

Animal Epigenomics workshop, PAG XXIV – San Diego, 12 Jan. 2016

Update on the French FAANG pilot project FR-AgENCODE

Elisabetta Giuffra

GABI, INRA, AgroParisTech, Université Paris Saclay - Jouy en Josas,
France

(INRA SelGen Metaprogramme)

Consortium



A	GABI (Coord.)	INRA, GA Div.	Elisabetta Giuffra, Michele Tixier-Boichard, Marie-Helene Pinard
A	VIM	INRA, SA Div.	Nicolas Bertho
B	PEGASE	INRA, GA Div.	Sandrine Lagarrigue
B	IGDR-CRNS- Univ.Rennes1	CNRS, Université Rennes 1, IGDR, Institut de Génétique et Développement de Rennes	Thomas Derrien
C	GENPHYSE	INRA, GA Div.	Sylvain Foissac, Hervé Acloque, Stephane Fabre
C	MIAT	INRA, MIA Div.	Christophe Klopp, Christine Gaspin
C	GenoToul	INRA, GA Div.	Diane Esquerré
D	URA	INRA, PHASE Div.	Joen Gautron
D	PRC	INRA, PHASE Div.	Xavier Druart
D	ISP	INRA, SA Div.	Pascale Quéré
E	WUR	Animal Breeding and Genomics Centre - Wageningen Univ.	Ole Madsen, Martien Groenen
F	RI	The Roslin Institute and Royal (Dick) School of Veterinary Studies - Univ. Edinburgh	Alan Archibald
G	EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute	Laura Clarke, Paul Flicek

Aims

❖ Realize a FAANG sample collection from 4 species



❖ Few target tissues for assays (chosen to benefit parallel ongoing research):

Liver (*hub tissue*) and two primary lymphoid cells (CD4+, CD8+)

❖ Optimization of 'new' methods and pipelines of analysis: whole transcriptome + a 'focus' on chromatin accessibility and interactome



Experimental animal facilities



National infrastructure of Biological Resource Centers for domestic animals - (<http://www.crb-anim.fr/>)



Elisabetta.giuffra@jouy.inra.fr - Animal Epigenomics workshop, PAG XXIV



Fr-AgENCODE: a FAANG sample collection

➤ 'all' tissues (snap-frozen) + preserved cells and crosslinked nuclei from some dissociated tissues (tot. 4900, *about 8_10 aliquots/tissue/animal*)



Large White INRA line

66 weeks female (n=2 females) and 27 weeks male (n=2 animals)



Holstein breed

200 weeks female (n=2 females) and 85 weeks male (n=2 animals)



Alpine dairy line

70 weeks female (n=2 females) and 27 weeks male (n=2 animals)



White Leghorn TEMLEG pure line

55 weeks female (n=2 females) and 55 weeks male (n=2 animals)

***Samples and metadata being entered in
BioSamples***

New pipelines (RNA-seq data)



Long-non coding RNAs repertoire in liver and adipose tissue in chicken

Sandrine Lagarrigue and collaborators

P0112 (+ talk at Poultry workshop)

Annotation of lncRNAs using FEELnc



- ⇒ Easy to use - does not need a species-specific lncRNA training set
- ⇒ Provides automatically a cutoff maximizing lncRNA & mRNA specificities
- ⇒ Provides automatically a classification of lncRNA based on the closest protein-coding genes

(developed by T. Derrien)

- **2954 lncRNA loci identified** (of which 90% are intergenic lncRNAs) **with similar features as human lncRNAs in terms of structure and expression**

On going in Fr-AgEncode:

Extend the analysis to more tissues (liver, CD4, CD8) **and to 4 species** (cattle, goat, pig & chicken)

Current progress on ATAC-seq and Hi-C

ATAC-seq (Kylie Munyard and collab.)

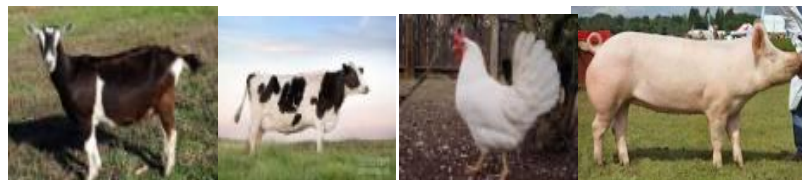
- Optimization carried out in pig: liver, spleen, CD3+CD4+ and CD3+CD8+ cells + GM12878 (human cell line)

Hi-C (Hervé Acloque and collab.)

- Optimization carried out in pig: cell lines (mouse fibroblasts and iPSCs), and fresh and snap frozen liver/muscle

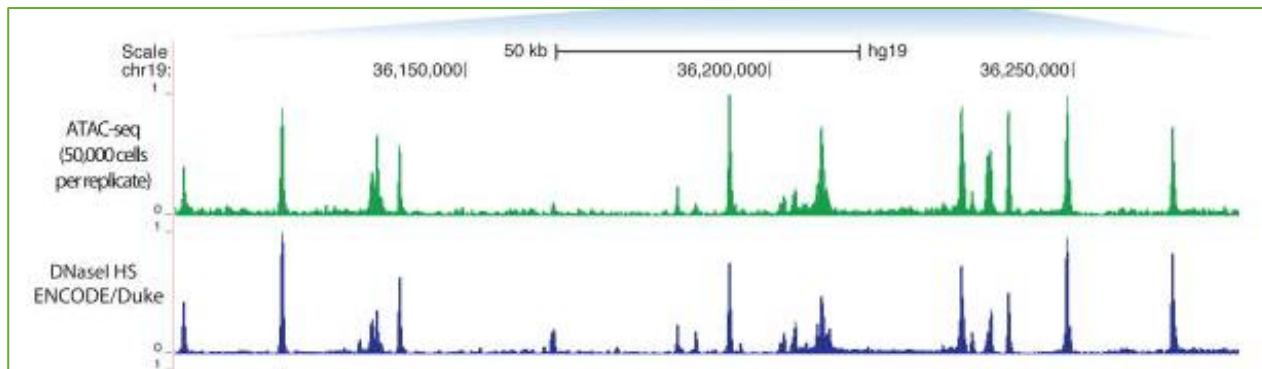
P0420 & P0421+ presentations at FAANG workshop

All sequencing runs (target samples of liver, blood cells/ 4 species) and bioinformatic analyses will take place in 2016 -



ATAC-seq

ATAC-seq correlates with various features of the genome:

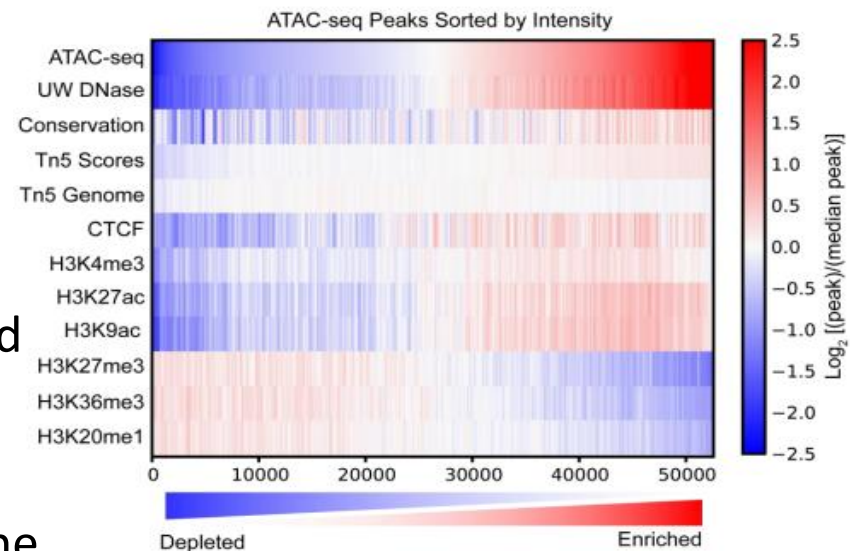


❖ Peak intensity strongly correlated with DNase hypersensitivity and CTCF

❖ Correlated with CTCF;

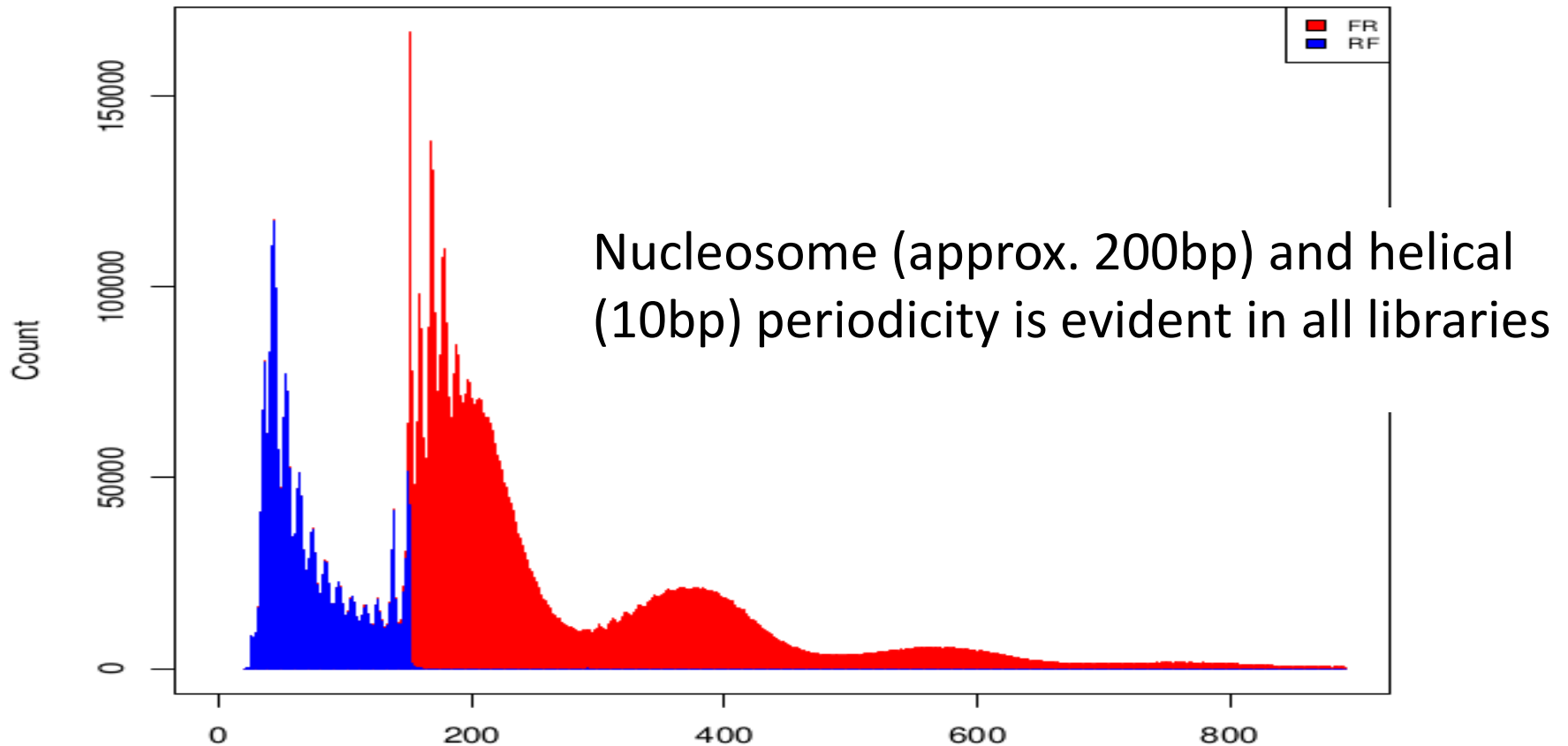
❖ correlated with histone marks associated with active chromatin

❖ anti-correlated with histone marks associated with inactive chromatin and gene bodies



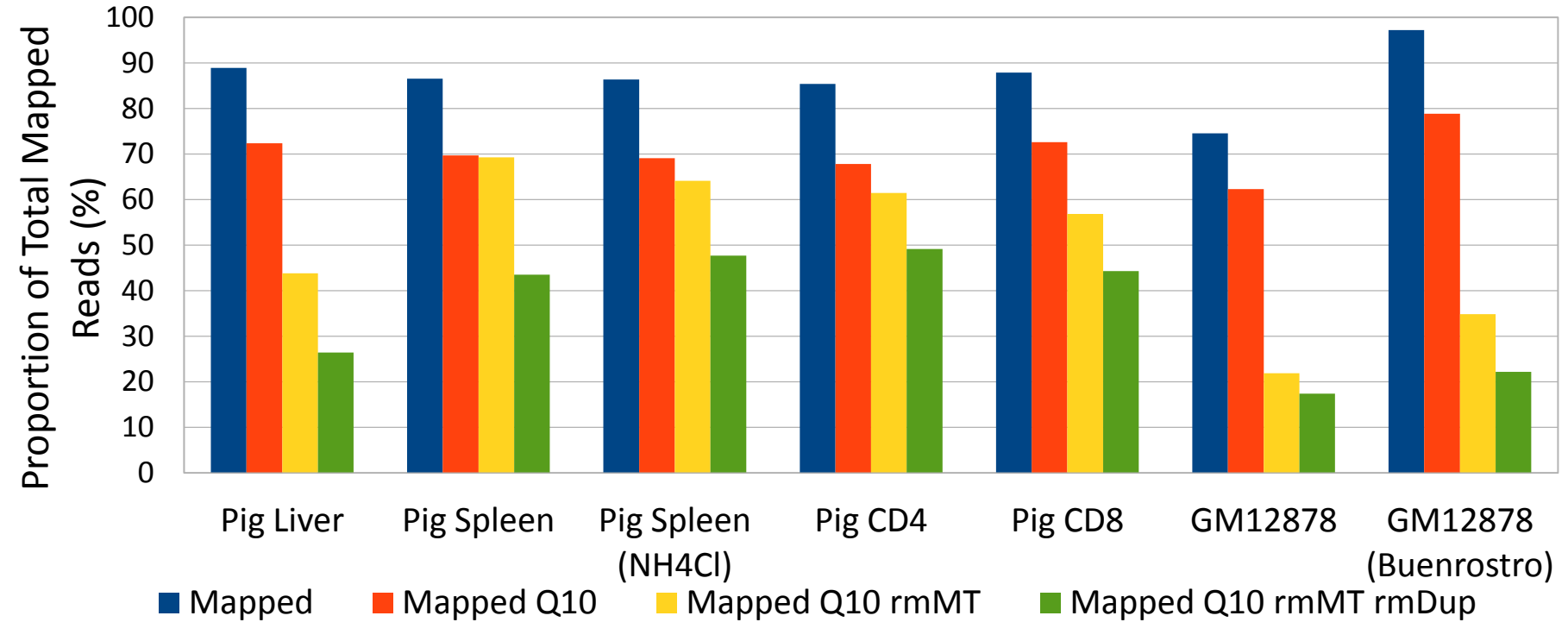
Buenrostro et al. Nat Methods. 2013, Curr Protoc Mol Biol. 2015

Preliminary Results: Pig Samples Library Fragment Size Distribution



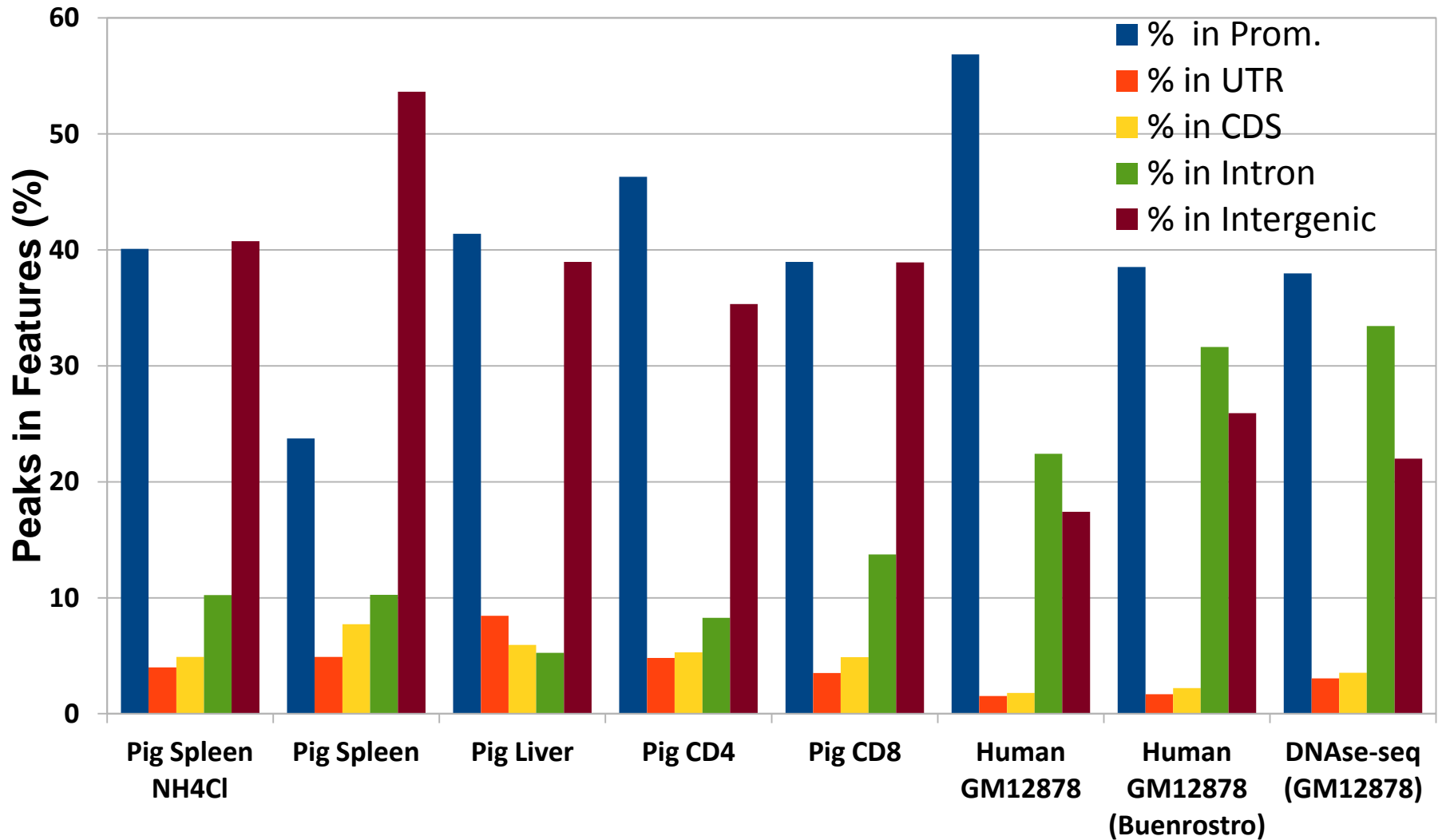


Mapping & Filtering



Sample Type	Liver	Spleen	Spleen (NH ₄ Cl)	CD4	CD8	GM12878	GM12878 (Buenrostro)
% MT reads in Q10 mapped	39.4	0.6	7.2	9.4	21.7	64.9	55.8

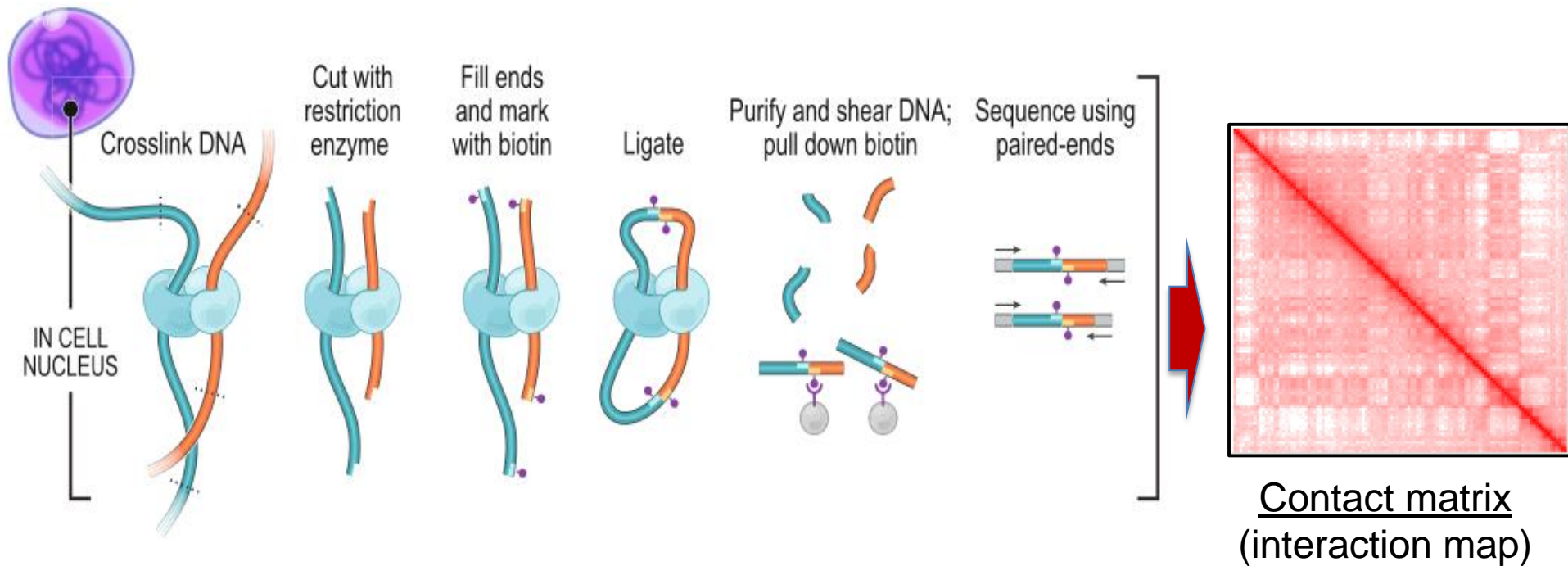
Distribution of Peaks in Genomic Features



Preliminary Results: Conclusions

- ATAC-seq library preparation was successful on all initial samples
 - Liver, spleen, CD4+ and CD8+ MeLiM pig primary cells/ tissues
 - NH_4Cl treatment had a large effect
- Results are consistent with published data
 - GM12878 vs. Buenrostro GM12878 and DNase-seq GM12878
 - Mitochondrial reads more problematic in cell lines
 - Evaluation of results ongoing
- Low coverage (probably) led to reduced peak numbers
 - Peaks in the human cell line were consistent with published data
 - Indication of different peak patterns in different tissues

Hi-C



Plenary lecture 11/01 - Lieberman-Aiden

- Protocol used: *Rao et al, Cell, 2014*
- Optimizing Ligation time and biotin removal at the extremities.

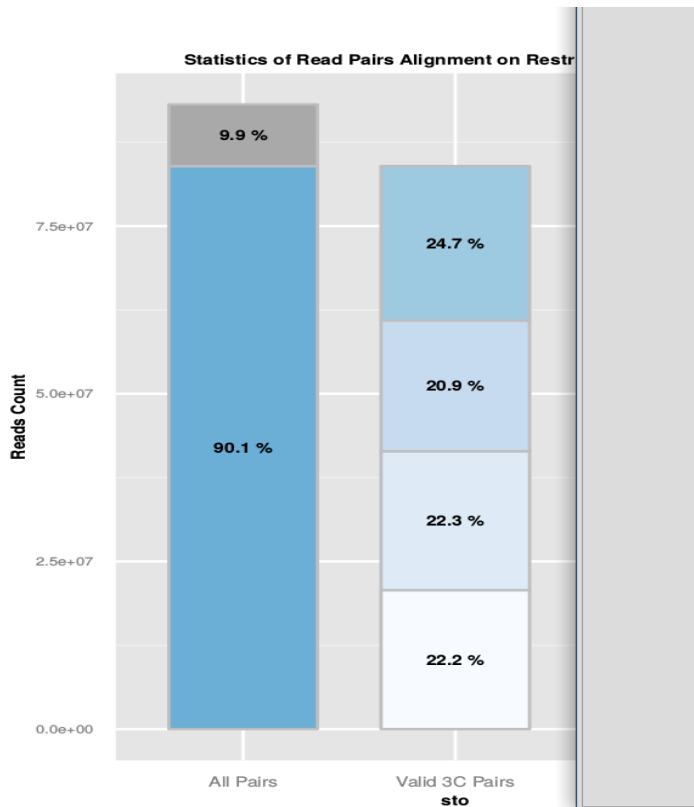
Hi-C: a virtually complete set of functional annotations in a single assay

Resolution	Over 25kb	25kb	5kb and under
Chromatin compartments (open and repressive)	yes	yes	yes
Chromatin sub-compartments (TAD and contact domains)	no	yes	yes
Chromatin Loops	no	no	yes
Promoter/enhancer detection	no	no	yes

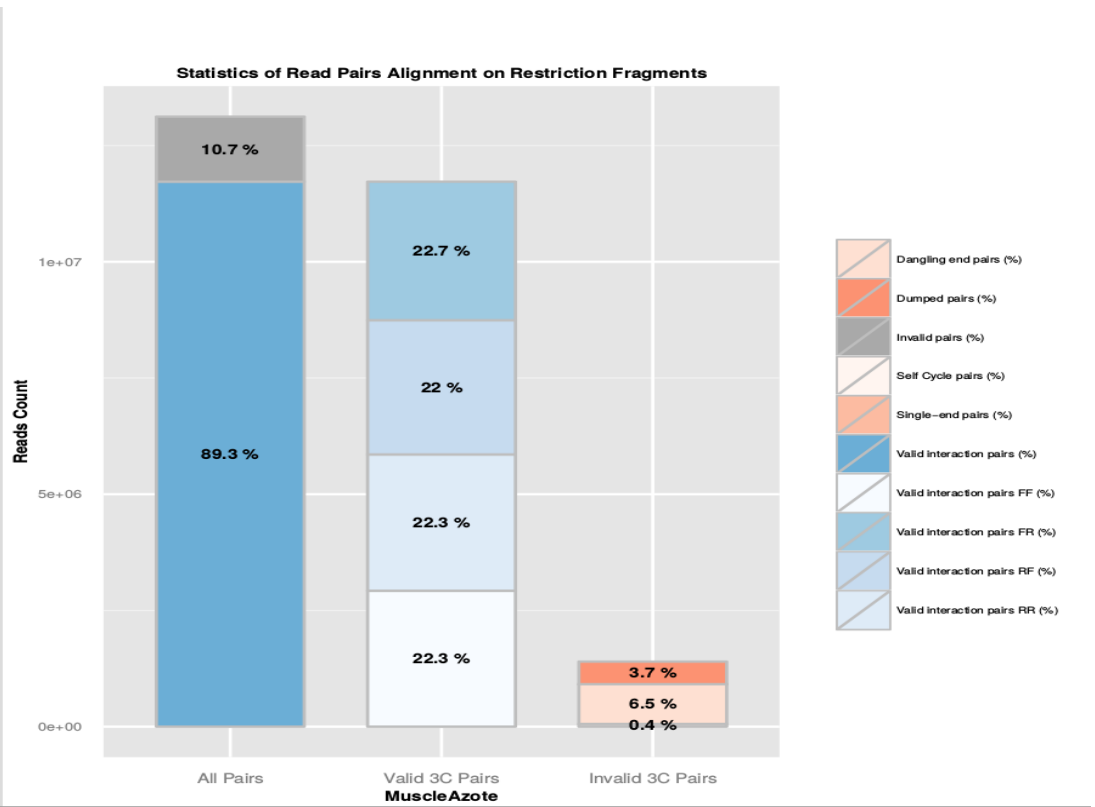
adapted from Rao et al. 2014

Hi-C: Increasing the yield of valid pairs of reads

mouse STO cells

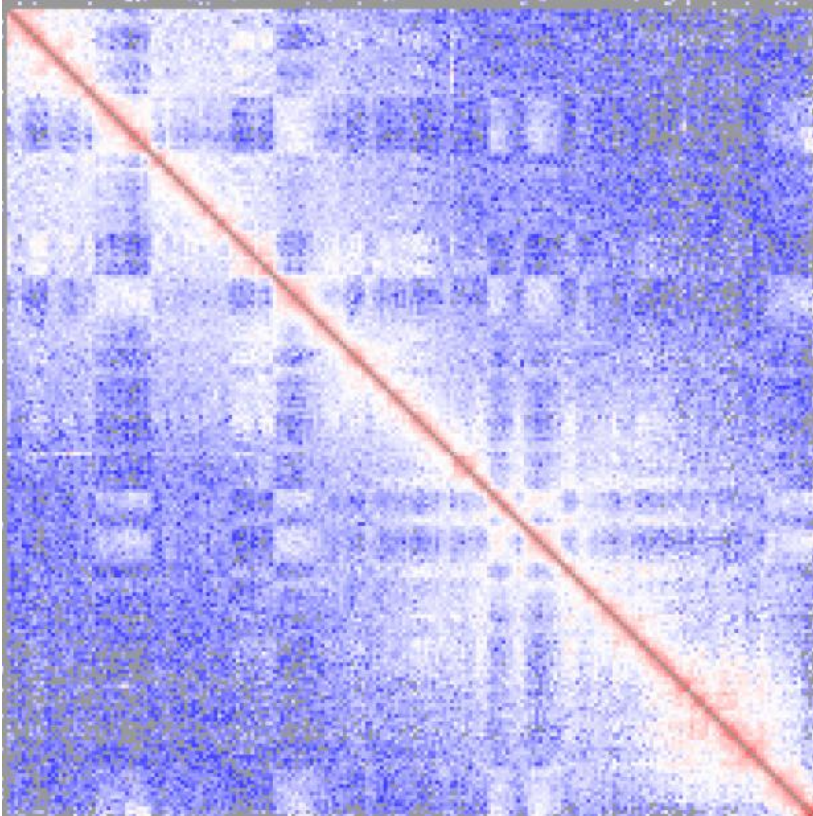


pig muscle

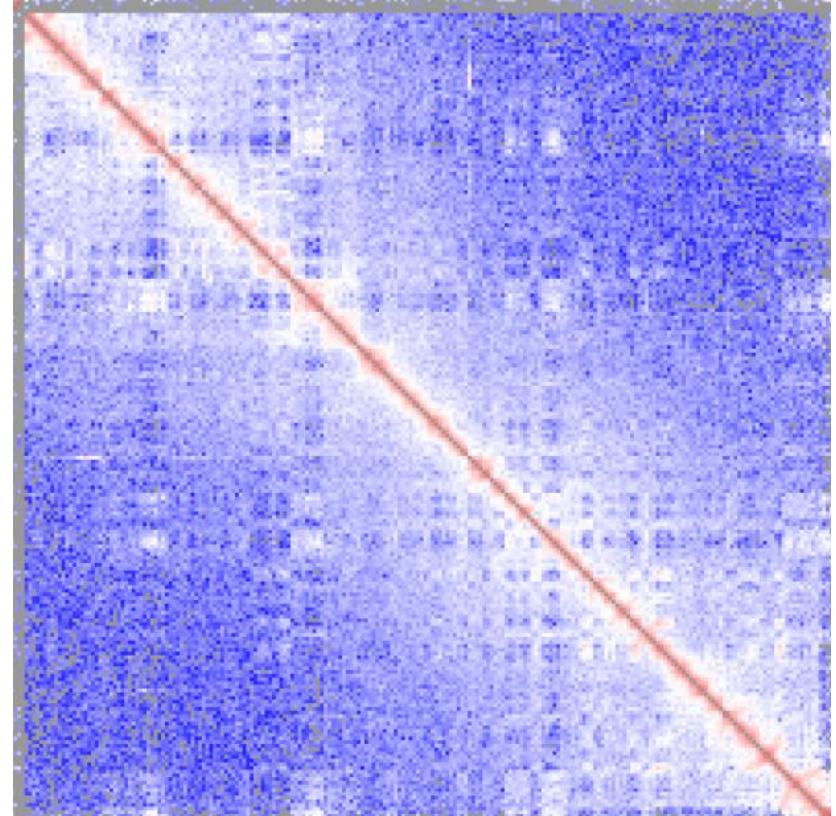


Hi-C: first low-resolution liver/muscle interaction maps

chr2: Liver



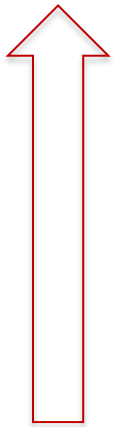
chr2: Muscle



P0421

Acknowledgements

The FAANG
colleagues



All INRA colleagues
who joined the fr-
AgENCODÉ
federative project

