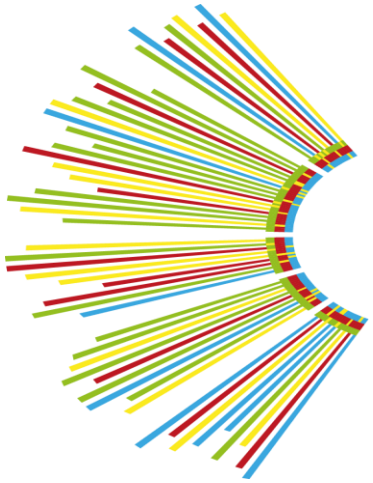
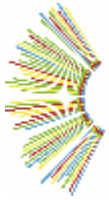


FAANG

Functional Annotation of Animal Genomes

B&DA Committee
Bioinformatics and Data Analysis

PAG January 2016



FAANG

Functional Annotation of Animal Genomes

B&DA – progress so far

- Aim: to define standard bfx pipelines for FAANG data
- Group has skyped many times
- We have a Wiki: hosted by EBI
- Sub-groups
  - RNA: RNA-Seq, miRNA-Seq, full length etc
  - ChIP: ChIP pull down of things bound to DNA
  - Methylation: pipelines to identify methylated regions
  - 3D structure: HiC and other methods
- Barriers: lack of data on which pipeline is best; seeking to compare pipelines; seeking suitable datasets

FAANG
Functional Annotation of Animal Genomes

Analysis issues

- Do we want fuzzy or hard pipelines?
  - If we don't use the same pipelines, when we compare data, we will just find pipeline differences
  - However, we don't want to take years finding the perfect pipeline (there isn't one anyway!)
  - Is a single pipeline enforceable?
  - Can we afford for it not to be?

FAANG
Functional Annotation of Animal Genomes

Other committees

- Are we communicating with the other FAANG committees in the right way?
  - Steering committee
    - Animals, Samples and Assays (ASA)
    - Bioinformatics and Data Analysis (B&DA)
    - Communication (COM)
    - Metadata and Data Sharing (M&DS)

FAANG
Functional Annotation of Animal Genomes

Stimulating collaborations

- How do we stimulate integration and development of future collaborative projects?
- How to we facilitate future activities?
- What do we need?
  - Funding?
  - Meetings?
  - Access to compute infrastructure?
  - Access to data?

FAANG
Functional Annotation of Animal Genomes

Impact

- What are translational impacts of FAANG work on animal production and human health by comparative genomics?
- If we do this project correctly, what will be possible in the future because of it?

- What is the **impact** of what we are doing?

FAANG
Functional Annotation of Animal Genomes

REPORTS FROM 4 SUB-COMMITTEES

FAANG
Functional Annotation of Animal Genomes

CHROMATIN STRUCTURE

FAANG
Functional Annotation of Animal Genomes

FAANG
Functional Annotation of Animal Genomes

B&DA Committee

Bioinformatics and Data Analysis

Chromatin Structure

PAG January 2016

Motivation

Gene expression can be regulated by modification of chromatin compaction and 3D distance between loci

=> interest in profiling chromatin structure & genome topology

Molecular assays

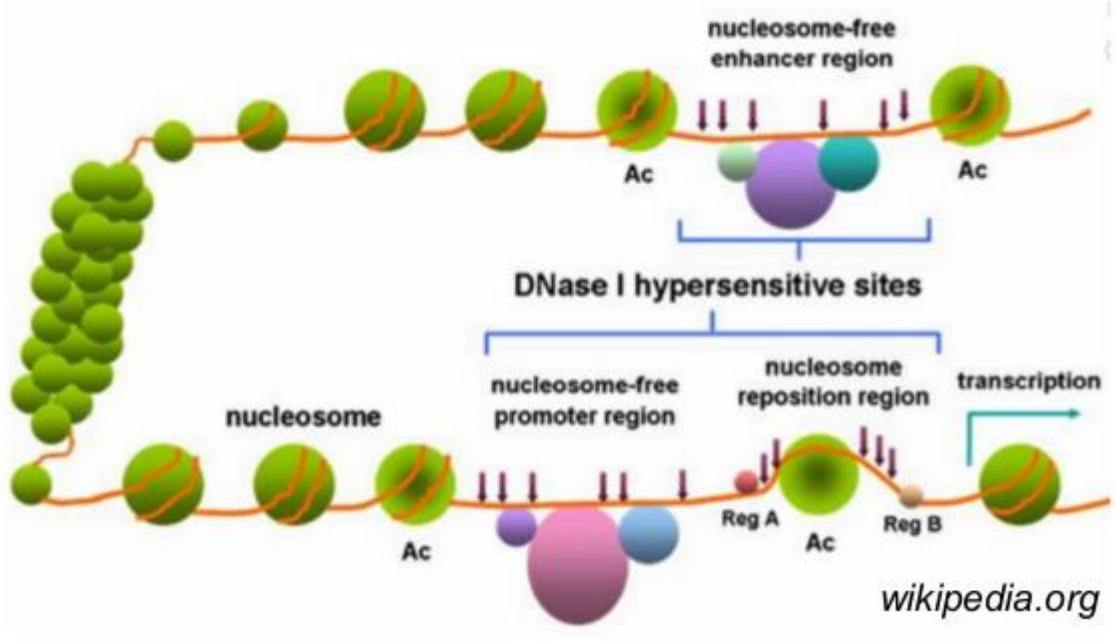
DNAse-seq, ATAC-seq, Hi-C

Activities of the subcommittee

List available tools and datasets, set up and test pipelines, define QC metrics and standard procedures

FAANG
Functional Annotation of Animal Genomes

DNAse-seq



Targets DNase I hypersensitive sites: open chromatin, regulatory elements, transcribed regions

FAANG
Functional Annotation of Animal Genomes

DNAse-seq

Analysis pipeline - ongoing work

- read mapping & trimming (bwa)
- remove duplicates and low-qual (picard, samtools)
- peak calling (MACS2, HotSpot)

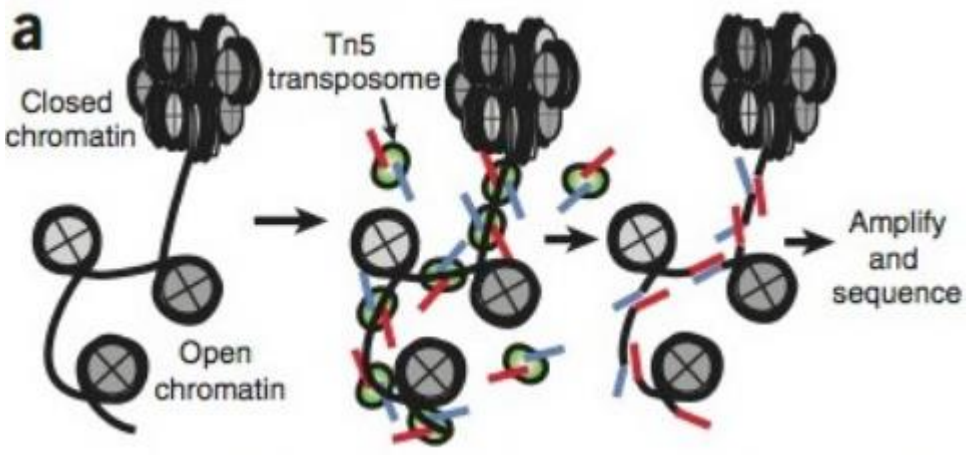
Reference datasets

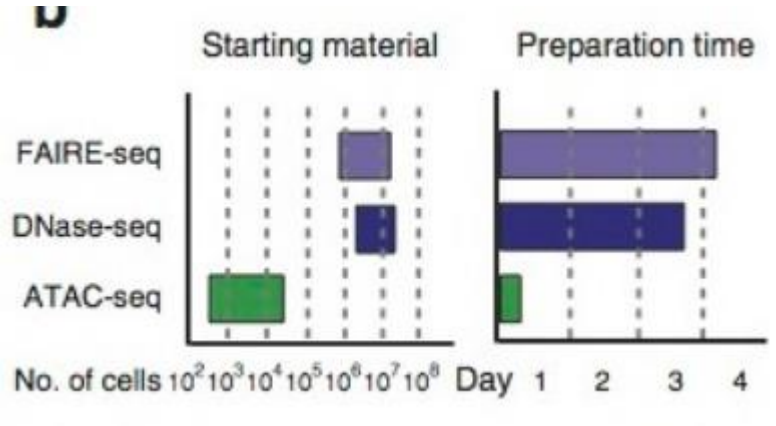
- livestock: none identified
- model organisms: several available from ENCODE, Blueprint, Fantom...

=> build/test pipeline on model organisms data first

FAANG
Functional Annotation of Animal Genomes

ATAC-seq





Buenrostro et al 2013

Transposase-mediated insertion of sequencing adapters in open chromatin

- simple protocol (no ligation)
- low requirements of initial material
- similar to DNAse-seq

FAANG
Functional Annotation of Animal Genomes

ATAC-seq

Analysis pipeline - ongoing work

- read trimming (trim_galore/cutadapt)
- read mapping (bowtie2)
- remove duplicates and low-qual (picard, samtools)
- peak calling (MACS2)

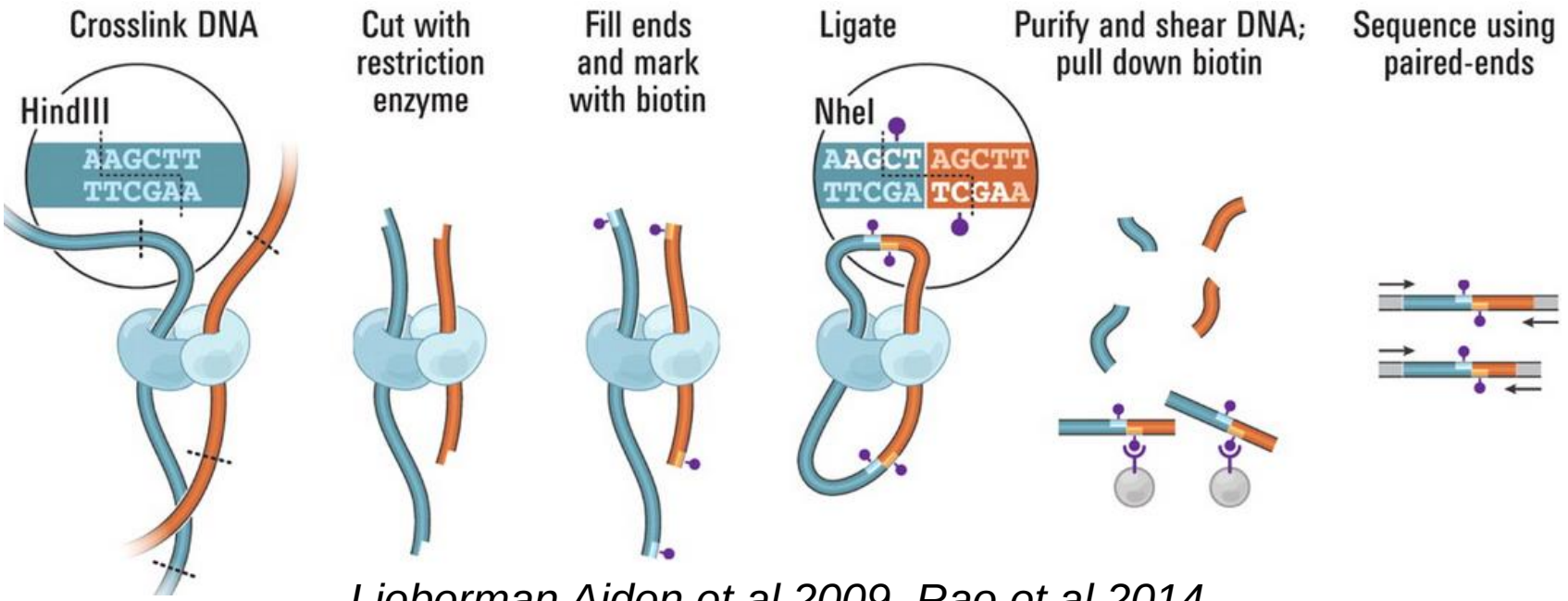
Reference datasets

- model organisms: human GM12877 cells
- (from Buenrostro et al 2013)
- livestock: none publicly available but several in production in pig, cattle, chicken and goat (FAANG French pilot project FR-AgENCODE)

=> preliminary results: poster P0420

FAANG
Functional Annotation of Animal Genomes

Hi-C



Lieberman Aiden et al 2009, Rao et al 2014

- genome-wide detection of proximal pairs of loci
- in 3D nuclear space
- generates contact matrix
- detects inter- and intra-chromosomal interactions

FAANG
Functional Annotation of Animal Genomes

Hi-C

Analysis pipeline - ongoing work

- read mapping and trimming (bowtie2, cutadapt)
- remove duplicates, low-qual and inconsistent mapping (samtools)
- generate and normalize contact matrix (ICE)
- identify Topologically Associated Domains
- workflow implementation: HiC-Pro, HiTC, HiFive

Reference datasets

- model organisms: human IMR90 (Dixon et al 2012), mouse CH12 (Rao et al 2014)
- livestock: none publicly available but several in production in pig, chicken, cattle and goat (FAANG French pilot project FR-AgENCODE)

FAANG
Functional Annotation of Animal Genomes

=> preliminary results: poster P0420

METHYLATION

FAANG
Functional Annotation of Animal Genomes

FAANG
Functional Annotation of Animal Genomes

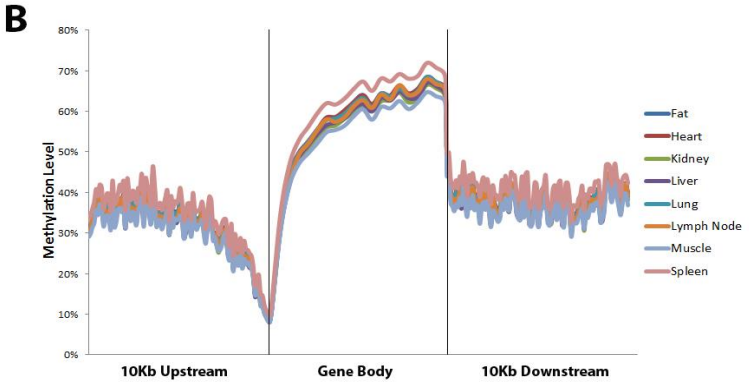
B&DA Committee

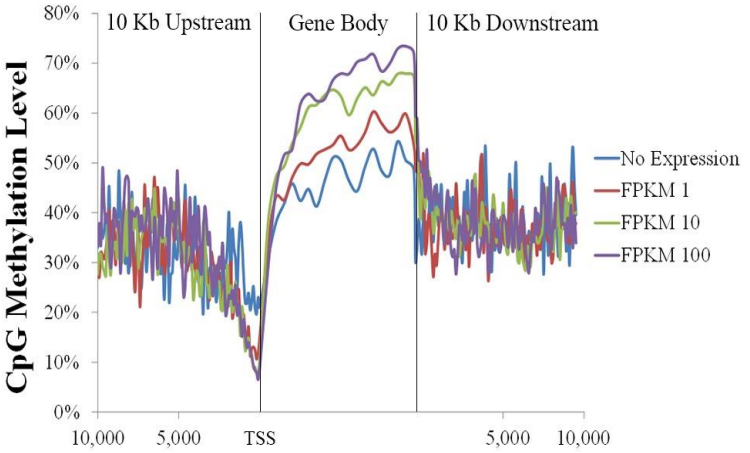
Bioinformatics and Data Analysis

Methylation

PAG January 2016

Presenter: Kyle Schachtschneider and Ole Madsen





FAANG
Functional Annotation of Animal Genomes

FAANG
Functional Annotation of Animal Genomes

Methylation

- **35 members Methylation group (January 2016)**
- **Aiming at one group meeting every month**
  - Decisions on pilot phase made
    - **Data**
    - **Tissue(s)**
    - **Species**
    - **Pipelines to test**
- **Needed? Data to analyse**

FAANG
Functional Annotation of Animal Genomes

Methylation

- PILOT FASE

- **DATA**
  - Whole genome bisulfite sequencing
  - Reduced representation bisulfite sequencing (RRBS)
    - Aiming at 30X for both types of data
- **Tissue(s)**
  - Liver from (healthy) adults (maybe also from embryo)
    - Minimum of three biological replicates
- **Species**
  - Chicken (bird) and a mammal (Cow, pig, sheep?)
- **Pipelines**
  - BLUEPRINT pipeline (provided by Simon Heath)
  - Toulouse pipeline (http://ng6.toulouse.inra.fr/, NG6)
  - Other(s)....?

Methylation

- PILOT FASE – info available on Confluence
- **Optimal fragment size/species RRBS (Toulouse)**
  - Genomes used:
    - capra_hircus: CHIR_1.0
    - gallus_gallus: Galgal4.80
    - mus_musculus: GRCm38.p4.81
    - sus_scrofa: Sscrofa10.2.80
    - bos_taurus: UMD3.1.80
  - Enzyme used:
    - Msp1
    - Taq1
    - Both (double digestion)
- **Software tools for Methylation analysis**

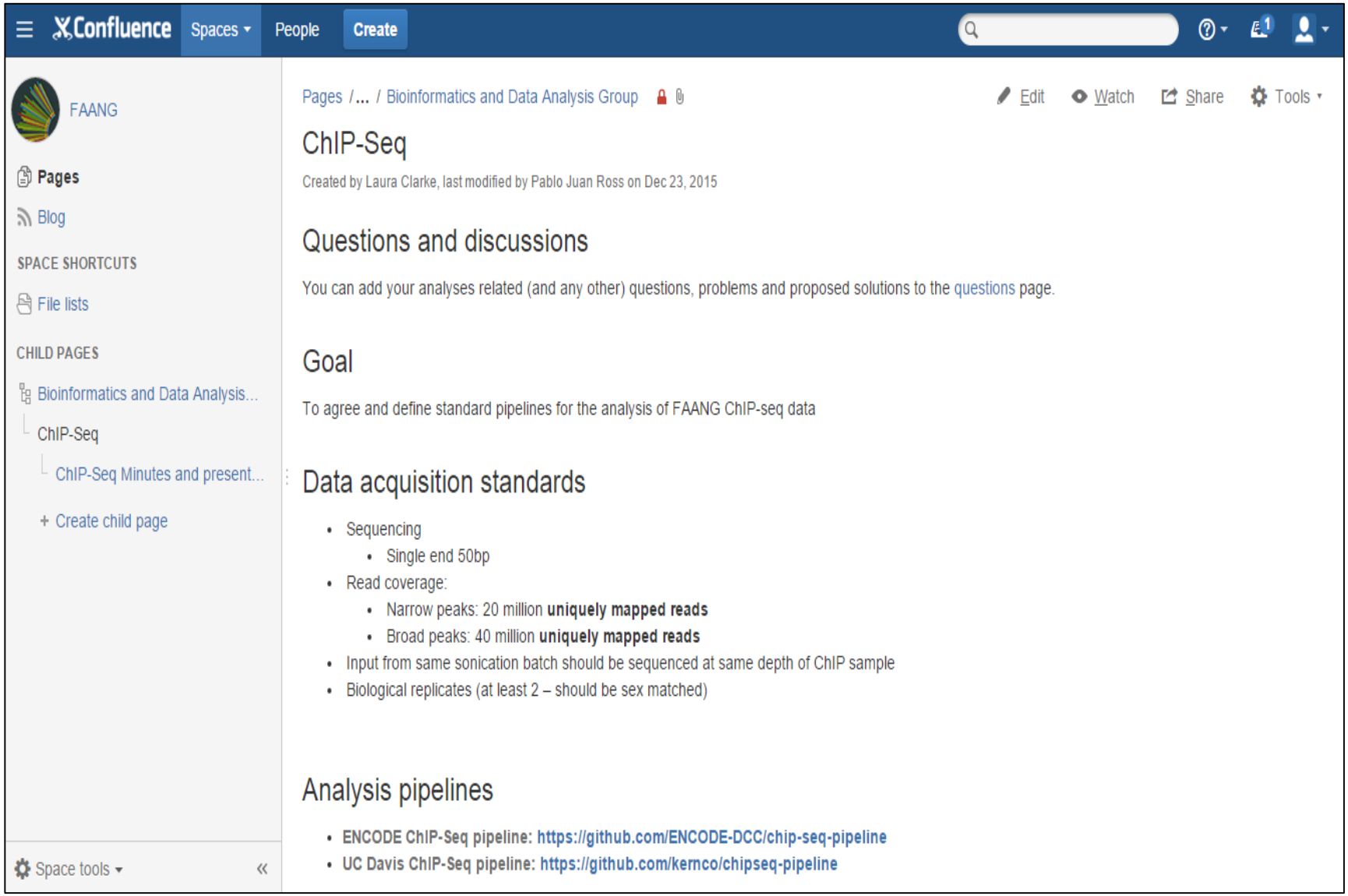
CHIP-SEQ

FAANG
Functional Annotation of Animal Genomes

ChIP-seq Data Analysis Standards

To agree and define standard pipelines for the analysis of FAANG ChIP-seq data

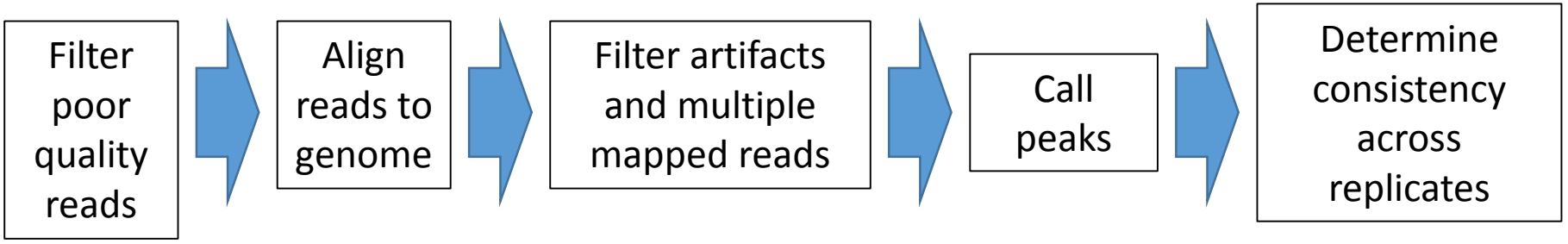
January 11th – FAANG Workshop at PAGXXIV


Confluence
Spaces
People
Create
FAANG
Pages
Blog
SPACE SHORTCUTS
File lists
CHILD PAGES
Bioinformatics and Data Analysis...
ChIP-Seq
ChIP-Seq Minutes and present...
+ Create child page
Space tools
Pages / ... / Bioinformatics and Data Analysis Group
Edit
Watch
Share
Tools
ChIP-Seq
Created by Laura Clarke, last modified by Pablo Juan Ross on Dec 23, 2015
Questions and discussions
You can add your analyses related (and any other) questions, problems and proposed solutions to the questions page.
Goal
To agree and define standard pipelines for the analysis of FAANG ChIP-seq data
Data acquisition standards
Sequencing
Single end 50bp
Read coverage:
Narrow peaks: 20 million uniquely mapped reads
Broad peaks: 40 million uniquely mapped reads
Input from same sonication batch should be sequenced at same depth of ChIP sample
Biological replicates (at least 2 – should be sex matched)
Analysis pipelines
ENCODE ChIP-Seq pipeline: https://github.com/ENCODE-DCC/chip-seq-pipeline
UC Davis ChIP-Seq pipeline: https://github.com/kernco/chipseq-pipeline

Data Acquisition Standards

- Sequencing
  - Single end 50 bp
- Read coverage:
  - Narrow peaks: 20 million **uniquely mapped reads**
  - Broad peaks: 40 million **uniquely mapped reads**
- Input from same sonication batch should be sequenced at similar depth of ChIP sample
- Biological replicates (at least 2 – should be sex matched)

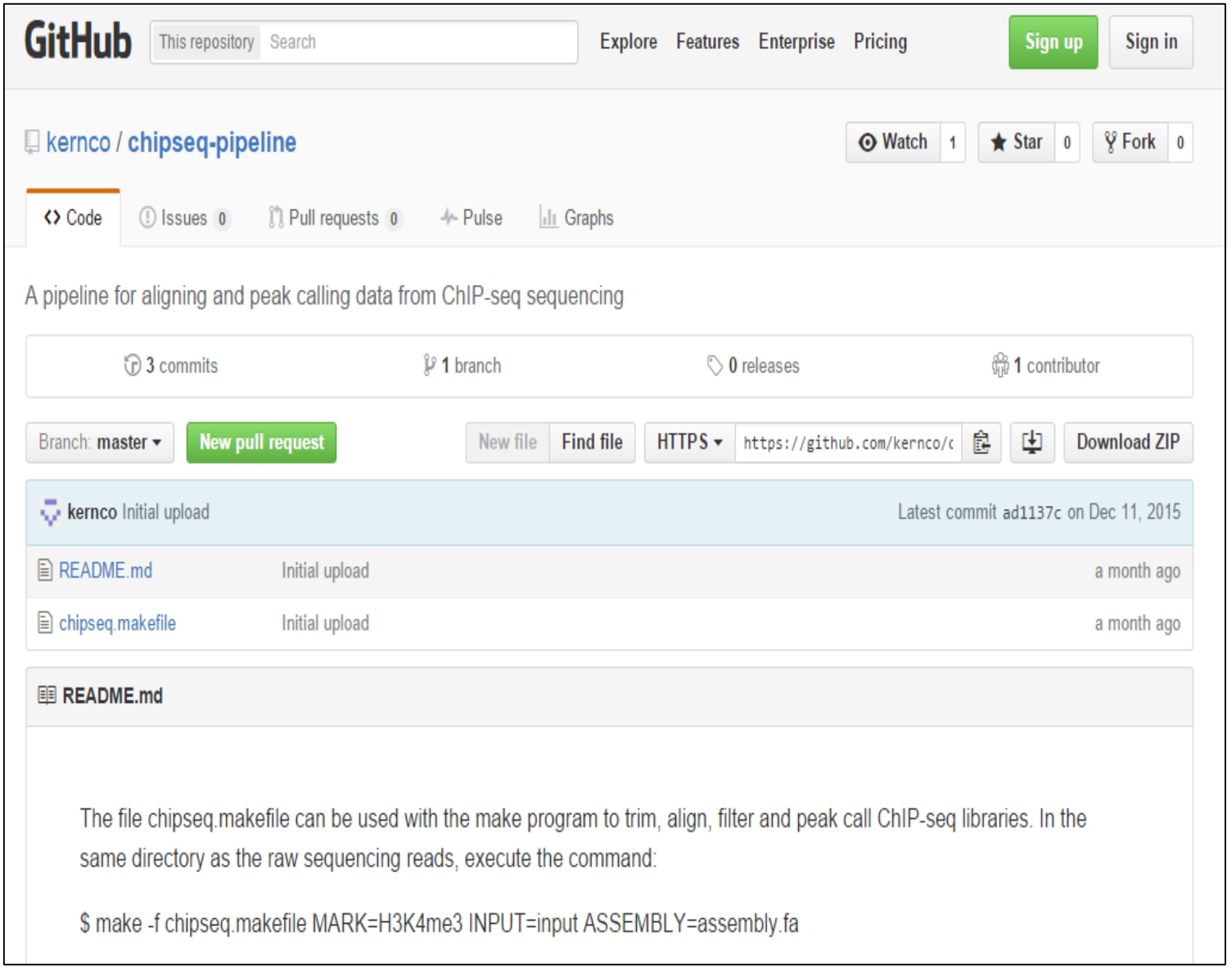
ChIP-seq analysis pipelines



- **ENCODE ChIP-Seq pipeline: https://github.com/ENCODE-DCC/chip-seq-pipeline**

- **UC Davis ChIP-Seq pipeline: https://github.com/kernco/chipseq-pipeline**

- **Your pipeline: https://github.com/PLEASE SUBMIT ASAP**


GitHub
This repository
Search
Explore
Features
Enterprise
Pricing
Sign up
Sign in
kernco / chipseq-pipeline
Watch 1
Star 0
Fork 0
Code
Issues 0
Pull requests 0
Pulse
Graphs
A pipeline for aligning and peak calling data from ChIP-seq sequencing
3 commits
1 branch
0 releases
1 contributor
Branch: master
New pull request
New file
Find file
HTTPS
https://github.com/kernco/c
Download ZIP
kernco Initial upload
Latest commit ad1137c on Dec 11, 2015
README.md
Initial upload
a month ago
chipseq.makefile
Initial upload
a month ago
README.md
The file chipseq.makefile can be used with the make program to trim, align, filter and peak call ChIP-seq libraries. In the same directory as the raw sequencing reads, execute the command:
$ make -f chipseq.makefile MARK=H3K4me3 INPUT=input ASSEMBLY=assembly.fa

UC Davis
pipeline: https://github.com/kernco/chipseq-pipeline

Make file: $ make -f chipseq.makefile MARK=H3K4me3 INPUT=input ASSEMBLY=assembly.fa

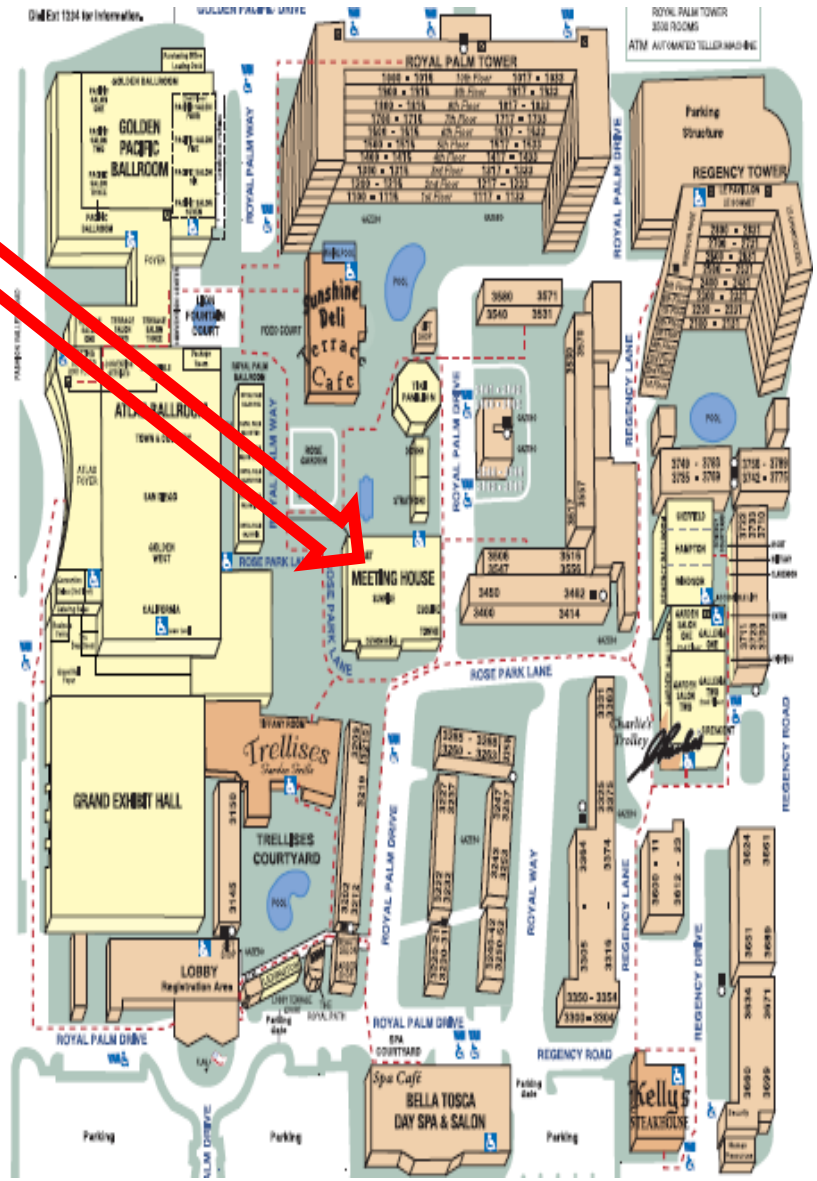
Requires 2 data files: ChIPed-sample.fastq and Input-sample.fastq; and a genome assembly.fa

- Trim raw reads (Illumina adapters and quality trimming-Trimmomatic)
- Align with BWA
- Mark duplicate alignments (picardtools)
- Peak calling using MACS2 (narrow and broad)
- IDR - Irreproducible Discovery Rate

ChIP-seq Analysis group meeting

- Tuesday 2PM – Devonshire Room

- Define a strategy to test ChIP-seq pipelines uploaded to the wiki



RNA

FAANG
Functional Annotation of Animal Genomes

Define bioinformatics pipelines for:

- Transcript discovery for animal genomes
  - location
  - strand
  - isoforms
  - function
- Quantify gene expression

FAANG
Functional Annotation of Animal Genomes

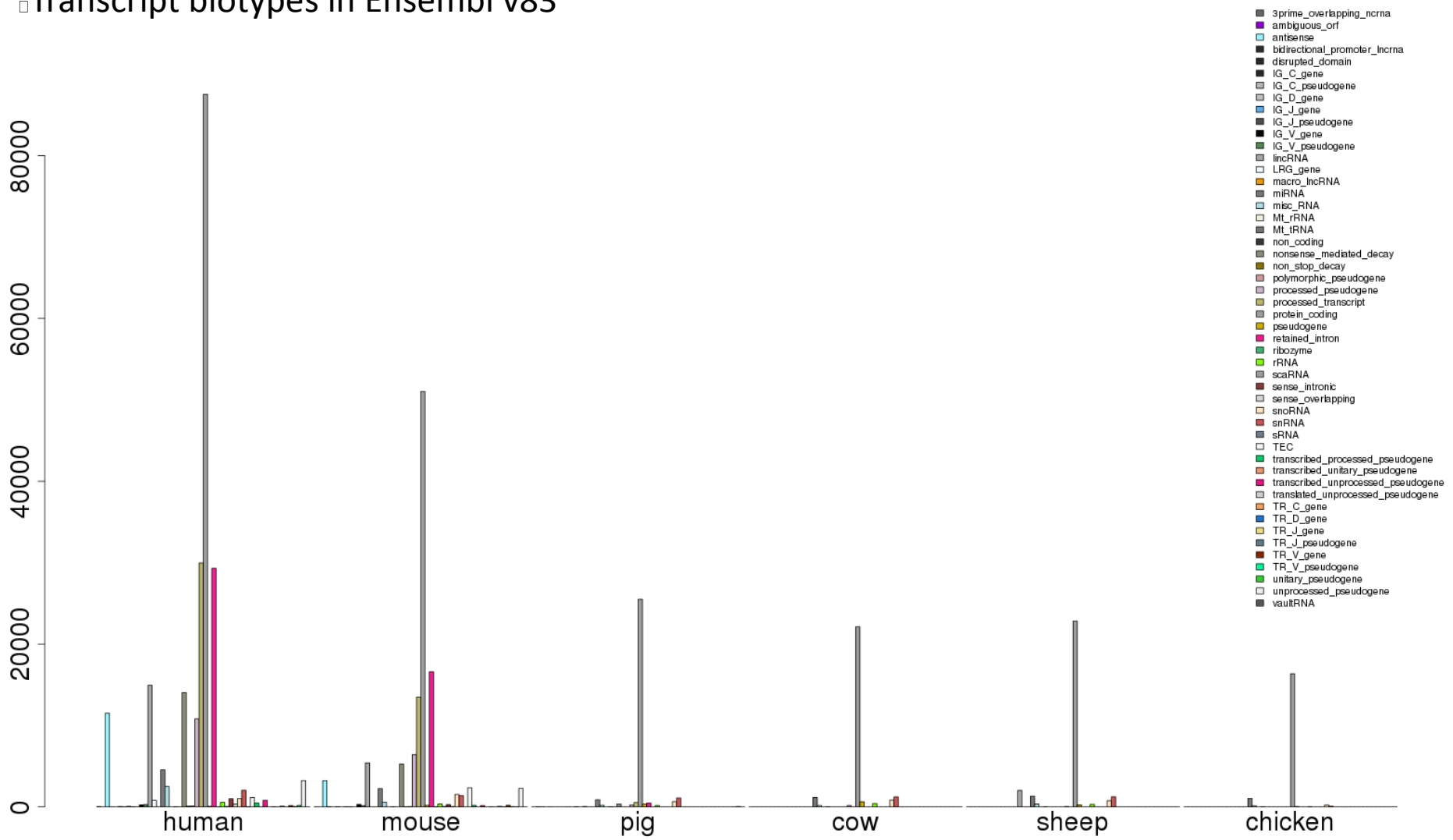
- Short read
  - RNA-seq
  - CAGE
  - PolyA-seq
- Long read (PacBio)

FAANG
Functional Annotation of Animal Genomes

- protein-coding
- pseudogene
- miRNA
- tRNA
- lncRNA
- snoRNA
- rRNA...

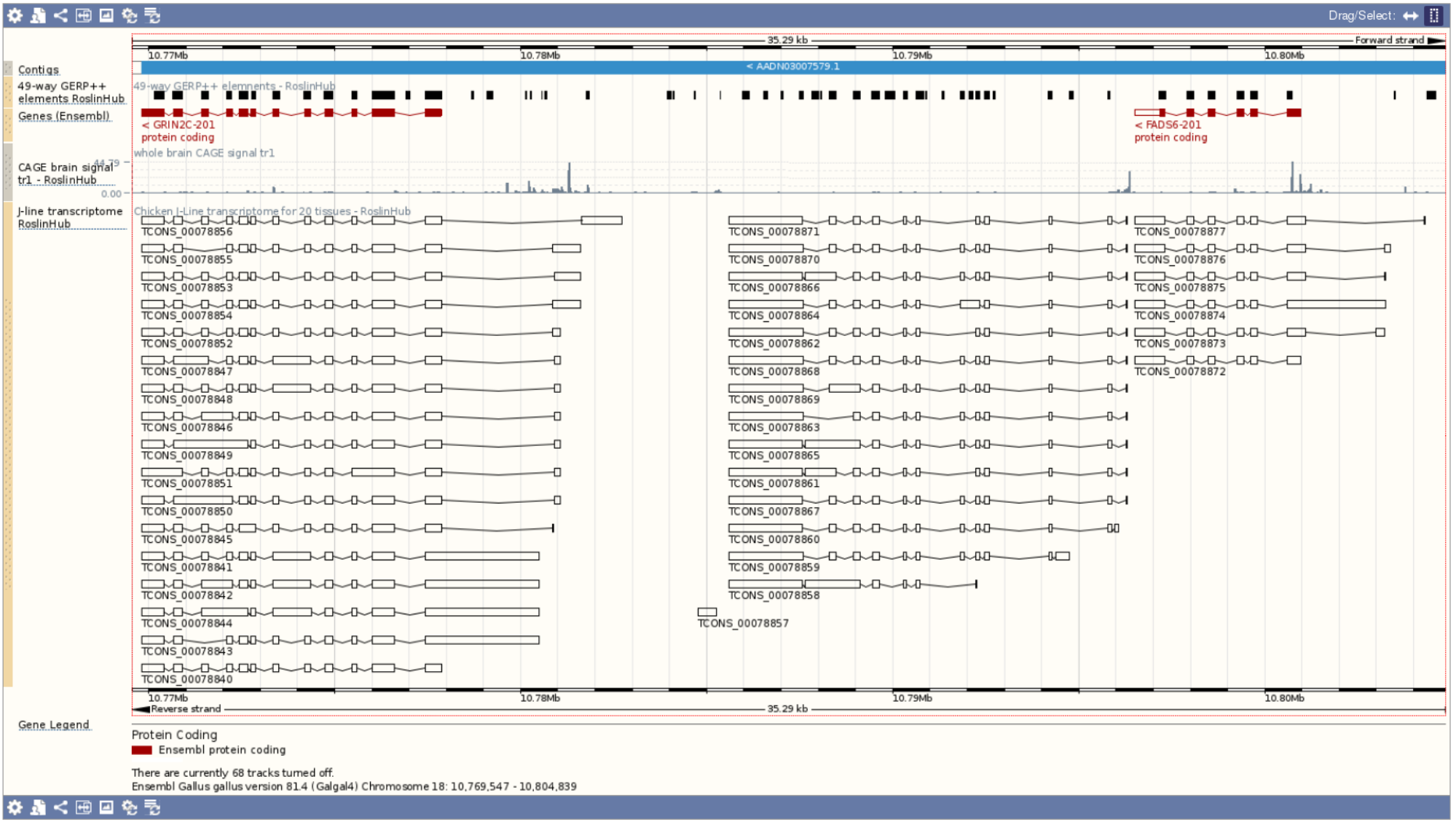
FAANG
Functional Annotation of Animal Genomes

Transcript biotypes in Ensembl v83



FAANG
Functional Annotation of Animal Genomes

FAANG
Functional Annotation of Animal Genomes



FAANG
Functional Annotation of Animal Genomes

- Data to benchmark
- Pipelines for transcript reconstruction
  - Reference guided (example in FAANG wiki)
  - De-novo (example in FAANG wiki)
  - Mapping PacBio long read data
- Pipelines for quantification
  - Performance comparison of software tools (e.g. Roberts & Watson, 2015) for RNA-seq
  - CAGE
- Pipelines for functional classification

FAANG
Functional Annotation of Animal Genomes

- Informal meeting tomorrow
- Venue: **Dover Room**
- Time: **9:30-11:30**

FAANG
Functional Annotation of Animal Genomes

Thank you!

Join us:

http://www.faang.org/groups?name=analysis

FAANG
Functional Annotation of Animal Genomes