NGS Data Analysis and Galaxy

University of Pretoria
Pretoria, South Africa
14-18 October 2013

Dave Clements, Emory University
http://galaxyproject.org/

Fourie Joubert, Burger van Jaarsveld
Bioinformatics & Computational Biology Unit
University of Pretoria
http://science.up.ac.za/

This Week

Monday	Welcome, Project Intro, Basic Galaxy Usage NGS QualityControl
Tuesday	RNA-Seq - Mapping and Transcript Prediction RNA-Seq: Differential expression and Alternative Pipelines; SNP & Variant Analysis
Wednesday	SNP & Variant Analysis Chip-Seq Analysis
Thursday	Genome Assembly Install your own Galaxy on Amazon Cloud
Friday	Customizing Galaxy, Galaxy Tool Shed, and Wrapping Tools for Galaxy

Wednesday Agenda

8:30	Welcome and Questions
8:45	SNP and Variant Analysis, Part III
10:30	Break
11:00	SNP and Variant Analysis, Part IV
12:30	Lunch
13:30	ChIP-Seq I
15:00	Break
15:30	ChIP-Seq II
16:30	Done

Wednesday Agenda

8:30	Welcome and Questions
8:45	SNP and Variant Analysis, Part III
10:30	Break
11:00	SNP and Variant Analysis, Part IV
12:30	Lunch
13:30	ChIP-Seq I
15:00	Break
15:30	ChIP-Seq II
16:30	Done

ChIP-Seq: FASTQ data and quality control

http://scriptogr.am/ohofmann

By Shannan Ho Sui

Create a new history

Shared Data → Data Libraries

→**ChIP-Seq Datasets**

Select everything in the Filtered Reads folder

Also grab genes_chr12.gtf from

UC Davis RNA-Seq Human library

ChIP-Seq Exercise: Mapping with Bowtie

Using Bowtie directly (could also use BWA)

NGS Mapping: → Map with Bowtie for Illumina

FASTQ file → H1hesc Nanog Rep1 chr12 qualityfiltered

Single End, hg19_chr12

Full Parameter List

Suppress all alignments for a read if ... → 1

ChIP-Seq Analysis: remove unmapped reads

SAM Tools → Filter SAM

- Click Add a new Flag
- Set Type to The read is unmapped
- Set flag state to No.

ChIP-Seq Analysis: Put mapped reads in BAM

- SAM Tools → SAM-to-BAM
- Also get the the control (already mapped for us)
  - Shared Data → Data Libraries → Aligned → Import H1hesc Input Rep1 Chr12 Mapped into current history

ChIP-Seq Analysis: Find Peaks

NGS: Peak Calling → MACS

Experiment name → MACS NanogRep1

Tag File → Nanog Rep1 BAM file

Control File → Input Rep1 BAM file

Tag Size → 36

Save shifted raw tag count → ... Save (leave res. at 10)

Check Perform the new peak detection method (futuredir)

ChIP-Seq Analysis: Visualize Results

Look at the HTML report dataset

Launch a Trackster visualization and bring in

- the called peaks
- the Treatment WIG
- the Control WIG
- the gene definitions

ChIP-Seq Analysis: Replicates

Shared Data → Data Libraries → ChIP-Seq Datasets → MACS Outputs

Import Peaks files for

- Nanog Rep 2
- Pou5f1 Rep 1
- Pou5f1 Rep 2

ChIP-Seq Analysis: Unify Replicates

Operate on Genomic Intervals → Concatenate

Concatenate Nanog Rep 1 and 2

Operate on Genomic Intervals → Cluster

Use default parameters

Rename the output dataset

Finally, add the Nanog Rep 2 Peaks, and the cluster output to your visualization

ChIP-Seq Analysis: Unify Replicates

Repeat for Pou5f1 replicates

Operate on Genomic Intervals → Concatenate

Concatenate Pou5f1 Rep 1 and 2

Operate on Genomic Intervals → Cluster

Use default parameters

Rename the output dataset

Finally, add the Pou5f1 Rep 1 & 2 Peaks, and the cluster output to your visualization

ChIP-Seq Analysis: Differential binding

Operate on Genomic Intervals → Subtract

- First dataset clustered → Pou5f1
- Second dataset clustered → Nanog
- Return → Intervals with no overlap

ChIP-Seq Mapping With MACS
Further reading & Resources

ChIP-Seq: FASTQ data and quality control

by Shannan Ho Sui

HAIB TFBS ENCODE collection

MACS Documentation

Model-based analysis of ChIP-Seq (MACS)

by Zhang *et al.*

Cistrome and Nebula Galaxy Servers

Nebula Tutorial

by Valentina Boeva

Wednesday Agenda

8:30	Welcome and Questions
8:45	SNP and Variant Analysis, Part III
10:30	Break
11:00	SNP and Variant Analysis, Part IV
12:30	Lunch
13:30	ChIP-Seq I
15:00	Break
15:30	ChIP-Seq II
16:30	Done

Thanks

Dave Clements

Galaxy Project
Emory University

clements@galaxyproject.org