FAANG workshop, PAG XXIV – San Diego, 11 Jan. 2016

Animals/Samples/Assays (ASA) Committee

Elisabetta Giuffra[1] & Huaijun Zhou[2]

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

[2]University of California, Davis - Davis, CA, USA

Animals/Samples/Assays (ASA) Committee

Aims:

- **Propose, develop and standardize** animal/tissue collection protocols, storage practices, and assay protocols for the FAANG community ('pilot projects').

- **Foster collaborations, avoid redundancies and enhance synergies**

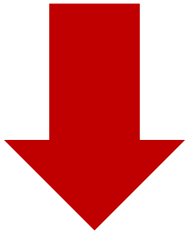
- **Presently: ~70 members**

- faang-samples@animalgenome.org

→ **faang.org (and FAANG Wiki)**

ASA Committee report: a focus on chromatin assays

- Process **tissues/primary cells** using **well-established and more informational assays** ***(Andersson et al. 2015; PMID: 25854118)***



- ***Most chromatin assays require optimization: soft vs. hard tissues, fresh vs. frozen, etc.***
- ***preliminary results of pilot projects: France and US***

ASA Committee report: 3 speakers

Andersson et al. 2015; PMID: 258

DNAse-seq and ChIP-seq analysis of farm animal tissues
Pablo Ross, UC-Davis

Core Assays

1. Transcribed loci:

❖**RNA-seq** (exhaustive catalogues of gene /ncRNAs expression)

2. Chromatin Accessibility and Architecture:

❖DNAse hypersensitivity **(DNase I-seq)** and possibly **ATAC-seq** (Transposase-Accessible Chromatin with high-throughput seq

3. Histone modification marks *(up*

❖ **H3K4me3** (promoters of active g
❖ **H3K27me3** (genes that have bee
❖ **H3K27Ac** (active regulatory elem
❖ **H3K4Me1(**regulatory elements a
enriched downstream of transcrip

Measuring Chromatin Accessibility using ATAC-seq: preliminary results on porcine tissues/cells
Kylie Munyard, INRA - P0420

Functional annotation of livestock genomes using Hi-C: preliminary results on porcine tissues/cells
Herve Acloque, INRA - P0421

Additional Assays

❖**DNA methylation**
❖**Chromosome conformation capture** (by **Hi-C**: genome-wide interactions in 3D)

DNase-seq and ChIP-seq analysis of farm animal tissues

Pablo J. Ross

Associate Professor

University of California Davis

UCDAVIS ANIMAL SCIENCE

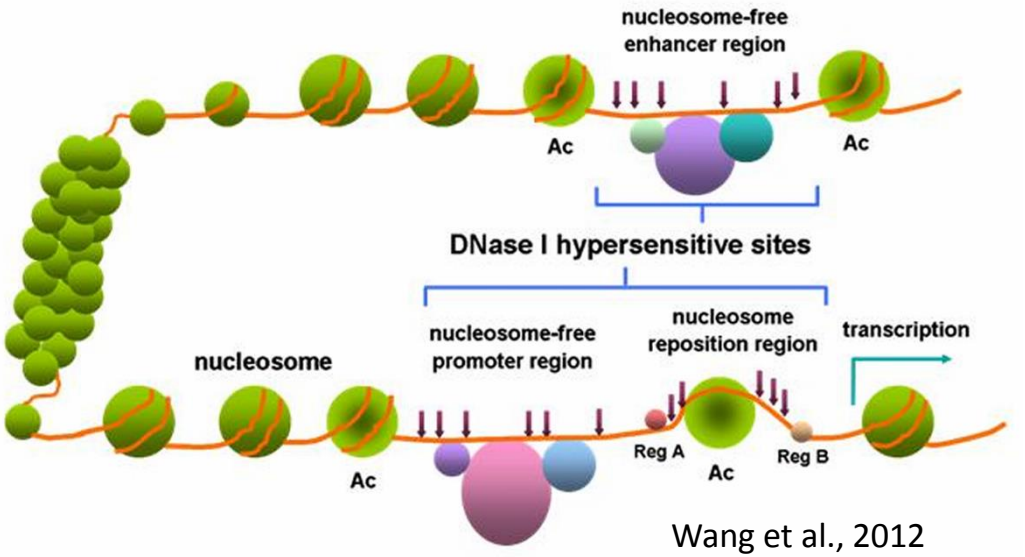
FAANG
Functional Annotation of Animal Genomes

DNase-seq

Profile DNase I hypersensitivity sites across the whole genome

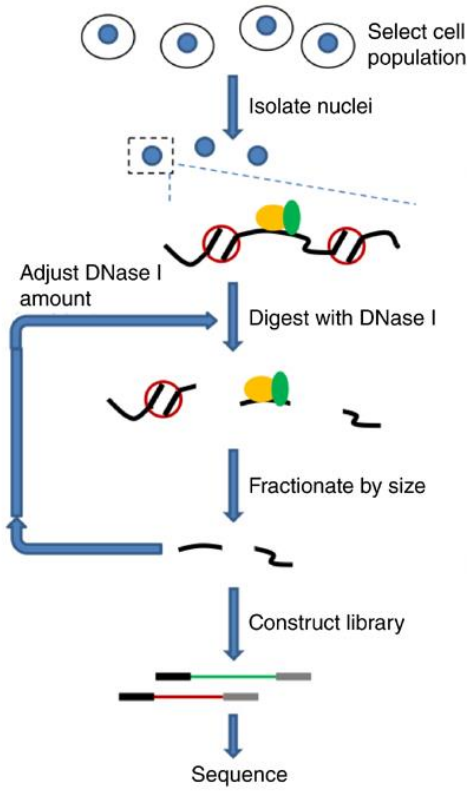
Active DNA elements (promoters and enhancers) are more accessible to DNase I digestion than the rest of the genome

DNase I sites = regulatory regions



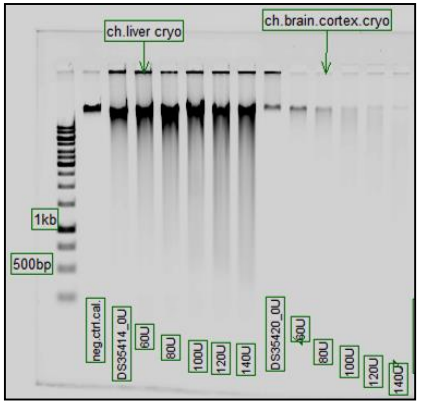
Wang et al., 2012

Workflow and critical considerations



Nuclear integrity and correct quantification is critical
Cell culture>tissue freshly isolated nuclei>frozen tissues
DNase I concentration is critical, tissue/sample dependent

Optimal nuclei isolation is tissue dependent



Libraries are sequenced at low level first, and amplified to >50 million reads/sample if good

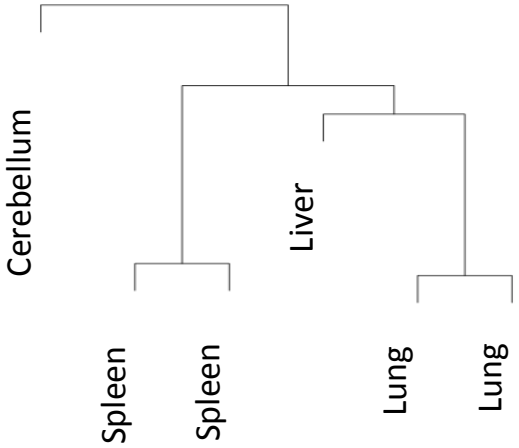
Results from UC Davis samples run at Stam Lab - University of Washington

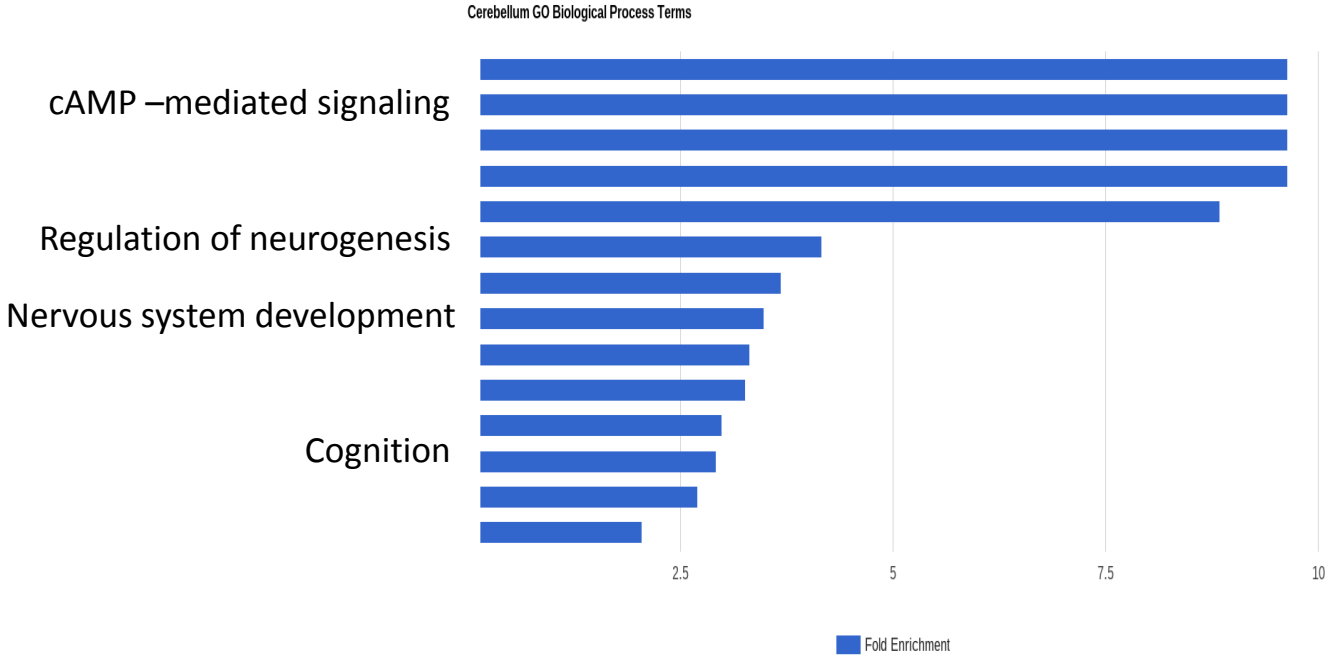
Chicken samples processed by Dnase-seq

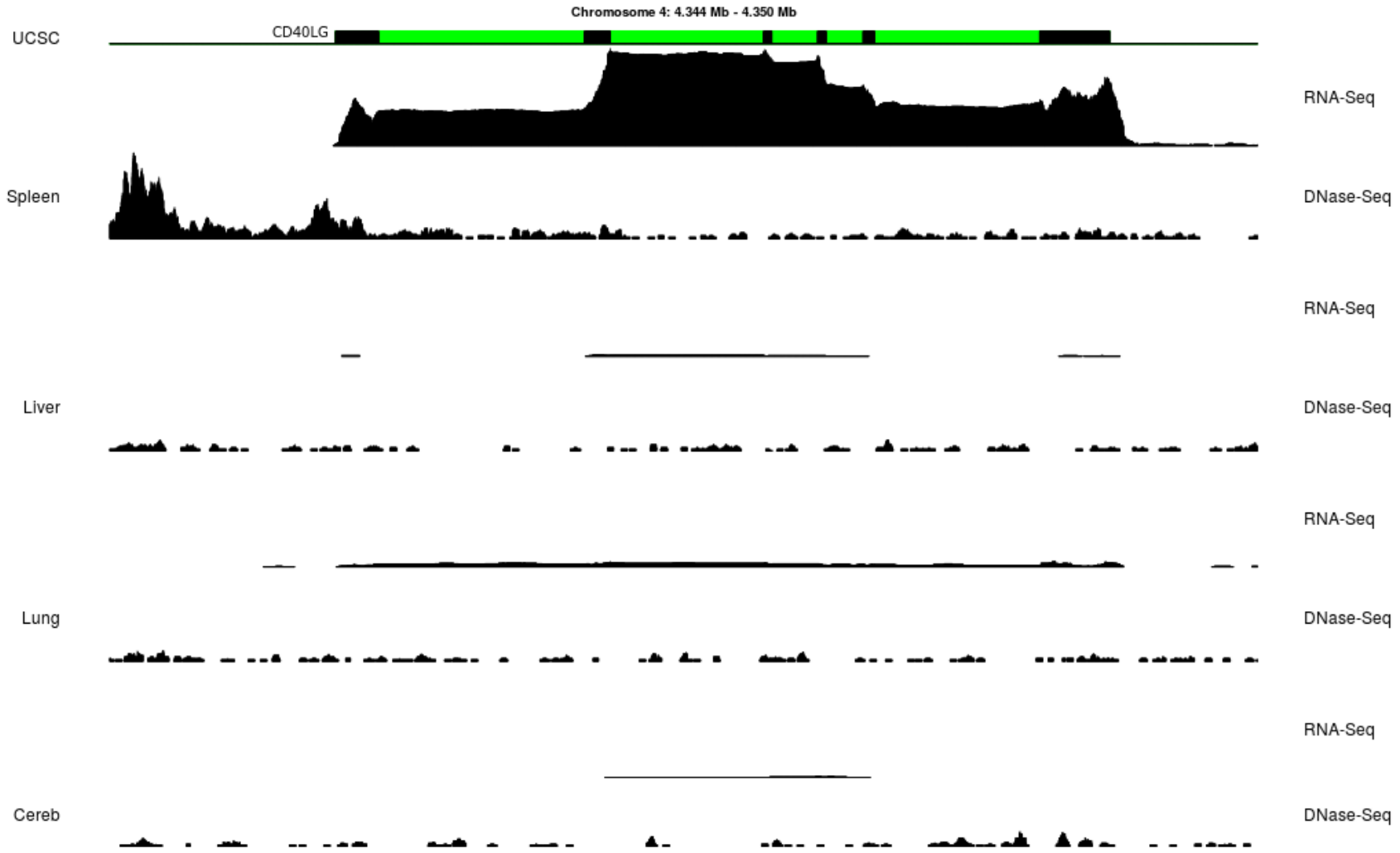
	Adipose	Cerebellum	Cortex	Hypothalamus	Liver	Lung	Muscle	Spleen
Bird A								
Bird B								
Reads	29,236,452	107,520,255	42,946,606	59,644,084	93,681,598	79,033,732	26,730,117	237,834,942

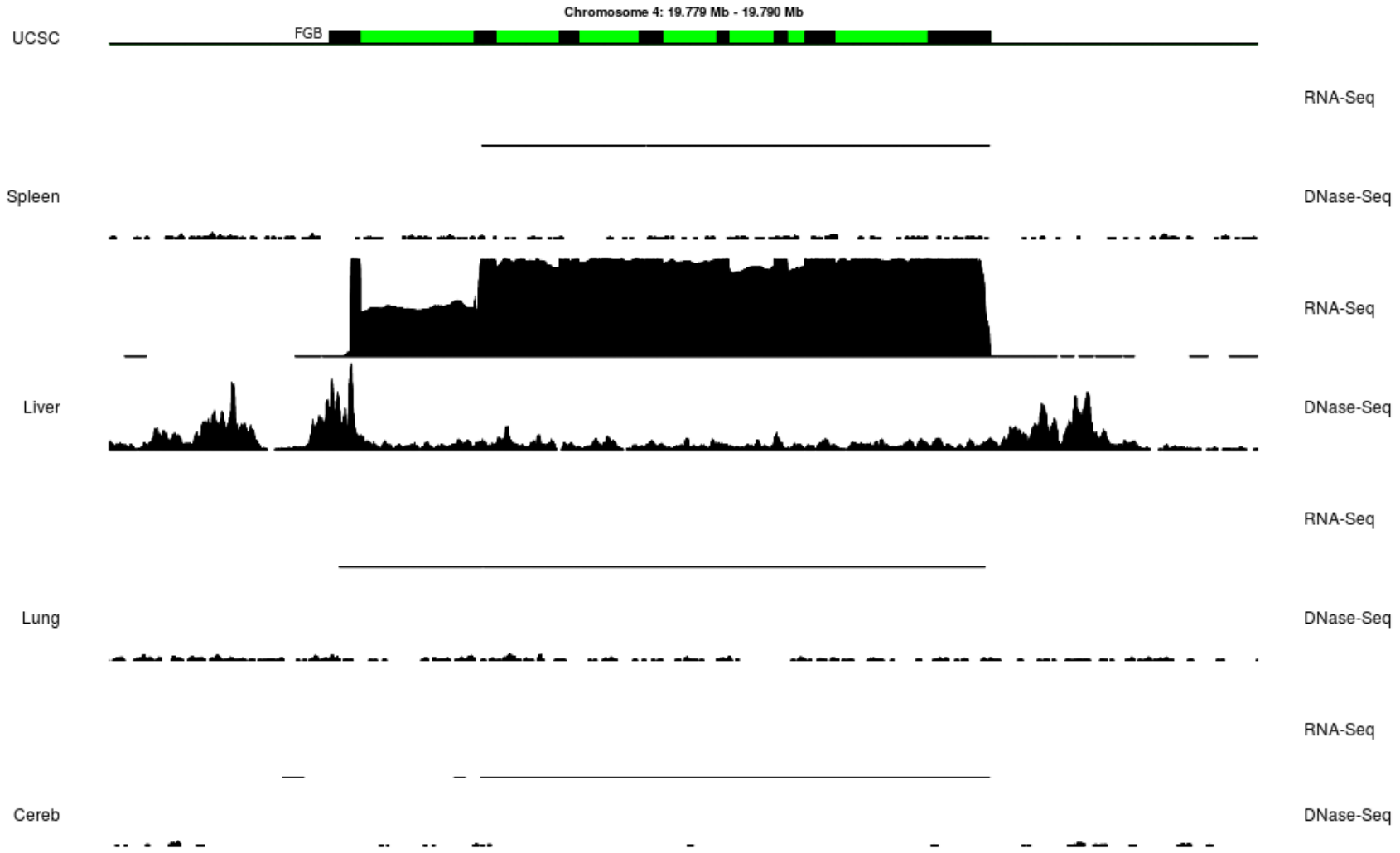
12 successful libraries
22 failed libraries

Hierarchical Clustering Based on DHS Peaks



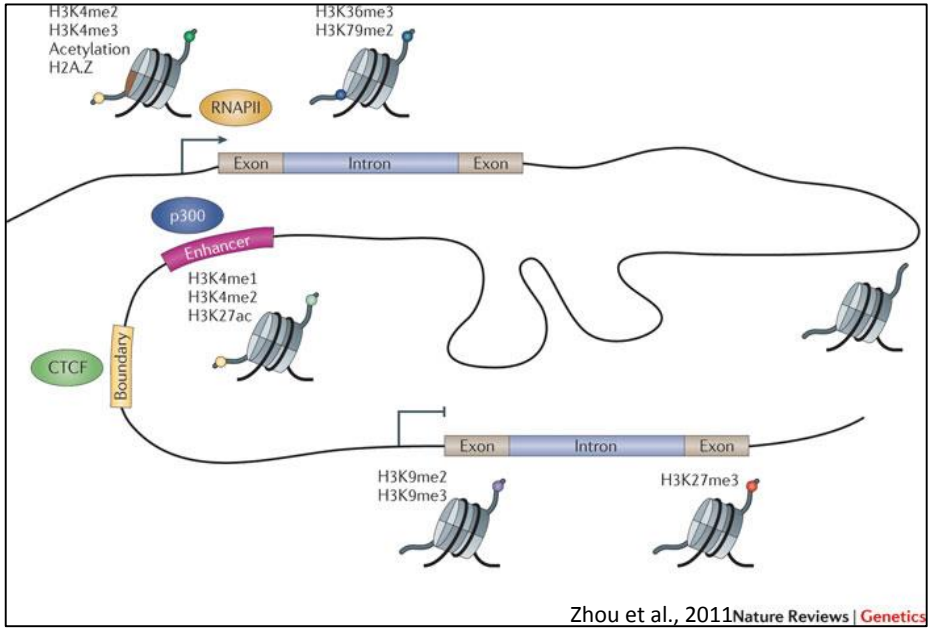



Chromosome 4: 4.344 Mb - 4.350 Mb
UCSC
CD40LG
RNA-Seq
Spleen
DNase-Seq
RNA-Seq
Liver
DNase-Seq
RNA-Seq
Lung
DNase-Seq
RNA-Seq
Cereb
DNase-Seq

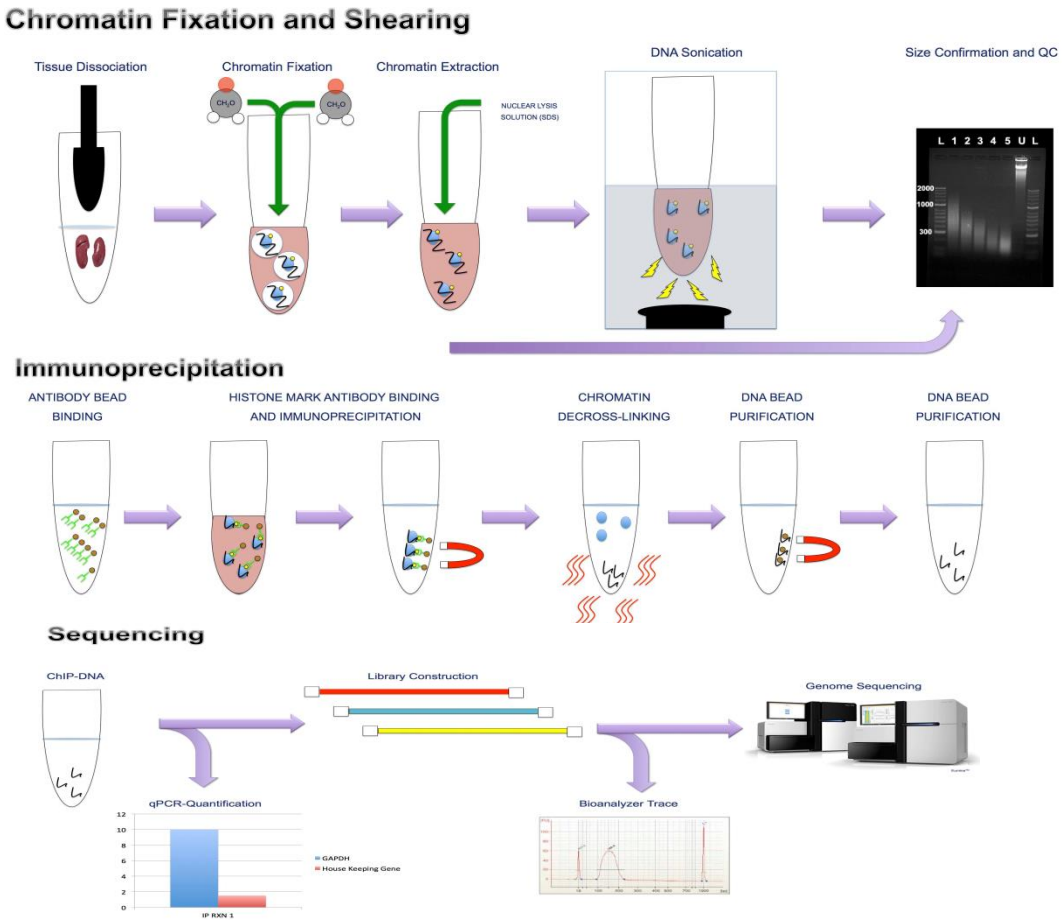

Chromosome 4: 19.779 Mb - 19.790 Mb
UCSC
FGB
RNA-Seq
Spleen
DNase-Seq
RNA-Seq
Liver
DNase-Seq
RNA-Seq
Lung
DNase-Seq
RNA-Seq
Cereb
DNase-Seq

ChIP-seq

Identify areas of the genome associated with specific proteins
Histone modifications demarcate functional elements in the genome
Mapping histone modifications = Mapping functional genomic areas



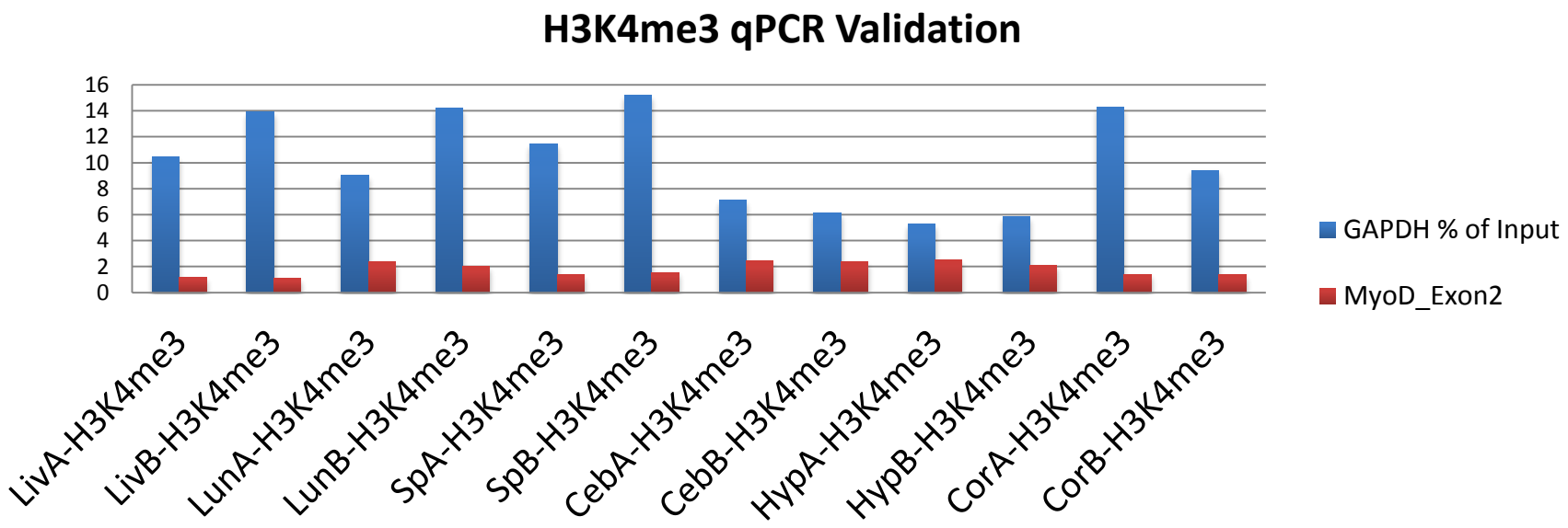
H3K4me3	**Active promoters**
H3K27me3	**Inactive chromatin**
H3K4me1	**Active enhancers**
H3K27ac	**Active promoters and enhancers**
CTCF	**Boundary element**



Diagnode IDeal ChIP-Seq Kit, Covaris S2 system. NEB library preparation kit

qPCR Confirmation of ChIP Enrichment-H3K4me3

- % recovery = $2^{(Ct_{input} - Ct_{sample})}$
- Expected >5% considered successful enrichment

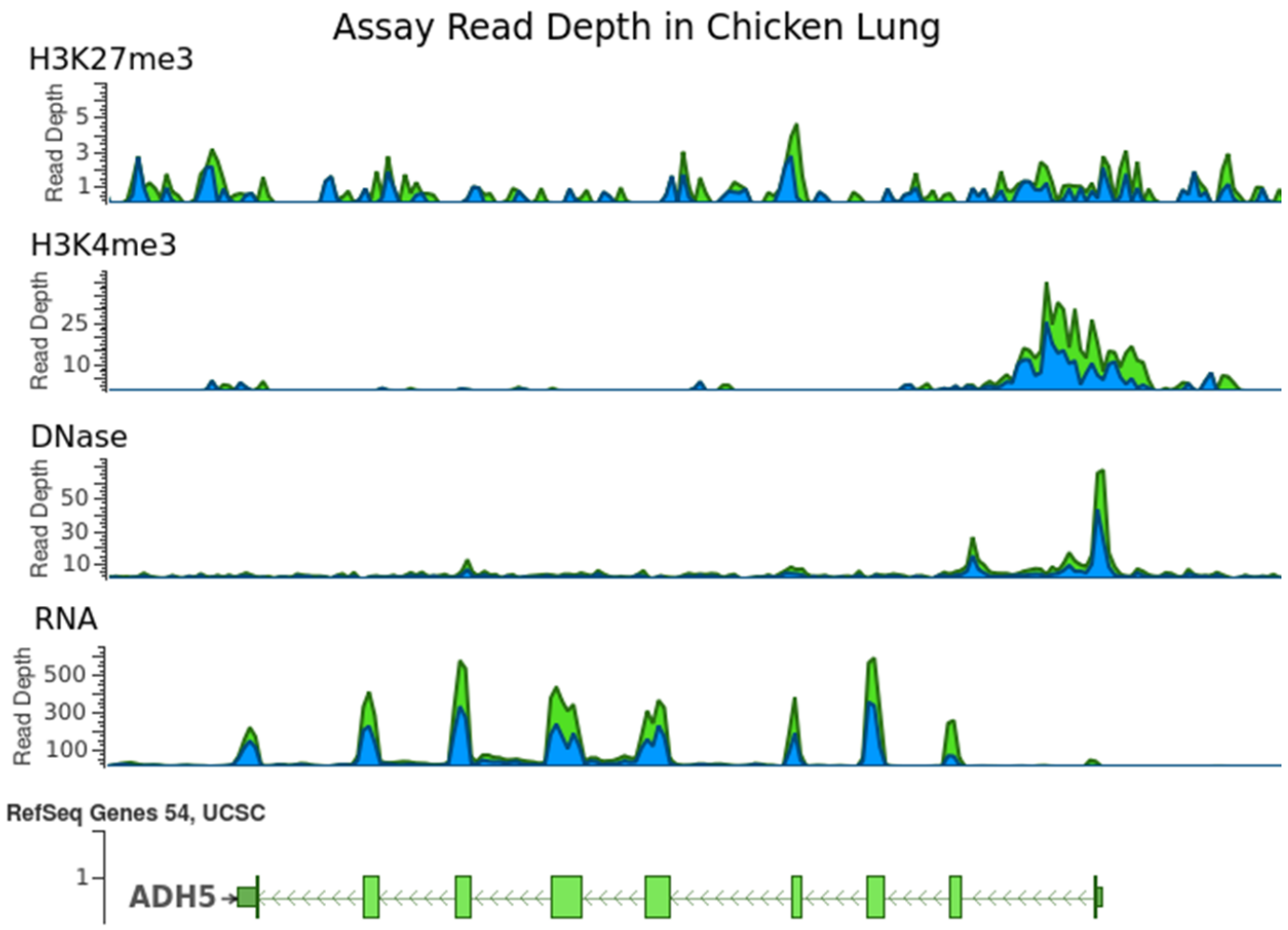


Raw ChIP-Seq Reads

	H3K4me3	H3K27me3	Control
Cerebellum	41,023,866	45,384,874	42,692,478
Cortex	49,805,512	47,458,518	76,985,076
Hypothalamus	64,159,608	71,047,968	83,411,580
Liver	49,061,112	76,310,510	66,759,934
Lung	44,510,256	42,373,494	60,945,834
Spleen	36,701,468	49,291,330	73,482,690

Aligned and Filtered ChIP-Seq Reads from Chicken

	H3K4me3	H3K27me3	Control
Cerebellum	24,331,529	26,132,801	27,935,330
Cortex	35,745,439	20,781,566	56,091,853
Hypothalamus	42,025,119	48,644,786	60,546,445
Liver	29,485,908	35,850,744	51,835,760
Lung	33,971,248	30,219,806	51,907,342
Spleen	28,553,718	37,290,417	44,115,980


Assay Read Depth in Chicken Lung
H3K27me3
Read Depth
5
3
1
H3K4me3
Read Depth
25
10
DNase
Read Depth
50
30
10
RNA
Read Depth
500
300
100
RefSeq Genes 54, UCSC
1
ADH5

Global Project Current Progress



RNA-Seq	Data Analysis		Data Analysis	Data Analysis		
DNase-Seq	Data Analysis		Tissues Collected	Tissues Collected		
H3K4me3	Data Analysis **6 / 8 Tissues**	IP Completed 2 / 8	Tissues Collected	IP Completed 2 / 8	Chromatin Sheared **6 / 8 Tissues**	
H3K27me3	Data Analysis **6 / 8 Tissues**	IP Completed 2 / 8	Tissues Collected	IP Completed 2 / 8	Chromatin Sheared **6 / 8 Tissues**	
H3K4me1	IP Completed		Tissues Collected	Tissues Collected		
H3K27ac	IP Completed		Tissues Collected	Tissues Collected		
CTCF	IP Completed		Tissues Collected	Tissues Collected		

Genome-wide identification and annotation of functional regulatory regions in livestock species

- H. Zhou, P. Ross, I. Korf, UC Davis
- Collaborators:
  - Poultry Genome Coordinators: M. Delany, H. Cheng
  - Cattle Genome Coordinators: J. Medrano, A. Van Eenennaam
  - Swine Genome Coordinators: C. Tuggle, C. Ernst
  - V. Leesburg, USDA ARS
  - Jim Kent, UCSC
  - Laura Clarke, Paul Flicek, EBI
  - Bin Ren, UCSD
- Postdocs and students
  - Colin Kern
  - Ying Wang
  - Perot Saelao
  - Michelle Halstead

Financial support:

USDA NIFA

grant #2015-67015-22940

pork checkoff

Aviagen

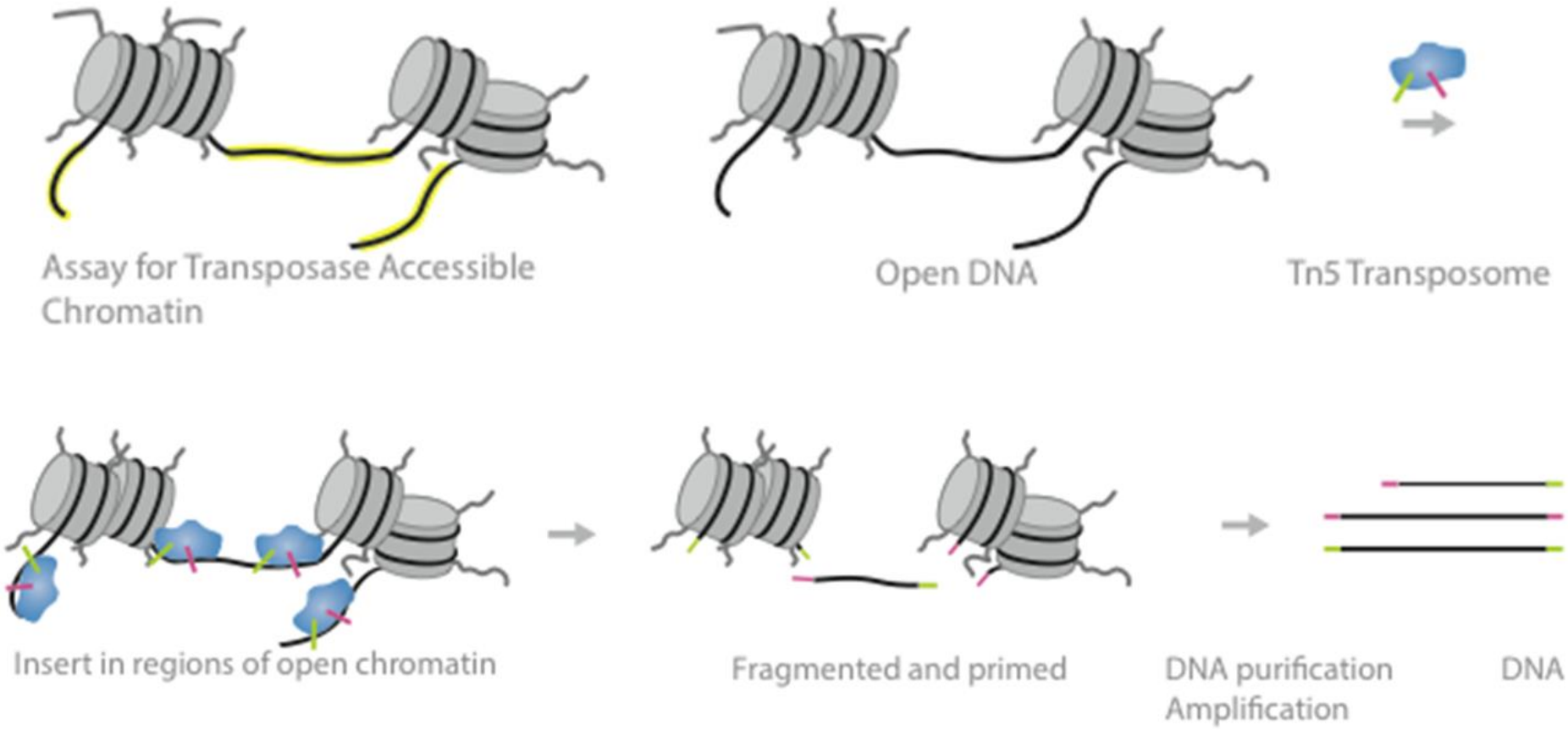
Poultry, Cattle, Swine Genome Coordination Funds

Measuring Chromatin Accessibility using ATAC-seq: preliminary results on Porcine tissues/Cells

- **ATAC-seq**: **A**ssay for **T**ransposase **A**ccessible **C**hromatin by **seq**uencing
- One important aspect of functional control of mammalian genomes is chromatin accessibility
- ATAC-seq is an emerging method for measuring chromatin accessibility
- It's advantages over other such methods (e.g. DNase-seq) are that:
  - It requires fewer cells (50,000)
  - It is fast and simple

Kornberg & Lorch (1992), Mellor (2005), Buenrostro et al. (2013)

The ATAC-seq Process



Adapted from: www.illumina.com/techniques

ATAC-seq

Samples

- Liver, spleen, CD3+CD4+, & CD3+CD8+ T-cells from 2 ♂ & 2 ♀ Large White pigs
- Liver, spleen, CD3+CD4+, & CD3+CD8+ T-cells from 1 Melanoma-bearing Libechov Minipig (MeLiM)
- GM12878 Human Lymphoblastoid cell line

ATAC-seq Libraries

- Transposition *in vitro* on fresh (liver, spleen) or thawed (CD4/8+) samples (Buenrostro *et al.* 2015, Curr. Prot. Mol. Biol. 109:21.2)
- Removal of primers & adaptors (AMPureXP)

Sequencing

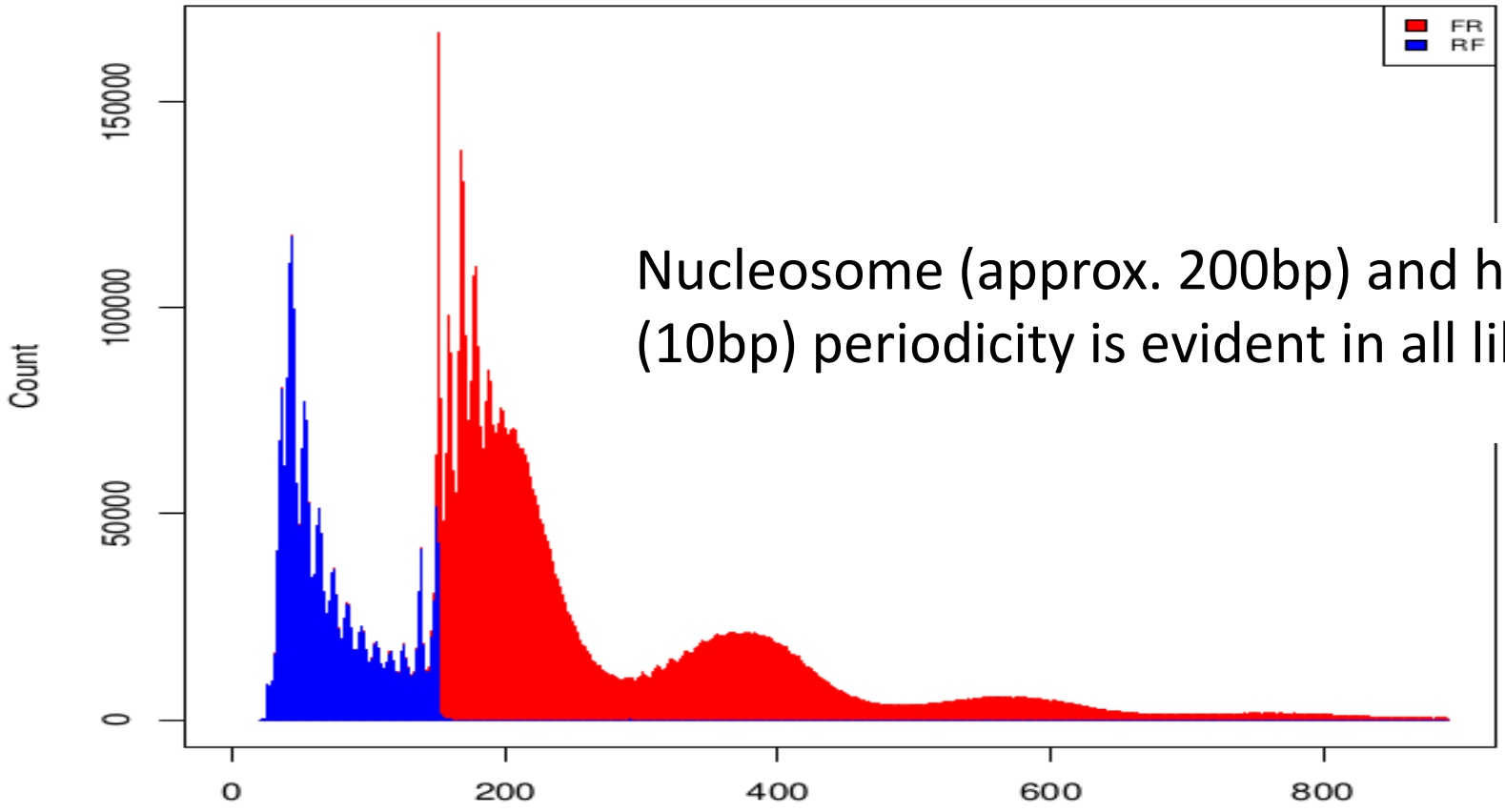
- Illumina HiSeq 3000, 150bp, PE
- 6 libraries (MeLiM: Liver, spleen x 2, CD3+CD4+, & CD3+CD8+ T-cells, & GM12878) pooled in one lane

Analysis Pipeline

- Adaptors trimmed (Trimgalore)
- Mapped to reference (Bowtie2)
- Duplicates Removed (Picard-Tools)
- Mitochondrial reads removed (Samtools)
- Peaks called (MACS2)

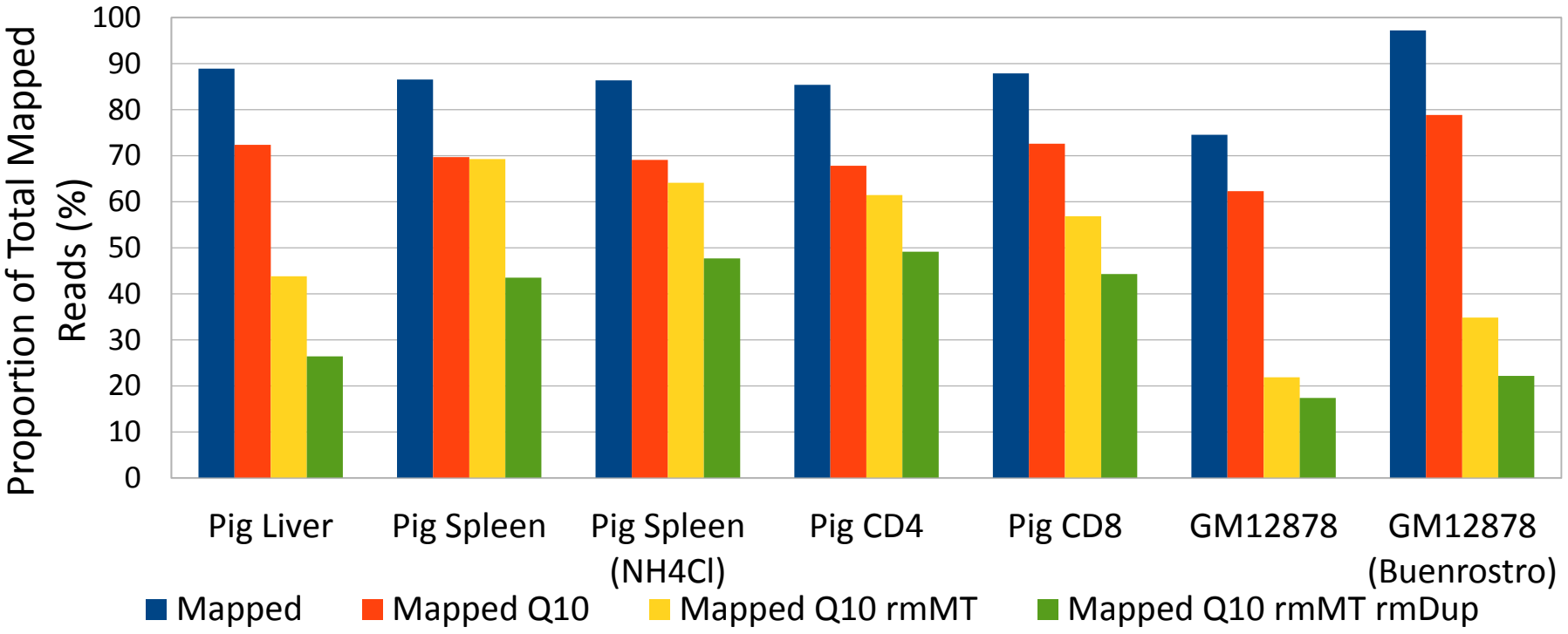
Preliminary Results: Pig Samples

Library Fragment Size Distribution



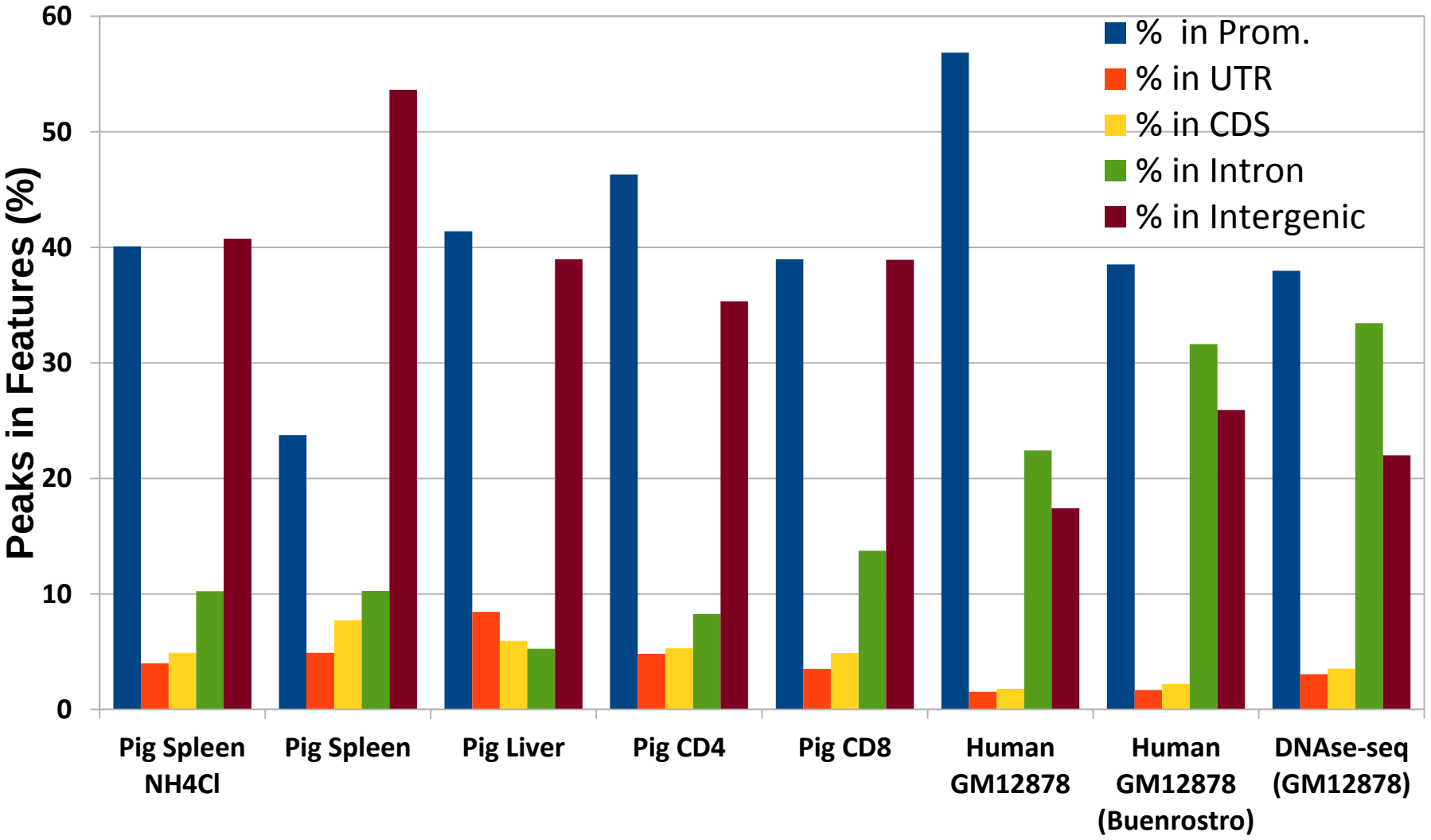
Nucleosome (approx. 200bp) and helical (10bp) periodicity is evident in all libraries

Mapping & Filtering



Sample Type	Liver	Spleen	Spleen (NH_4Cl)	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

The Fr-AgENCODE ATAC-seq team

Roadmap for 2016:

1st semester:

Additional sequencing of these libraries

Complete analysis

Complete libraries & sequencing for target samples:

- pigs, chicken and goats
- Liver, CD4+, CD8+ cells

2nd semester:

Start data analysis on target samples & integration with Hi-C and whole transcriptome data

Molecular Biology

Elisabetta Giuffra

Adeline Goubil

Diane Esquerré

Blood cells (for pig)

Silvia Vincent-Naulleau

Bioinformatics

Kylie Munyard

Sylvain Foissac

Sarah Djebali

Fr-AgENCODE coordination

Elisabetta Giuffra & Sylvain Foissac

Hi-C for Fr-AgenCode: Profiling the Nuclear Spatial Conformation of Livestock Genomes in Cell Lines and Liver Cells

Brief reminder of the Hi-C proposal within the Fr-AgEncode Project:

4 Species (chicken, pig, cattle, goat)
4 Animals/species (2 males and 2 females)
1 Tissue for each animal: Liver

▶ 16 different maps of DNA-DNA interactions

A minimal resolution of 100kb for each map (with at least 1000 contacts for each 1Mb bin)
For a 3Gb mammalian genome, to reach a 50-100kb resolution: 70M of paired-reads/library

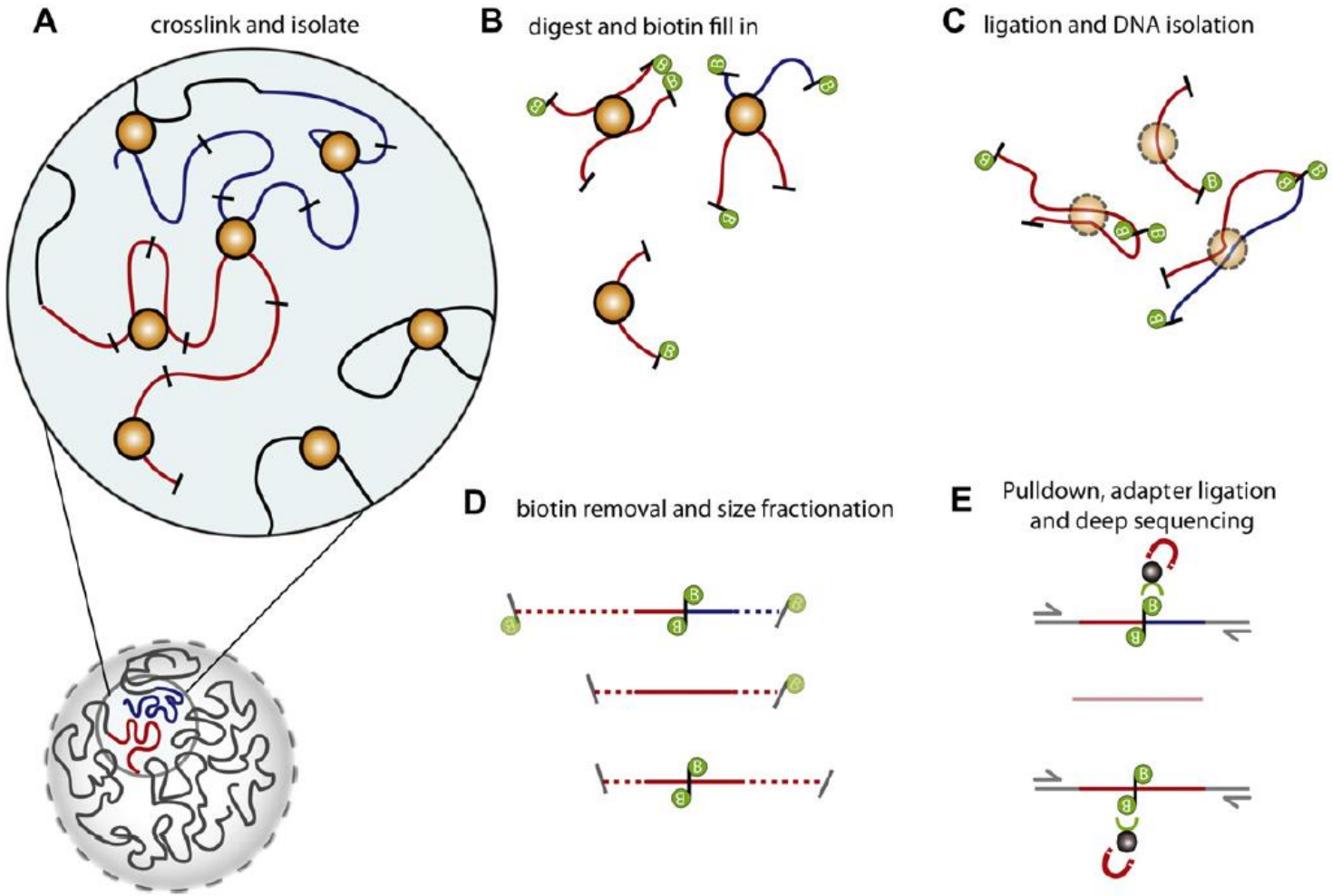
▶ Protocol adapted from *in situ* Hi-C developed by the group of Lieberman-Aiden, Rao et al. 2014

Why did we decide to include Hi-C assays for Fr-AgenCode?

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

The Hi-C Technology



Belton et al. 2012

Hi-C for Fr-AgenCode: tested tissues

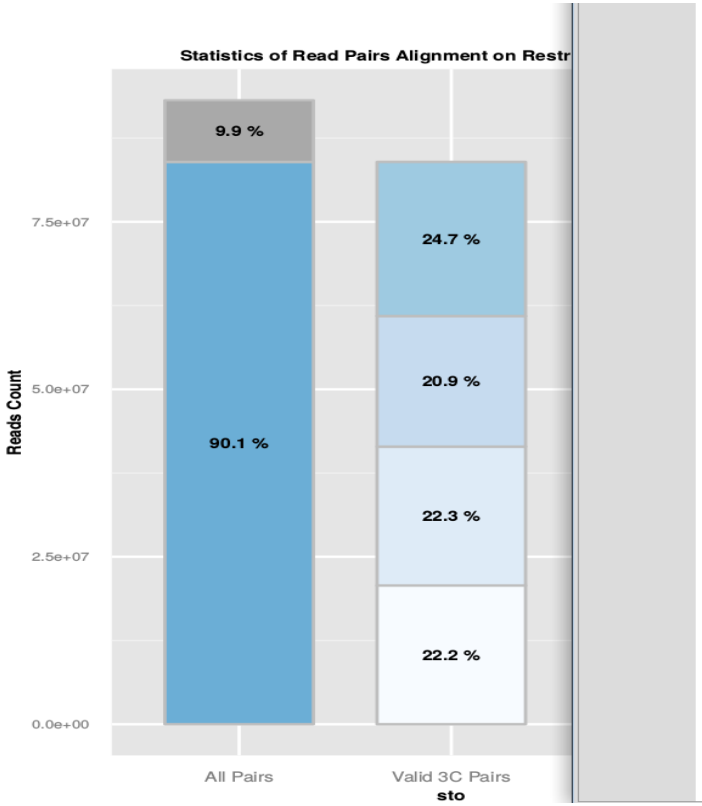
Working on :

Cell lines (mouse fibroblasts and pig iPSCs)
Pig tissues: fresh and snap frozen liver
fresh and snap frozen skeletal muscle

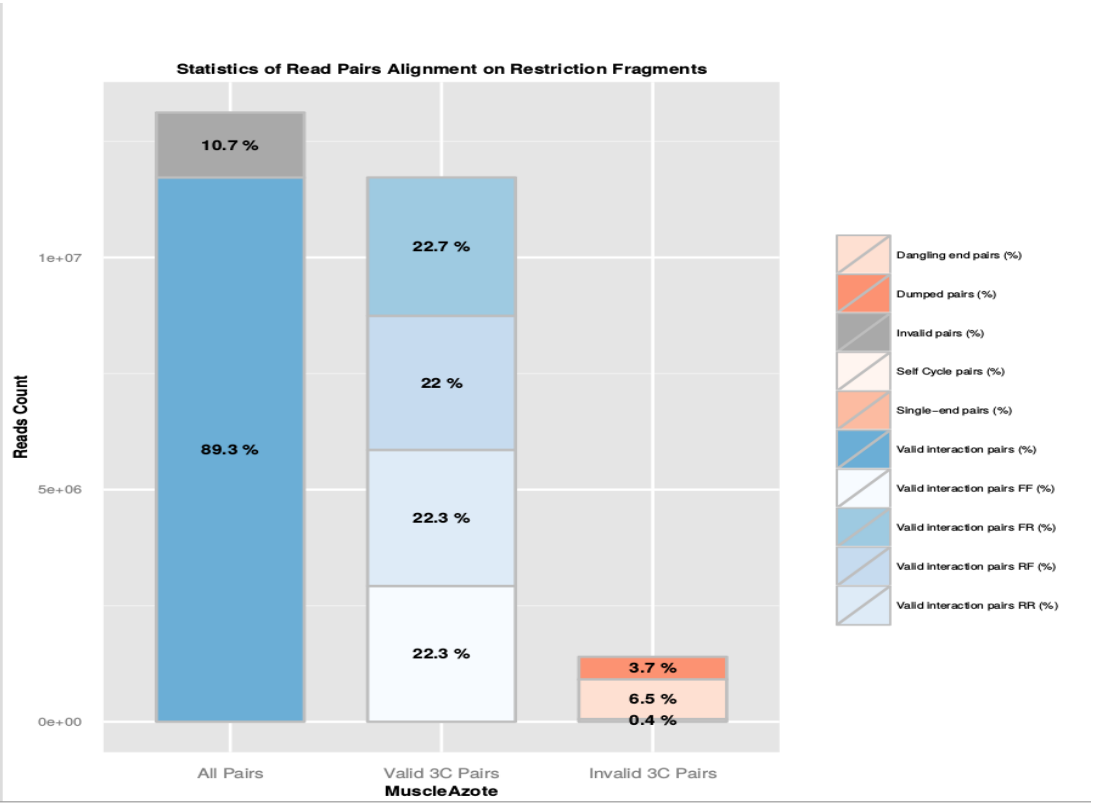
on going: chicken and cow fresh liver

Hi-C for Fr-AgenCode:
Increasing the yield of valid pairs of reads

mouse STO cells



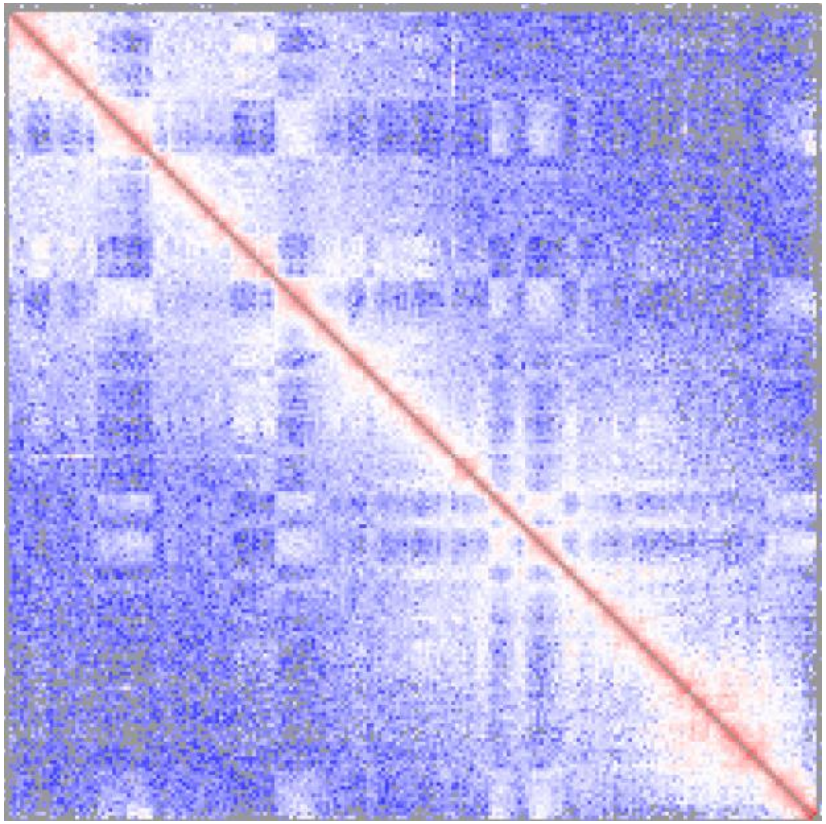
pig muscle



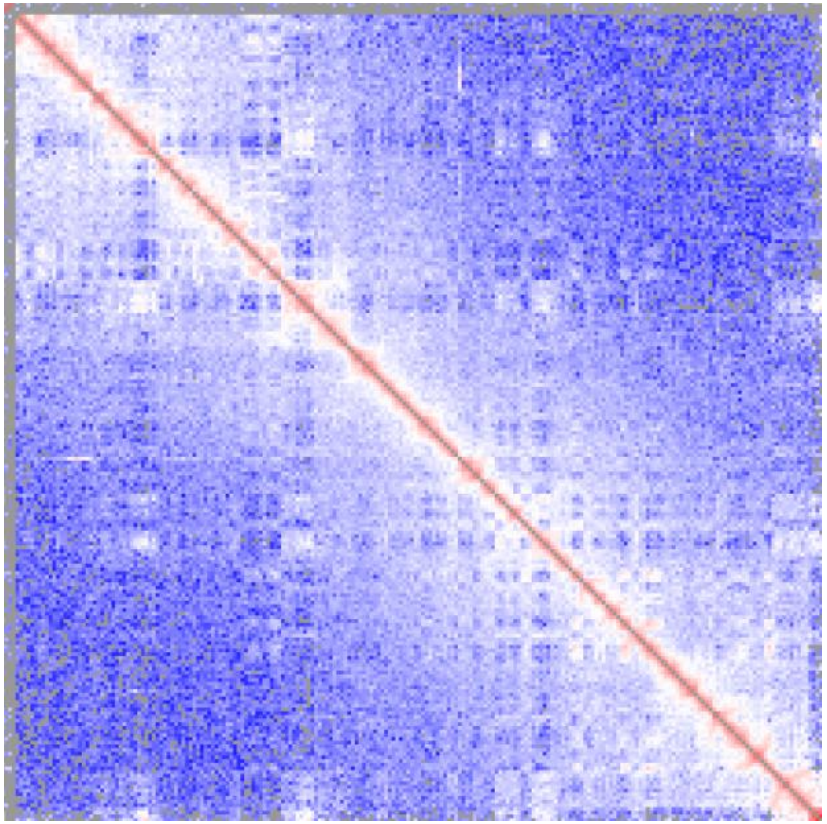
Optimizing Ligation time and biotin removal at the extremities.

Hi-C for Fr-AgenCode:
First low-resolution liver/muscle interaction maps

chr2: Liver



chr2: Muscle



Poster 0421: Sylvain Foissac

The Fr-AgENCODE Hi-C team

Roadmap for 2016:

1st semester:

16 Liver libraries and sequencing for cows, pigs, chicken and goats

Comparison of pig muscle and liver 50kb resolution maps
(INTERPIG project, INRA)

2nd semester:

Analysis and comparison of liver maps

Molecular biology
Florence Mompart
Diane Esquerré
Maria Marti

Bioinformatics
Sylvain Foissac
Sarah Djebali

Biostatistics
David Robelin
Magali San Cristobal
Matthias Zytnicki

Fr-AgENCODE coordination
Elisabetta Giuffra & Sylvain Foissac