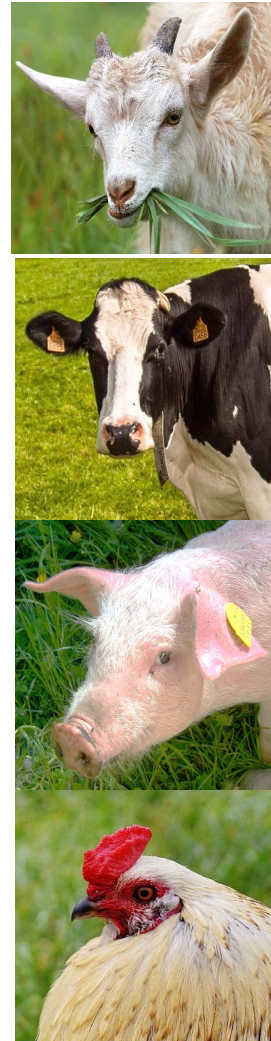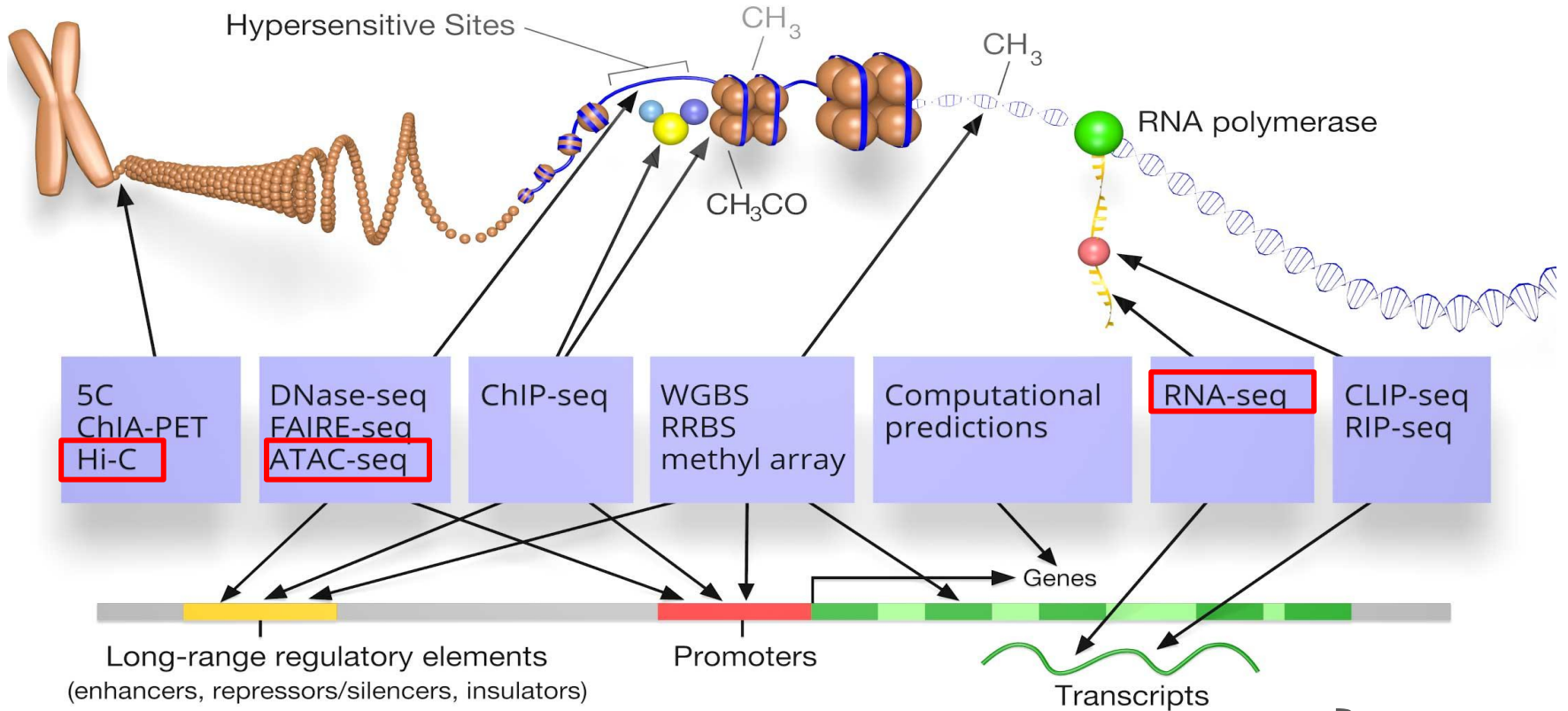# An update on the FAANG pilot project
FR-AgENCODE

S. Foissac, S. Djebali, **Andrea Rau**, S. Lagarrigue,
H. Acloque, E. Giuffra, and the FR-AgENCODE group

INRA - {Jouy en Josas, Toulouse, Rennes} - France

FR-AgENCODE data

Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

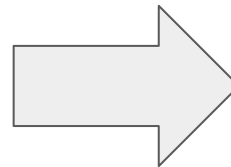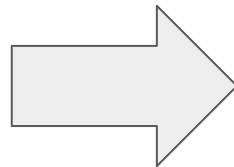# Overview of FR-AgENCODE analyses

Multi-species single data analyses:

- Expression (**RNA-seq**)
- Open chromatin (**ATAC-seq**)
- 3D chromatin conformation (**Hi-C**)

$\Rightarrow$ Quantification, normalization, differential analysis, clustering, ...

Multi-species integrative analyses:

- Expression + open chromatin + 3D chromatin conformation
- Expression + open chromatin

RNA-seq identifies many novel transcripts

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 90	**26,740**	16,100	**60.2**	**65,539**	**8,296**
Capra hircus	CHIR_ARS 1 / NCBI	**53,266**	34,442	**64.7**	**38,197**	**1,224**
Gallus gallus	GalGal 5 / Ensembl 90	**38,118**	22,180	**58.2**	**34,852**	**3,504**
Sus scrofa	SScrofa 11.1 / Ensembl 90	**49,448**	29,786	**60.2**	**39,032**	**4,778**

* with TPM ≥ 0.1 in ≥ 2 samples / ** with at least 2 exons and classified by FEELnc

Directed, 2 x 150bp, 100 million read pairs / sample
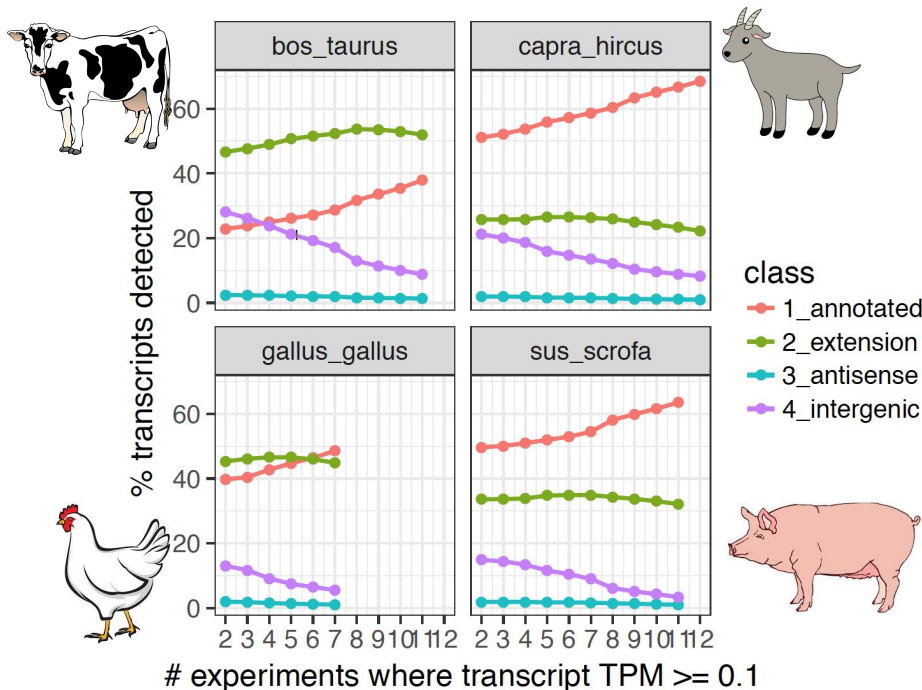
Mapped reads:
- Exonic: 50-80%
- Intronic: 10-20%
- Intergenic: 20% for cattle/chicken, 10% for goat/pig

4

Differentially expressed genes reflect underlying biology

Two per-gene (TPM > 0.1 in at least 2 samples) differential models:
- **Tissue effect**, blocking on individuals:

  `expression ~ individual + tissue`

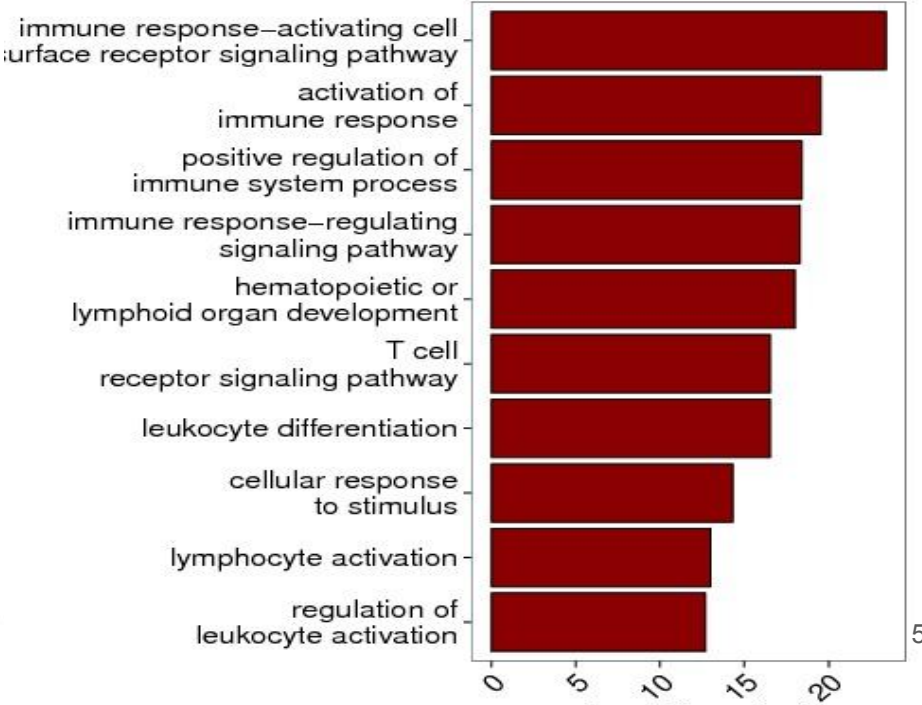- **Tissue & tissue-specific sex effects**, blocking on (nested) individuals:

  `expression ~ sex + tissue:sex + individual:sex`

- Generally large differences for liver vs CD4/CD8, smaller differences for CD4 vs CD8 or tissue-specific ♂ vs ♀
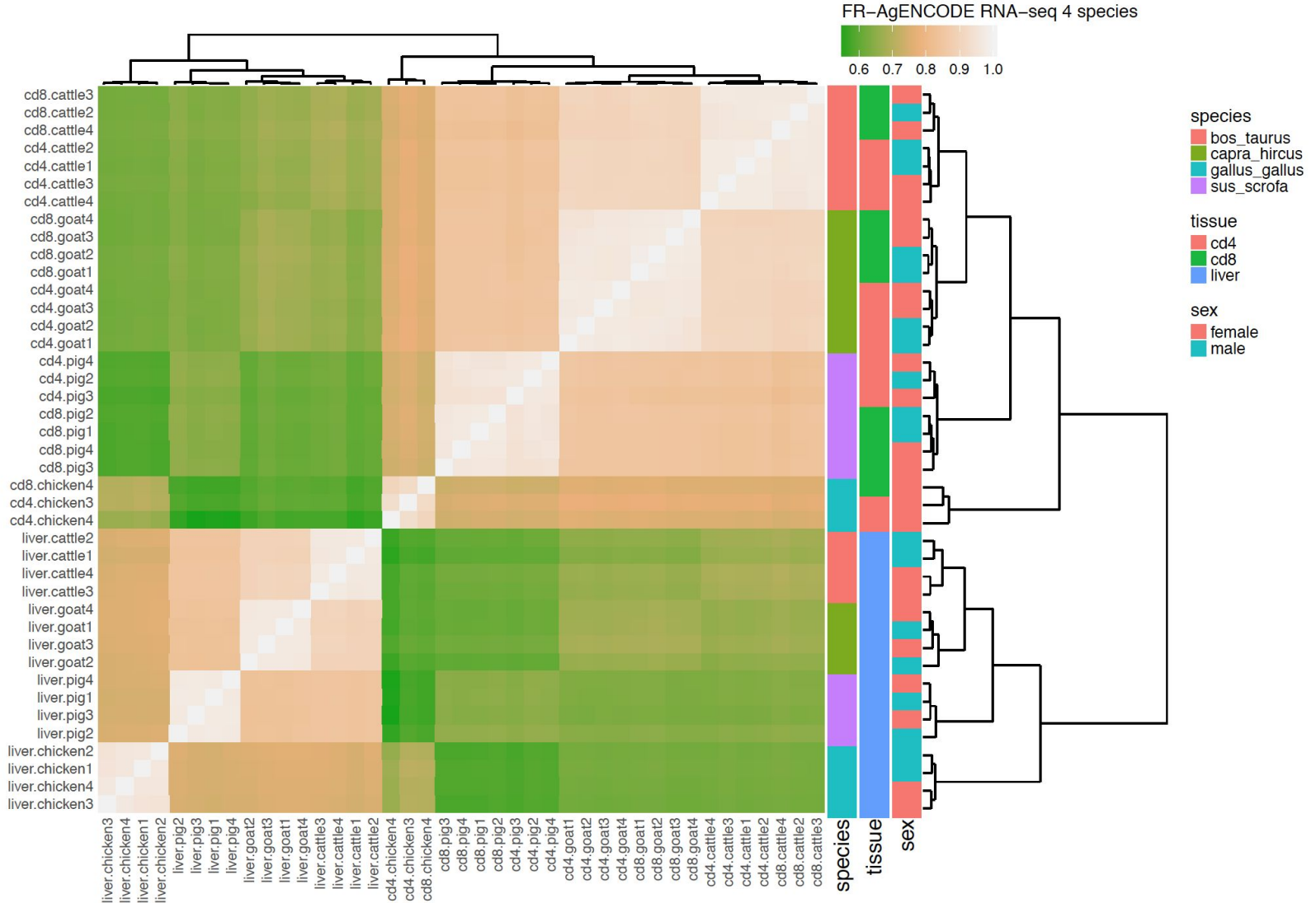- GO (BP) enrichment for genes consistently DE across species:

Over-expressed in **liver**
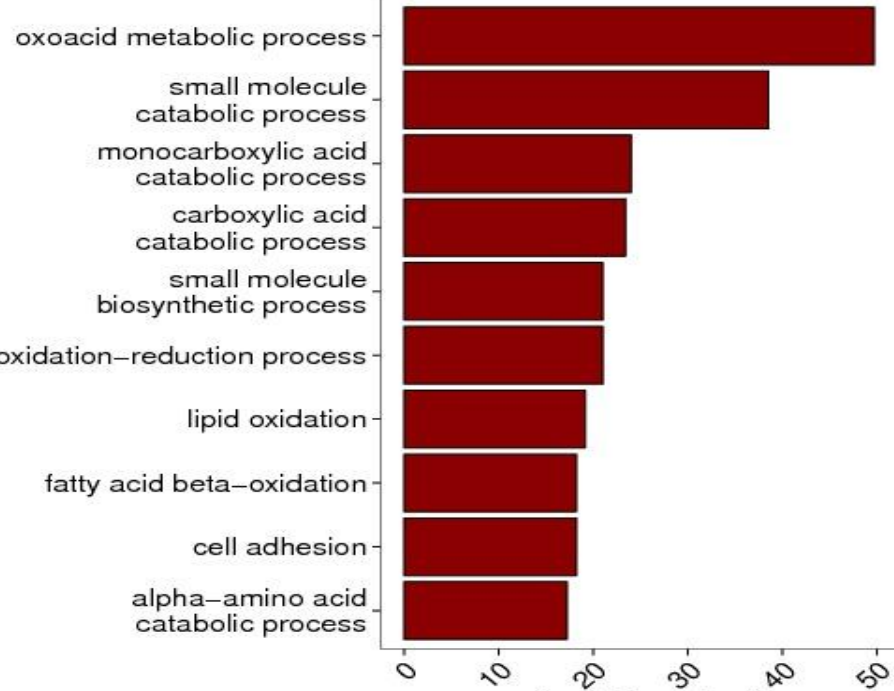
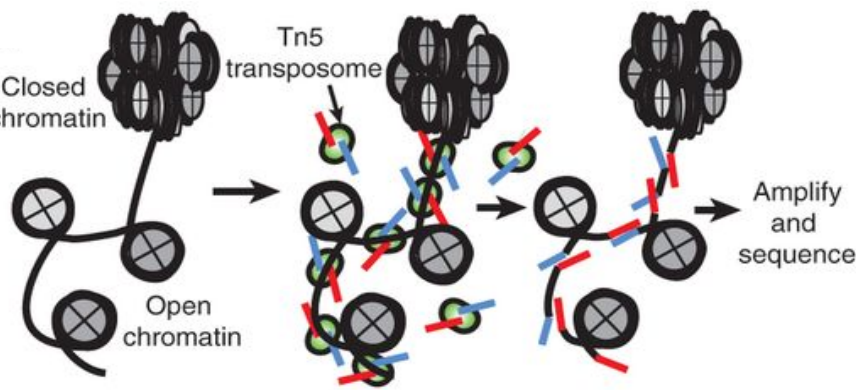Over-expressed in **immune cells**

Cross-species & cross-tissue transcriptome clustering

- Liver clearly separated from CD4/CD8 cells, and species cluster within cell types (liver vs T-cells)

# ATAC-seq identifies pertinent regions of open chromatin

[*Buenrostro et al., Nature Methods, 2013*]

| Species	Number of ATAC-seq peaks	Genome size (bp)	ATAC-seq peak coverage	
			# bp	% of genome
Bos taurus	104,986	2,670,422,299	80,562,624	3.02
Capra hircus	74,806	2,922,813,246	57,043,999	1.95
Gallus gallus	119,894	1,230,258,557	51,000,066	4.15
Sus scrofa	149,334	2,501,912,388	106,645,814	4.26
* 50 million read pairs / sample				

- 3 major peak classes: **promoter regions** (TSS +/- 1kb), **intronic**, **intergenic**
- Significant proportion of peaks in ATAC-seq reads found in promoter region (36-66%)
- For all species, peaks close to TSS (+/- 5kb) are most ubiquitous open chromatin
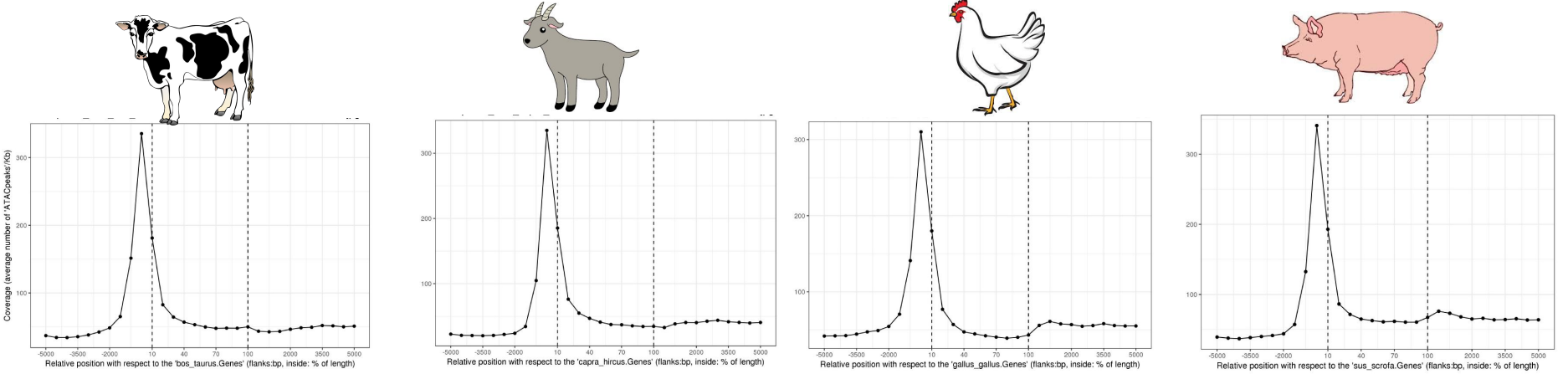
ATAC-seq sample clustering:
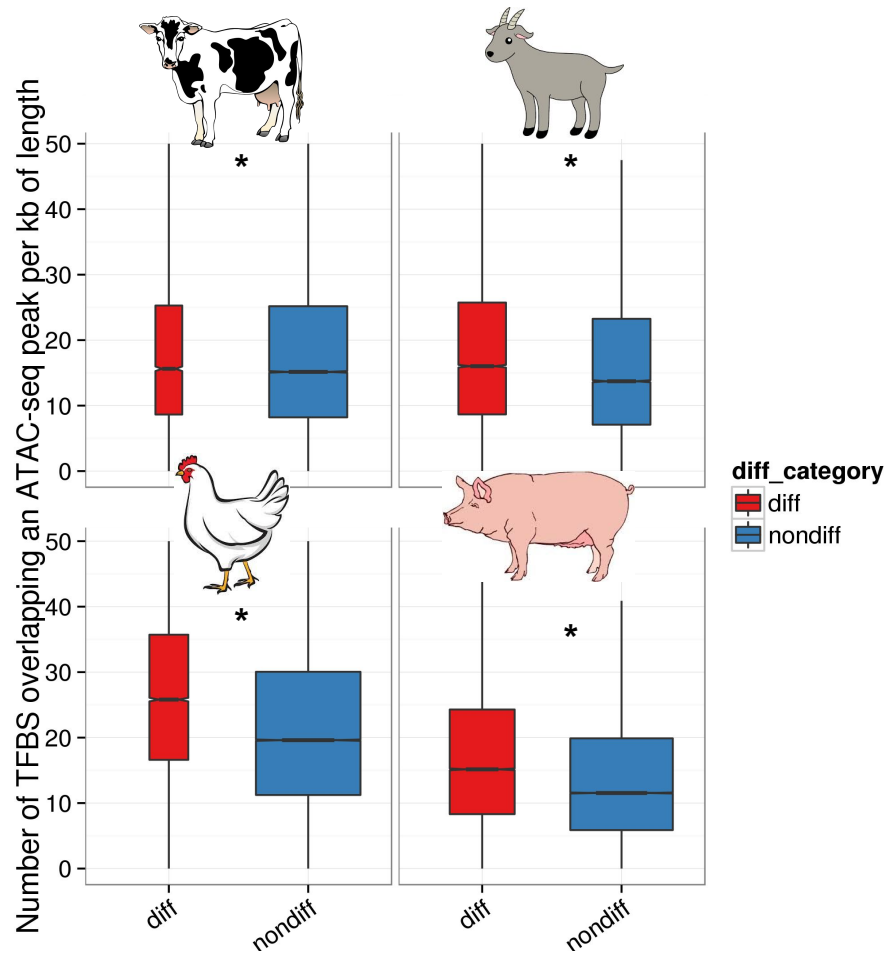## Liver vs immune cells, ♂ vs ♀ in immune cells
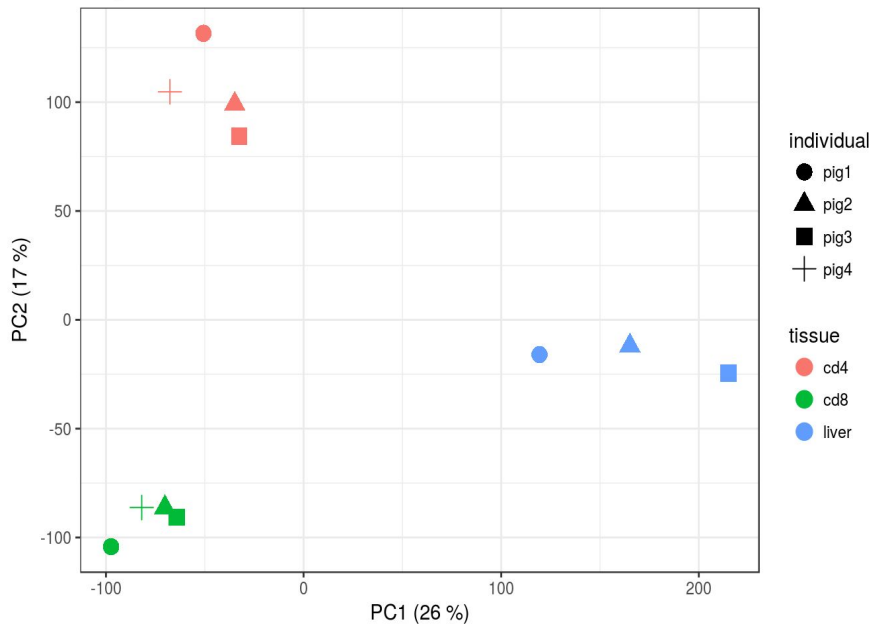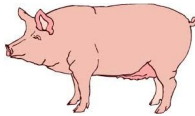
- PC1: liver versus CD4/CD8
- PC2 : CD4 vs CD8
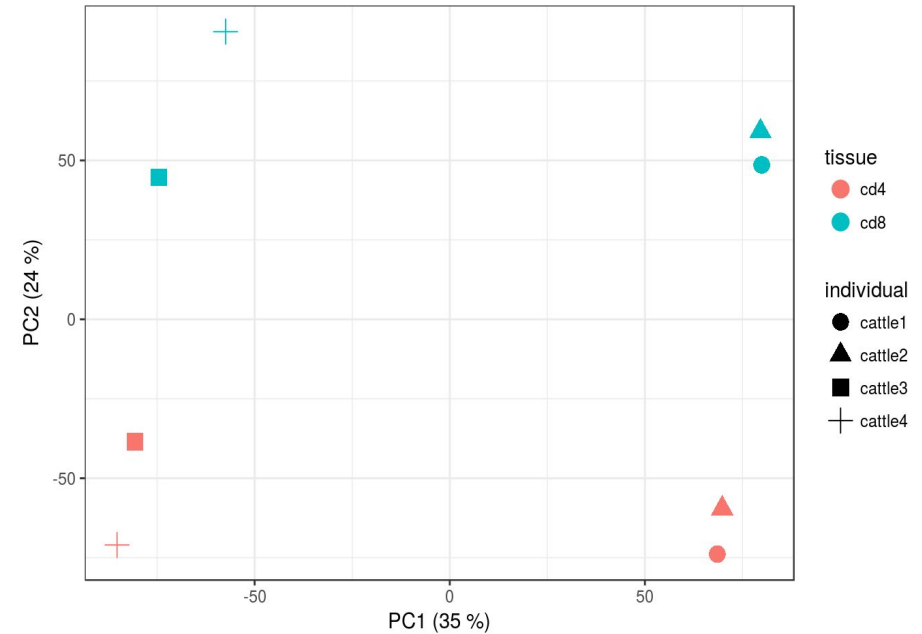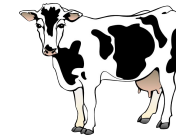
- PC1: males vs females
(note: absence of liver samples)
- PC2: CD4 vs CD8

8

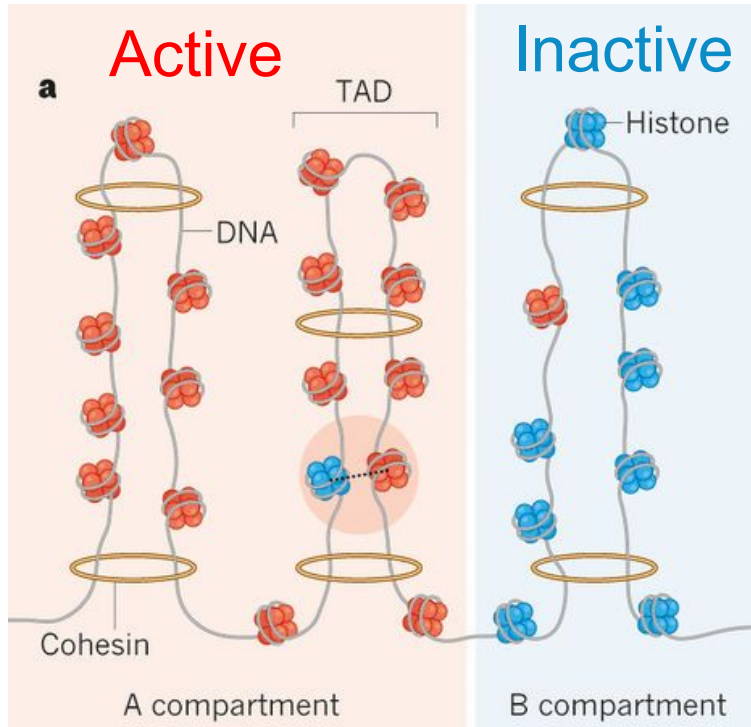# Differential ATAC-seq peaks are more likely to be regulatory

Between-tissue **differential** ATAC-seq peaks have a higher **TFBS density** than nondifferential peaks
(Wilcoxon test, p-value $< 10^{-15}$)

→ Differential ATAC-seq peaks are more likely to have a **regulatory** role
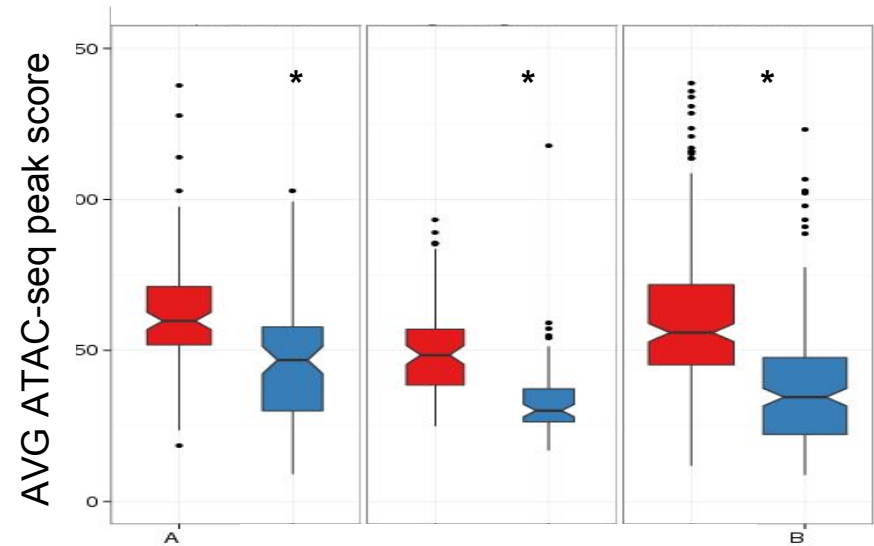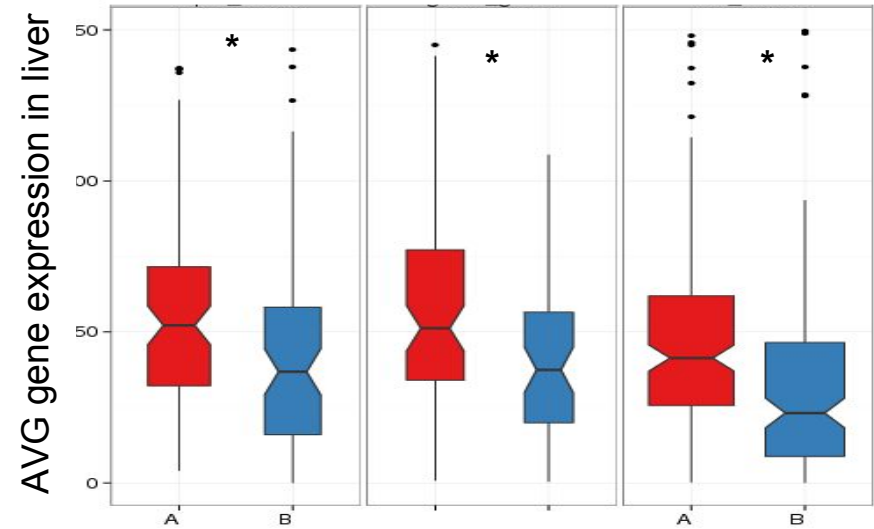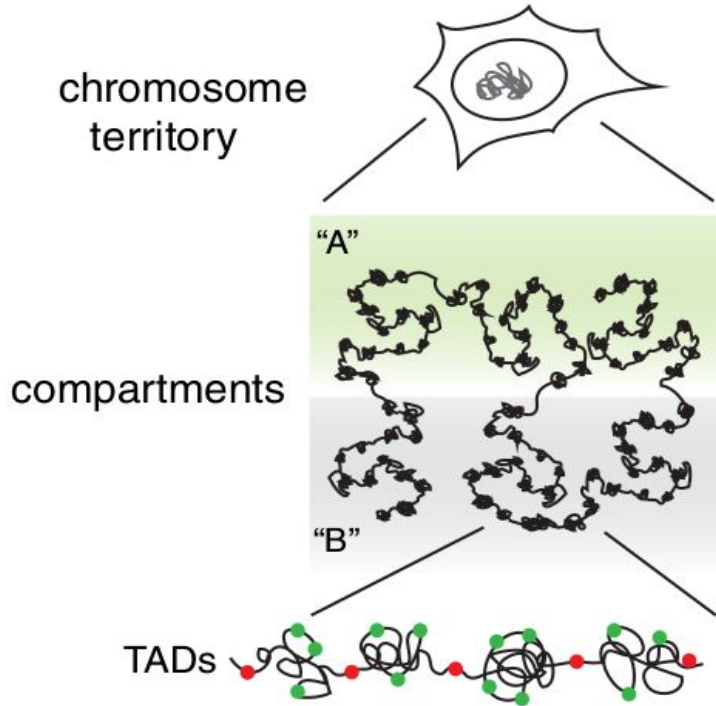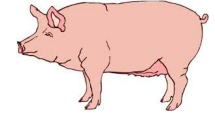
Hi-C for 3D genomic structures

Active

Inactive

Several levels for 3D conformation:
- TADs (Toplogically associating domains)
- Compartments
- Enhancer-promoter interactions

- Liver x 4 animals x 3 species (all but cattle)
- 180 million read pairs / sample

- Predicted CTCF binding sites peak at TAD boundaries ✓

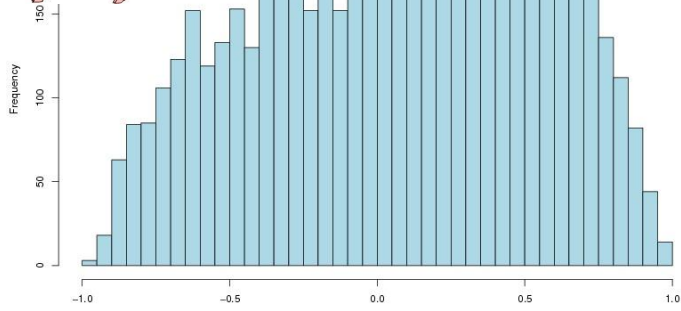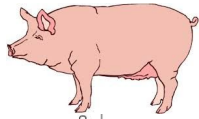AVG coverage
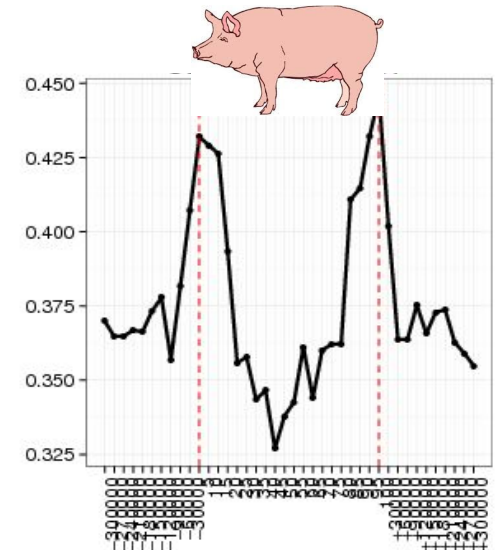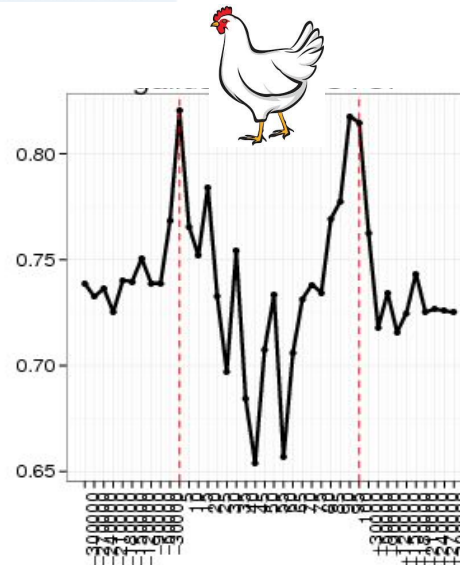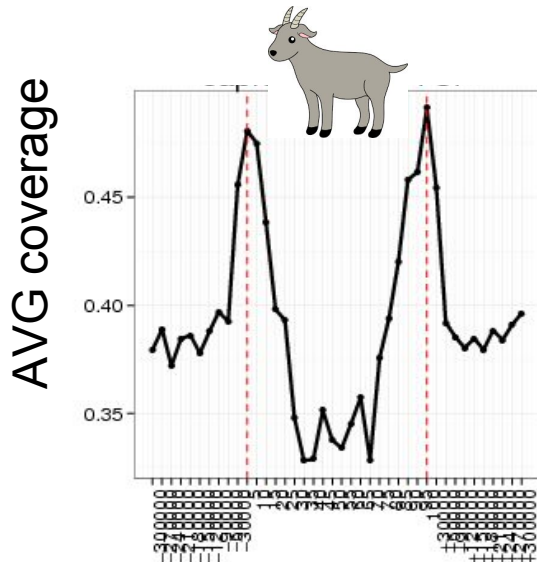
RNA-seq + ATAC-seq + Hi-C consistency

A compartments: open, expressed
B compartments: closed, repressed
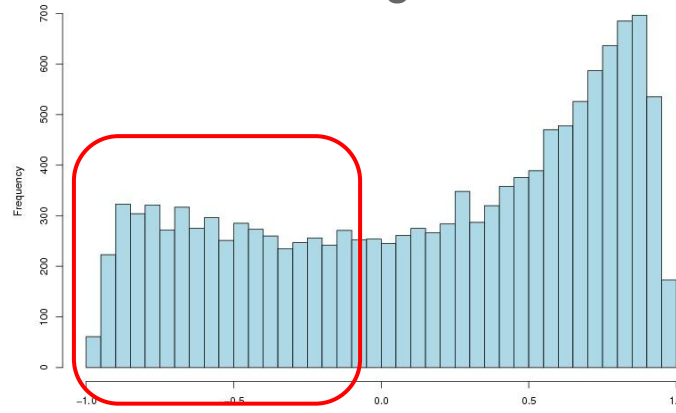
Global consistency between expression, chromatin accessibility, and chromatin conformation!

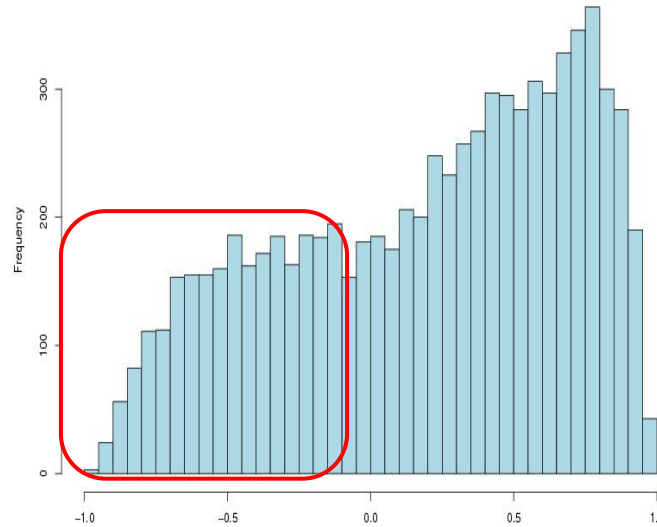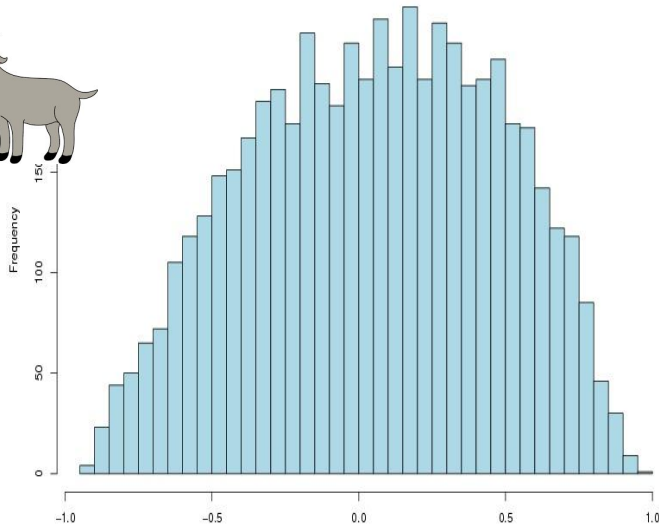# Exploring correlation between promoter accessibility and gene expression

Non-DE genes | DE genes

Positive AND negative regulatory mechanisms?

- Pattern unique to proximal promoter region (TSS +/- 1kb)
- Promoter peaks negatively correlated with gene expression tend to be wider

Next steps & future work

- More **integrative** analyses:
  - Enhancer / gene relationships
  - Small RNAs vs long RNAs / open chromatin / HiC TADs
- More **comparative** analyses:
  - Evolution of functional elements, in particular regulatory
  - Human/mouse element projection to livestock species

- bioRxiv preprint to be submitted in coming weeks
- More tissues, functional validation and link to genotype/phenotype (G/P) data:
  - **H2020 project proposal** (co-coordinated by E. Giuffra & H. Acloque)
  - ...

Fr-AgENCODE group

- **Management:**
  - Elisabetta Giuffra ⎤
  - Sylvain Foissac ⎦ Coordinators
  - Sandrine Lagarrigue
  - Marie-Hélène Pinard

- **Sampling and assays:**
  - Hervé Acloque
  - Cécile Berri
  - Fany Blanc
  - Françoise Drouet
  - Diane Esquerre
  - Stéphane Fabre
  - Joel Gautron
  - Adeline Goubil
  - Fabrice Laurent
  - Florence Mompart
  - Sophie Pollet
  - Pascale Queré
  - Michèle Tixier-Boichard
  - Gwenola Tosser-Klopp
  - Silvia Vincent-Nailleau

- **Data analysis:**
  - Philippe Bardou
  - Cédric Cabau
  - Thomas Derrien
  - Sarah Djebali-Quelen
  - Christine Gaspin
  - Ignacio Gonzalez
  - Christophe Klopp
  - Sylvain Marthey
  - Maria Marti
  - Kylie Munyard
  - Kévin Muret
  - David Robelin
  - Magali San Cristobal
  - Nathalie Villa-Vialaneix
  - Matthias Zytnicki

Thanks for your attention!