

Sequencing-Data from the Public Domain Bioinformatics for Wetlab-Scientists

GRADE Brain Workshop

Vera Laub
PhD candidate
Edinger Institute
17th of May, 2022

Agenda

Day 1: ChIP-seq

1. Introduction
2. Genomics Basics
3. ChIP-seq
4. Public Repositories
5. Data Visualization: IGV
6. Peak Annotation:
ChIPseeker

Day 2: RNA-seq

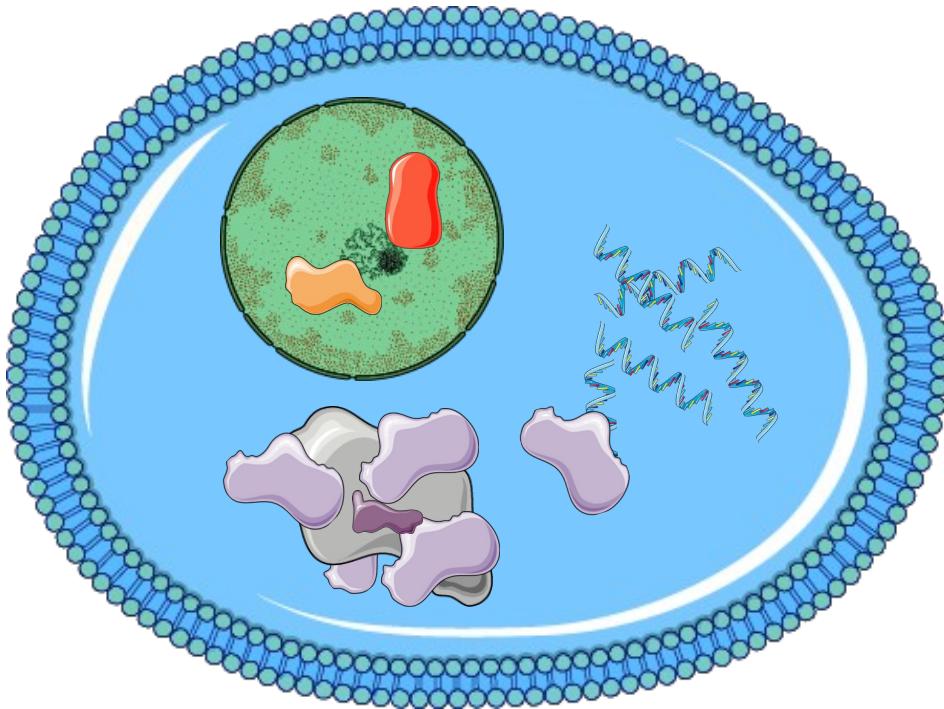
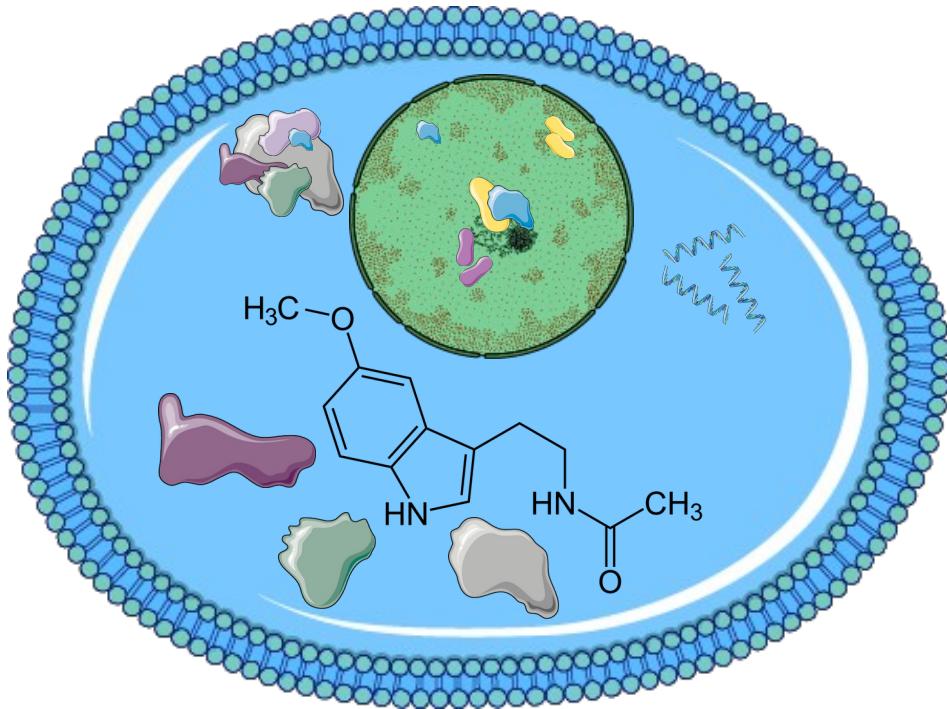
1. Introduction and Recap
2. RNA-seq
3. Data intersection: Venny
4. GO term analysis
 1. PANTHER
 2. STRING
5. Network analysis:
STRING

Coming back from your Vacation

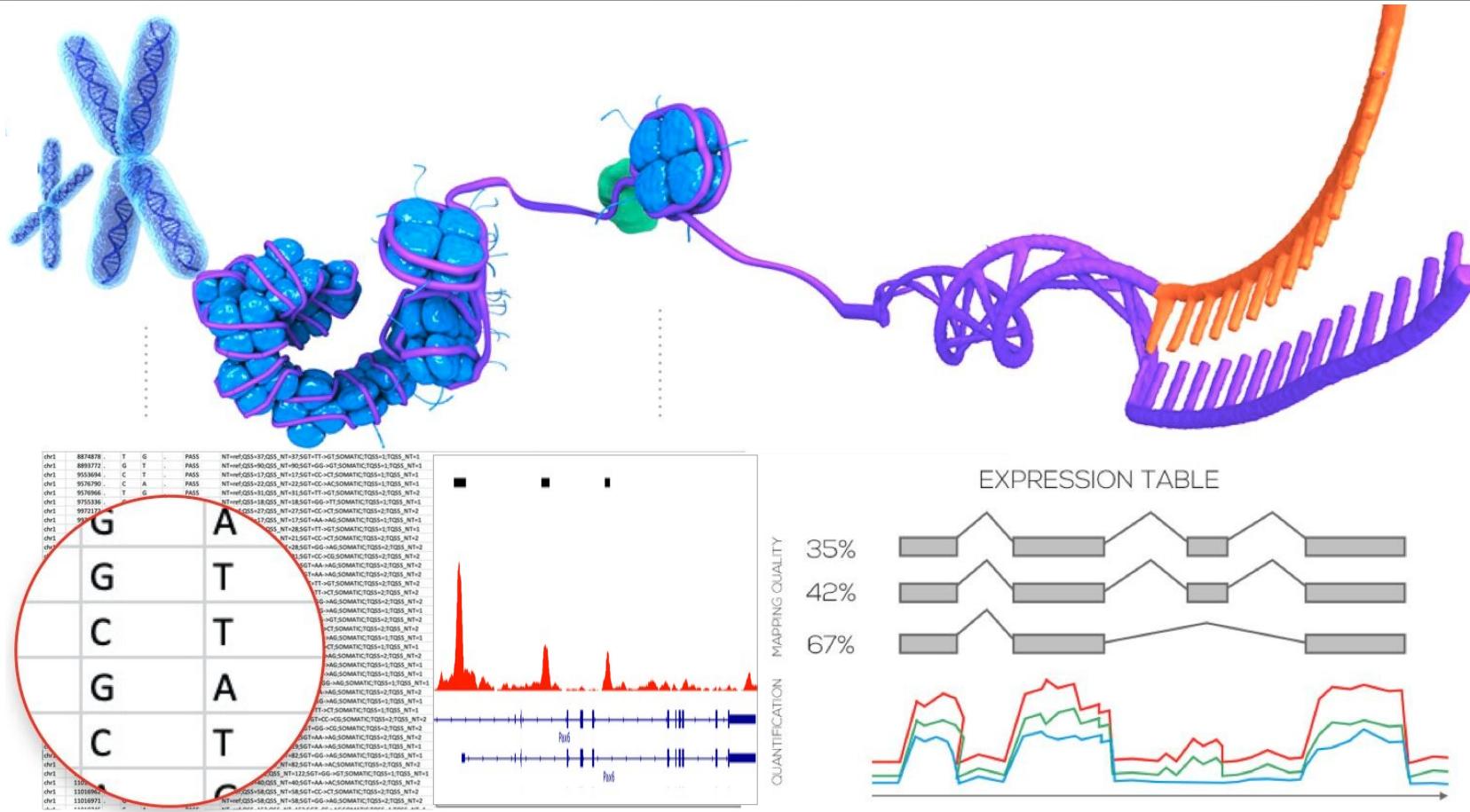


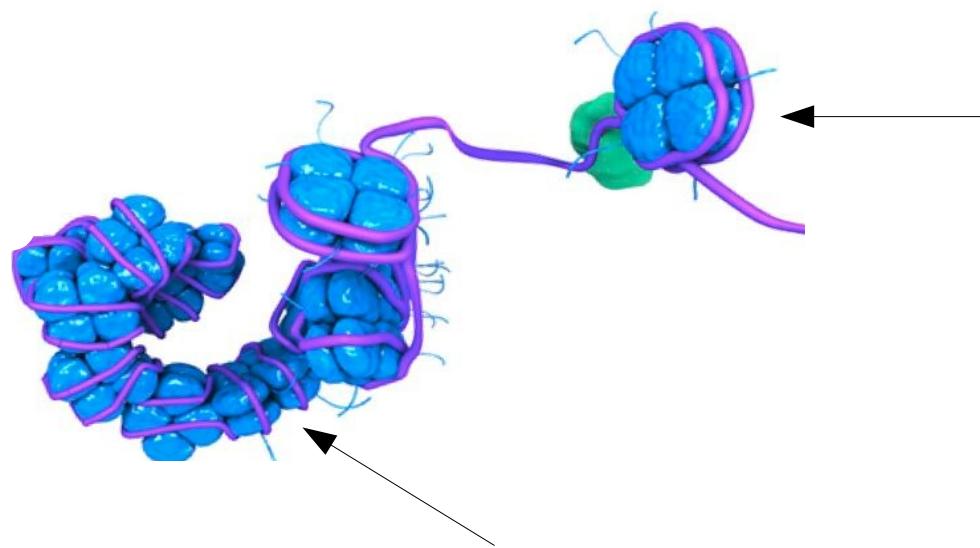


Circadian Rhythm on Molecular Level?



What is Genomics?





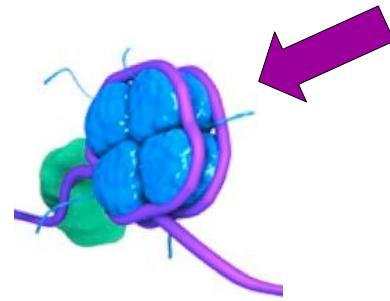
Euchromatin

- Loosely packed
- Transcription possible

Heterochromatin

- Inaccessible to most Transcription Factors
- Densely packed by histones

Histones: Post-translational Modifications

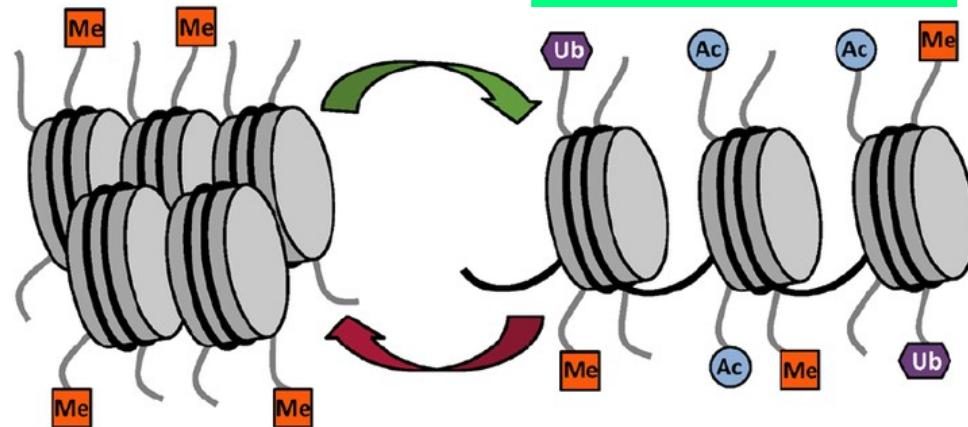


amino acid residue type of modification

H3K27me3

histone 3

position of the amino acid from the N terminal



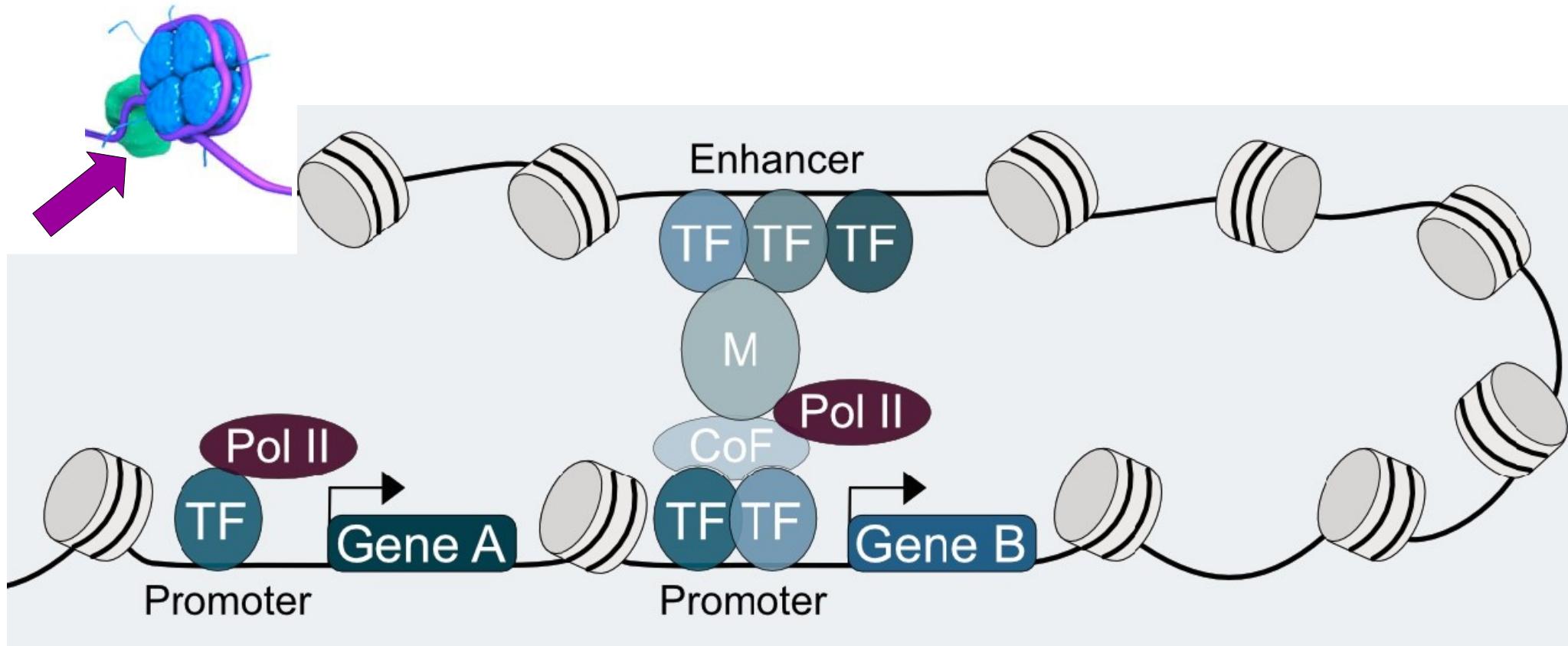
Transcription active

- H3K4me3
- H3K36me3
- H3K27ac

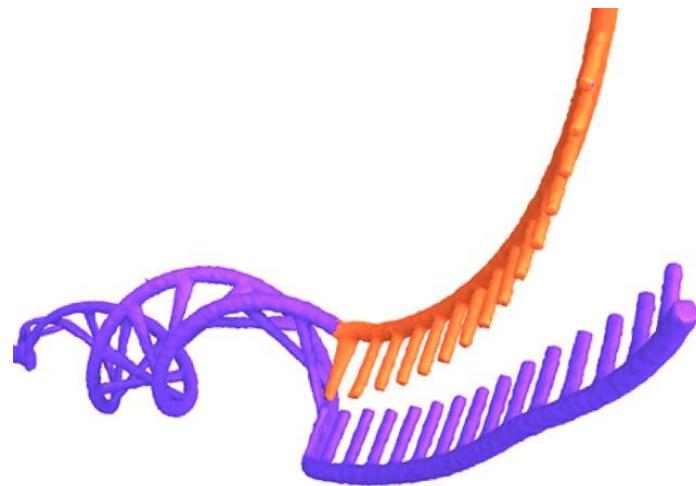
Transcription repressed

- H3K9me3
- H3K27me3

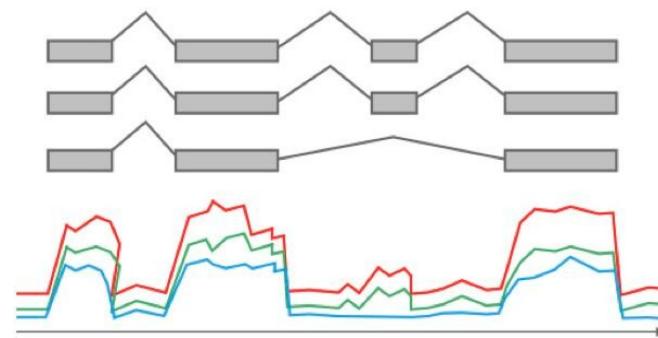
Transcription Machinery

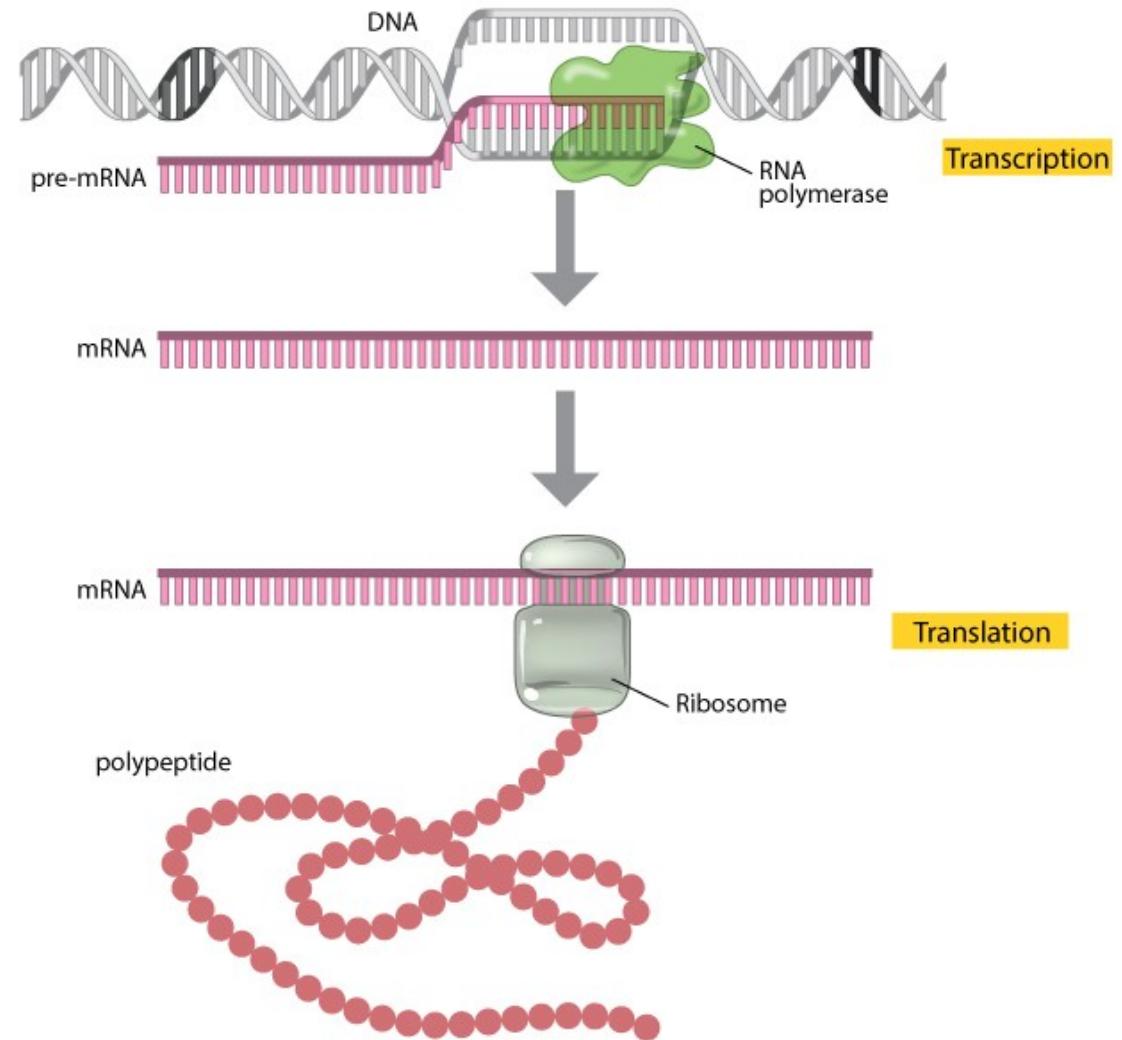
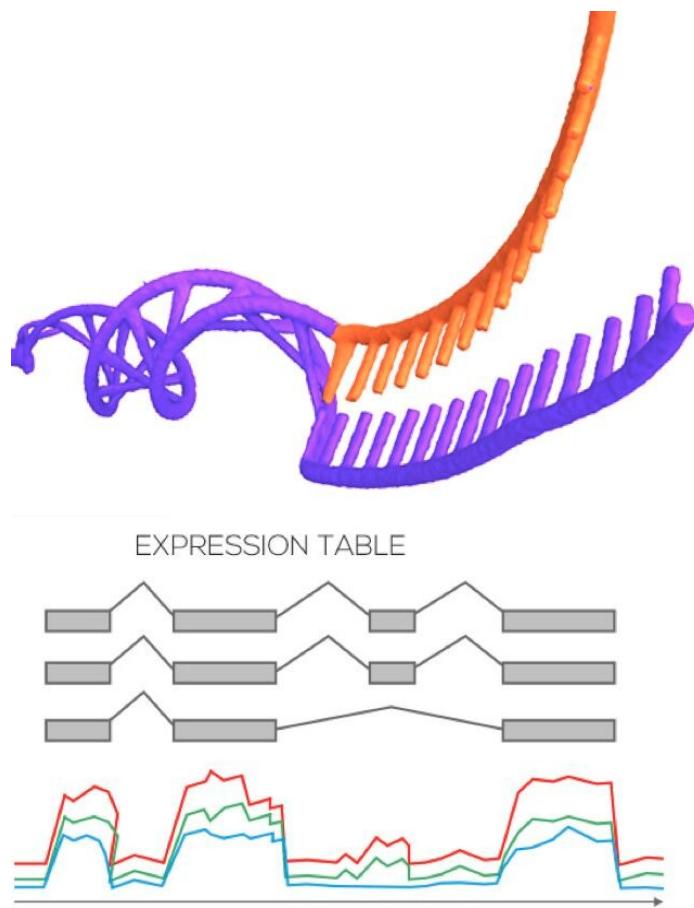


Transcription and Translation

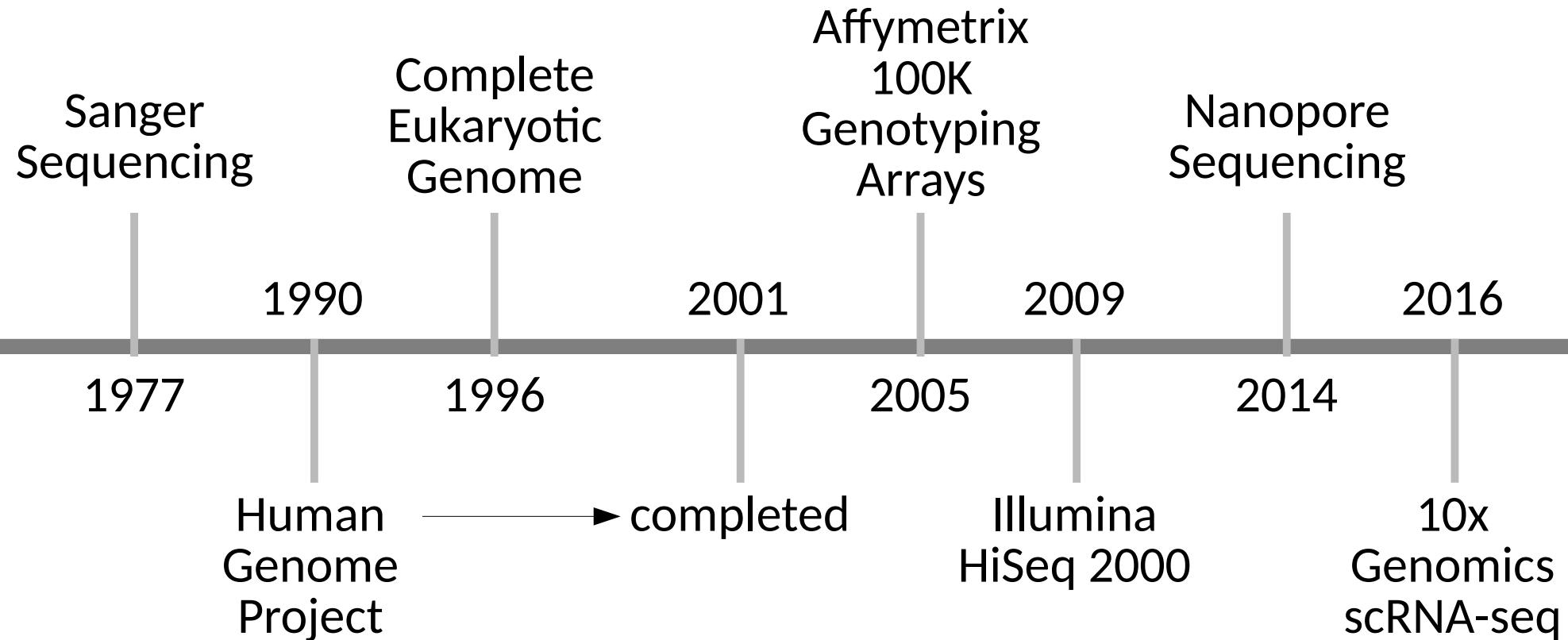


EXPRESSION TABLE

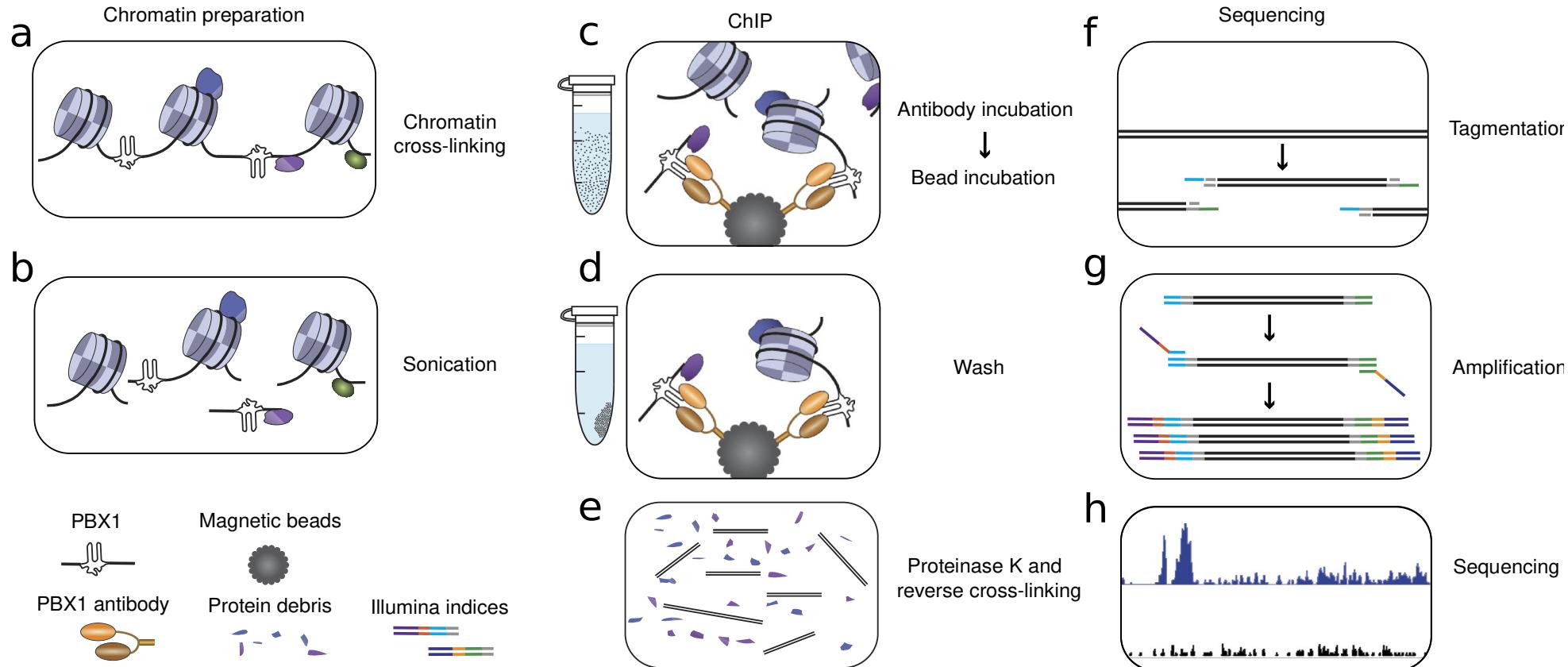




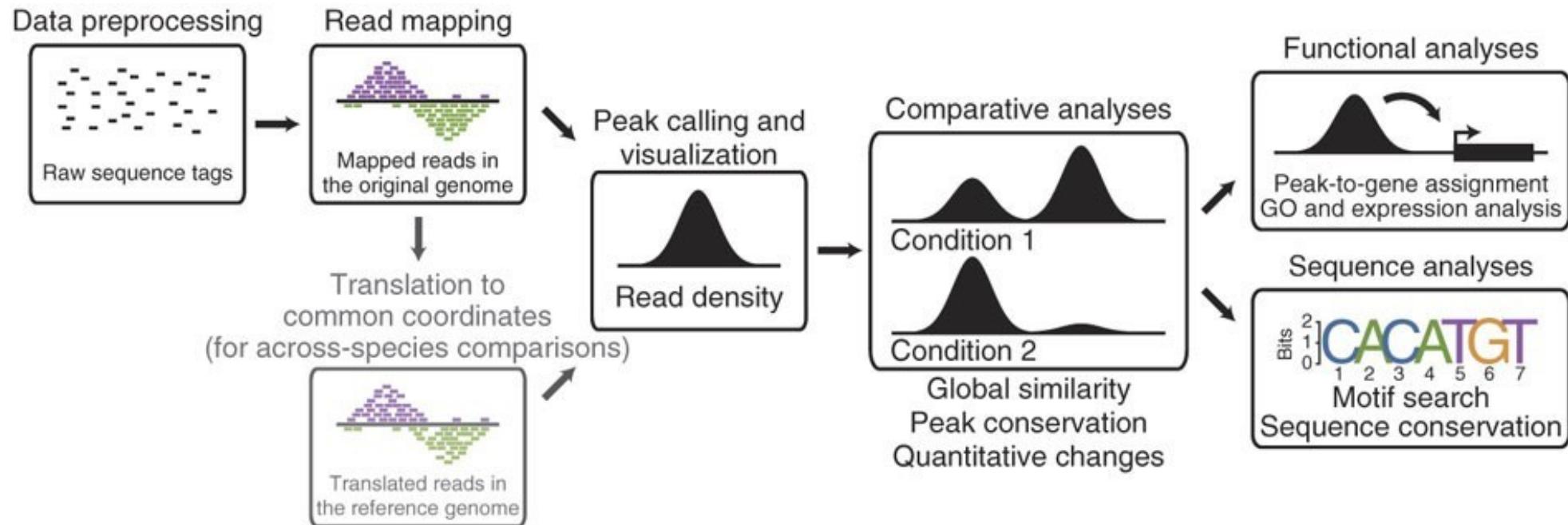
Sequencing Technologies



ChIP-sequencing: Workflow



ChIP-sequencing: Analysis



How are Molecular Interactions studied?

DNA

- ATAC-seq
- Methylome analysis

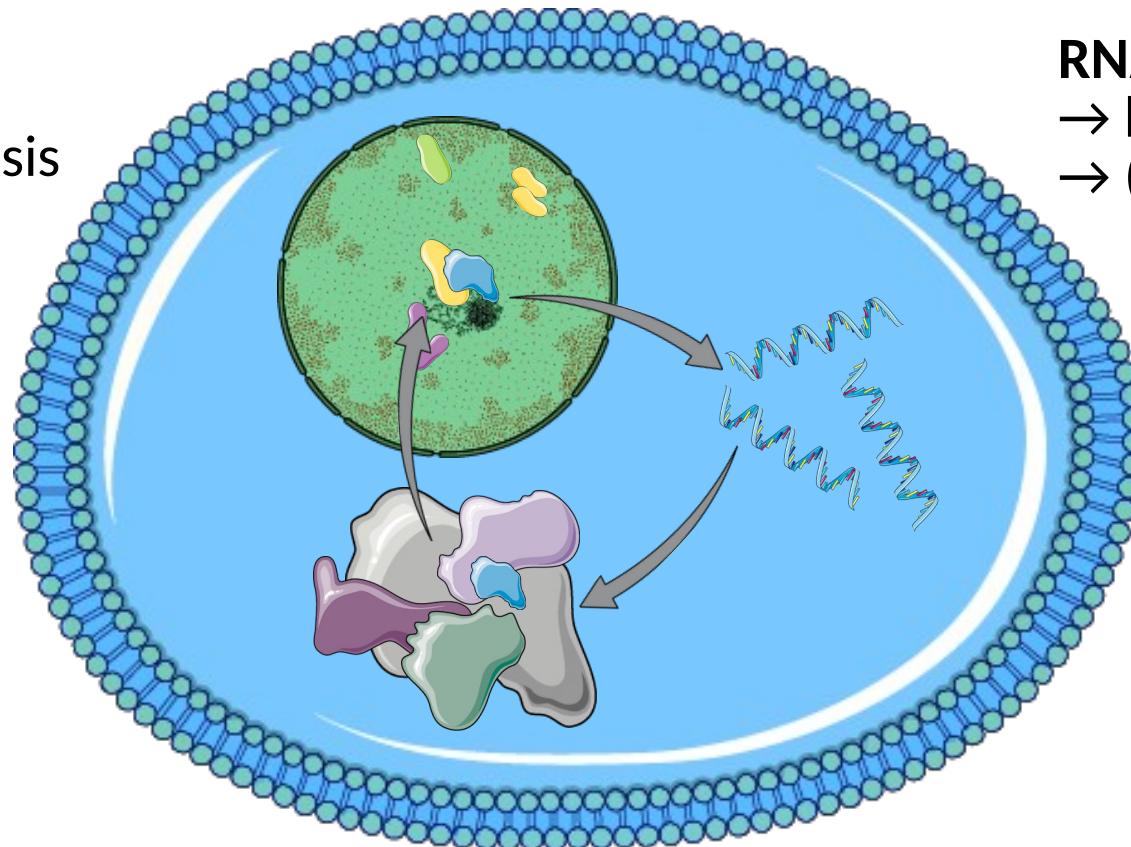
→ ChIP-seq

Proteins

- Mass-Spec
- WB, FACS etc.

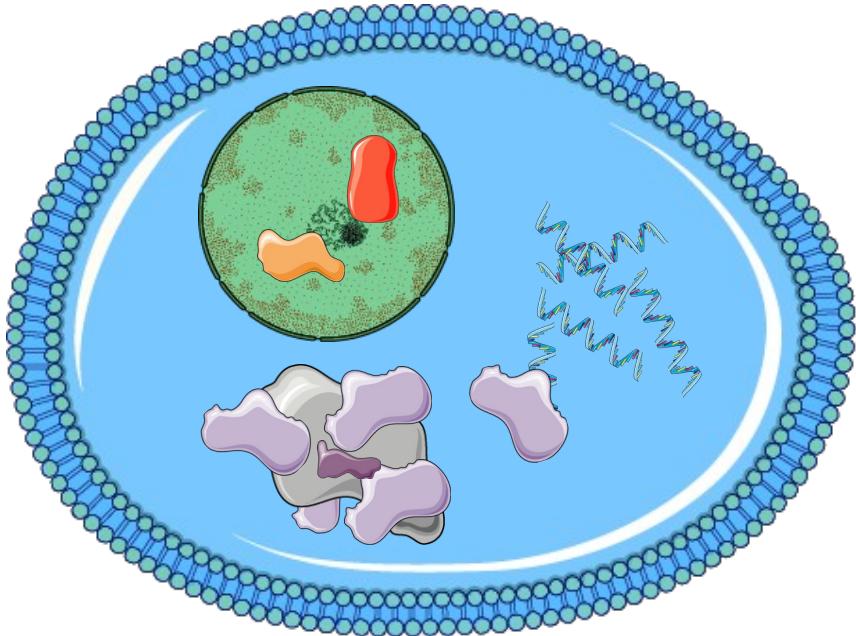
RNA

- bulk RNA-seq
- (sc)RNA-seq



Time for a break...

Case Study: Circadian Rhythm and Gene Expression



Circadian Clock Effektors:

- BMAL1 (=ARNTL)
- TIMELESS
- CLOCK
- ...

Have you used datasets from the
public domain for your research?

Public Repositories

RESEARCH ARTICLE

Targeting Glioblastoma Stem Cells through Disruption of the Circadian Clock



- most journals require authors to **deposit data** in public repository such as GEO

→ accession number

Data and Software Availability

All newly generated raw sequencing data are available on GEO through the accession number **GSE134974**. All data from external sources have been referenced in Methods.

Practical Session

Open your Browser and follow along!
Use links in the Handout/ Chat.

Task: Retrieval of ChIP-seq dataset

Accessing Data from Public Repository

Scope: Self Format: HTML Amount: Quick GEO accession: GO

enter accession number

Status	Public on Oct 24, 2019
Title	Targeting Glioblastoma Stem Cells through Disruption of the Circadian Clock [ChIP-Seq]
Organism	Homo sapiens
Experiment type	Genome binding/occupancy profiling by high throughput sequencing
Summary	<p>Glioblastomas are highly lethal cancers, containing self-renewing glioblastoma stem cells (GSCs). Here, we show that GSCs, differentiated glioblastoma cells (DGCs), and normal brain cultures all displayed robust circadian rhythms, yet GSCs alone displayed exquisite dependence on core clock transcription factors, BMAL1 and CLOCK, for optimal cell growth. Downregulation of BMAL1 or CLOCK in GSCs induced cell cycle arrest and apoptosis. Chromatin immunoprecipitation revealed BMAL1 preferentially bound at metabolic genes in GSCs, associated with differences in active chromatin regions compared to NSCs. Targeting BMAL1 or CLOCK attenuated mitochondrial metabolic function and reduced expression of the tricarboxylic acid (TCA) cycle enzymes. Small molecule agonists of two independent BMAL1::CLOCK negative regulators, the Cryptochromes and REV-ERBs, downregulated stem cell factors and reduced GSC growth. Combination of Cryptochrome and REV-ERB agonists induced synergistic anti-tumor efficacy. Collectively, GSCs coopt circadian regulators beyond canonical circadian circuitry to promote stemness maintenance and metabolism, offering novel therapeutic paradigms.</p>
Overall design	BMAL1 ChIP-sequencing was performed on two glioma stem cell and two normal neural stem cell samples. Two replicates were performed for each sample
Contributor(s)	Dong Z , Zhang G , Qu M , Gimple R , Wu Q , Qiu Z , Prager B , Wang X , Kim L , Morton A , Dixit D , Zhou W , Huang H , Li B , Zhu Z , Bao S , Mack S , Chavez L , Kay S , Rich J
Citation(s)	Dong Z, Zhang G, Qu M, Gimple RC et al. Targeting Glioblastoma Stem Cells through Disruption of the Circadian Clock. <i>Cancer Discov</i> 2019 Nov;9(11):1556-1573. PMID: 31455674
Submission date	Jul 26, 2019
Last update date	Oct 24, 2019
Contact name	Jeremy Rich
E-mail(s)	jerich@ucsd.edu
Organization name	University of California, San Diego

Title

Project

Study Design

Publication

Correspondent Researcher

Platforms (1) [GPL20301](#) Illumina HiSeq 4000 (Homo sapiens)

Samples (12) [GSM3982994](#) GSC_387_Rep1
[More...](#)
[GSM3982995](#) GSC_387_Rep2
[GSM3982996](#) 387_Input

This SubSeries is part of SuperSeries:

[GSE134974](#) Targeting Glioblastoma Stem Cells through Disruption of the Circadian Clock

Relations

BioProject [PRJNA556900](#)
SRA [SRP216541](#)

Download family

[SOFT formatted family file\(s\)](#)

Format

SOFT [?](#)

[MINiML formatted family file\(s\)](#)

MINiML [?](#)

[Series Matrix File\(s\)](#)

TXT [?](#)

Supplementary file	Size	Download	File type/resource
GSE134972_3565_BMAL1_R1_R2_Merged.sorted.rmdup.bam.rpkm.bw	31.5 Mb	(ftp)(http)	BW
GSE134972_387_BMAL1_R1_R2_Merged.sorted.rmdup.bam.rpkm.bw	18.3 Mb	(ftp)(http)	BW
GSE134972_BMAL1_PeaksGainedinGSC.bed.gz	117.7 Kb	(ftp)(http)	BED
GSE134972_BMAL1_PeaksGainedinNSC.bed.gz	82.4 Kb	(ftp)(http)	BED
GSE134972_ENSA_BMAL1_R1_R2_Merged.sorted.rmdup.bam.rpkm.bw	45.0 Mb	(ftp)(http)	BW
GSE134972_RAW.tar	10.9 Mb	(http) (custom)	TAR (of NARROWPEAK)
GSE134972_hNP1_BMAL1_R1_R2_Merged.sorted.rmdup.bam.rpkm.bw	39.4 Mb	(ftp)(http)	BW

Sequencing Platform

Available samples

Data Retrieval

Option 1:

→ [download here](#)

Option 2: ChIP-Atlas

ChIP-Seq Repository: ChIP-Atlas



京都大学
KYOTO UNIVERSITY

IN COLLABORATION WITH



GSM ID of interest
(provided in paper or as subset of GSE dataset in GEO accession viewer)

▲ ChIP-Atlas

Peak Browser

Target Genes

Colocalization

Enrichment Analysis

Publications

Docs

GSM3982997

Go

Advanced

Find By ID

▲ ChIP-Atlas

An integrative, comprehensive database to explore public Epigenetic dataset, including ChIP-Seq, DNase-Seq, ATAC-Seq, and Bisulfite-Seq data: ChIP-Atlas covers almost all public data archived in Sequence Read Archive of NCBI, EBI, and DDBJ with over 224,000 experiments.

[Watch movie introduction](#)

SRX6602363

GSM3982997: GSC 3565 Rep1; Homo sapiens; ChIP-Seq

[Visualize ▾](#)[Analyze ▾](#)[Download ▾](#)[Link Out ▾](#)

Sample information curated by ChIP-Atlas

Antigen

Antigen Class	TFs and others
Antigen	ARNTL

information on sample/experiment

Cell type

Cell type Class	Neural
Cell type	Glioma stem cells
NA	NA

Attributes by original data submitter

Sample

source_name	Patient Derived Glioma Stem Cell
tumor type	Glioblastoma
antibody	BMAL1

Sequencing Platform

instrument_model

Illumina HiSeq 4000

information on used sequencing platform

Read processing pipeline log

Where can I get the processing logs?

hg19

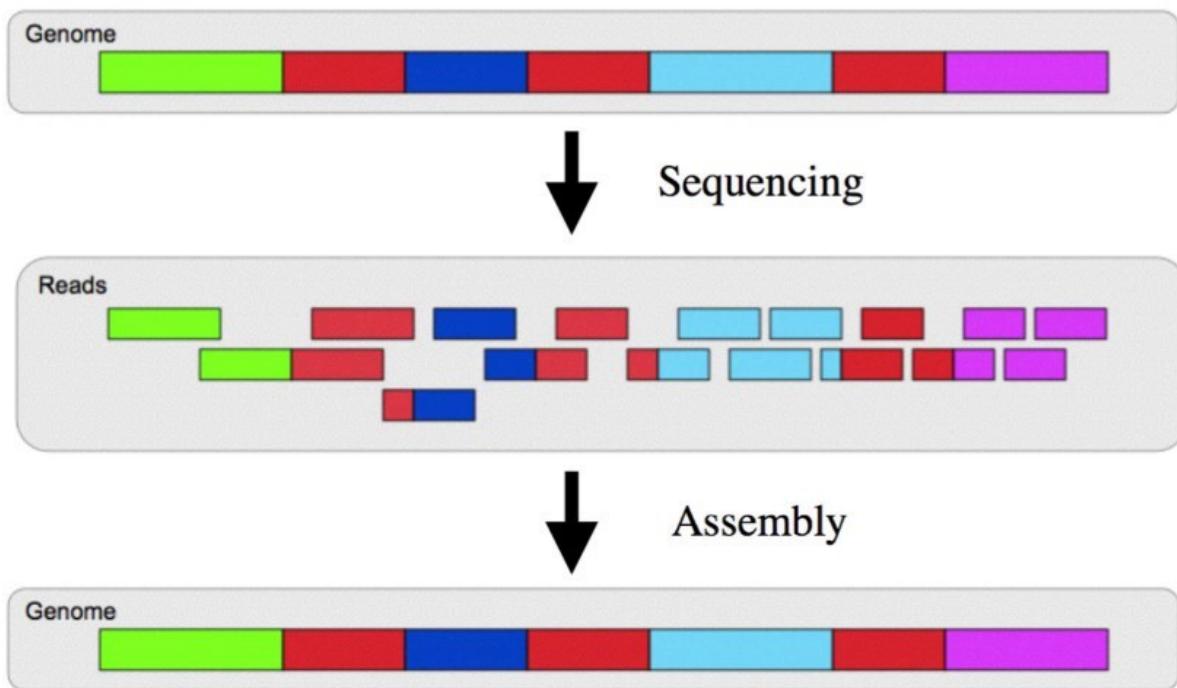
Number of total reads 15778959
Reads aligned (%) 23.9
Duplicates removed (%) 68.1
Number of peaks 15704 (qval < 1E-05)

hg38

Number of total reads 15778959
Reads aligned (%) 24.6
Duplicates removed (%) 67.3
Number of peaks 15870 (qval < 1E-05)

information on peak calling relative to 2 assemblies

Genome Assemblies



- Processed data are always **annotated to a specific assembly**
e.g. hg19
= human genome, version 19
- **LiftOver:** tool to transfer from one assembly/ reference genome to another, also **between species!**
→ tutorial in extra slides

click to download

SRX6602363

GSM3982997: GSC 3565 Rep1; Homo sapiens; ChIP-Seq

Visualize ▾

Analyze ▾

Download ▾

Link Out ▾

Sample information curated by ChIP-Atlas

Antigen

Antigen Class	TFs and others
Antigen	ARNTL

Cell type

Cell type Class	Neural
Cell type	Glioma stem cells
NA	NA

[Download ▾](#)[Link Out ▾](#)

For hg19

BigWig

Peak-call (q < 1E-05)

Peak-call (q < 1E-10)

Peak-call (q < 1E-20)

For hg38

BigWig

Peak-call (q < 1E-05)

Peak-call (q < 1E-10)

Peak-call (q < 1E-20)

download

- BigWig: continuous, genome-wide signal
- Peak-call: enriched peaks only (.bed format)

Download hg38 ARNTL ChIP-seq data
(GSM3982997):

- BigWig
- Peak call q < 1E-05
- Peak call q < 1E-20

[Tutorial movies ▾](#)

Visualizes protein binding on given genomic loci with [IGV genome browser](#)

[H.sapiens \(hg38\)](#)[H. sapiens \(hg19\)](#)[M. musculus \(mm10\)](#)[M. musculus \(mm9\)](#)[R. norvegicus \(rn6\)](#)[D. melanogaster \(dm6\)](#)[D. melanogaster \(dm3\)](#)[C. elegans \(ce11\)](#)[C. elegans \(ce10\)](#)[S. cerevisiae \(sacCer3\)](#)

1. Experiment type

[ChIP: Histone \(16612\)](#)[ChIP: RNA polymerase \(2099\)](#)[ChIP: TFs and others \(15155\)](#)[ChIP: Input control \(8078\)](#)[ATAC-Seq \(27318\)](#)[DNase-seq \(1642\)](#)[Bisulfite-Seq \(15000\)](#)

specify experiment type

▲ ChIP-Atlas: Peak Browser

Tutorial movies ▾

Visualizes protein binding on given genomic loci with [IGV genome browser](#)

H.sapiens (hg38)

H. sapiens (hg19)

M. musculus (mm10)

M. musculus (mm9)

R. norvegicus (rn6)

D. melanogaster (dm6)

D. melanogaster (dm3)

C. elegans (ce11)

C. elegans (ce10)

S. cerevisiae (sacCer3)

1. Experiment type

ChIP: Histone (16612)

ChIP: RNA polymerase (2099)

ChIP: TFs and others (15155)

ChIP: Input control (8078)

ATAC-Seq (27318)

DNase-seq (1642)

Bisulfite-Seq (15000)

2. Cell type Class

Kidney (1095)

Liver (404)

Lung (614)

Muscle (102)

Neural (548)

Others (309)

Pancreas (98)

Placenta (17)

Pluripotent stem cell (1042)

specify cell type

▲ ChIP-Atlas: Peak Browser

Tutorial movies ▾

Visualizes protein binding on given genomic loci with [IGV genome browser](#)

H.sapiens (hg38)

H. sapiens (hg19)

M. musculus (mm10)

M. musculus (mm9)

R. norvegicus (rn6)

D. melanogaster (dm6)

D. melanogaster (dm3)

C. elegans (ce11)

C. elegans (ce10)

S. cerevisiae (sacCer3)

1. Experiment type

- ChIP: Histone (16612)
- ChIP: RNA polymerase (2099)
- ChIP: TFs and others (15155)**
- ChIP: Input control (8078)
- ATAC-Seq (27318)
- DNase-seq (1642)
- Bisulfite-Seq (15000)

2. Cell type Class

- Kidney (1095)
- Liver (404)
- Lung (614)
- Muscle (102)
- Neural (548)**
- Others (309)
- Pancreas (98)
- Placenta (17)
- Pluripotent stem cell (1042)

3. Threshold for Significance ⓘ

- 50
- 100
- 200
- 500

specify threshold of
peak calling

▲ ChIP-Atlas: Peak Browser

Tutorial movies ▾

Visualizes protein binding on given genomic loci with [IGV genome browser](#)

H.sapiens (hg38)

H. sapiens (hg19)

M. musculus (mm10)

M. musculus (mm9)

R. norvegicus (rn6)

D. melanogaster (dm6)

D. melanogaster (dm3)

C. elegans (ce11)

C. elegans (ce10)

S. cerevisiae (sacCer3)

1. Experiment type

- ChIP: Histone (16612)
- ChIP: RNA polymerase (2099)
- ChIP: TFs and others (15155)**
- ChIP: Input control (8078)
- ATAC-Seq (27318)
- DNase-seq (1642)
- Bisulfite-Seq (15000)

2. Cell type Class

- Kidney (1095)
- Liver (404)
- Lung (614)
- Muscle (102)
- Neural (548)**
- Others (309)
- Pancreas (98)
- Placenta (17)
- Pluripotent stem cell (1042)

3. Threshold for Significance ⓘ

- 50
- 100
- 200
- 500

ChIP Antigen (optional)

type to search

- All
- ALKBH1 (1)
- ARID1A (1)
- ARID2 (3)
- ARNTL (8)**
- ASCL1 (1)
- ATF3 (1)
- ATF4 (12)

specify protein of
interest
OR...

▲ ChIP-Atlas: Peak Browser

Tutorial movies ▾

Visualizes protein binding on given genomic loci with [IGV genome browser](#)

H.sapiens (hg38)

H. sapiens (hg19)

M. musculus (mm10)

M. musculus (mm9)

R. norvegicus (rn6)

D. melanogaster (dm6)

D. melanogaster (dm3)

C. elegans (ce11)

C. elegans (ce10)

S. cerevisiae (sacCer3)

1. Experiment type

- ChIP: Histone (16612)
- ChIP: RNA polymerase (2099)
- ChIP: TFs and others (15155)**
- ChIP: Input control (8078)
- ATAC-Seq (27318)
- DNase-seq (1642)
- Bisulfite-Seq (15000)

2. Cell type Class

- Kidney (1095)
- Liver (404)
- Lung (614)
- Muscle (102)
- Neural (548)**
- Others (309)
- Pancreas (98)
- Placenta (17)
- Pluripotent stem cell (1042)

3. Threshold for Significance ⓘ

- 50
- 100
- 200
- 500

ChIP Antigen (optional)

- type to search
- All
- ALKBH1 (1)
- ARID1A (1)
- ARID2 (3)
- ARNTL (8)**
- ASCL1 (1)
- ATF3 (1)
- ATF4 (12)

Cell type (optional)

- type to search
- BT-16 (4)
- BT142 (2)
- BT168 (4)
- CHLA-90 (2)
- CHP-134 (7)
- CLB-Ga (5)
- COG-N-415 (1)
- D283 Med (9)

OR cell type of interest

▲ ChIP-Atlas: Peak Browser

[Tutorial movies ▾](#)

Visualizes protein binding on given genomic loci with [IGV genome browser](#)

[H.sapiens \(hg38\)](#)[H. sapiens \(hg19\)](#)[M. musculus \(mm10\)](#)[M. musculus \(mm9\)](#)[R. norvegicus \(rn6\)](#)[D. melanogaster \(dm6\)](#)[D. melanogaster \(dm3\)](#)[C. elegans \(ce11\)](#)[C. elegans \(ce10\)](#)[S. cerevisiae \(sacCer3\)](#)

1. Experiment type

- [ChIP: Histone \(16612\)](#)
- [ChIP: RNA polymerase \(2099\)](#)
- ChIP: TFs and others (15155)**
- [ChIP: Input control \(8078\)](#)
- [ATAC-Seq \(27318\)](#)
- [DNase-seq \(1642\)](#)
- [Bisulfite-Seq \(15000\)](#)

2. Cell type Class

- Kidney (1095)
- Liver (404)
- Lung (614)
- Muscle (102)
- Neural (548)**
- Others (309)
- Pancreas (98)
- Placenta (17)
- Pluripotent stem cell (1042)

3. Threshold for Significance ⓘ

- 50
- 100
- 200
- 500

ChIP Antigen (optional)

-
- All
- ALKBH1 (1)
- ARID1A (1)
- ARID2 (3)
- ARNTL (8)**
- ASCL1 (1)
- ATF3 (1)
- ATF4 (12)

Cell type (optional)

-
- BT-16 (4)
- BT142 (2)
- BT168 (4)
- CHLA-90 (2)
- CHP-134 (7)
- CLB-Ga (5)
- COG-N-415 (1)
- D283 Med (9)

[View on IGV](#)

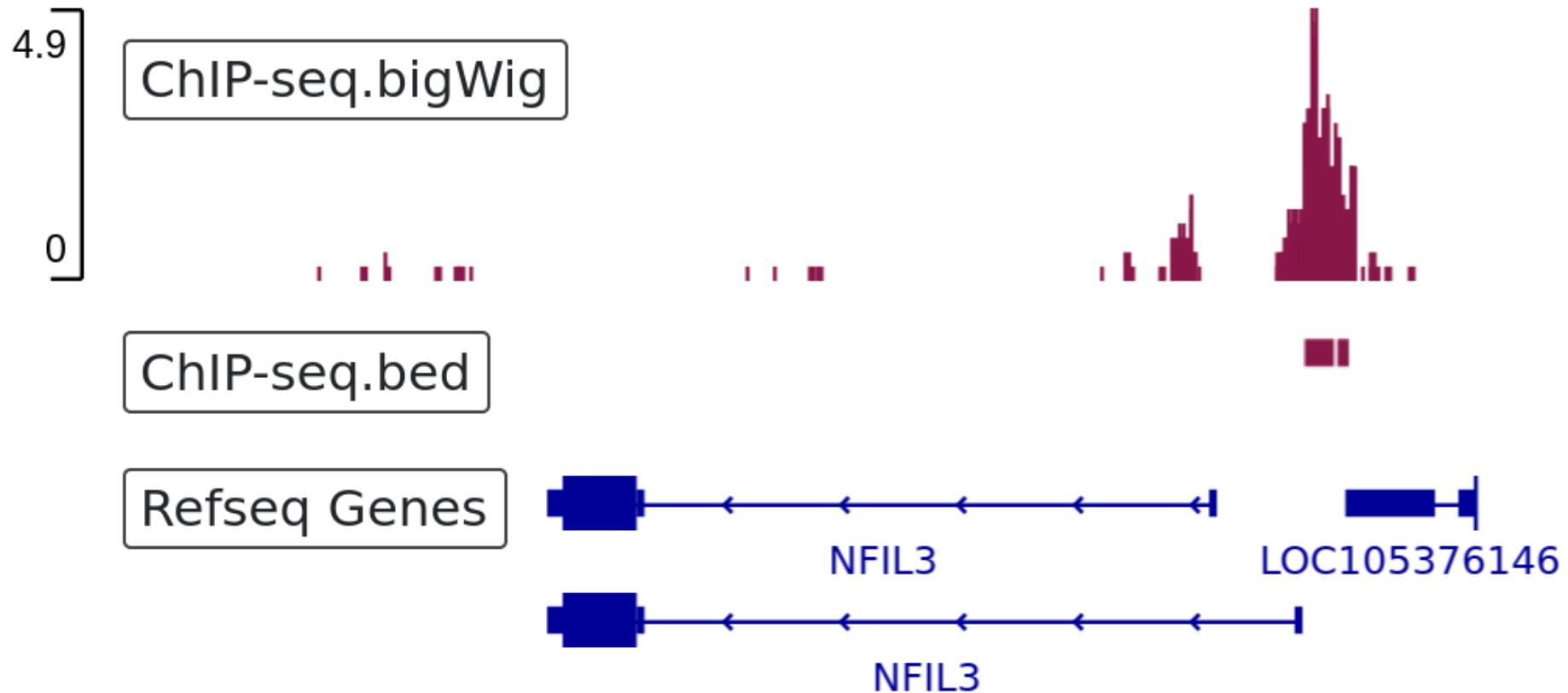
Error connecting to IGV?

[Download BED file](#)

Download

Time for a break...

Data Visualization



Data visualization with IGV

- IGV = Integrated Genomics Viewer
- Desktop and online versions
- Data from public domain (e.g. ENCODE) can be visualized alongside own data
- Visualization can be customized according to your needs and exported as .svg (vector graphic)

[Genome](#)[Tracks](#)[Session](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#)[Help](#)[Local File ...](#)[Dropbox !\[\]\(0a0b609dcde5224e7480a4f51bb76cda_img.jpg\) ...](#)[Google Drive !\[\]\(42c374edbfc1163de9dfb6a29a515bf4_img.jpg\) ...](#)[URL ...](#)[Human \(T2T CHM13-v2.0\)](#)[Human \(T2T CHM13-v1.1\)](#)[Human \(GRCh38/hg38\)](#)[Human \(hg38 1kg/GATK\)](#)[Human \(GRCh37/hg19\)](#)[Human \(hg18\)](#)[Mouse \(GRCm39/mm39\)](#)[Mouse \(GRCm38/mm10\)](#)[Mouse \(NCBI37/mm9\)](#)[Rat \(rn7\)](#)[Rat \(RGCS 6.0/rn6\)](#)[Gorilla \(Kamilah_GGO_v0/gorGor6\)](#)[Gorilla \(gorGor4.1/gorGor4\)](#)[Chimp \(panTro6\) \(panTro6\)](#)[Chimp \(panTro5\) \(panTro5\)](#)

igv.org

[Cursor Guide](#)[Track Labels](#)

5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 22 | X | Y

Zoom in to see features

click to chose appropriate
genome for the data you want
to visualize

[UC San Diego](#)[BROAD
INSTITUTE](#)

Genome ▾

Tracks ▾

Session ▾

Share

Bookmark

Save SVG

Circular View ▾

Help ▾

IGV

hg38

all

Local File ...

Dropbox ⚡ ...

Google Drive Google Sheets ...

URL ...

Annotations ...

ENCODE Signals ...

ENCODE Other ...

1

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

22

X

Y



Cursor Guide

Track Labels

Zoom in to see features



click to upload your own data
in .bed or .bigWig/.bw format

[Genome](#) ▾[Tracks](#) ▾[Session](#) ▾[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#) ▾[Help](#) ▾

IGV

hg38

all

▼

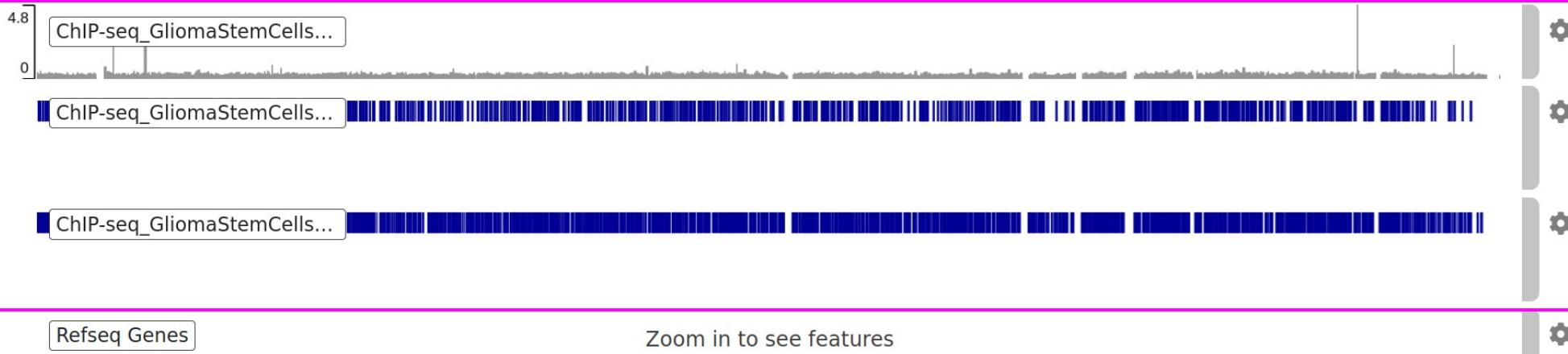
all



Cursor Guide

Track Labels

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 22 | X | Y

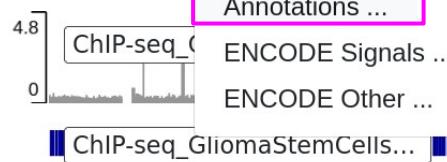


[Genome ▾](#)[Tracks ▾](#)[Session ▾](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View ▾](#)[Help ▾](#)

IGV

hg38

all

[Local File ...](#)[Dropbox ...](#)[Google Drive ...](#)[URL ...](#)[Annotations ...](#)[ENCODE Signals ...](#)[ENCODE Other ...](#)[ChIP-seq_GliomaStemCells...](#)[ChIP-seq_GliomaStemCells...](#)[Refseq Genes](#)

Zoom in to see features

click to add **annotations** (availability depends on chosen genome)

Genome ▾

Tracks ▾

Session ▾

Share

Bookmark

Save SVG

Circular View ▾

Help ▾

IGV

hg38

all

Annotations



Search Guide Track Labels

Gencode v28 Basic

Ensembl v90

UCSC sno/miRNA

GWAS

Common SNPs (150)

Repeat Masker

GC %

CpG Islands

Phastcons (20 way)

Annotations from the [UCSC Genome Browser](#), [GENCODE](#), and Ensembl

Cancel

OK

Refseq Genes

Zoom in to see features

[Genome](#)[Tracks](#)[Session](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#)[Help](#)

IGV

hg38

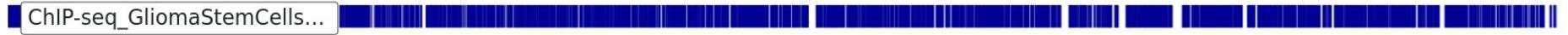
all

▼

all

[Cursor Guide](#)[Track Labels](#)

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 22 | X | Y

[Refseq Genes](#)[Zoom in to see features](#)

[Genome](#)[Tracks](#)[Session](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#)[Help](#)

IGV

hg38

all

all

[Cursor Guide](#)[Track Labels](#)

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 22 | X | Y



Filename: ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363.bw



CpG Islands

click to change settings

Refseq Genes

Zoom in to see features



[Genome](#)[Tracks](#)[Session](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#)[Help](#)

IGV

hg38

all

all



Cursor Guide

Track Labels

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 22 | X | Y



ChIP-seq_GliomaStemCells

ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363.bw



ChIP-seq_GliomaStemCells...



Filename: ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363.bw



CpG Islands

Refseq Genes

Zoom in to see features

- Set track name
- Set track height
- Set track color
- Unset track color
- Set alt color
- Flip y-axis
- Set data range
- Log scale
- Autoscale
- Remove track

[Genome](#)[Tracks](#)[Session](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#)[Help](#)

IGV

hg38

all

▼

all

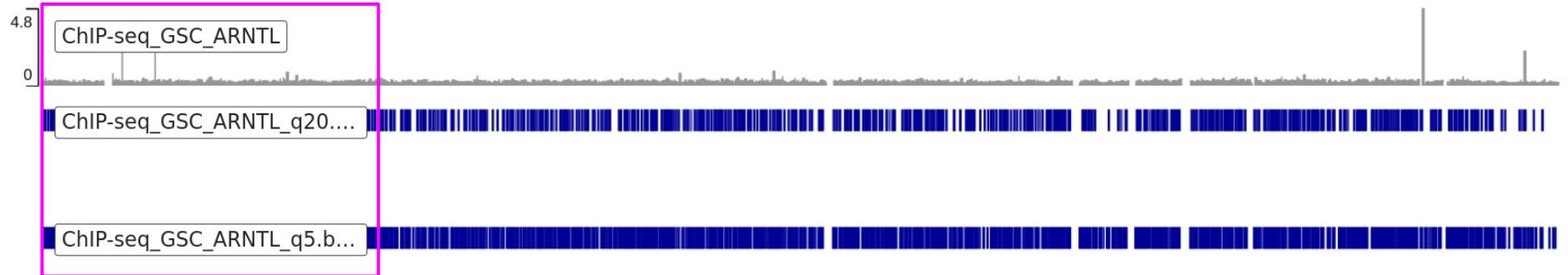


Cursor Guide

Track Labels

enter gene of interest

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 22 | X | Y



CpG Islands

Refseq Genes

Zoom in to see features

[Genome](#)[Tracks](#)[Session](#)[Share](#)[Bookmark](#)[Save SVG](#)[Circular View](#)[Help](#)

IGV

hg38

chr1

chr1:7,759,885-7,844,953

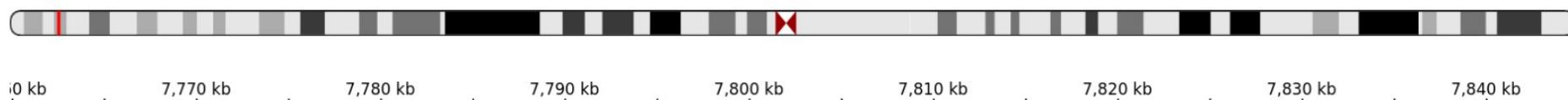


85 kb

Cursor Guide

Center Line

Track Labels

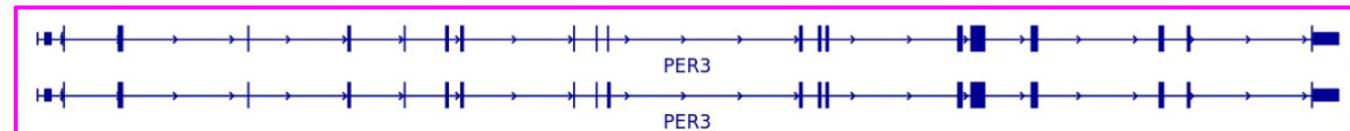
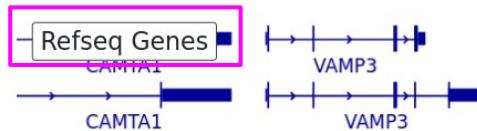
[ChIP-seq_GSC_ARNTL_q20....](#)

SRX6602363.20_peak_12

[ChIP-seq_GSC_ARNTL_q5.b...](#)

SRX6602363.05_peak_78

SRX6602363.05_peak_79



gene of interest

IGV

hg38

chr1

chr1:7,759,885-7,844,953

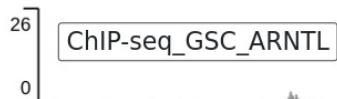
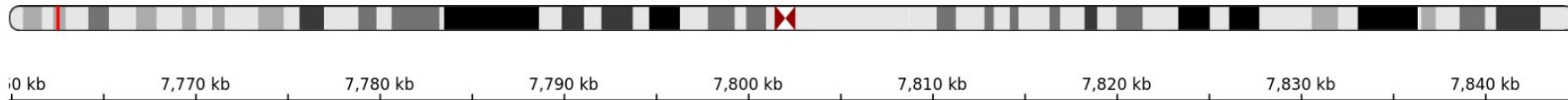


85 kb

Cursor Guide

Center Line

Track Labels



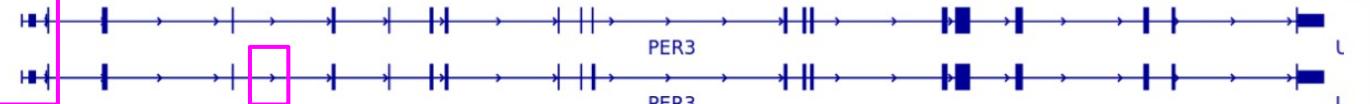
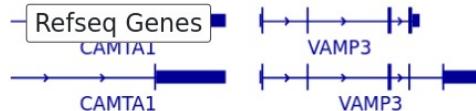
SRX6602363.20_peak_12

ChIP-seq peaks in
promotor-region



SRX6602363.05_peak_78

SRX6602363.05_peak_79



Direction of gene

- + strand >
- - strand <

IGV

hg38

chr1



chr1:7,759,885-7,844,953

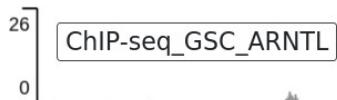
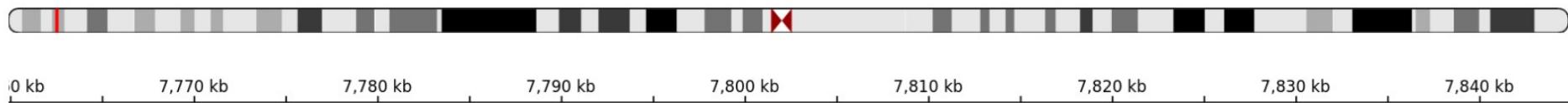


85 kb

Cursor Guide

Center Line

Track Labels

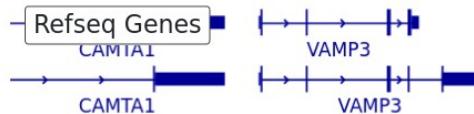


SRX6602363.20_peak_12



SRX6602363.05_peak_78

SRX6602363.05_peak_79



customize track appearance

- Set track name
- Set track height
- Set track color
- Unset track color
- Set alt color
- Flip y-axis
- Set data range
- Log scale
- Autoscale
- Remove track

Genome ▾ Tracks ▾ Session ▾ Share Bookmark Save SVG Circular View ▾ Help ▾

IGV hg38

chr1

chr1:7,760,661-7,845,730

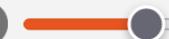


85 kb

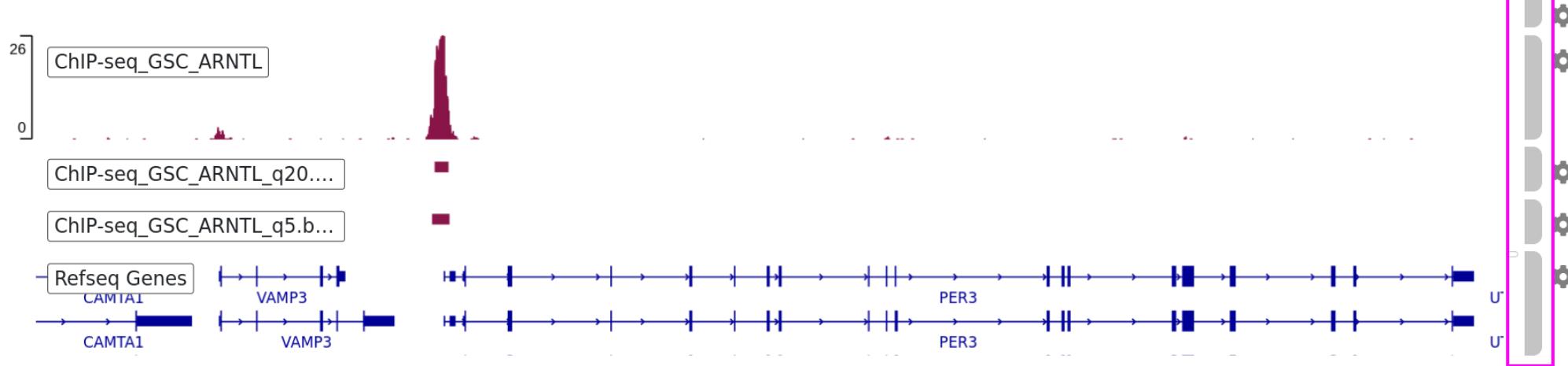
Cursor Guide

Center Line

Track Labels



kb
7,770 kb 7,780 kb 7,790 kb 7,800 kb 7,810 kb 7,820 kb 7,830 kb 7,840 kb



drag/drop to change order of tracks

zoom in/out

Genome ▾

Tracks ▾

Session ▾

Share

Bookmark

Save SVG

Circular View ▾

Help ▾

IGV

hg38

chr1

chr1:7,759,885-7,844,953

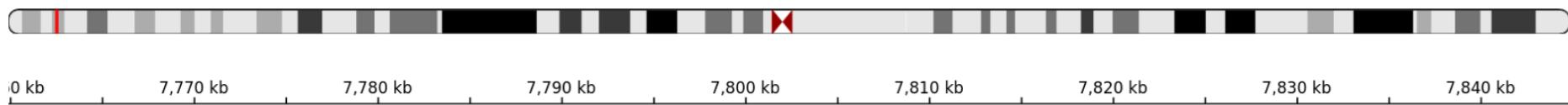


85 kb

Cursor Guide

Center Line

Track Labels

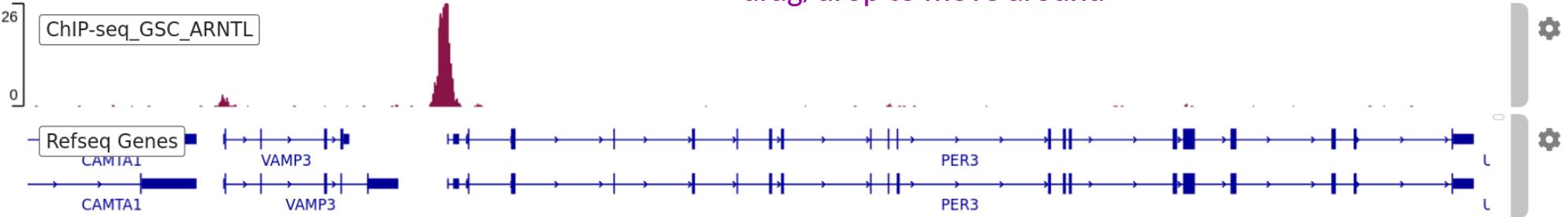


ChIP-seq_GSC_ARNTL_q20....

ChIP-seq_GSC_ARNTL_q5.b...

ChIP-seq_GSC_ARNTL

click on track and
drag/drop to move around



Task: Visualize ChIP-seq dataset in IGV

Breakout room 1: Anna, Felix, Finn

Breakout room 2: Leon, Lorenzo, Miles, Minju

Breakout room 3: Rebecca, Sara, Tobias

Assess ARNTL binding to

- OLIG2
- CLOCK
- TIMELESS
- Your favorite gene

Which problems did you encounter?

What did you find out about the binding of ARNTL?

Time for a break...

Peak annotation to ChIPseeker

1. Uploading data to Galaxy (.bed)
2. Retrieval of Annotation Reference file
3. Uploading of Annotation Reference to Galaxy (.gtf)
4. Peak Annotation with ChIPseeker

The screenshot shows the Galaxy web interface. On the left, there is a sidebar titled "Tools" with a search bar and an "Upload Data" button. Below the search bar is a list of categories: "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", "BED", "VCF/BCF", "Nanopore", and "Convert Formats". The "GENERAL TEXT TOOLS" and "GENOMIC FILE MANIPULATION" sections are highlighted with a light gray background. On the right, there is a list of bullet points describing Galaxy's features:

- open source, web-based platform for data intensive biomedical research
- collection of an abundance of bioinformatic tools

At the top right, there are icons for "Using 0%" and other user information.

The screenshot shows the Galaxy web interface with a sidebar on the left containing a list of tools categorized under 'Tools' and 'GENERAL TEXT TOOLS'. The main area displays a bulleted list of features, and the top right corner shows a 'History' panel.

Galaxy

Workflow Visualize Shared Data ▾ Help ▾ Login or Register

Using 0%

Tools

search tools (x)

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

History

search datasets (?x)

Unnamed history

(empty)

This history is empty. You can load your own data or get data from an external source

- open source, web-based platform for data intensive biomedical research
- collection of an abundance of bioinformatic tools
- data can be uploaded and analysis trajectory saved

Practical Session

Open your Browser and follow along!
Use links in the Handout/ Chat.

Task: Create Galaxy account, and
upload ChIP-seq data (.bed files)

 Galaxy

[Workflow](#) [Visualize](#) [Shared Data ▾](#) [Help ▾](#) [Login or Register](#)   

Using 0%

Tools

search tools (x)

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

click to create account

History

search datasets (? x)

Unnamed history

(empty)

This history is empty. You can load your own data or get data from an external source

Tools



search tools



Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

upload data

manage history

History

+



rename history

Unnamed history

(empty)



This history is empty. You can load your own data or get data from an external source

Uploading Data to Galaxy

The screenshot shows the Galaxy web interface. On the left, there's a sidebar with various tools and categories like 'Tools', 'Upload Data', 'Get Data', 'Send Data', etc. The main area has a title 'Download from web or upload from disk' with tabs for 'Regular', 'Composite', 'Collection', and 'Rule-based'. A large dashed box contains the text 'drag and drop data here...' and a 'Drop files here' placeholder with a paper icon. Below this is a section with 'or browse' text and a pink arrow pointing to a 'Choose local files' button. At the bottom, there are dropdown menus for 'Type (set all)' (Auto-detect) and 'Genome (set all)' (unspecified (?)), and buttons for 'Choose local files', 'Choose remote files', 'Paste/Fetch data', 'Start', 'Pause', 'Reset', and 'Close'. The status bar at the top right says 'Using 26%'.

Galaxy

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

Workflow Visualize Shared Data Help User

Using 26%

Download from web or upload from disk

Regular Composite Collection Rule-based

drag and drop data here...

Drop files here

or browse

Type (set all): Auto-detect

Genome (set all): unspecified (?)

Choose local files Choose remote files Paste/Fetch data Start Pause Reset Close

Choose local files



Download from web or upload from disk



Regular

Composite

Collection

Rule-based

You added 2 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
ChIP-seq_GliomaStc	1.3 MB	Auto-de...	unspecified (?)		0%
ChIP-seq_GliomaStc	228.5 KB	Auto-de...	unspecified (?)		0%

click start to upload

Type (set all):

Auto-detect

Genome (set all):

unspecified (?)

Choose local files

Choose remote files

Paste/Fetch data

Start

Pause

Reset

Close

Download from web or upload from disk

Regular

Composite

Collection

Rule-based

Name	Size	Type	Genome	Settings	Status
ChIP-seq_GliomaSt... ChIP-seq_GliomaSt...	1.3 MB 228.5 KB	Auto-de... Auto-de...	unspecified (?) unspecified (?)	⚙️ ⚙️	100% 100% ✓ ✓
ChIP-seq_GliomaSt... ChIP-seq_GliomaSt...	1.3 MB 228.5 KB	Auto-de... Auto-de...	unspecified (?) unspecified (?)	⚙️ ⚙️	100% 100% ✓ ✓

Type (set all):

Auto-detect



Genome (set all):

unspecified (?)

then close
window Choose local files Choose remote files Paste/Fetch data

Start

Pause

Reset

 Close

Galaxy's job concurrency limits and maximum allowed run times have been decreased in order to ensure enough resources are available to students participating in the GTN Smörgåsbord.

Tools



bedtools intersect

**Upload Data****Show Sections**

bedtools Compute both the depth and breadth of coverage of features in file B on the features in file A (bedtools coverage)

bedtools Intersect intervals find overlapping intervals in various ways

bedtools SpacingBed reports the distances between features

bedtools RandomBed generate random intervals in a genome

bedtools Convert from BAM to FastQ

bedtools WindowBed find overlapping intervals within a window around an interval

bedtools ComplementBed Extract intervals not represented by an interval file

bedtools NucBed profile the nucleotide content of intervals in a FASTA file

bedtools GroupByBed group by common cols and summarize other cols

bedtools ExpandBed replicate lines based on lists of values in columns

bedtools MakeWindowsBed make interval windows across a genome

bedtools SubtractBed remove intervals based on overlaps

bedtools MultiCovBed counts coverage from multiple BAMs at specific

History



search datasets



Bioinfo MolMed

7 shown

(empty)



Grey: waiting in line

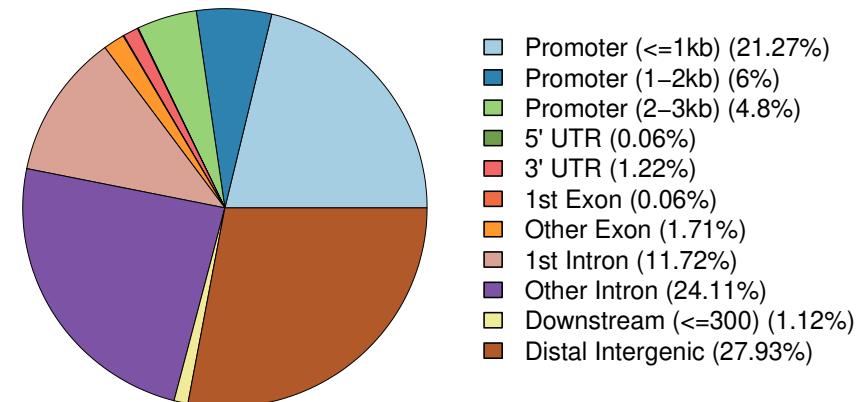
Orange: currently uploading

Green: upload successful

7: Tgif2_mm10.bed	
6: Tgif1_mm10.bed	
5: Smad3_mm10.bed	
4: Nanog_mm10.bed	
3: Hdac1_mm10.bed	
2: GSM1335481_Tgif1.peaks.txt	
1: ChIP-Seq_PBX1_allpeaks_aNSWT_mm10.bed	

- Bioconductor package for annotating ChIP-seq data
- **position and strand information of nearest genes** are reported, in addition to the **distance from the peak to the TSS** of its nearest gene
- **genomic region** of the peak is reported in the annotation column. Since some annotations may overlap for a peak, ChIPseeker adopts the following priority in genomic annotation:
 - Promoter
 - 5' UTR
 - 3' UTR
 - Exon
 - Intron
 - Downstream
 - Intergenic
- produces plots to help users **visualize** annotation

- **Inputs:**
 - ChIP-seq peaks in .bed, Interval or tabular format
 - .gtf file for annotation
- **Outputs:**
 - file of annotated peaks in Interval or Tabular format
 - PDF of plots
- Optionally, you can choose to output:
 - R script used by this tool
 - RData file



Open your Browser and follow along!
Use links in the Handout/ Chat.

Tasks:

1. Retrieve .gtf file and upload to Galaxy
2. Perform peak annotation with
ChIPseeker

[Human](#)[Mouse](#)[How to access data](#)[FAQ](#)[Documentation](#)[About us](#)

HUMAN

GENCODE 40 (11.04.22)



MOUSE

GENCODE M29 (11.04.22)



The goal of the GENCODE project is to identify and classify all gene features in the human and mouse genomes with high accuracy based on biological evidence, and to release these annotations for the benefit of biomedical research and genome interpretation.



Human

Release 21 (GRCh38)

- [Statistics of this release](#)
- [More information about this assembly](#) (including patches, scaffolds and haplotypes)

GTF / GFF3 files

click to download

Content	Regions	Description	Download
Comprehensive gene annotation	CHR	<ul style="list-style-type: none">• It contains the comprehensive gene annotation on the reference chromosomes only• This is the main annotation file for most users	GTF GFF3
Comprehensive gene annotation	ALL	<ul style="list-style-type: none">• It contains the comprehensive gene annotation on the reference chromosomes, scaffolds, assembly patches and alternate loci (haplotypes)• This is a superset of the main annotation file	GTF GFF3
Long non-coding RNA gene annotation	CHR	<ul style="list-style-type: none">• It contains the comprehensive gene annotation of lncRNA genes on the reference chromosomes• This is a subset of the main annotation file	GTF GFF3

Download from web or upload from disk

Regular

Composite

Collection

Rule-based

Name

Size

Type

Genome

Settings

Status



gencode.v21.annota

1.1 GB

Auto-de...



unspecified (?)



100%



upload annotation file to Galaxy
→ since it's quite large, this might take a while

Type (set all): Auto-detect

Genome (set all): unspecified (?)

Choose local files

Choose remote files

Paste/Fetch data

Start

Pause

Reset

Close

The screenshot shows the Galaxy web interface. At the top, there is a dark header bar with the Galaxy logo on the left and a 'Workflow' button on the right. Below the header, the word 'Tools' is displayed in bold black text. To the right of 'Tools' are two icons: a star and a three-line menu. A search bar is positioned below 'Tools'. Inside the search bar, the text 'ChIPseeker' is highlighted with a pink rectangular border. To the right of the search bar is a close button (an 'X'). Below the search bar are three buttons: 'Upload Data' (with an upward arrow icon), 'Show Sections' (with an eye icon), and a large button labeled 'ChIPseeker for ChIP peak annotation and visualization' (also highlighted with a pink rectangular border). To the right of this large button, the text 'click to use tool' is written in purple.

Galaxy

Workflow

Tools

ChIPseeker

Upload Data

Show Sections

ChIPseeker for ChIP peak annotation and visualization

click to use tool

ChIPseeker for ChIP peak annotation and visualization (Galaxy Version 1.18.0+galaxy1)



Peaks file



1: ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363_q5.bed



A peaks file in BED format.

choose file you want to assess

Peaks file has header?



No

If this option is set to Yes, the tool will assume that the peak file has column headers in the first row. Default: No

Annotation source

Use a GTF from history



Select a GTF to use for annotation source.

enter uploaded reference (.gtf file)

Select a history GTF



13: gencode.v21.annotation.gtf



Output Format

Tabular (tab-separated)

Output PDF of plots?



Default: Yes

Output Rscript?



If this option is set to Yes, the Rscript used to annotate the IDs will be provided as a text file in the output. Default: No

Output RData file?



Output all the data used by R to construct the tables and plots, can be loaded into R. Default: No

Advanced Options



Email notification



Send an email notification when the job completes.

✓ Execute

customize settings and execute
(depending on server capacity, takes about 10 min)

What it does

info on input and output files

ChIPseeker is a Bioconductor package for annotating ChIP-seq data analysis. Peak Annotation is performed by the `annotatePeak` function. The position and strand information of nearest genes are reported, in addition to the distance from the peak to the TSS of its nearest gene. Users can define the TSS (transcription start site) region under **Advanced Options**, by default the TSS region is defined from -3kb to +3kb. The genomic region of the peak is reported in the annotation column. Since some annotations may overlap for a peak, ChIPseeker adopts the following priority in genomic annotation:

- Promoter
- 5' UTR
- 3' UTR
- Exon
- Intron
- Downstream
- Intergenic

ChIPseeker also produces plots to help users visualise the overlaps in annotation for peaks, for example, the `vennpie` and `upsetplot`. See the [ChIPseeker vignette](#) for more information.

Inputs

A peaks file in BED, Interval or Tabular format e.g from MACS2 or DiffBind. Note that there is an option to specify if the input peaks file has a header row. No header row is assumed by default, which is usually the case for BED format e.g. MACS narrowpeak, however other formats e.g. MACS tabular format, may contain a header row.

Example:

Chromosome	Start	End	Name	Score	Strand
18	394599	396513	DiffBind	0	.
18	111566	112005	DiffBind	0	.
18	346463	347342	DiffBind	0	.
18	200012	400282	DiffBind	0	.

History



search datasets



GRADE Brain Bioinfo

5 shown

1.69 MB



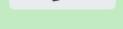
5: ChIPseeker on data
3 and data 1: Plots



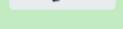
4: ChIPseeker on data
3 and data 1: Annotated Peaks



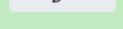
genCODE.v40.annotation.gtf



ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM39
82997_SRX6602363_q20.bed



ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM39
82997_SRX6602363_q5.bed



chr4	17577053	17577604	SRX6602363.05_peak_10972	673	.	18.83639	71.70705	67.34354	359
chrX	2691128	2691429	SRX6602363.05_peak_15415	79	.	6.88961	10.82879	7.97018	131
chr7	43112887	43113282	SRX6602363.05_peak_13205	159	.	8.4873	19.28815	15.99564	222
chr7	43113403	43113590	SRX6602363.05_peak_13206	139	.	7.881	17.18586	13.98485	89
chr7	43249060	43249382	SRX6602363.05_peak_13207	80	.	7.24585	10.91063	8.04712	206
chr7	43432192	43432373	SRX6602363.05_peak_13208	130	.	9.8054	16.2572	13.09869	108
chr7	1929733	1929929	SRX6602363.05_peak_12964	155	.	8.24376	18.84119	15.5674	44
chr7	1951020	1951453	SRX6602363.05_peak_12965	157	.	7.58448	19.05674	15.77405	307
chr7	1990928	1991172	SRX6602363.05_peak_12966	120	.	7.01091	15.12464	12.02185	168
chr7	2035682	2035940	SRX6602363.05_peak_12967	113	.	6.64433	14.43678	11.36954	100
chr7	2069674	2069868	SRX6602363.05_peak_12968	155	.	7.26164	18.86575	15.59039	69
chr7	2104311	2104516	SRX6602363.05_peak_12969	213	.	5.92053	24.87837	21.37656	100
chr7	2105933	2106192	SRX6602363.05_peak_12970	177	.	5.33321	21.14242	17.7747	102
chr7	2106389	2106683	SRX6602363.05_peak_12971	238	.	6.1479	27.40546	23.82549	142
chr7	2161368	2162069	SRX6602363.05_peak_12972	330	.	11.13269	36.90112	33.07427	272
chr7	2232471	2232652	SRX6602363.05_peak_12973	78	.	6.17093	10.73531	7.88228	114
chr7	2232790	2233176	SRX6602363.05_peak_12974	119	.	7.71367	15.00052	11.90353	258
chr17	38869667	38870191	SRX6602363.05_peak_5933	355	.	12.62834	39.44969	35.56734	343
chr17	48107433	48107849	SRX6602363.05_peak_6110	167	11 020RR	20 04851	16 7247	160	

info from input dataset

Promoter (<=1kb)
Intron (ENST00000395891.4/ENSG)
Intron (ENST00000265854.9/ENSG)
Intron (ENST00000265854.9/ENSG)
Intron (ENST00000265854.9/ENSG)
Intron (ENST00000265854.9/ENSG)
Promoter (1-2kb)
Promoter (1-2kb)
Intron (ENST00000265854.9/ENSG)
Promoter (<=1kb)
Promoter (<=1kb)
Promoter (<=1kb)
Promoter (<=1kb)

annotation

14: ChIPseeker on data 13			
and data 1: Annotated Peaks			
15,765 lines			
format: tabular, database: ?			
>> preparing features information... 2022-05-09 07:45:41			
>> identifying nearest features... 2022-05-09 07:45:43			
>> calculating distance from peak to TSS... 2022-05-09 07:45:45			
>> assigning genomic annotation... 2022-05-09 07:45:45			
>> assigning			
1 2 3 4			
Chrom Start End SRX6602363.05_pe			
chr1 169893350 169893823 SRX6602363.05_pe			
chr6 143511281 143511802 SRX6602363.05_pe			
chr6 53544417 53545192 SRX6602363.05_pe			
chr6 53548032 53548348 SRX6602363.05_pe			

download and open
in spreadsheet

ChIPseeker: Output

Screenshot of LibreOffice Calc showing the "Text Import" dialog box. The dialog is set to "Separated by" and has the character set set to "UTF-16". The separator field contains a vertical bar (|). A pink arrow points from the text "insert ‘|’ as separator" to the separator field.

Untitled 1 - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation Sa 10 B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA AE

A1

1 Chrom Start End Comment

2 chr3 108145913 108146375 peak_60655 chr3 108145604 108146450 12

3 chr16 18811330 18812645 peak_36207 chr16 18810943 18811897 25

4 chr11 108342728 108344232 peak_17200 chr11 108342957 108344127 44

5 chr11 121236375 121238082 peak_18202 chr11 121236302 121237889 9.

6 chr13 5860523 5864620 peak_22829 chr13 5860601 5862371 11

7 chr4 120405443 120405769 peak_66531 chr4 120404846 120406054 7.

8 chr9 57520339 57522476 peak_92941 chr9 57520827 57521572 11

Help Cancel OK

Sheet1

Find Formatted Display Match Case

Average: Sum: 0

insert “|” as separator

A	B	C	D	K	L	M	N	O	P	Q	R	S	T	U
1	Chrom	Start	End	SRX6602363.05_peak_1	annotation	geneChr	geneStart	geneEnd	geneLength	geneStrand	transcriptId	distanceTo	geneName	genelid
2	chr1	169893350	169893823	SRX6602363.05_peak_1013	Promoter (<=1kb)	1	169849631	169893952	44322	2	ENST0000009	120	SCYL3	ENSG0000000457.11
3	chr6	143511281	143511802	SRX6602363.05_peak_12792	Promoter (<=1kb)	6	143494811	143511690	16880	2	ENST0000009	9	FUCA2	ENSG0000001036.11
4	chr6	53544417	53545192	SRX6602363.05_peak_12530	Promoter (<=1kb)	6	53497341	53545129	47789	2	ENST0000009	9	GCLC	ENSG0000001084.8
5	chr6	53548032	53548348	SRX6602363.05_peak_12531	Promoter (2-3kb)	6	53497346	53545129	47784	2	ENST0000009	-290	GCLC	ENSG0000001084.8
6	chr6	41073124	41073315	SRX6602363.05_peak_12448	Promoter (<=1kb)	6	41072983	4107950	24968	1	ENST0000009	14	NYFA	ENSG0000001167.12
7	chr6	41073585	41073831	SRX6602363.05_peak_12449	Promoter (<=1kb)	6	41072983	4107950	24968	1	ENST0000009	60	NYFA	ENSG0000001167.12
8	chrX	65534682	65534863	SRX6602363.05_peak_15613	Promoter (<=1kb)	23	65512582	65534754	22173	2	ENST0000009	9	LAS1L	ENSG0000001497.14
9	chr7	92245969	92246486	SRX6602363.05_peak_13482	Promoter (<=1kb)	7	92246234	92401384	155151	1	ENST0000009	9	ANKIB1	ENSG0000001629.7
10	chr7	92133734	92134905	SRX6602363.05_peak_13480	Promoter (<=1kb)	7	92112151	92134530	22380	2	ENST0000009	9	CYP51A1	ENSG0000001630.13
11	chr4	17577053	17577604	SRX6602363.05_peak_10972	Promoter (<=1kb)	4	17577192	17670968	30777	1	ENST0000009	9	LAP3	ENSG0000002549.10
12	chrX	2691128	2691429	SRX6602363.05_peak_15415	Promoter (<=1kb)	23	2691179	2741309	50131	1	ENST0000009	9	CD99	ENSG0000002586.15
13	chr7	43112887	43113282	SRX6602363.05_peak_13205	Promoter (<=1kb)	7	43112645	43562052	449408	1	ENST0000009	248	HECW1	ENSG0000002746.12
14	chr7	43113403	43113590	SRX6602363.05_peak_13206	Promoter (<=1kb)	7	43113741	43438763	325023	1	ENST0000009	-15	HECW1	ENSG0000002746.12
15	chr7	43249060	43249382	SRX6602363.05_peak_13207	Promoter (<=1kb)	7	43249137	43311993	62857	1	ENST0000009	9	HECW1	ENSG0000002746.12
16	chr7	43432192	43432373	SRX6602363.05_peak_13208	Intron (ENST00000395891.4/ENSG0000002746.12, intron 8 of 29)	7	43396673	43440597	43925	1	ENST0000009	35520	HECW1	ENSG0000002746.12
17	chr7	1929733	1929929	SRX6602363.05_peak_12964	Intron (ENST00000265854.9/ENSG00000002822.13, intron 15 of 16)	7	1898200	1940549	42350	2	ENST0000009	10620	MAD1L1	ENSG0000002822.13
18	chr7	1951020	1951453	SRX6602363.05_peak_12965	Intron (ENST00000265854.9/ENSG00000002822.13, intron 14 of 16)	7	1898200	1940549	42350	2	ENST0000009	-1042	MAD1L1	ENSG0000002822.13
19	chr7	1990928	1991172	SRX6602363.05_peak_12966	Intron (ENST00000265854.9/ENSG00000002822.13, intron 12 of 16)	7	1816082	1980791	164710	2	ENST0000009	-10139	MAD1L1	ENSG0000002822.13
20	chr7	2035682	2035940	SRX6602363.05_peak_12967	Intron (ENST00000265854.9/ENSG00000002822.13, intron 10 of 16)	7	1957629	2080013	122385	2	ENST0000009	44078	MAD1L1	ENSG0000002822.13
21	chr7	2069674	2069868	SRX6602363.05_peak_12968	Intron (ENST00000265854.9/ENSG00000002822.13, intron 9 of 16)	7	1957629	2080013	122385	2	ENST0000009	10145	MAD1L1	ENSG0000002822.13
22	chr7	2104311	2104516	SRX6602363.05_peak_12969	Intron (ENST00000265854.9/ENSG00000002822.13, intron 9 of 16)	7	2014506	2107708	93203	2	ENST0000009	3192	MAD1L1	ENSG0000002822.13
23	chr7	2105933	2106192	SRX6602363.05_peak_12970	Promoter (1-2kb)	7	2014506	2107708	93203	2	ENST0000009	1516	MAD1L1	ENSG0000002822.13
24	chr7	2106389	2106683	SRX6602363.05_peak_12971	Promoter (1-2kb)	7	2014506	2107708	93203	2	ENST0000009	10265	MAD1L1	ENSG0000002822.13
25	chr7	2161368	2162069	SRX6602363.05_peak_12972	Intron (ENST00000265854.9/ENSG00000002822.13, intron 8 of 16)	7	2069194	2175385	106192	2	ENST0000009	13316	MAD1L1	ENSG0000002822.13
26	chr7	2232471	2232652	SRX6602363.05_peak_12973	Promoter (<=1kb)	7	1815794	2232945	417152	2	ENST0000009	2908	MAD1L1	ENSG0000002822.13
27	chr7	2232790	2233176	SRX6602363.05_peak_12974	Promoter (<=1kb)	7	1815794	2232945	417152	2	ENST0000009	9	MAD1L1	ENSG0000002822.13
28	chr17	38869667	38870191	SRX6602363.05_peak_5933	Promoter (<=1kb)	17	38869859	38921769	51911	1	ENST0000009	9	LASP1	ENSG0000002834.15
29	chr17	48107433	48107849	SRX6602363.05_peak_6110	Promoter (<=1kb)	17	48107549	48119008	11460	1	ENST0000009	9	SNX11	ENSG0000002919.12
30	chr12	8949518	8949714	SRX6602363.05_peak_2789	Promoter (<=1kb)	12	8940363	8949636	9274	2	ENST0000009	9	M6PR	ENSG0000003056.5
31	chr12	8949833	8950166	SRX6602363.05_peak_2790	Promoter (<=1kb)	12	8940365	8949955	9591	2	ENST0000009	9	M6PR	ENSG0000003056.5
32	chrX	118117679	118117022	SRX6602363.05_peak_15726	Promoter (<=1kb)	23	117897813	118116798	218986	2	ENST0000009	9	KLHL13	ENSG0000003096.11
33	chr2	72143123	72143439	SRX6602363.05_peak_7906	Promoter (<=1kb)	2	72129393	72143084	13692	2	ENST0000009	-40	CYP26B1	ENSG0000003137.6
34	chr7	8261651	8262312	SRX6602363.05_peak_13049	Promoter (<=1kb)	7	8113184	8262281	149098	2	ENST0000009	9	ICA1	ENSG0000003147.15
35	chr16	90019417	90019853	SRX6602363.05_peak_5536	Promoter (<=1kb)	16	90004865	90019473	14609	2	ENST0000009	9	DBNDD1	ENSG0000003249.11
36	chr2	201780654	201781063	SRX6602363.05_peak_8431	Promoter (<=1kb)	2	201721990	201780956	58967	2	ENST0000009	9	ALS2	ENSG0000003393.12
37	chr2	201116409	201116746	SRX6602363.05_peak_8427	Promoter (<=1kb)	2	201116407	201139877	23471	1	ENST0000009	9	CFLAR	ENSG0000003402.17
38	chr8	17496765	17497353	SRX6602363.05_peak_13990	Promoter (<=1kb)	8	17497088	17570573	73486	1	ENST0000009	9	SLC7A2	ENSG0000003989.14
39	chr7	127588169	127588743	SRX6602363.05_peak_13681	Promoter (<=1kb)	7	127588345	127591705	3361	1	ENST0000009	9	ARF5	ENSG0000004059.8
40	chr17	28371382	28371843	SRX6602363.05_peak_5827	Promoter (<=1kb)	17	28371662	28404049	32388	1	ENST0000009	9	SARM1	ENSG0000004139.11
41	chr17	28357428	28357792	SRX6602363.05_peak_5826	Promoter (<=1kb)	17	28347177	28357522	10346	2	ENST0000009	9	POLDIP2	ENSG0000004142.9
42	chr3	129577448	129577656	SRX6602363.05_peak_10528	Promoter (<=1kb)	3	129574336	129578548	4213	2	ENST0000009	892	PLXND1	ENSG0000004399.10
43	chr3	129606167	129606672	SRX6602363.05_peak_10529	Promoter (<=1kb)	3	129555175	129606818	51644	2	ENST0000009	146	PLXND1	ENSG0000004399.10
44	chr3	129606792	129607105	SRX6602363.05_peak_10530	Promoter (<=1kb)	3	129555175	129606818	51644	2	ENST0000009	9	PLXND1	ENSG0000004399.10
45	chr1	33036730	33037022	SRX6602363.05_peak_315	Promoter (<=1kb)	1	33007984	33036870	28887	2	ENST0000009	9	AK2	ENSG0000004455.14

visualized in plots

will be used in second part of workshop for overlap with RNA-seq

ChIPseeker: Output

15: ChIPseeker on data 13 and data 1: Plots

20.4 KB

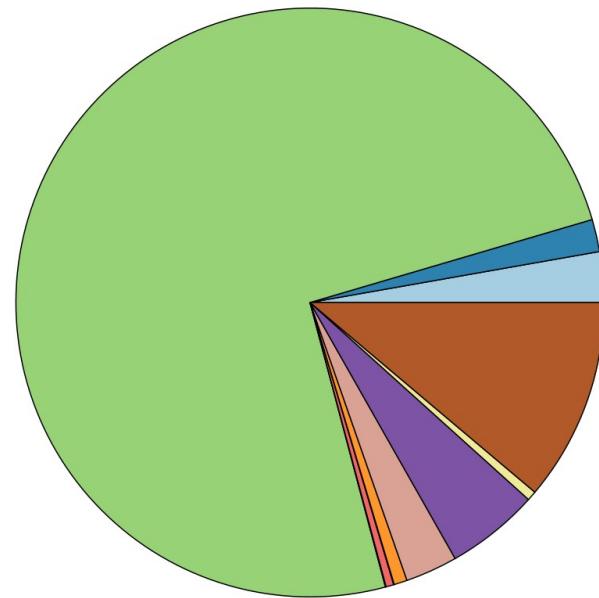
format: **pdf**, database: ?

```
>> preparing features information...
2022-05-09 07:45:41
>> identifying nearest features...
2022-05-09 07:45:43
>> calculating distance from peak to
TSS... 2022-05-09 07:45:45
>> assigning genomic annotation...
2022-05-09 07:45:45
>> assigning
```

[Image in pdf format](#)



- Promoter (1-2kb) (2.77%)
- Promoter (2-3kb) (1.78%)
- Promoter (<=1kb) (74.58%)
- 5' UTR (0.04%)
- 3' UTR (0.43%)
- 1st Exon (0.03%)
- Other Exon (0.72%)
- 1st Intron (2.87%)
- Other Intron (5.11%)
- Downstream (<=300) (0.53%)
- Distal Intergenic (11.15%)

click to download plots as .pdf
→ several modes of representation

Practical Session

Task: Repeat Peak annotation with ChIPseeker using the dataset with more stringent threshold ($q < 1E-20$)

Practical Session

What problems did you encounter?
What did you find out about the
binding of ARNTL to genomic regions?

Which questions remain open?

Extra Slides

Annotation Conversion: LiftOver



Lift Genome Annotations

This tool converts genome coordinates and genome annotation files between assemblies. The input data can be pasted into the text box, or uploaded from a file. If a pair of assemblies cannot be selected from the pull-down menus, a direct lift between them is unavailable. However, a sequential lift may be possible. Example: lift from Mouse, May 2004, to Mouse, Feb. 2006, and then from Mouse, Feb. 2006 to Mouse, July 2007 to achieve a lift from mm5 to mm9.

Original Genome:

Mouse

Original Assembly:

July 2007 (NCBI37/mm9)

New Genome:

Mouse

New Assembly:

Dec. 2011 (GRCm38/mm10)

species of
original data

assembly of
original data

species of
output data

assembly of
output data

Minimum ratio of bases that must remap:

0.95

leave at
default
values

BED 4 to BED 6 Options

Allow multiple output regions:

0

Minimum hit size in query:

0

Minimum chain size in target:

BED 12 Options

Min ratio of alignment blocks or exons that must map:

1

If thickStart/thickEnd is not mapped, use the closest mapped base:

Paste in data ([BED](#) or chrN:start-end formats):

this is where you enter your data....

Submit

Clear

Or upload data from a file ([BED](#) or chrN:start-end in plain text format):

[Browse...](#)

No file selected.

[Submit File](#)

Annotation Conversion: LiftOver

1. open peak data (.bed format) with Text editor
2. copy-paste peak coordinates in LiftOver (or browse)

Paste in data ([BED](#) or chrN:start-end formats):

```
chr1 4797331 4798448 7.19 79 + 4798251 4798251 4798251
chr1 4809404 4810115 9.84 34 + 4809720 4809720 4809720
chr1 4824788 4826020 14.20 70 + 4825108 4825108 4825108
chr1 4845400 4846273 5.56 35 + 4845985 4845985 4845985
chr1 4847329 4848698 8.94 104 + 4847815 4847815 4847815
chr1 5007828 5009145 10.15 70 + 5008872 5008872 5008872
chr1 5012642 5013389 8.94 45 + 5013031 5013031 5013031
chr1 5072995 5073758 43.48 132 + 5073191 5073191 5073191
chr1 5222901 5223395 14.63 34 + 5223136 5223136 5223136
chr1 6132783 6133410 13.82 30 + 6133104 6133104 6133104
chr1 6204227 6205664 11.98 136 + 6204707 6204707 6204707
```

Submit Clear

Or upload data from a file ([BED](#) or chrN:start-end in plain text format):

3. submit job, view and download results

Results

Successfully converted 16390 records. [View Conversions](#)

Conversion failed on 10 records. [Display failure file](#) [Explain failure messages](#)

click to download

Practical Session

1. Go to ChIP-Atlas and look for ARNTL data in a neural context from mouse
2. Download and visualize the dataset on IGV (different session)
3. Use LiftOver to convert the dataset to hg38