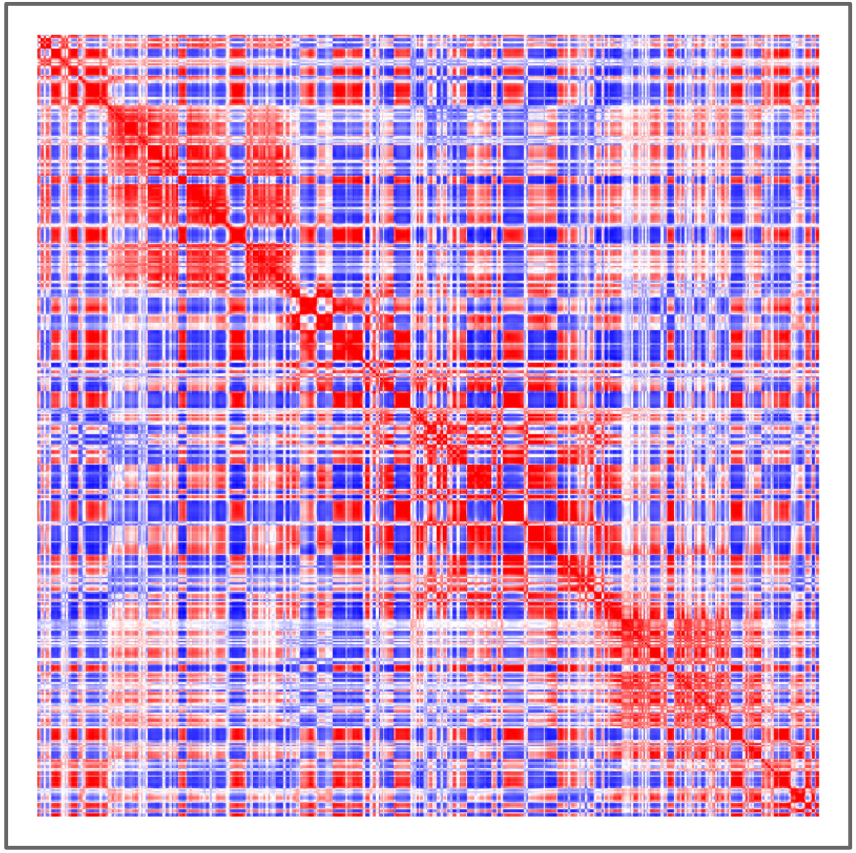

pig, chr1, norm2

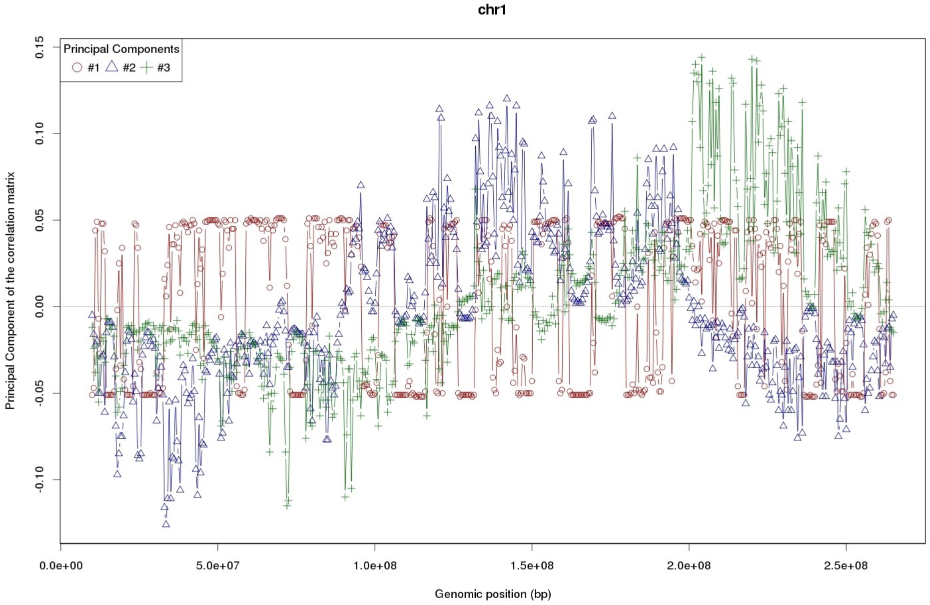

pig, chr1, pearson

Hi-C results : finding A/B compartments

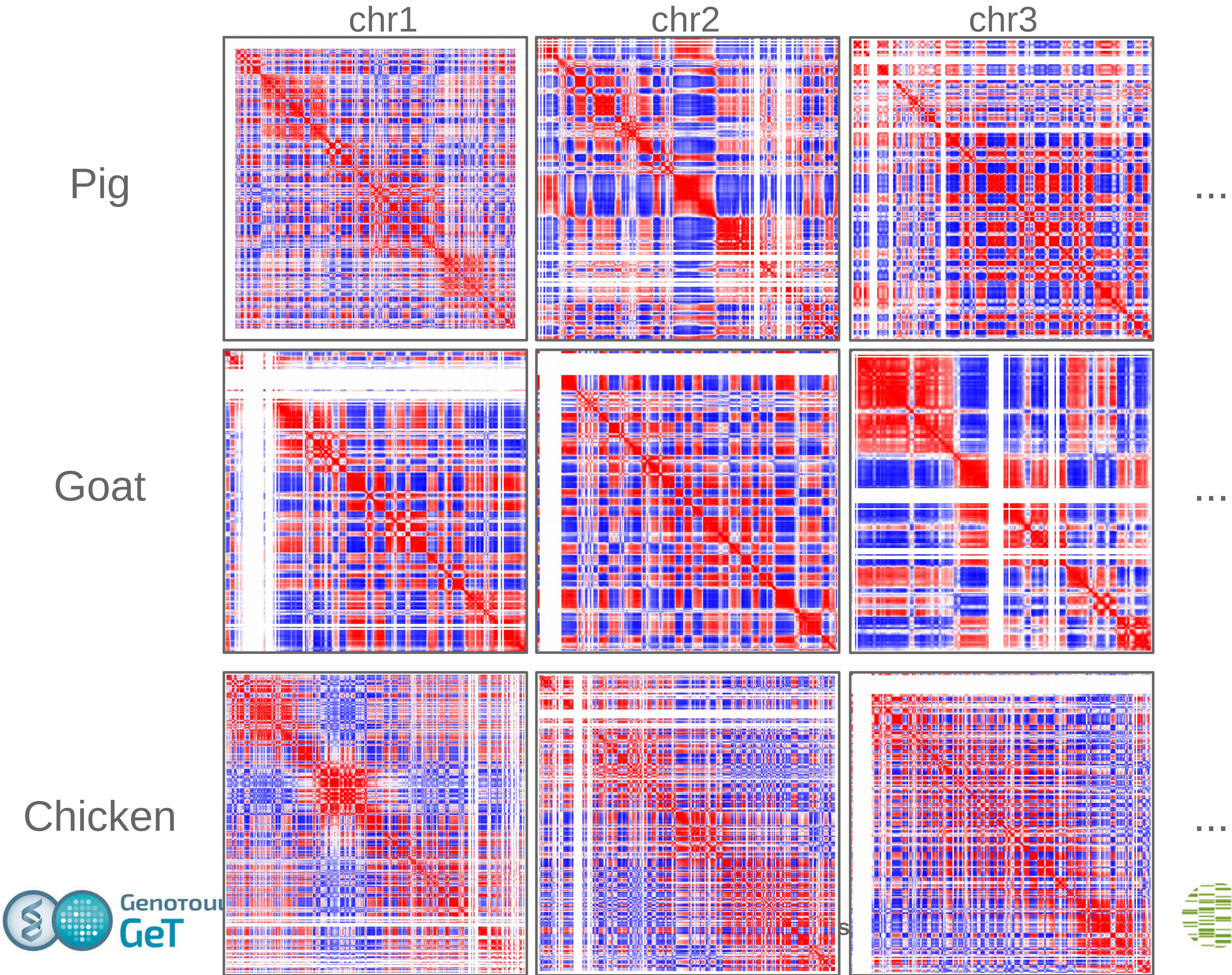
- Principal Component Analysis => 1^{st} PC sign

pig, chr1, pearson

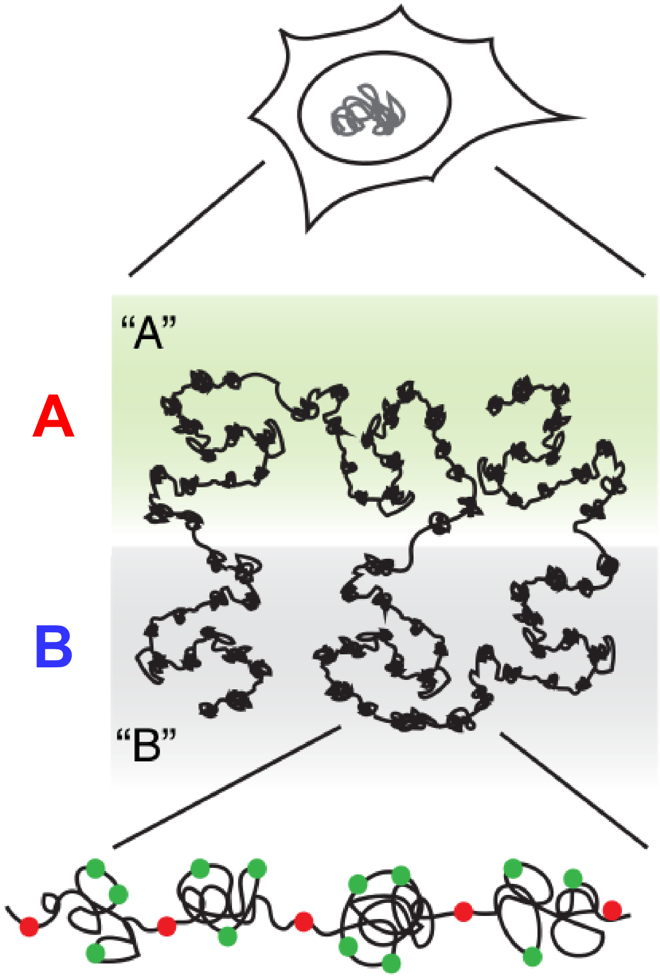




Hi-C results : finding A/B compartments



Results consistency: Hi-C + RNA-seq + ATAC-seq

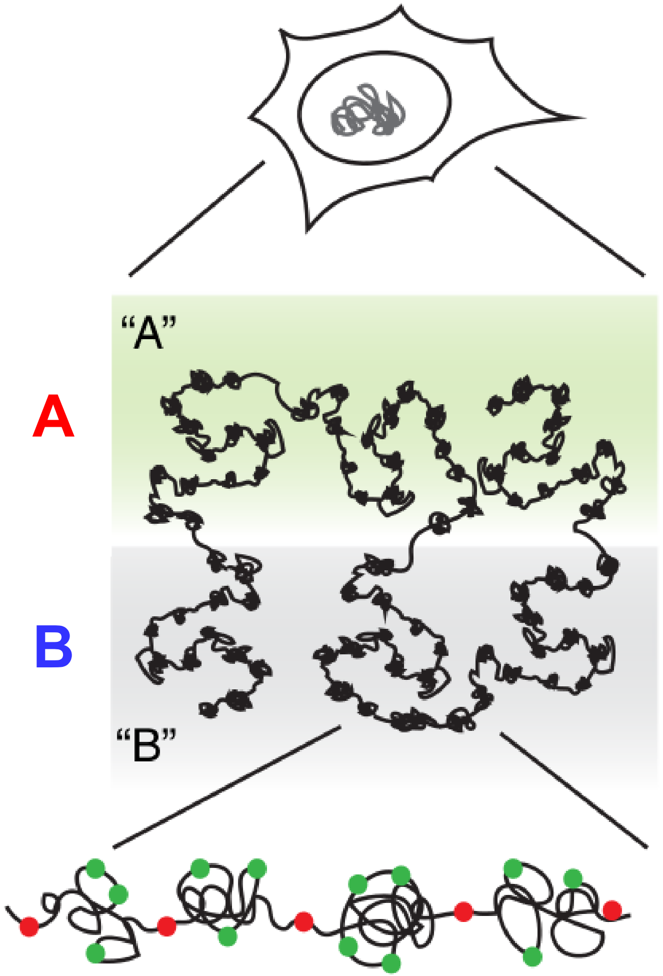


3D nuclear organization & features

Compartments
- **A**: open, expressed
- **B**: closed, repressed

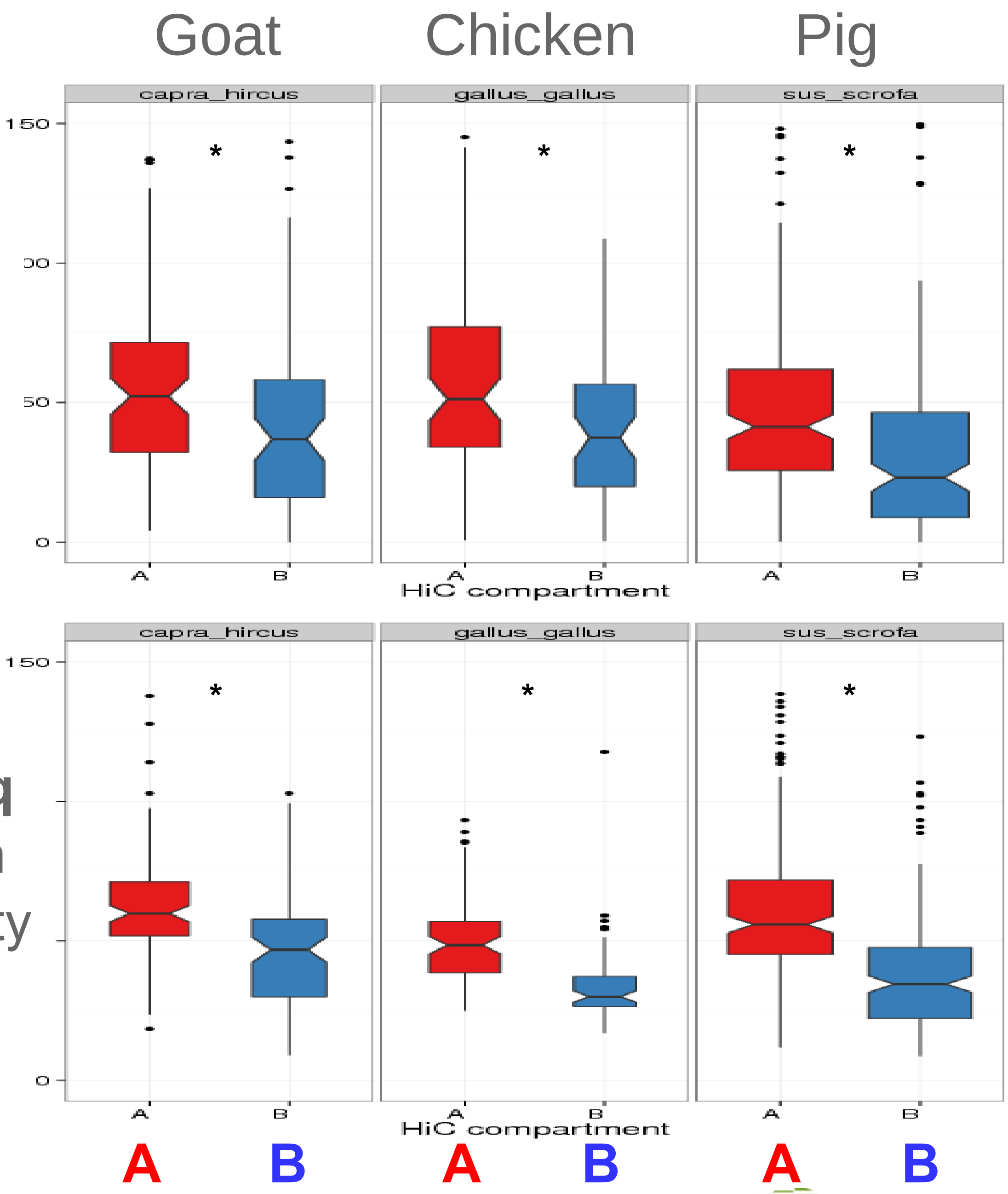
?

Results consistency: Hi-C + RNA-seq + ATAC-seq



RNA-seq
gene
expression

ATAC-seq
chromatin
accessibility



Conclusion

- Livestock genomes need a better annotation
- The 3D organization of the genome can be profiled with Hi-C from tissue samples in livestock species
- Chromosomes are structured in A/B compartments
- Hi-C results are consistent with RNA-seq and ATAC-seq data: compartments A > B for gene expression and chromatin accessibility
- The FR-AgENCODE project is a valuable contribution to the functional annotation effort of the FAANG consortium

Acknowledgements

FR-AgENCODE

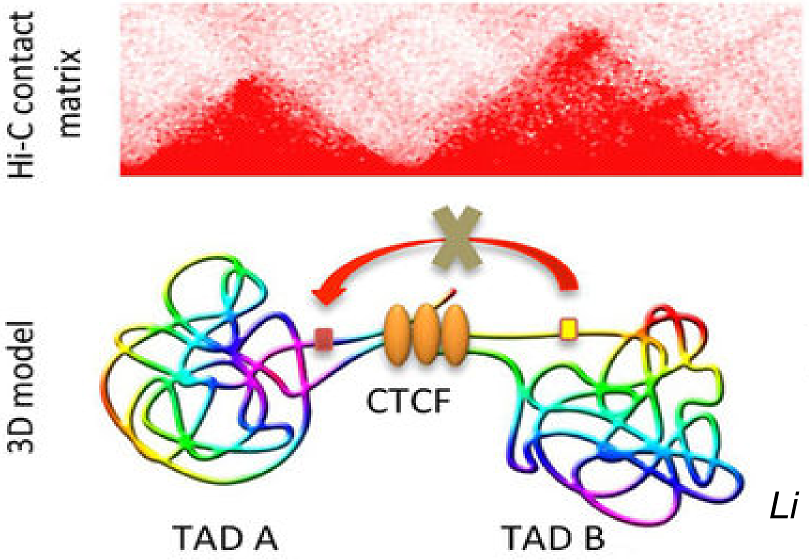
Acloque H.
Bardou P.
Blanc F.
Cabau C.
Crisci E.
Derrien T.
Djebali S.
Drouet F.
Esquerre D.
Fabre S.
Foissac, S
Gaspin C.
Giuffra E.
Gonzalez I.
Goubil A.
Klopp C.
Lagarrigue S.

Laurent F.
Marthey S.
Marti M.
Mompart F.
Munyard K.
Muret K.
Pinard-Van der Laan MH.
Pollet S.
Queré P.
Rau A.
Robelin D.
San Cristobal M.
Tixier-Boichard M.
Tosser-Klopp G.
Villa-Vialaneix N.
Vincent-Naulleau S.
Zytnicki M. [...]

Hi-C @ Toulouse

Maria Marti
Martine Yerle
Yvette Lahbib
Florence Mompart
Laurence Liaubet
Nathalie Villa-Vialaneix
Pierre Neuvial
Alain Vignal
Matthias Zytnicki
David Robelin
Hervé Acloque
Sarah Djebali

Predicted CTCF binding sites peak at Topologically Associating Domain (TAD) boundaries



Li et al, Scientific Reports, 2016