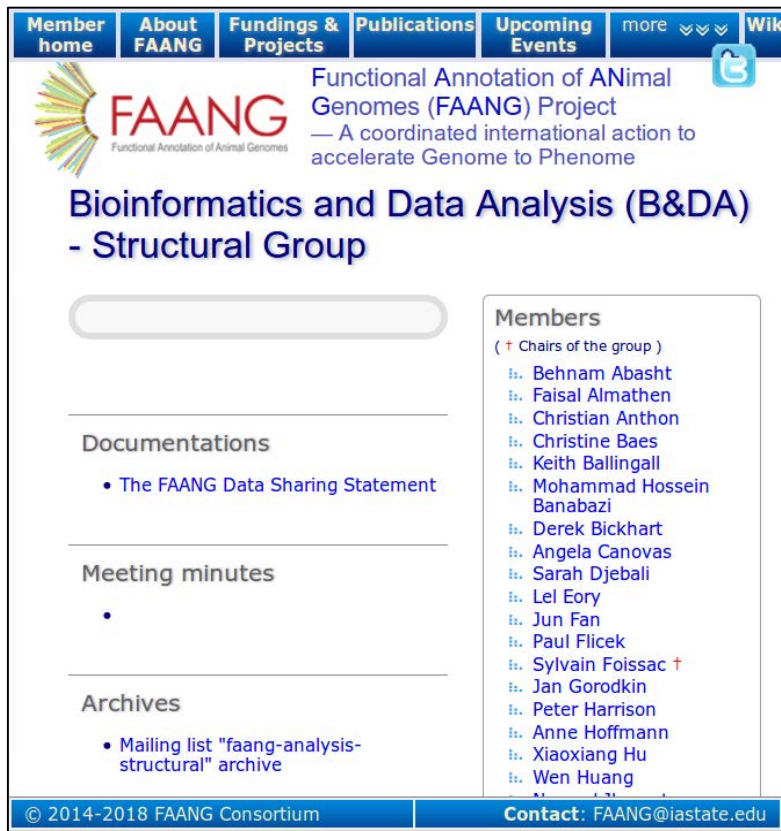


Bioinformatics and Data Analysis

- RNA Analysis working group – Lel Eory
- Methylation working group – Ole Madsen
- Structural working group – Sylvain Foissac
- DNA binding working group – Pablo Ross

- Group organization
 - 53 registered members (including S. Foissac & J. Reecy)
 - Teleconf meetings
 - Reports on FAANG confluence website: www.ebi.ac.uk/seqdb/confluence



Member home | About FAANG | Fundings & Projects | Publications | Upcoming Events | more | Wik

FAANG Functional Annotation of Animal Genomes (FAANG) Project
— A coordinated international action to accelerate Genome to Phenome

Bioinformatics and Data Analysis (B&DA) - Structural Group

Members

- Behnam Abasht
- Faisal Almuthen
- Christian Anthon
- Christine Baes
- Keith Ballingall
- Mohammad Hossein Banabazi
- Derek Bickhart
- Angela Canovas
- Sarah Djebali
- Lel Eory
- Jun Fan
- Paul Flicek
- Sylvain Foissac †
- Jan Gorodkin
- Peter Harrison
- Anne Hoffmann
- Xiaoxiang Hu
- Wen Huang

Documentations

- The FAANG Data Sharing Statement

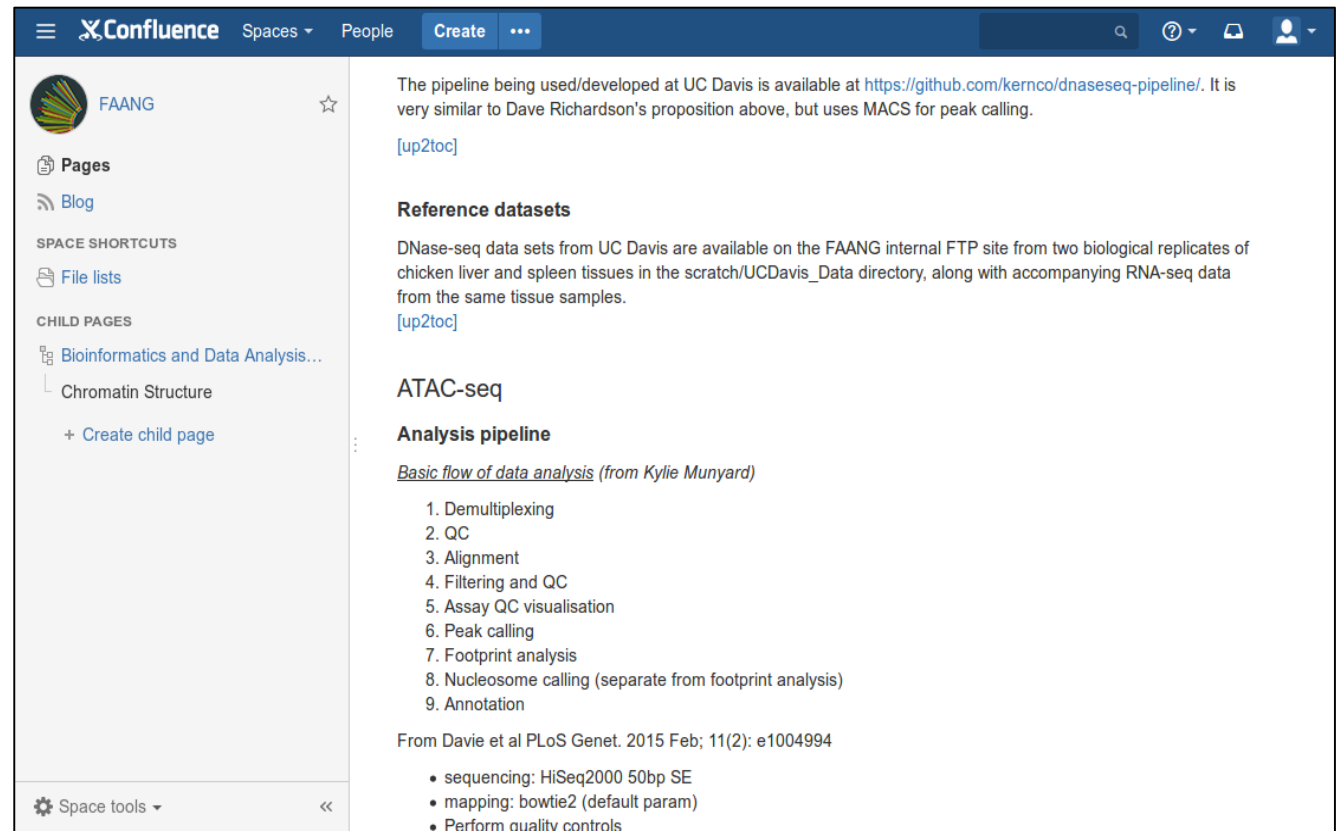
Meeting minutes

-

Archives

- Mailing list "faang-analysis-structural" archive

© 2014-2018 FAANG Consortium | Contact: FAANG@iastate.edu



Confluence Spaces People Create

FAANG

Pages

Blog

SPACE SHORTCUTS

File lists

CHILD PAGES

- Bioinformatics and Data Analysis...
 - Chromatin Structure
 - + Create child page

The pipeline being used/developed at UC Davis is available at <https://github.com/kernco/dnaseq-pipeline/>. It is very similar to Dave Richardson's proposition above, but uses MACS for peak calling.

[up2toc]

Reference datasets

DNase-seq data sets from UC Davis are available on the FAANG internal FTP site from two biological replicates of chicken liver and spleen tissues in the scratch/UCDavis_Data directory, along with accompanying RNA-seq data from the same tissue samples.

[up2toc]

ATAC-seq

Analysis pipeline

Basic flow of data analysis (from Kylie Munyard)

- Demultiplexing
- QC
- Alignment
- Filtering and QC
- Assay QC visualisation
- Peak calling
- Footprint analysis
- Nucleosome calling (separate from footprint analysis)
- Annotation

From Davie et al PLoS Genet. 2015 Feb; 11(2): e1004994

- sequencing: HiSeq2000 50bp SE
- mapping: bowtie2 (default param)
- Perform quality controls

Space tools

Methylation

PAG 2018 updates

Ole Madsen, University of Wageningen

Methylation Pipelines

- The EBI Pipeline:
<https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/Bisulfite+Sequencing+%28BS%29+pipeline>
- Toulouse pipeline WorkflowBS:
<https://github.com/FAANG/faang-methylation/tree/master/workflowbs>
- INRA Paris pipeline: link to poster
https://www.ebi.ac.uk/seqdb/confluence/download/attachments/32192334/RRBS_pipeline_poster.pdf?version=1&modificationDate=1483369795000&api=v2

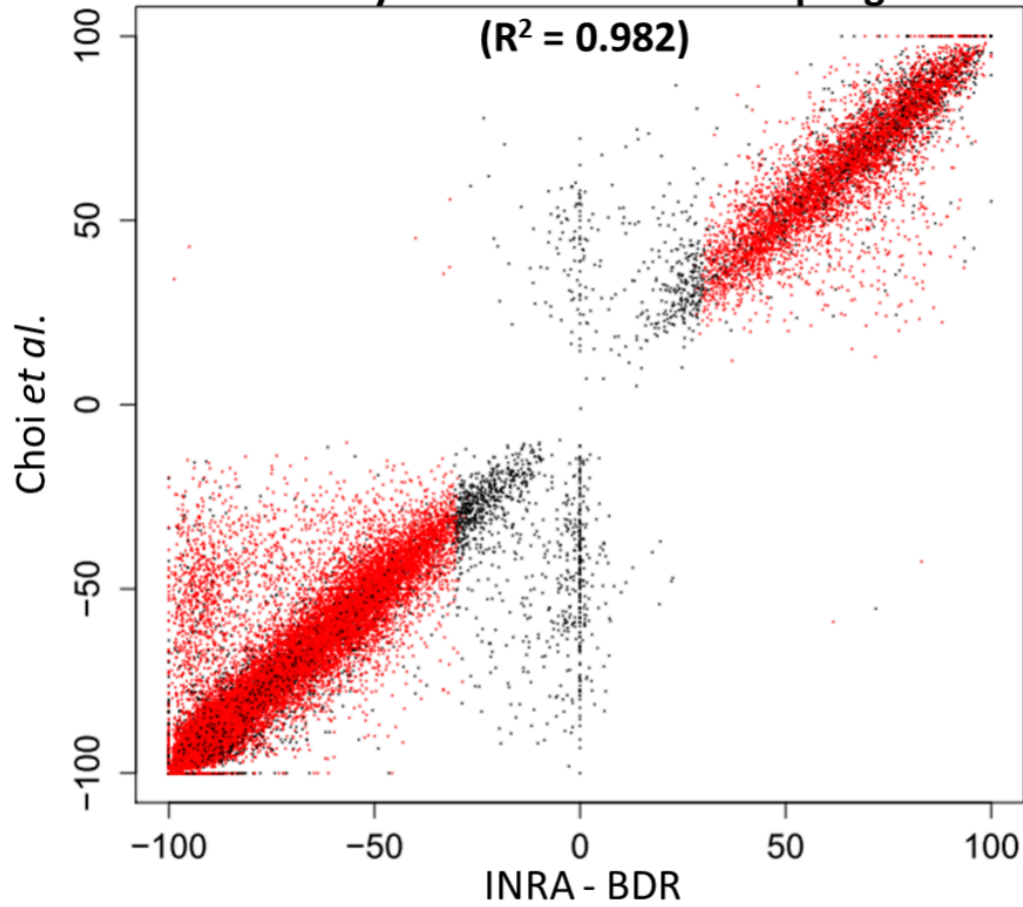
Methylation Pipelines

For future analysis/next steps:

We (Wageningen University) recently made RRBS and WGBS from the same samples which will be used to compare the two methods (preliminary results indicate they seem to complement each other).

Conclusion

Difference in methylation between macrophages and muscle



- : Difference of methylation between the two conditions for all tested CpGs
- : Difference of methylation between the two conditions for DMCs ($q < 0.001$)

Even if alignment softwares and parameters have a major influence on the set of CpGs selected for further analyses, the results obtained on the commonly found CpGs are in good correlation (see graph on the left).

RNA

PAG 2018 updates

Lel Rory, Roslin

Currently 140 members.

Linked with non-coding RNA subgroup led by Jan Gorodkin (at University of Copenhagen), with 56 member within lncRNA group

Web:

<https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/RNA-Seq>

Email:

faang-analysis-rna@animalgenome.org

faang-analysis-lncrna@animalgenome.org

Aims:

- To improve annotation.
- Identify functional categories of novel genes.
- Quantify expression.
- Study gene interactions.

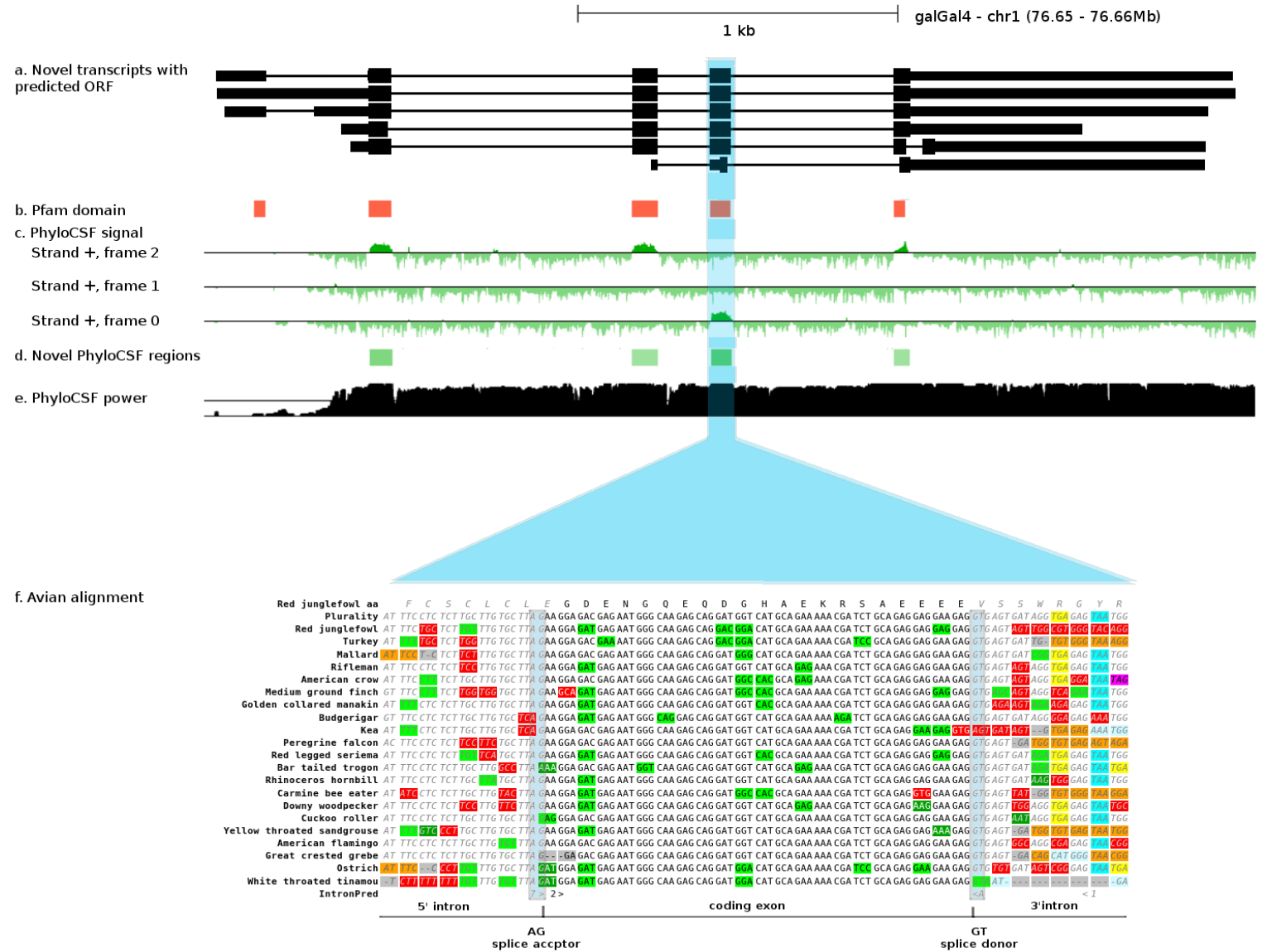
By:

- Contributing to experimental assay requirements and bioinformatics pipeline specifications.
- Benchmarking pipelines.
- Running assays: RNA-seq, small RNA-seq, PacBio Iso-Seq, CAGE.
- Providing reference datasets: for pig and chicken some of these from Roslin pilot projects.

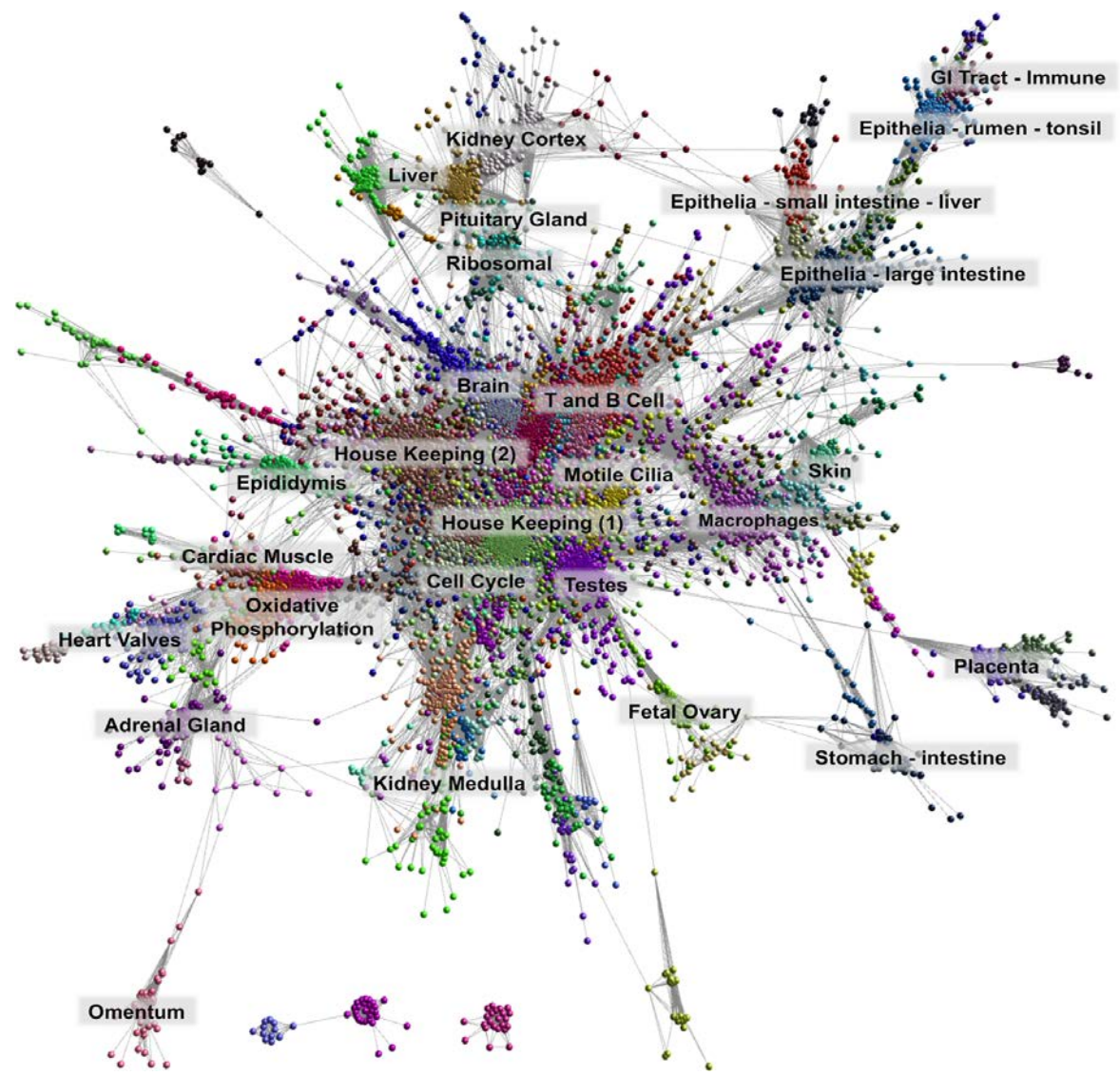
Pipelines:

- RNA-seq: quality control (FastQC), adapter trimming (cutadapt), assembly (Ensembl, HISAT2/StringTie2 or STAR/Cufflinks)
- PacBio Iso-Seq: Iso-Seq pipeline
- micro RNA-seq: read mapping (bowtie2), miRdeep2
- Protein coding potential: FEELnc, PhyloCSF, Pfam, BLAST
- CAGE: demultiplexing, trimming, quality control, mapping (bwa), promoter and expression analysis Paraclu/CAGEr
- Gene expression: RSEM, kallisto
- Gene interaction: Biolayout/Cytoscape

Example of novel transcripts from merged PacBio Iso-Seq and RNA-seq models.



Gene network from the sheep atlas project



Clark et al. 2017, PLoS Genetics



DNA Binding

PAG 2018 updates

Pablo Ross, University of California - Davis

ChIP-seq data Analysis

<https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/ChIP-Seq>

- **ChIP-Seq data acquisition standards**
 - Sequencing: Single end 50bp
 - Read coverage: Narrow peaks: 20 million; Broad peaks: 40 million uniquely mapped reads
 - Input from same sonication batch
 - Biological replicates (at least 2 – same sex)
- **Software tools for analysis and parameter**
 - Quality Control and Trimming
 - Read Mapping
 - Peak calling
- **Available Pipelines**
 - ENCODE ChIP-Seq pipeline: <https://github.com/ENCODE-DCC/chip-seq-pipeline>
 - UC Davis ChIP-Seq pipeline: <https://github.com/kernco/chipseq-pipeline>
 - EpiDB pipeline: <https://github.com/ercfritz/epidb/blob/master/epidb.load.chipseq.pl>

ChIP-seq data Analysis

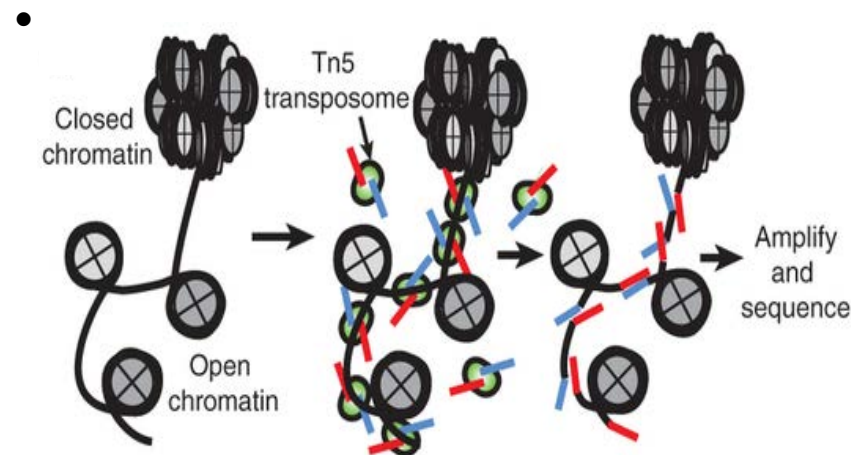
- Currently investigating
 - ChIP-seq data quality metrics
 - Usable fragments (aligned, quality filtered, deduplicated)
 - Non-Redundant Fraction (NRF)
 - PCR Bottlenecking Coefficient (PBC)
 - Cross-Strand Correlation
 - Data reproducibility
 - IDR (mostly for TF)
 - Looking for good alternatives for Histone Marks
- Panned activities
 - Test a common set of parameters (initially defined by the ENCODE pipeline) under different conditions using a common sample dataset and consider modifications as needed.
 - Test the selected parameters in datasets from different animal species
 - Incorporate the selected parameters into easily accessible pipelines implemented in different platforms such as GitHub, CyVerse, Galaxy, etc.

Chromatin structure and 3D genomics

PAG 2018 updates

Sylvain Foissac, INRA

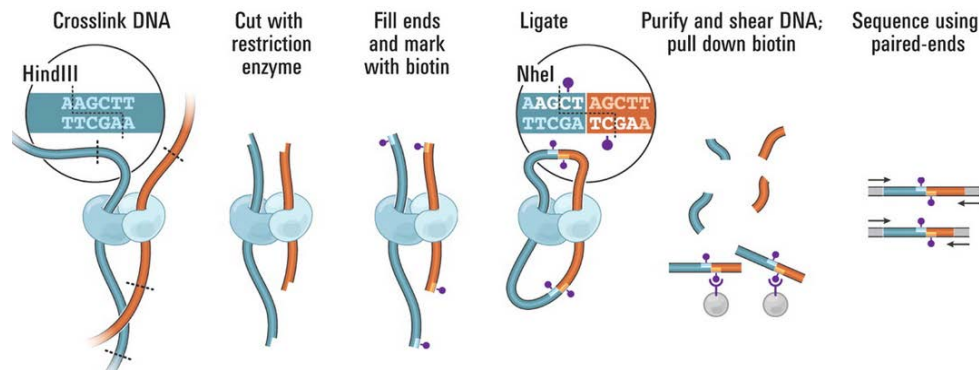
- Protocol & Pipelines



Buenrostro et al, 2013

ATAC-seq

- Read trimming (trimgalore)
- Read mapping (Bowtie2)
- PCR duplicate removal (samtools)
- Mitochondrial read removal (samtools)
- Peak calling (MACS2)
- Peak merging (bedtools)
- Peak quantification (samtools)
- LOESS normalization (csaw)
- Differential analysis (edgeR)



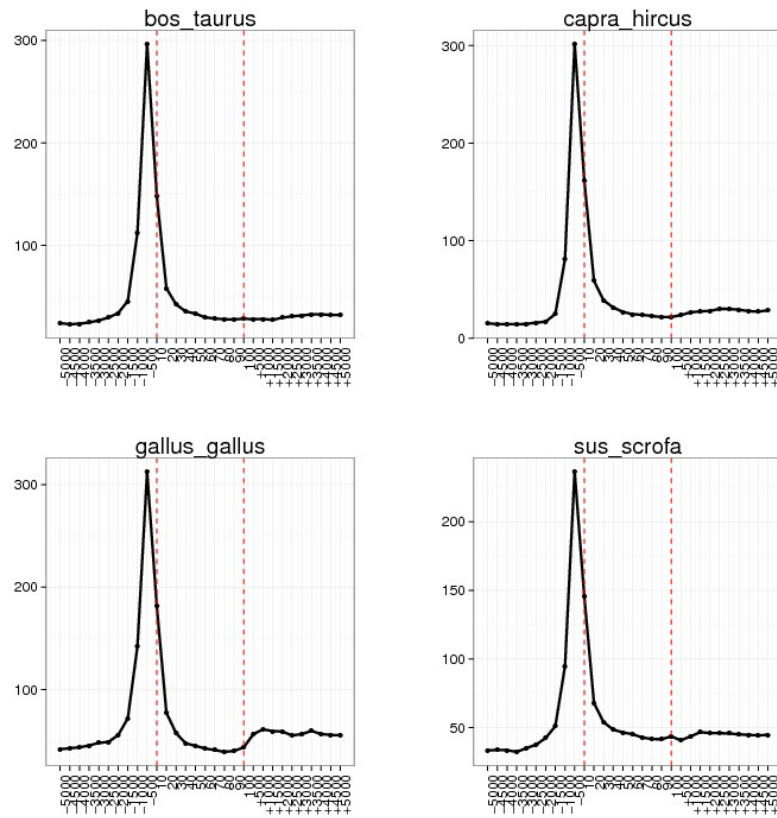
Rao et al, 2014

Hi-C

- Read trimming (cutadapt)
- Read mapping (Bowtie2)
- Invalid pairs filtering (samtools)
- Contact matrix generation (HiC-Pro)
- Matrix balancing normalization (ICE)
- TAD calling (armatus)
- A/B compartments calling (HiT-C)
- Visualization (juicebox)

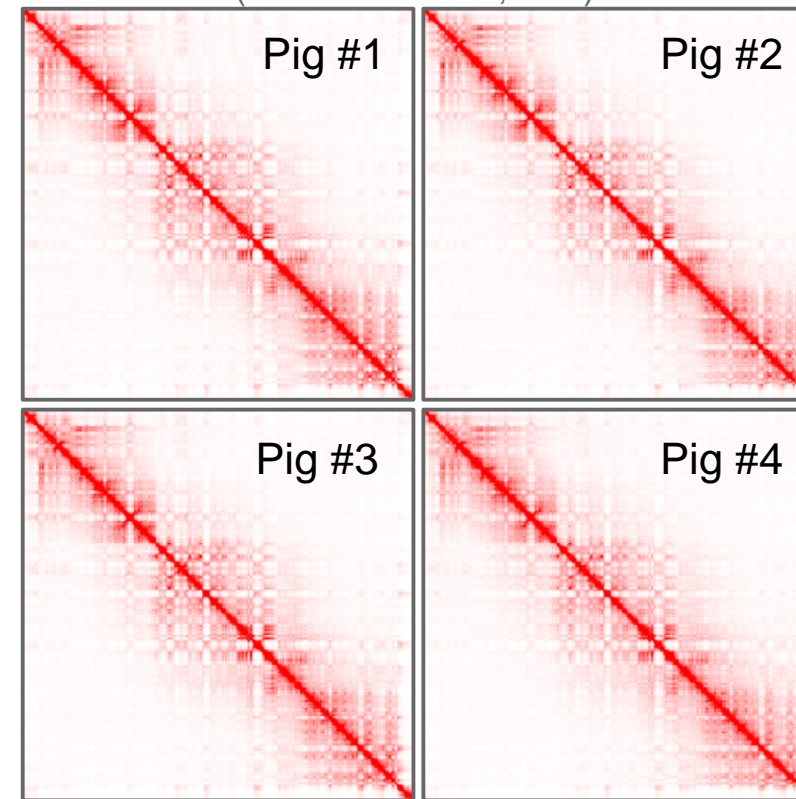
- Results: on Hi-C & ATAC-seq data from FR-AgENCODE pilot project

ATAC-seq gene coverage per species



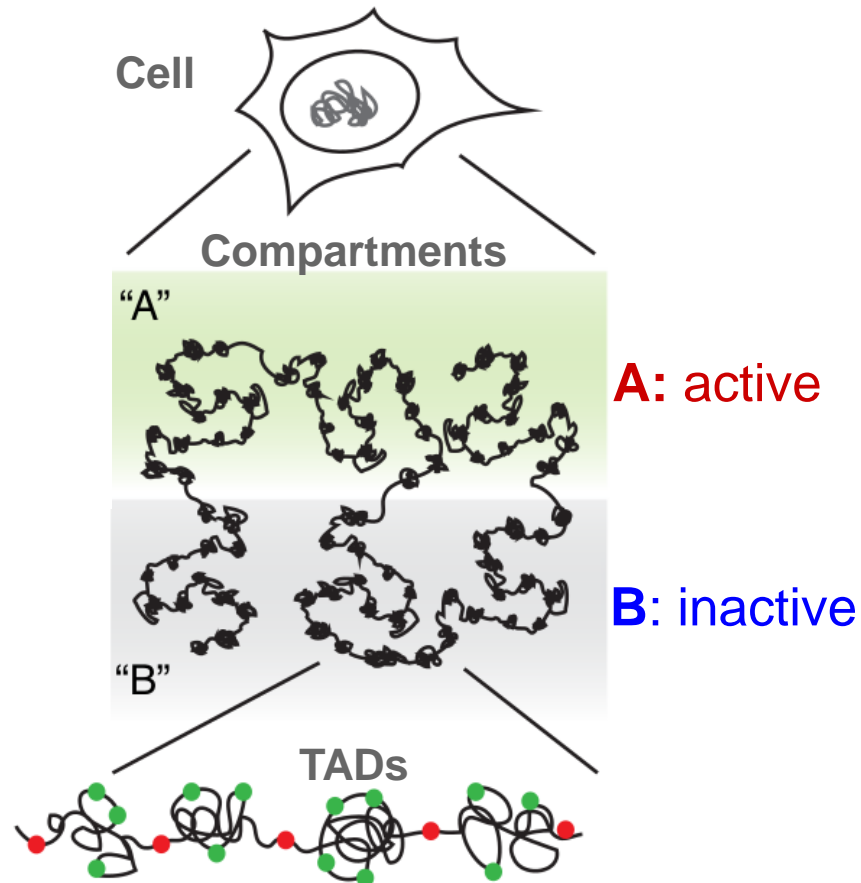
ATAC-seq signal peaks at gene starts (TSS)

Hi-C interaction matrices per replicate
(*Sus scrofa* chr1, liver)

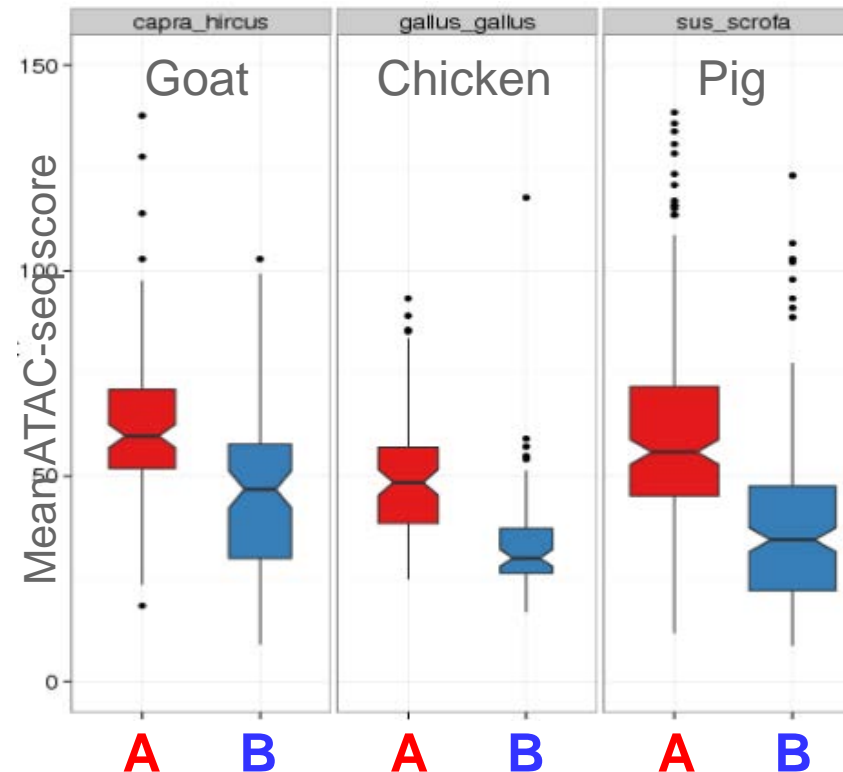


Hi-C signal is resolutive and reproducible

- Results: integrative analysis



Chromatin accessibility (from ATAC-seq) in A vs. B compartments (from Hi-C)



Chromatin is more accessible in active compartments
ATAC-seq and Hi-C pipelines generate consistent results
 => FR-AgENCODE manuscript in prep.