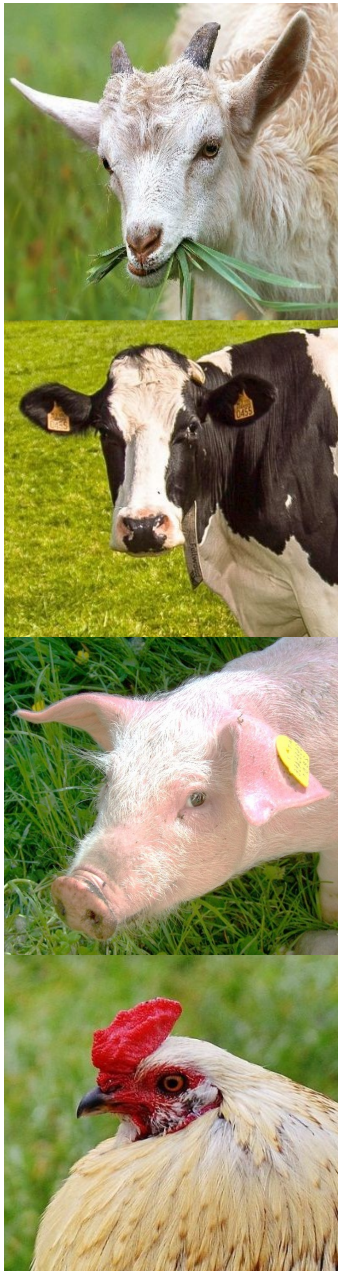
Integrative analysis of transcriptomes and chromatin accessibility regions reveals regulatory mechanisms involved in pig immune and metabolic functions

[FAANG pilot project “FR-AgENCODE”]

Sarah Djebali, Kylie Munyard, Nathalie Villa-Vialaneix, Cedric Cabau, Andrea Rau, Elisa Crisci, Thomas Derrien, Christophe Klopp, Matthias Zytnicki, Sandrine Lagarrigue, Hervé Acloque, Sylvain Foissac, Elisabetta Giuffra

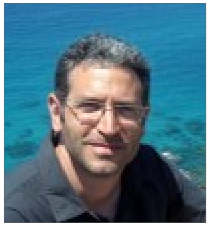
sylvain.foissac@inra.fr



36th International Society for Animal Genetics Conference – 16th-21st July 2017 – UCD – Dublin, Ireland

The FR-AgENCODE project

Elisabetta Giuffra



Sylvain Foissac

FR-AgENCODE: a French pilot project of the FAANG action

Goal: improve the functional annotation of animal genomes

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

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)

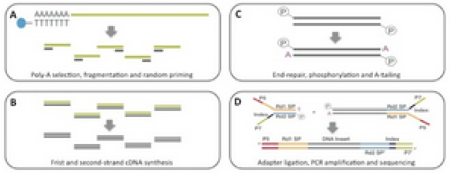
4,115 BioSamples entries available at INRA biorepository

Sequencing assays on 3 target tissues

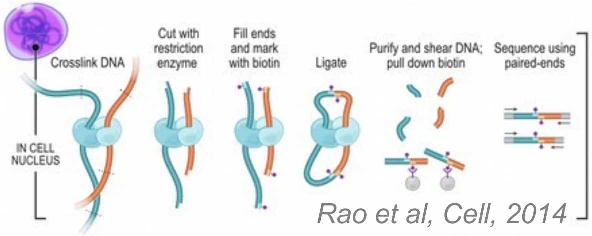
Transcriptome & chromatin structure profiling: polyA+ RNA-seq (130M RP/lib), small RNA-seq (40MR/lib), Hi-C (130M RP/lib), ATAC-seq (50M RP/lib)

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

Illumina Tru-Seq RNA-seq protocol

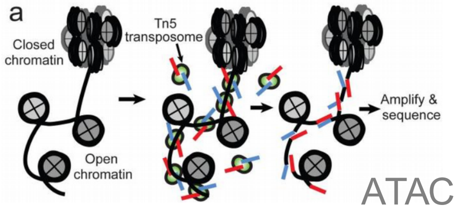


RNA-seq: transcriptome



Rao et al, Cell, 2014

Hi-C: chromosome conformation

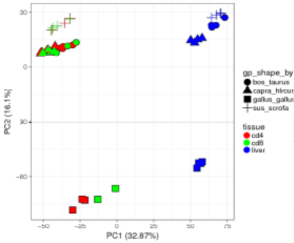


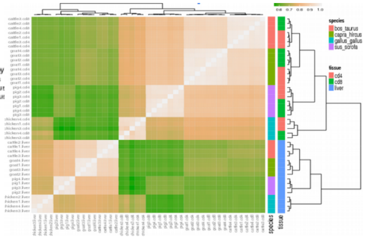
Buenrostro et al, 2013

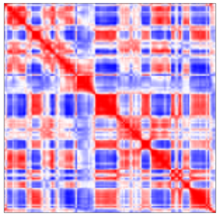
ATAC-seq: chromatin accessibility

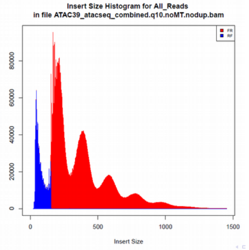
Data analysis

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









sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

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)

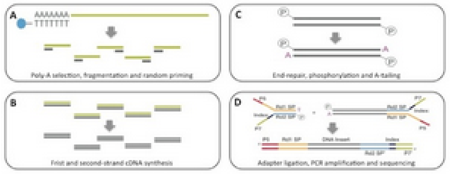
4,115 BioSamples entries available at INRA biorepository

Sequencing assays on 3 target tissues

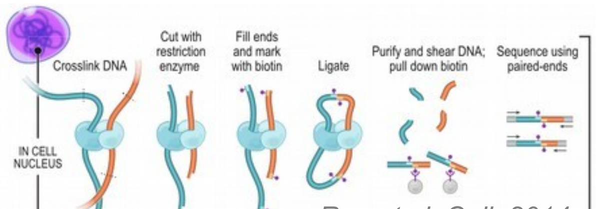
Transcriptome & chromatin structure profiling: polyA+ RNA-seq (130M RP/lib), small RNA-seq (40MR/lib), Hi-C (130M RP/lib), ATAC-seq (50M RP/lib)

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

Illumina Tru-Seq RNA-seq protocol

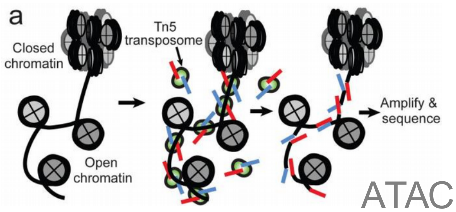


RNA-seq: transcriptome



Rao et al, Cell, 2014

Hi-C: chromosome conformation

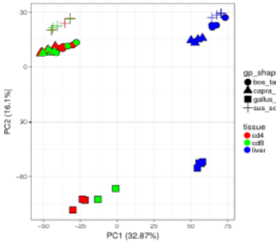


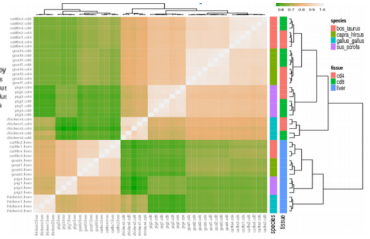
Buenrostro et al, 2013

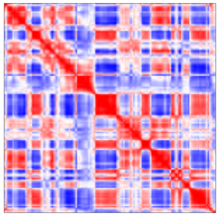
ATAC-seq: chromatin accessibility

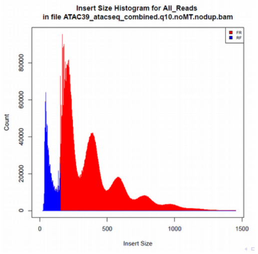
Data analysis

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









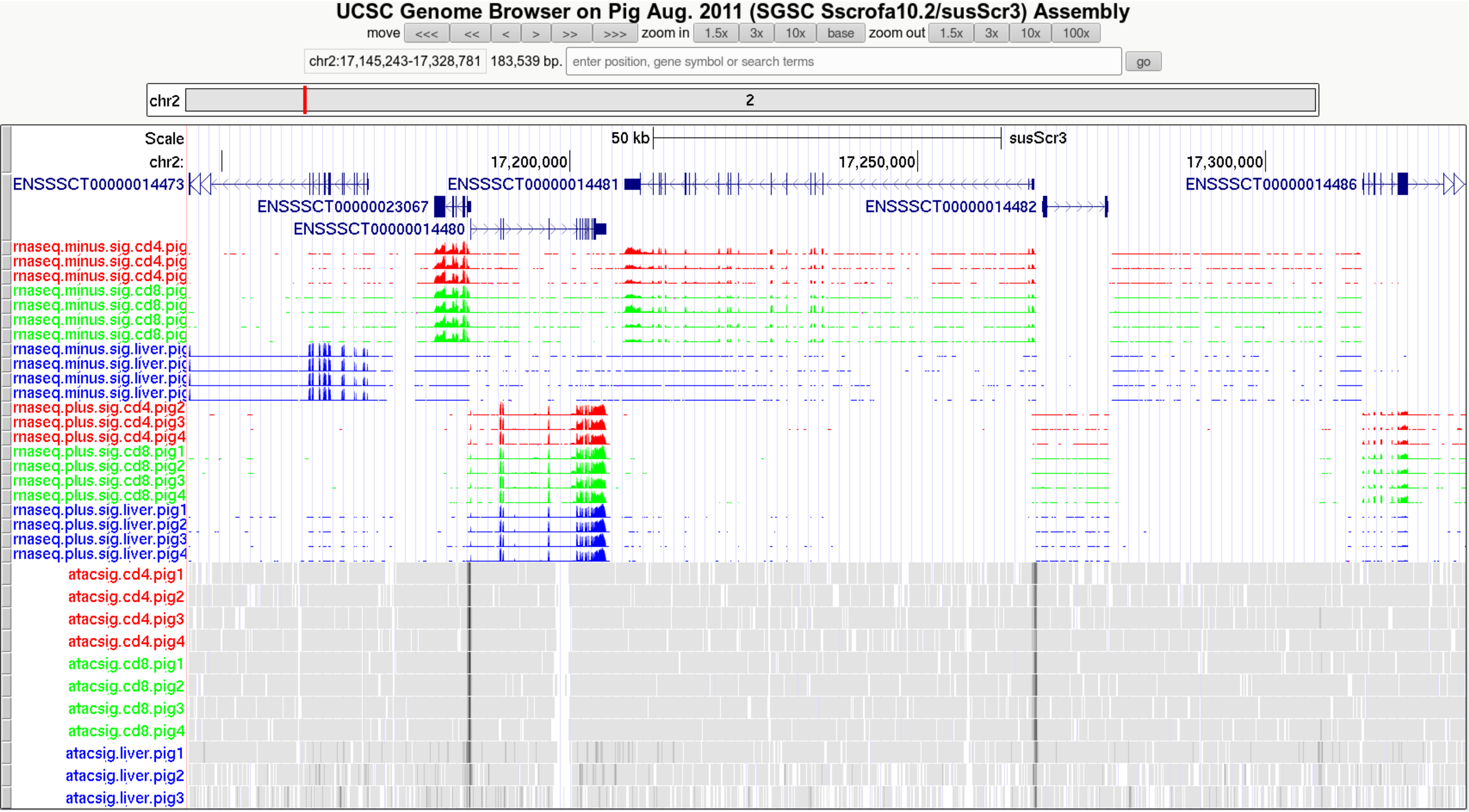
sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

From reads to functional annotation

	Average # reads per library	Mapped on Sscrofa v10.2	Type of annotated element	# of annotated elements
RNA-seq	252M	89.5%	Transcripts, genes	18,746 known RNAs 57,978 novel RNAs (including 6,581 lncRNAs)
ATAC-seq	102M	85.5%	Open chromatin regions ("peaks")	120,914 peaks
Hi-C	320M	82.6%	Topologically Associated Domains, A&B compartments	12,086 TADs 386 compartments

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Browsing the data

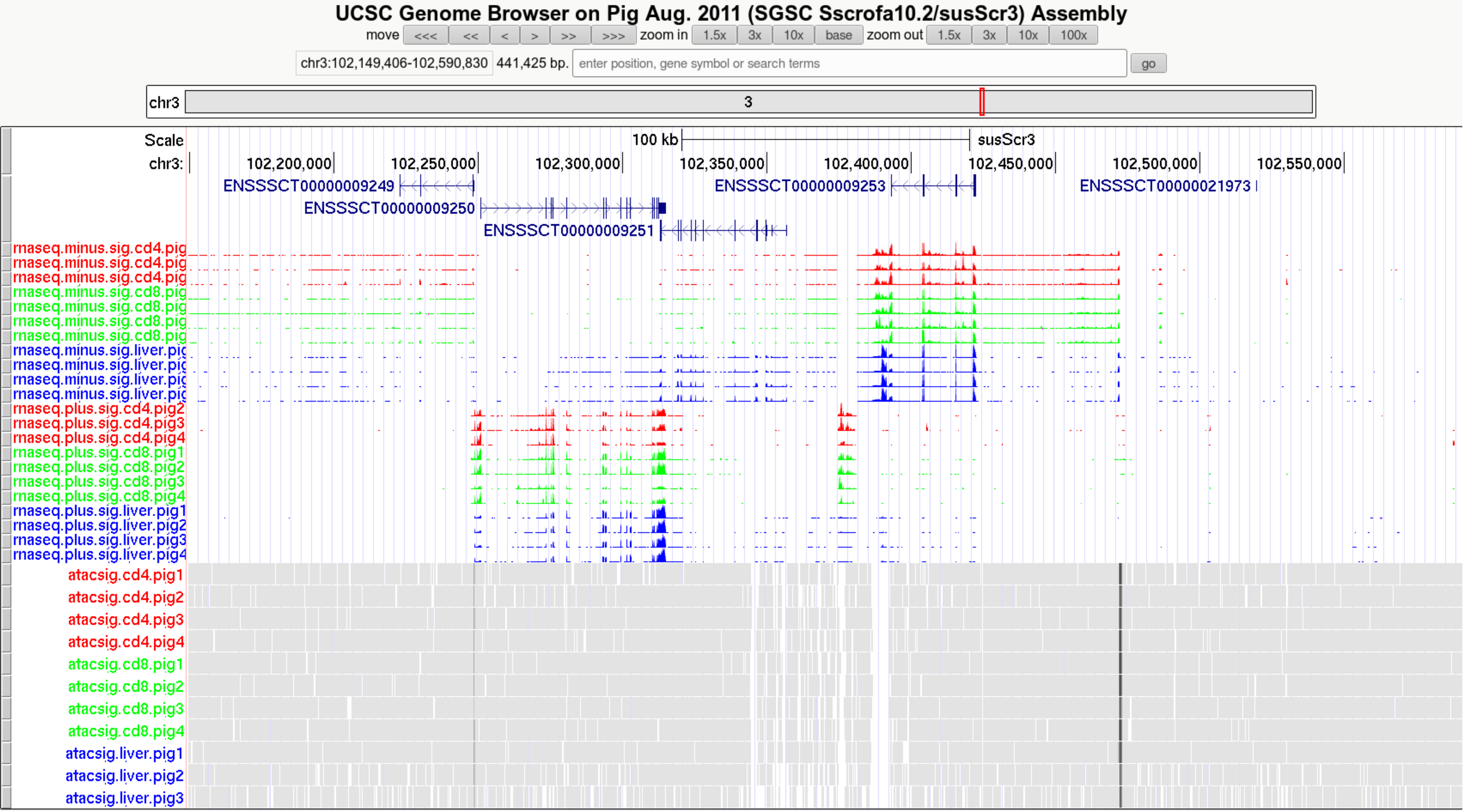


UCSC Genome Browser on Pig Aug. 2011 (SGSC Sscrofa10.2/susScr3) Assembly

chr2:17,145,243-17,328,781 183,539 bp.

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Browsing the data

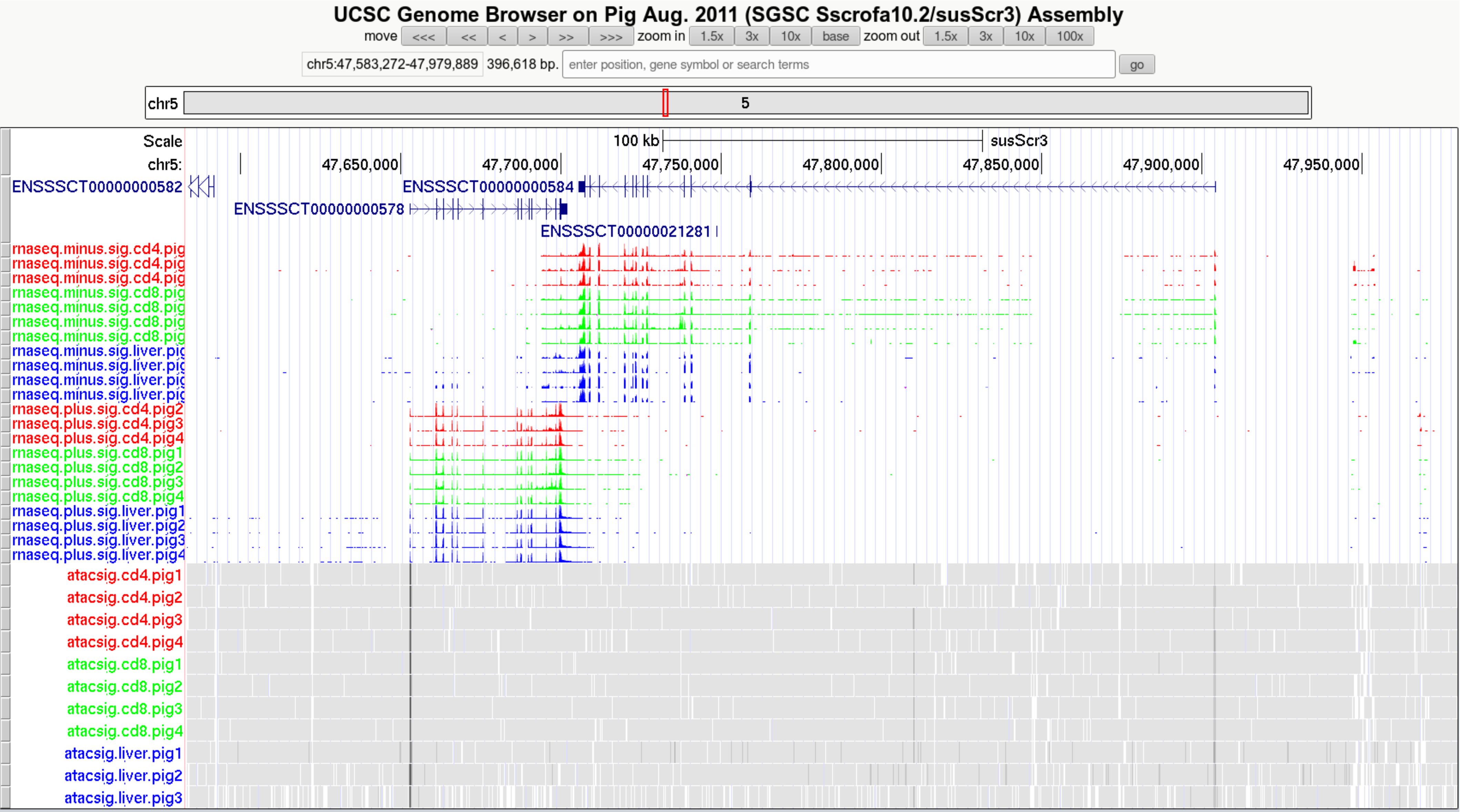


UCSC Genome Browser on Pig Aug. 2011 (SGSC Sscrofa10.2/susScr3) Assembly

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

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

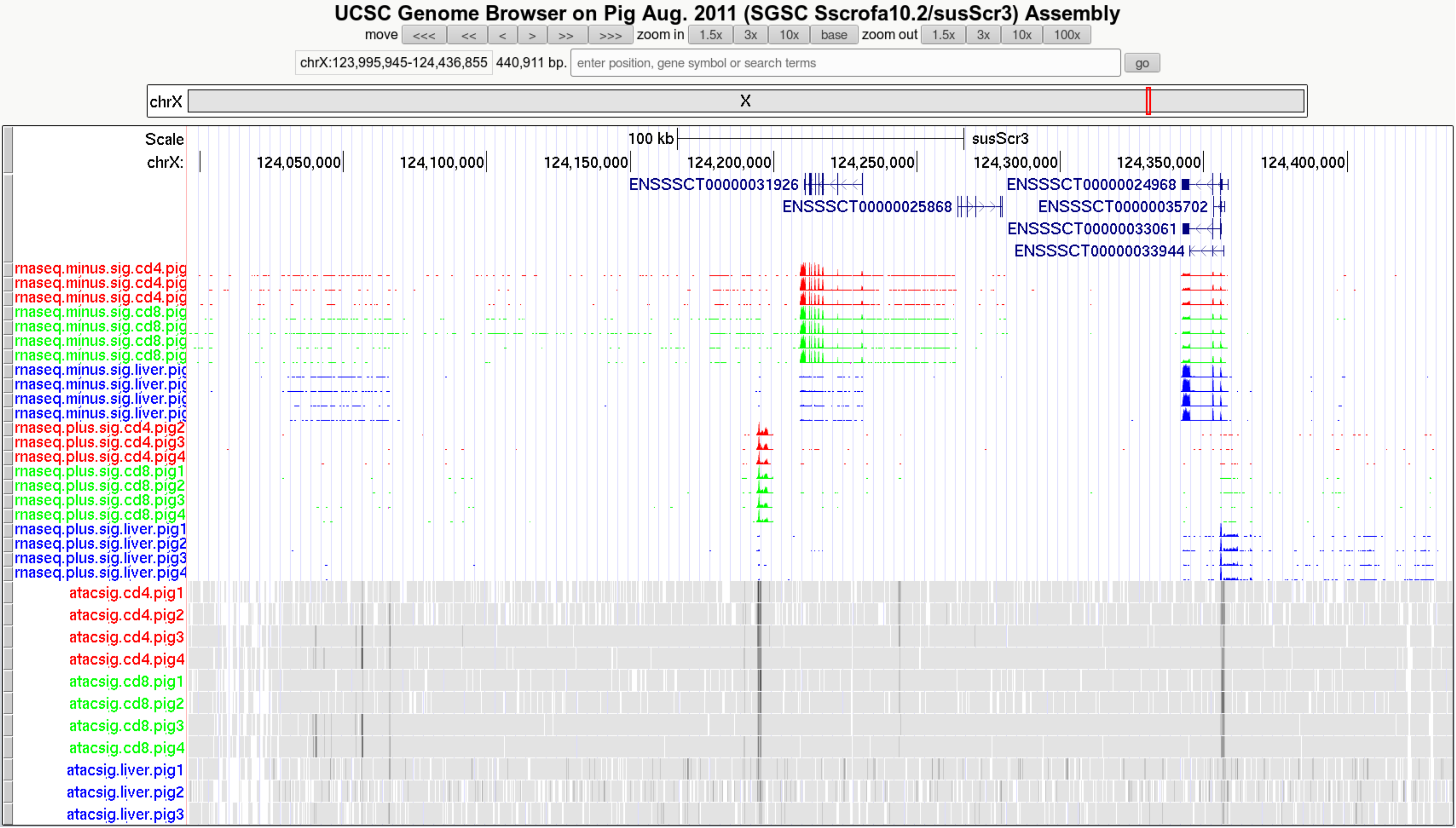
Browsing the data



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

chr5:47,583,272-47,979,889 396,618 bp.

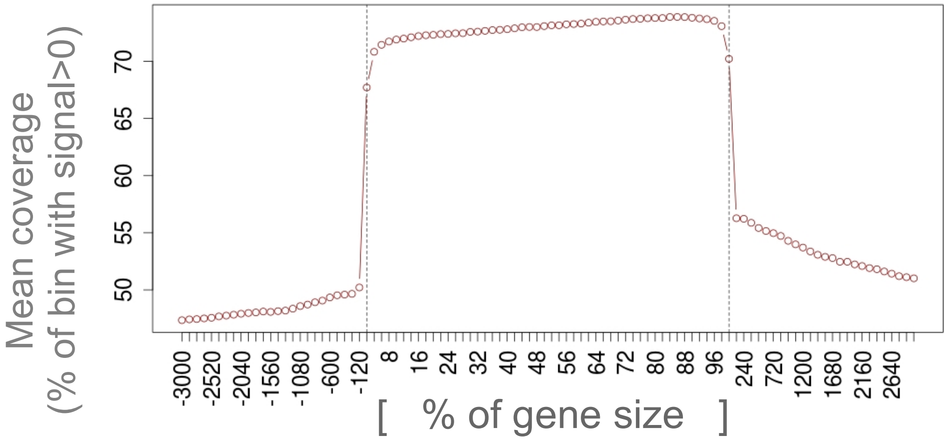
Browsing the data

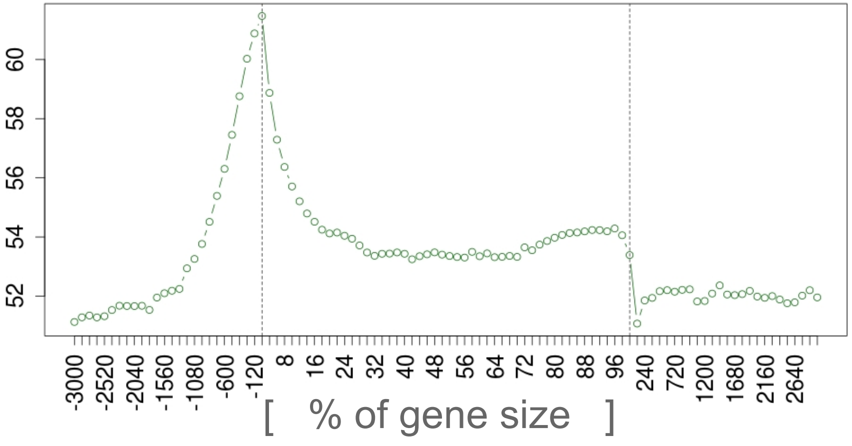


UCSC Genome Browser on Pig Aug. 2011 (SGSC Sscrofa10.2/susScr3) Assembly

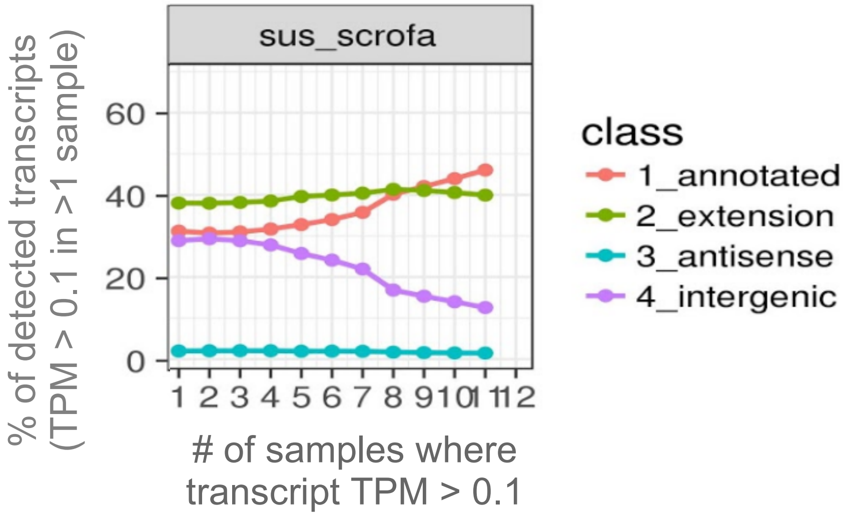
chrX:123,995,945-124,436,855 440,911 bp.

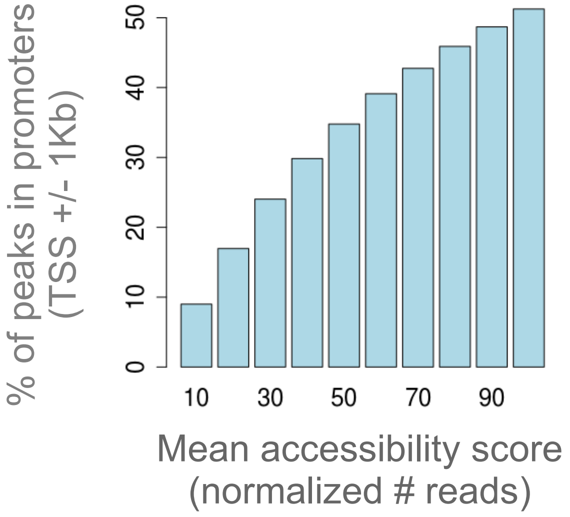
Profiling expression and accessibility


RNA-seq


ATAC-seq

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



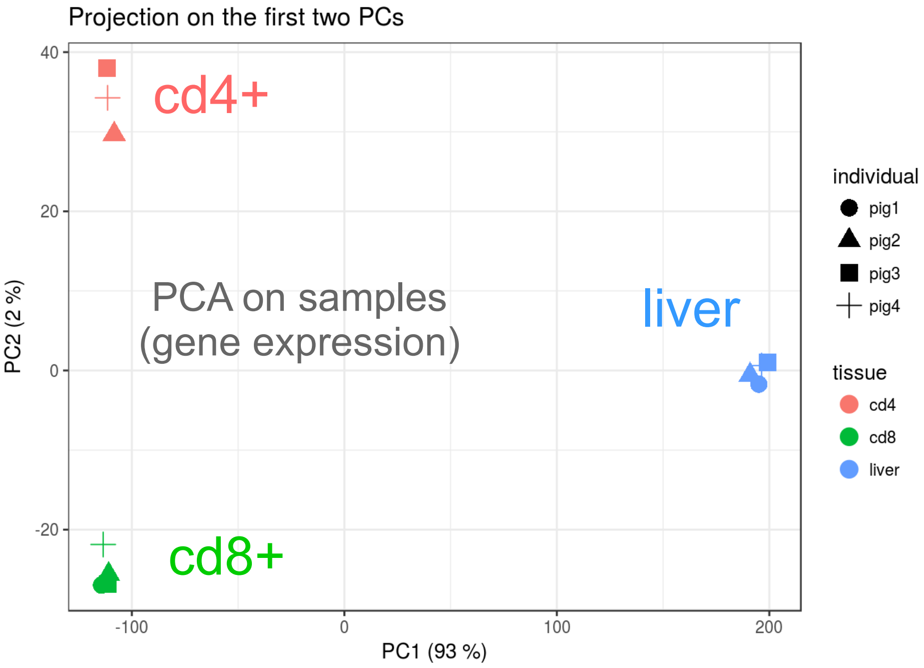


Annotated transcripts are more expressed than novel ones
ATAC-seq reads and peaks accumulate in promoters

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

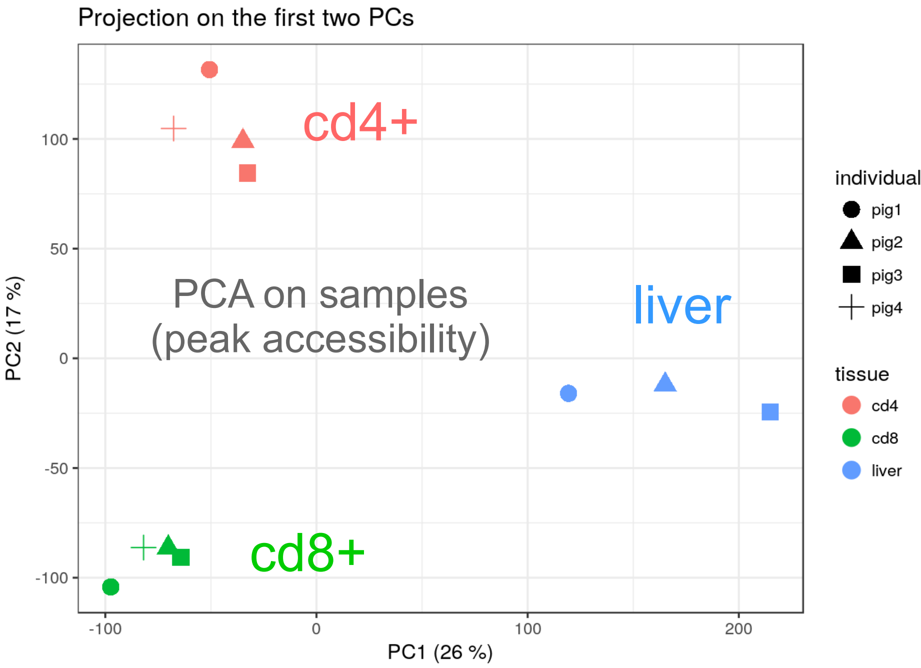
Profiling expression and accessibility

RNA-seq

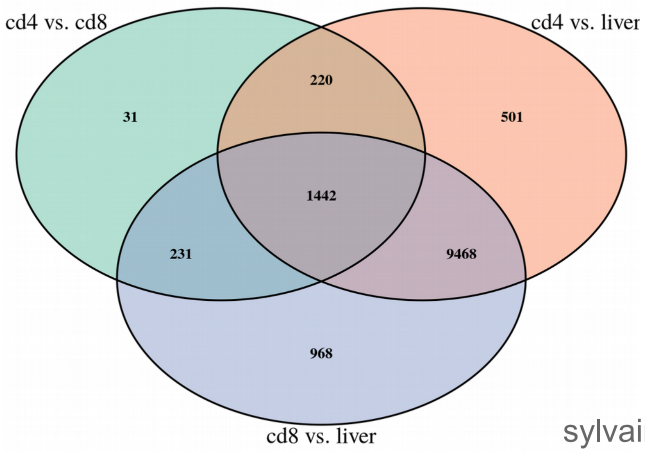


Filtering: TPM>0.1 in min 2 samples
Normalization: TMM (edgeR)
N=15,928

ATAC-seq



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



Differential gene expression

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

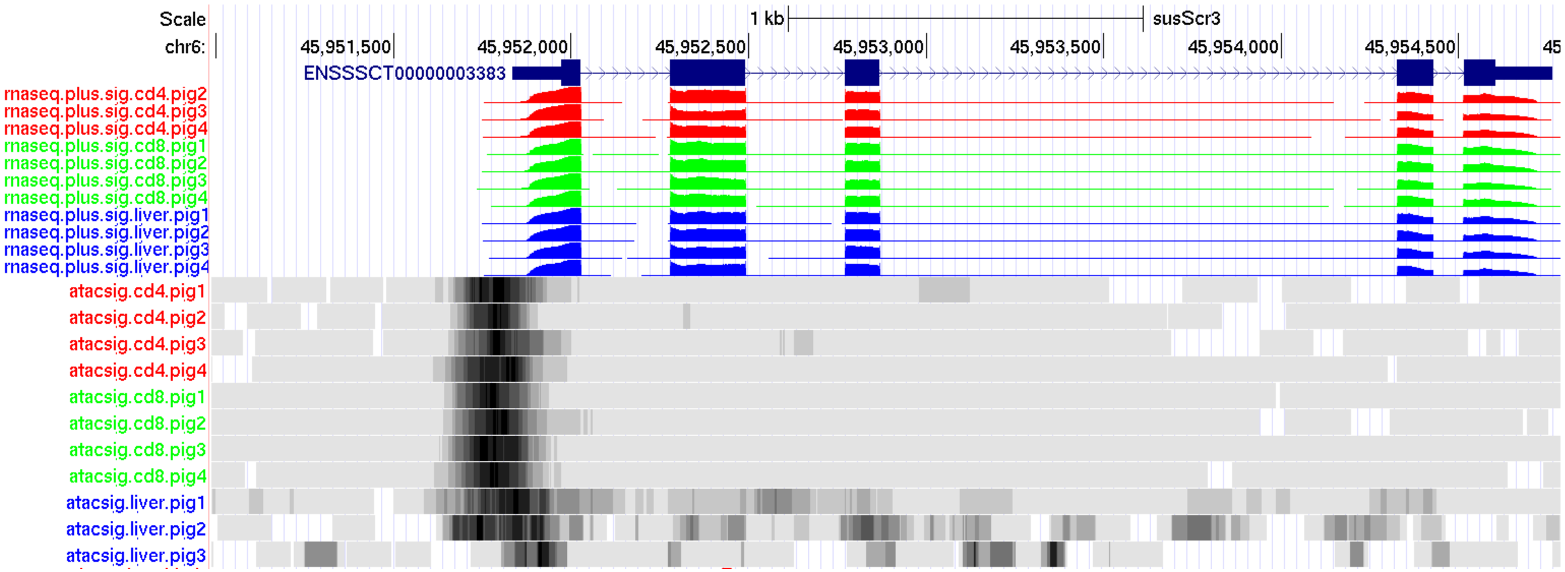
alpha−amino acid catabolic process

lipid metabolic process

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Integrating RNA-seq & ATAC-seq

Any correlation between promoter accessibility and gene expression?

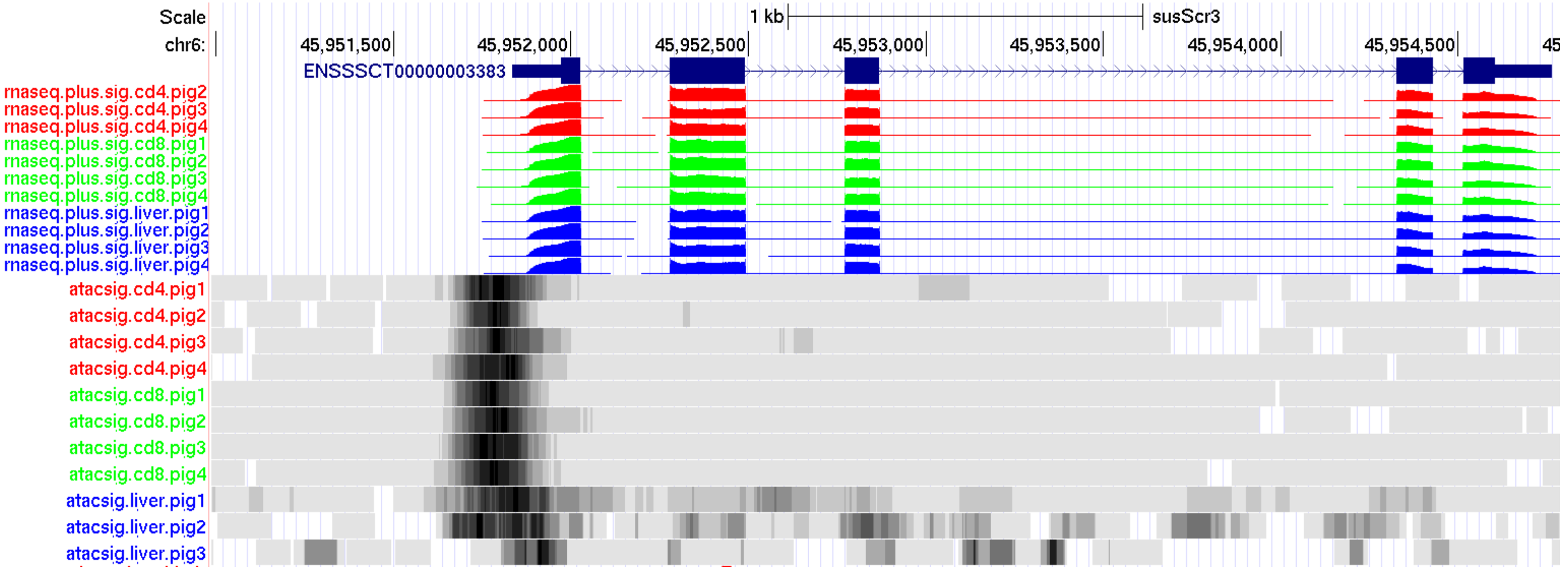


Gene A

ATAC-seq **RNA-seq**

Integrating RNA-seq & ATAC-seq

Any correlation between promoter accessibility and gene expression?



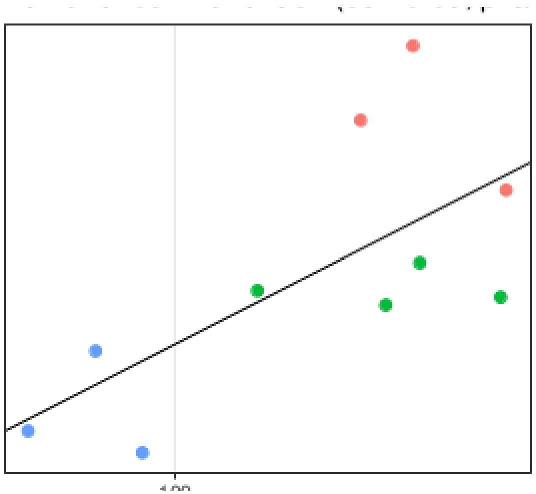
Gene A

ATAC-seq	RNA-seq
liver.pig1	liver.pig1
liver.pig2	liver.pig2
...	...
cd4.pig1	cd4.pig1
cd4.pig2	cd4.pig2
...	...
cd8.pig1	cd8.pig1
...	...

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Integrating RNA-seq & ATAC-seq

Any correlation between promoter accessibility and gene expression?



Correlation: 0.68

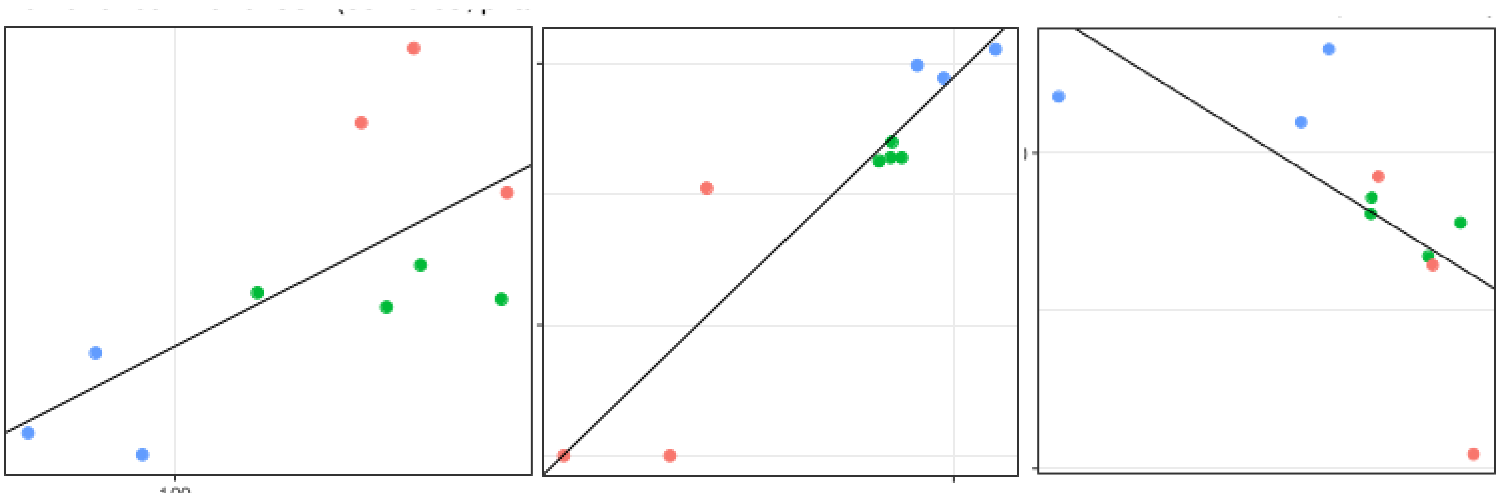
Gene A

ATAC-seq	RNA-seq
liver.pig1	liver.pig1
liver.pig2	liver.pig2
...	...
cd4.pig1	cd4.pig1
cd4.pig2	cd4.pig2
...	...
cd8.pig1	cd8.pig1
...	...

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Integrating RNA-seq & ATAC-seq

Any correlation between promoter accessibility and gene expression?



[...]

Correlation: **0.68** **0.92** **-0.66**

Gene A **Gene B** **Gene C**

ATAC-seq	RNA-seq
liver.pig1	liver.pig1
liver.pig2	liver.pig2
...	...
cd4.pig1	cd4.pig1
cd4.pig2	cd4.pig2
...	...
cd8.pig1	cd8.pig1
...	...

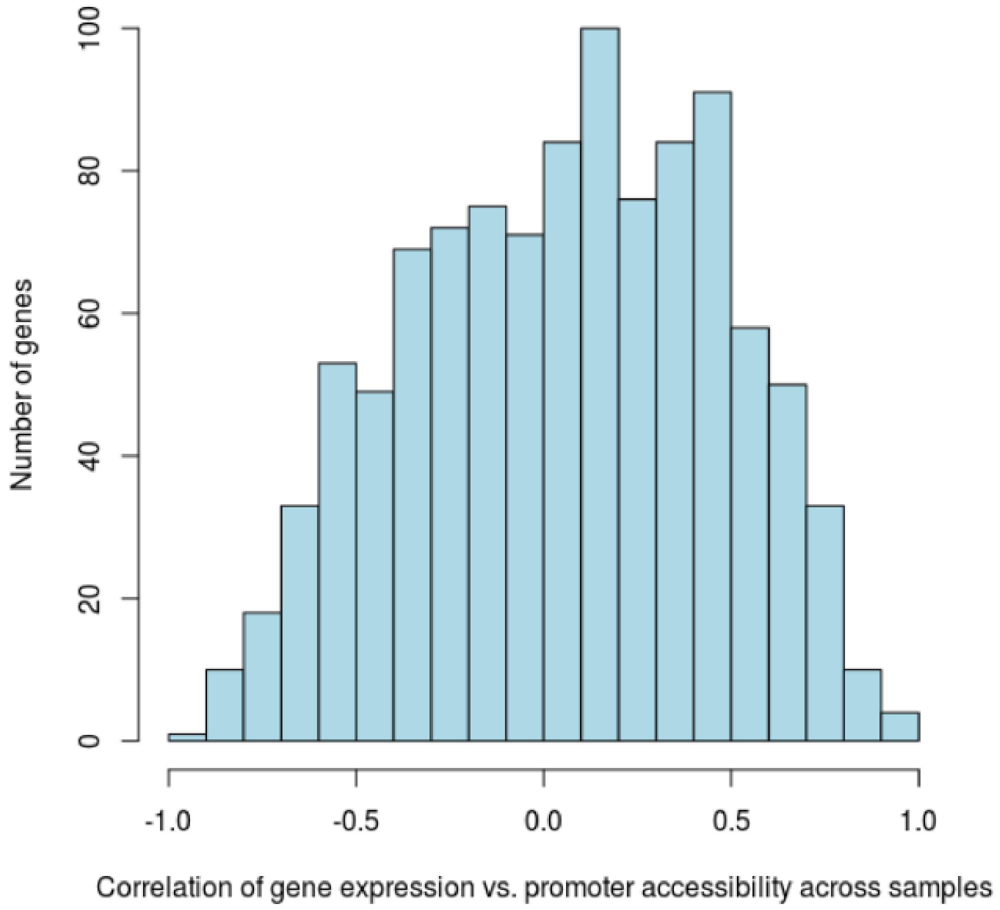
=> Distribution of correlations...

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Integrating RNA-seq & ATAC-seq

Any correlation between promoter accessibility and gene expression?

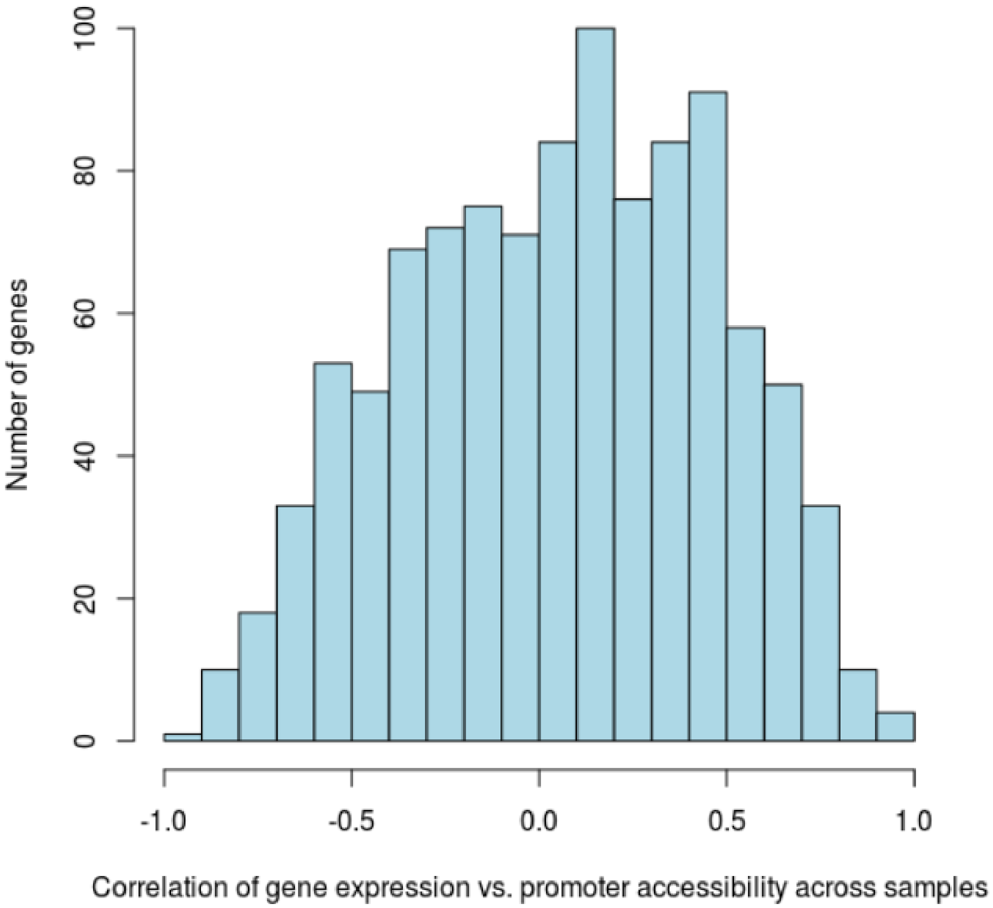
Non differentially expressed genes



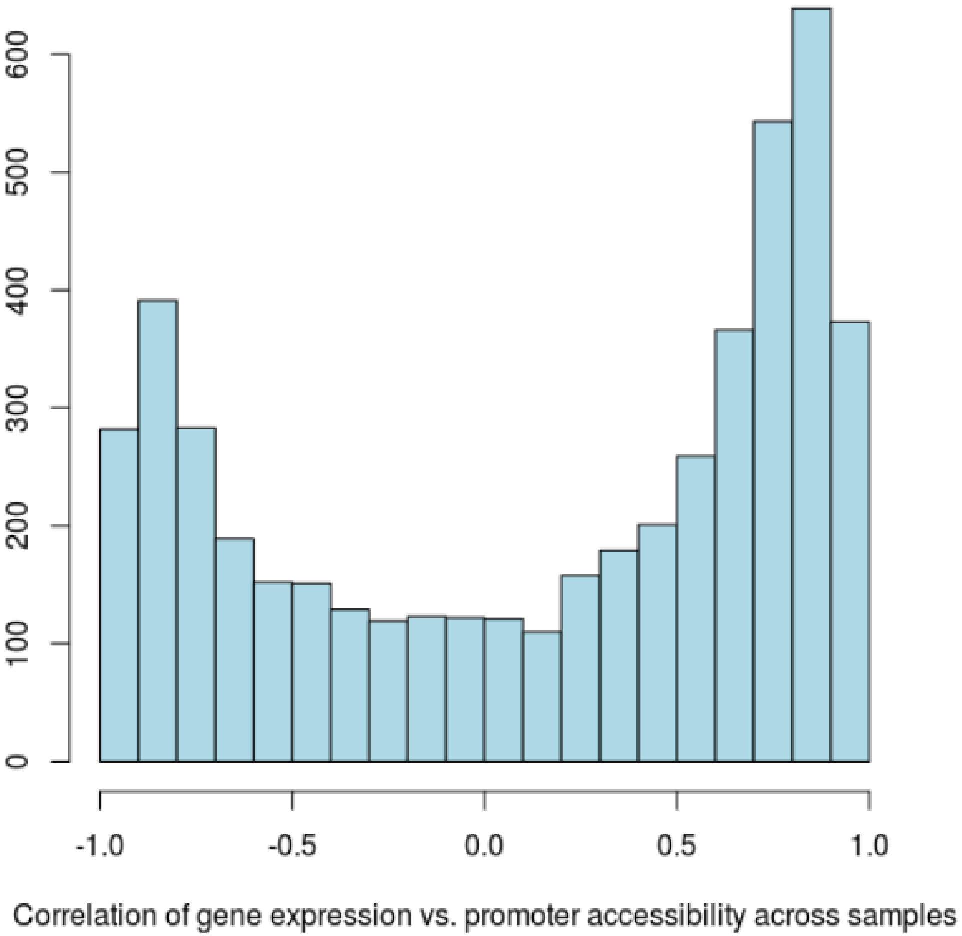
Integrating RNA-seq & ATAC-seq

Any correlation between promoter accessibility and gene expression?

Non differentially expressed genes



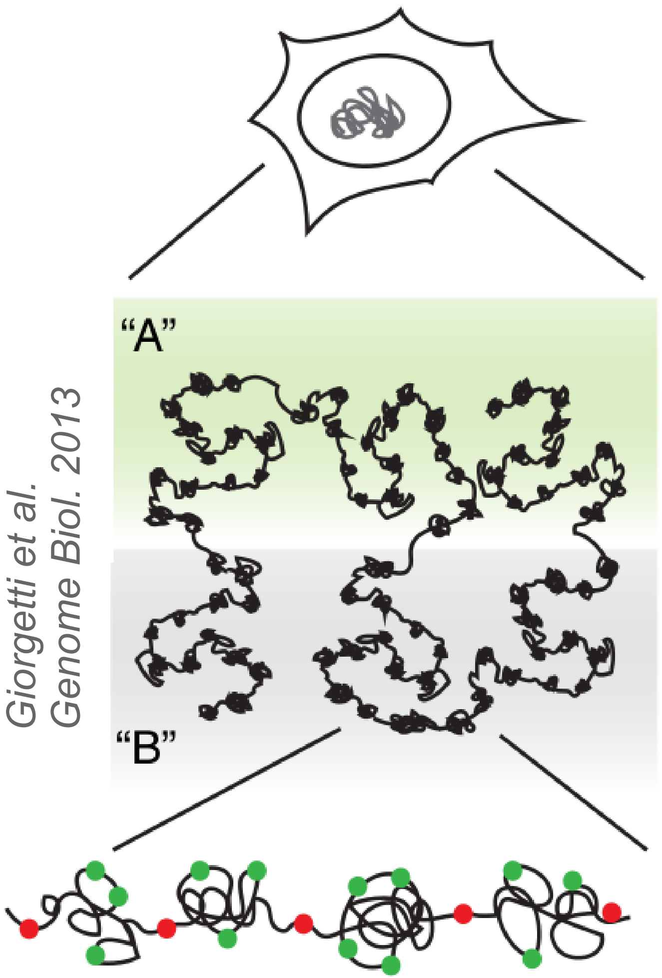
Differentially expressed genes



<u>Positive and negative regulatory mechanisms might be involved</u>

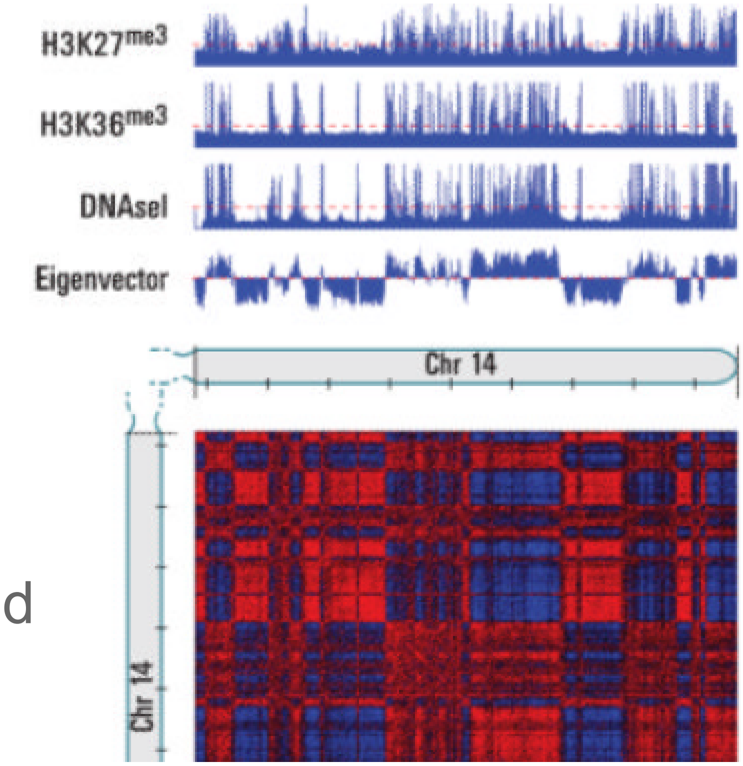
sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Profiling chromosome conformation



Giorgetti et al.
Genome Biol. 2013

3D nuclear organization & features

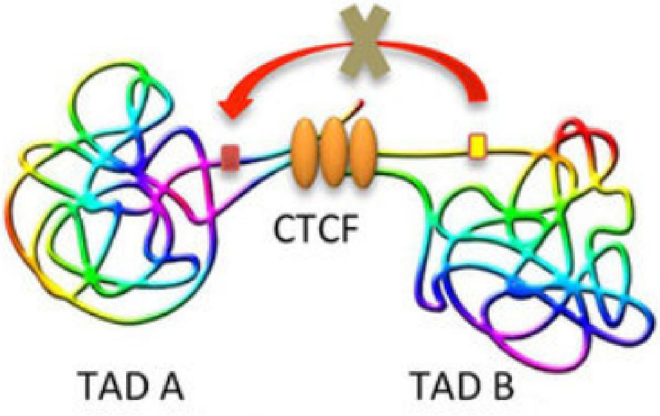


Lieberman-Aiden et al. Science, 2009

Compartments

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

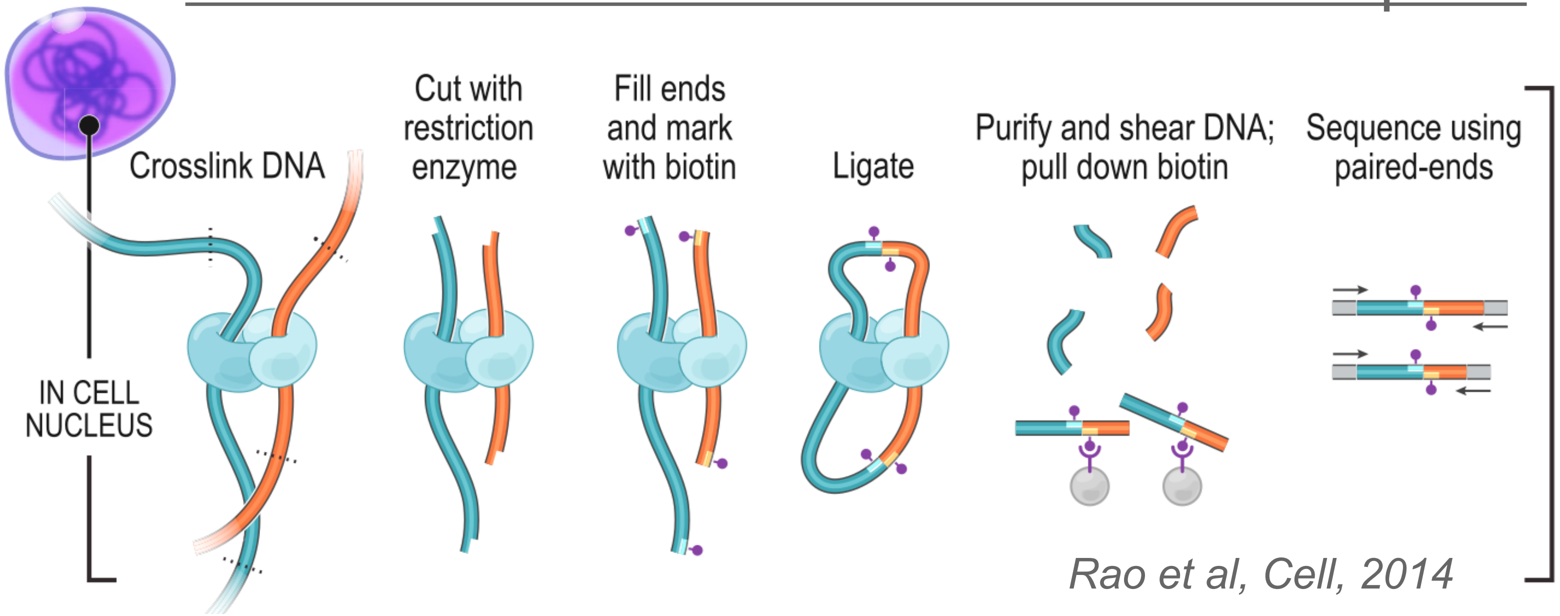
TADs (Topological Associated Domains):
- flanked by CTCF



Li et al, Sc. Reports, 2016

Profiling chromosome conformation

In situ Hi-C: chromosome conformation capture

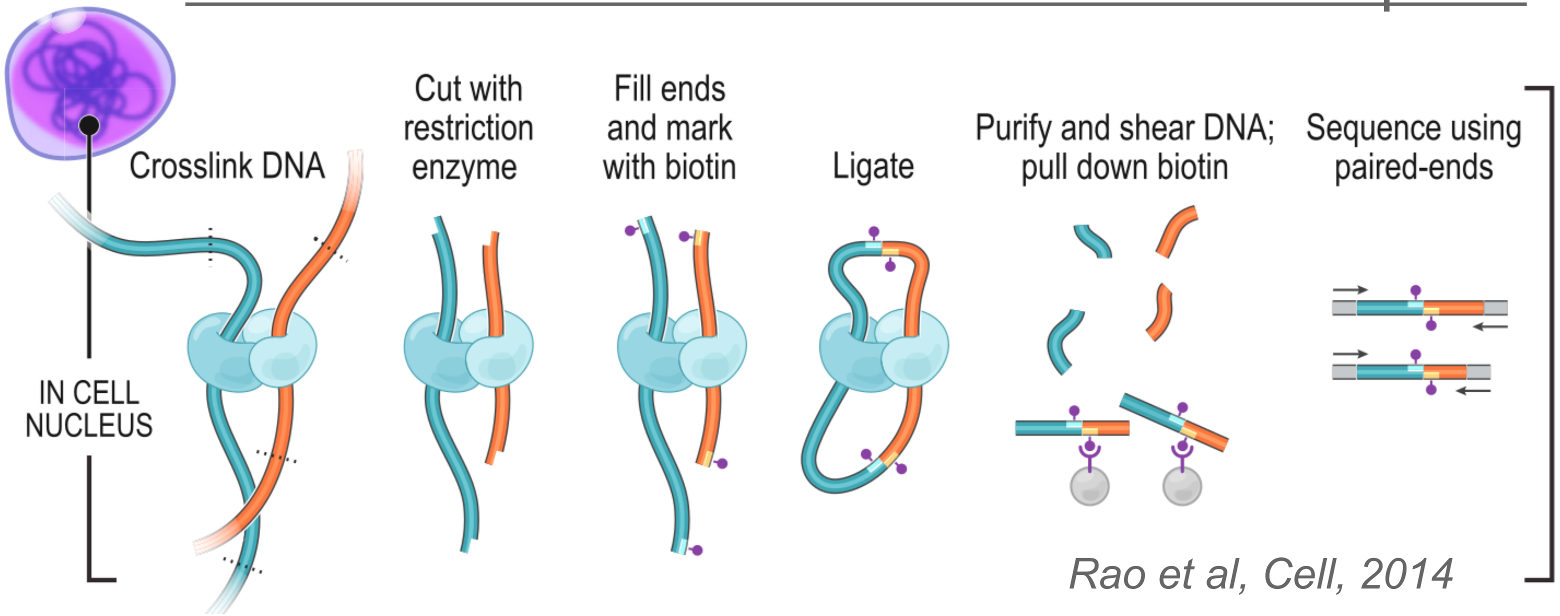


Rao et al, Cell, 2014

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Profiling chromosome conformation

In situ Hi-C: chromosome conformation capture



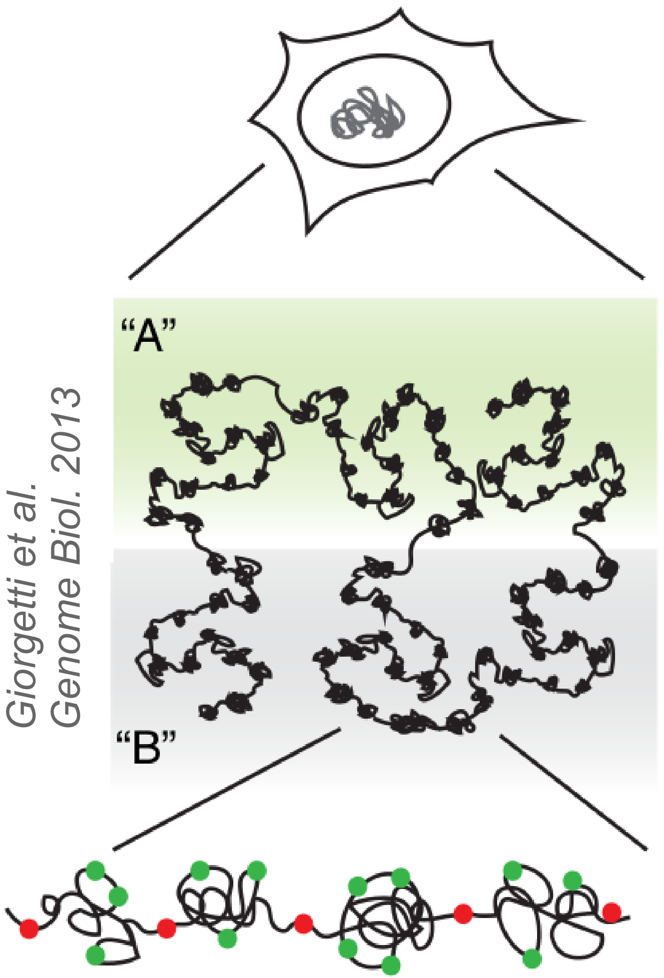
Rao et al, Cell, 2014

Data analysis

Pipeline
- Trim reads (religation site)
- Map on reference genome
- Discard inconsistent pairs
- Build contact matrix
- Normalize contact matrix
- Generate html report
- Find **TADs**
- Find **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 & 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)

Profiling chromosome conformation

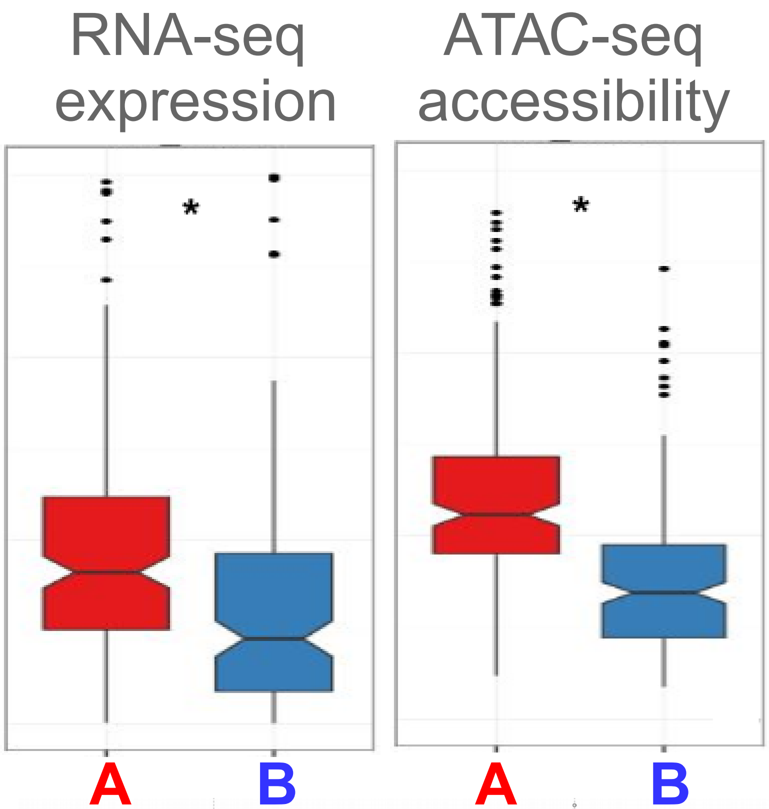


Giorgetti et al.
Genome Biol. 2013

3D nuclear organization & features

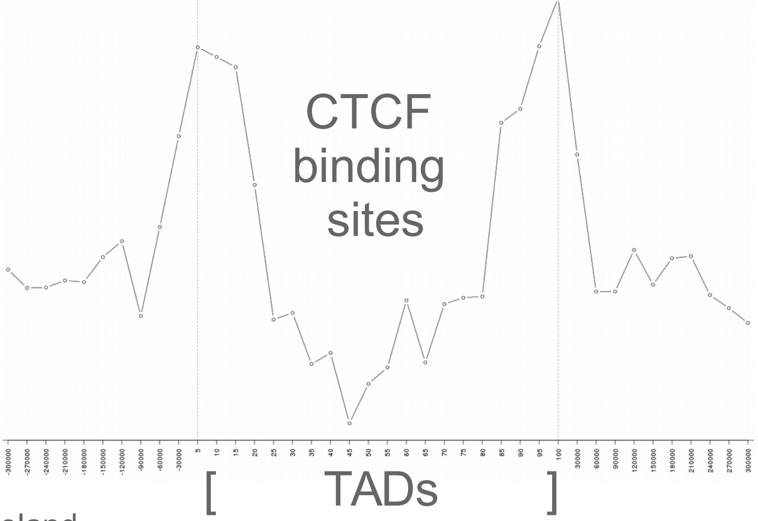
Compartments

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



TADs (Topological Associated Domains):

- flanked by CTCF



Conclusion / perspectives

Conclusion

- ✓ FR-AgENCODE: contribution to the functional annotation of livestock genomes
- ✓ Differential analyses of gene expression and chromatin accessibility in liver vs. T cells: metabolism and immunity
- ✓ Integrative analysis of RNA-seq & ATAC-seq data: positive and negative regulatory mechanisms of gene expression
- ✓ Results consistency across levels of genome organization: structure/function

On the way

- ✓ Small RNA-seq (<200nt): miRNAs, snoRNAs, tRFs...
- ✓ Deeper ATAC-seq
- ✓ Ugrade from v10.2 to v11

stay tuned...

sylvain.foissac@inra.fr - ISAG 2017, Dublin, Ireland.

Acknowledgements

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.