NGS Data Analysis and Galaxy

University of Cape Town
Cape Town, South Africa
21-25 October 2013

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

Gerrit Botha
Computational Biology
University of Cape Town
http://cbio.uct.ac.za/

CBIO Computational Biology @ UCT

AFRICAN CENTRE FOR GENE TECHNOLOGIES

Galaxy

amazon web services™

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

South Africa Galaxy Workshop Tour

This Week

Monday	Welcome, Project Intro, Basic Galaxy Usage NGS QualityControl
Tuesday	RNA-Seq - Mapping and Transcript Prediction RNA-Seq: Differential expression and Alternative tools
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

9:00 Welcome and Questions

9:15 SNP and Variant Analysis, Part I

11:00 Break

11:30 SNP and Variant Analysis, Part II

13:00 Lunch

14:00 ChIP-Seq I

15:30 Break

16:00 ChIP-Seq II

17:00 Done

ChIP-Seq: FASTQ data and quality control

http://scriptogr.am/ohofmann

By Shannan Ho Sui

H3ABioNet

Look at two transcription factor proteins, Pou5f1 and Nanog, in H1hesc cell lines.

Both are involved in self-renewal of undifferentiated embryonic stem cells.

ChIP-Seq Analysis: Get the Data

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

Bowtie settings to use → 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

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 → H1hesc_Input_Rep1_Chr12_Mapped BAM file

Tag Size → 36

Leave MFOLD → 32

Save shifted raw tag count ... → Save (leave resolution 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 peak files

Operate on Genomic Intervals → Cluster

Use default parameters

Rename the output dataset

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 Peak files

Operate on Genomic Intervals → Cluster

Use default parameters

Rename the output dataset

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

9:00	Welcome and Questions
9:15	SNP and Variant Analysis, Part III
11:00	Break
11:30	SNP and Variant Analysis, Part IV
13:00	Lunch
14:00	ChIP-Seq I
15:30	Break
16:00	ChIP-Seq II
17:00	Done

Thanks

Dave Clements

Galaxy Project
Emory University

clements@galaxyproject.org