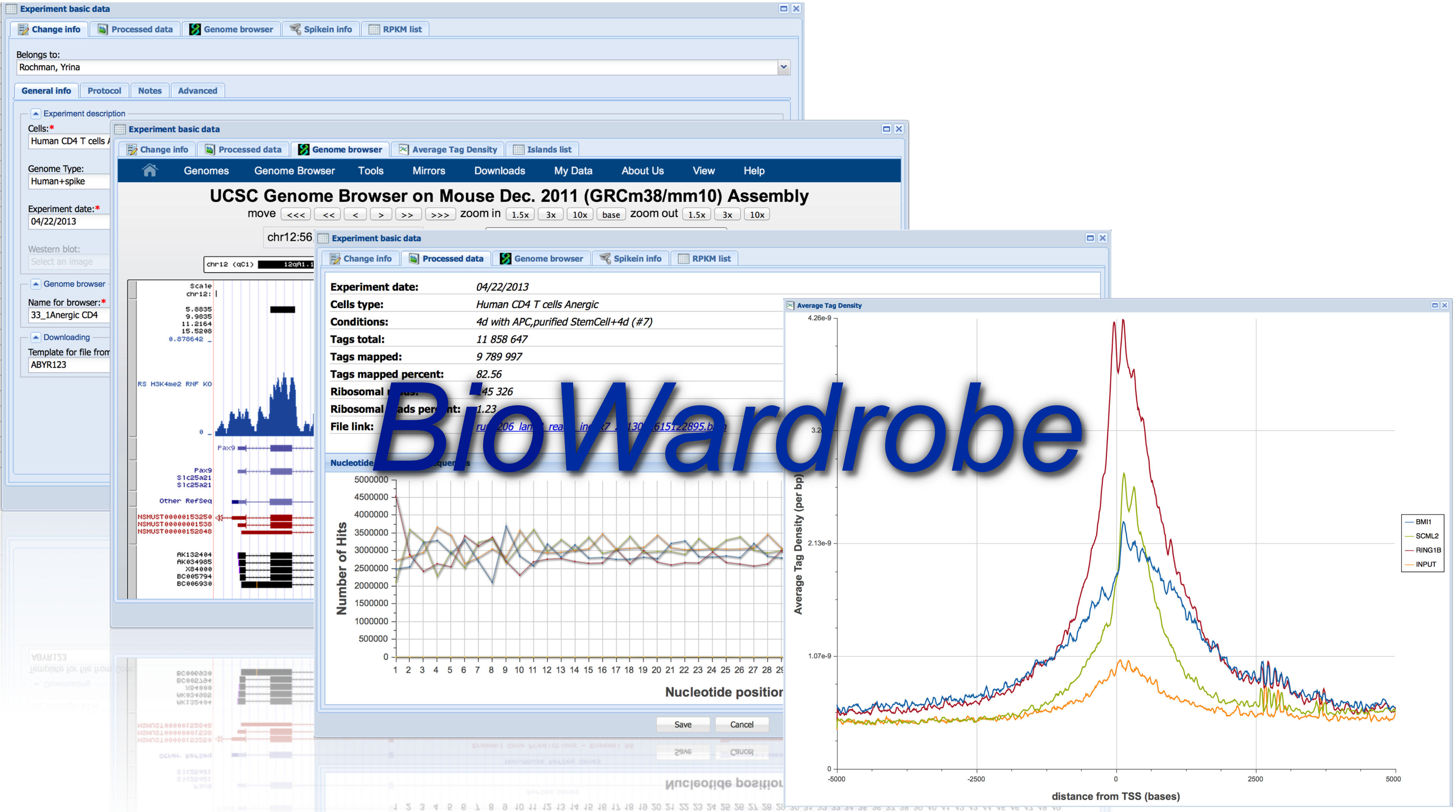


Belongs to: Rochman, Yrina

Cells:* Human CD4 T cells A

Genome Type: Human+spike

Experiment date:* 04/22/2013

Name for browser:* 33_1Anergic CD4

Template for file from: ABYR123

| | |
|---|---|
| **Experiment date:** | *04/22/2013* |
| **Cells type:** | *Human CD4 T cells Anergic* |
| **Conditions:** | *4d with APC,purified StemCell+4d (#7)* |
| **Tags total:** | *11 858 647* |
| **Tags mapped:** | *9 789 997* |
| **Tags mapped percent:** | *82.56* |
| **Ribosomal** | *45 326* |
| **Ribosomal ads per nt:** | *1.23* |
| **File link:** | *ru 206 lan rea in 7 130 615122895.b* |

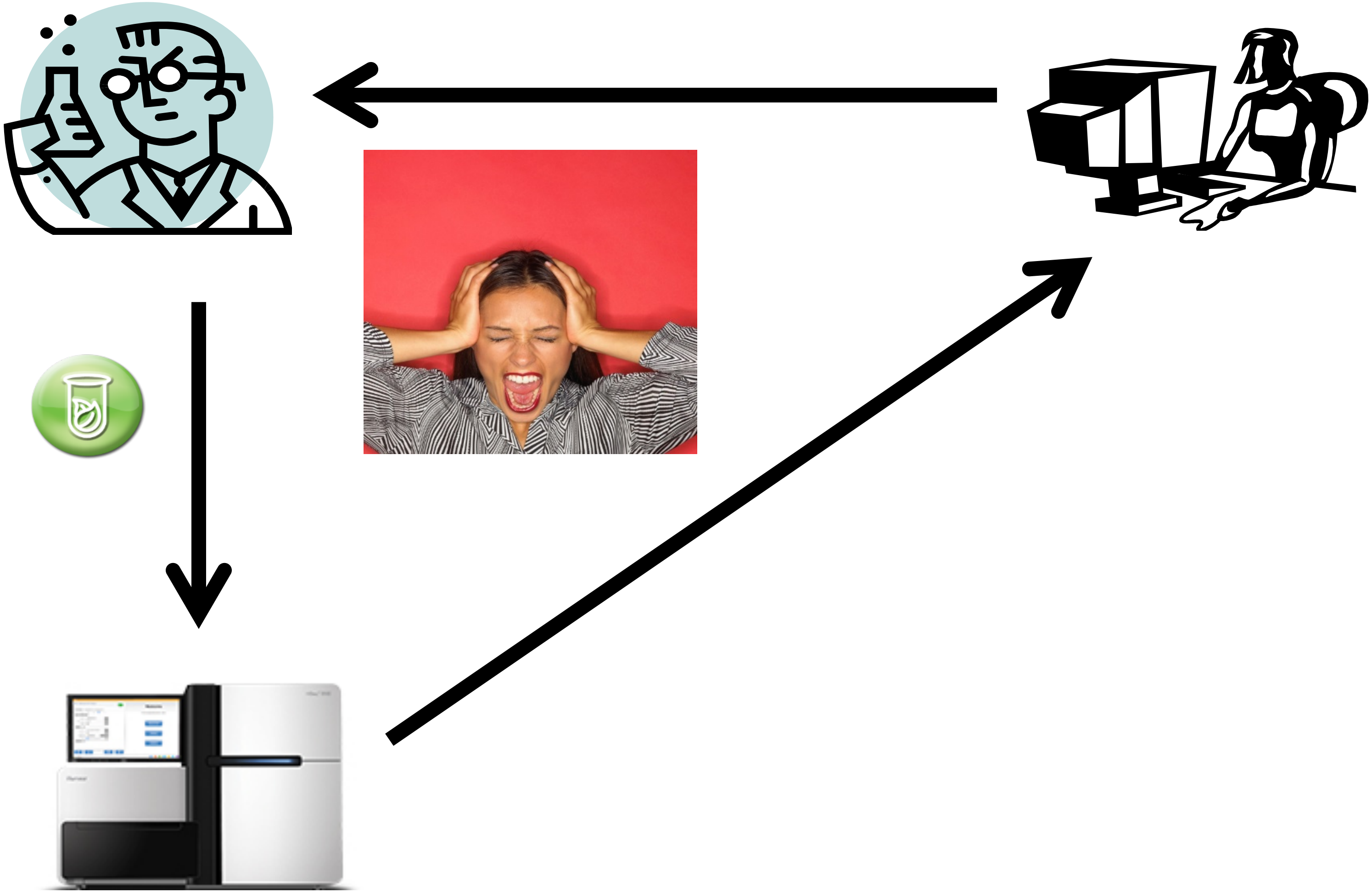
# BioWardrobe

So easy, even a full professor can do data analysis!

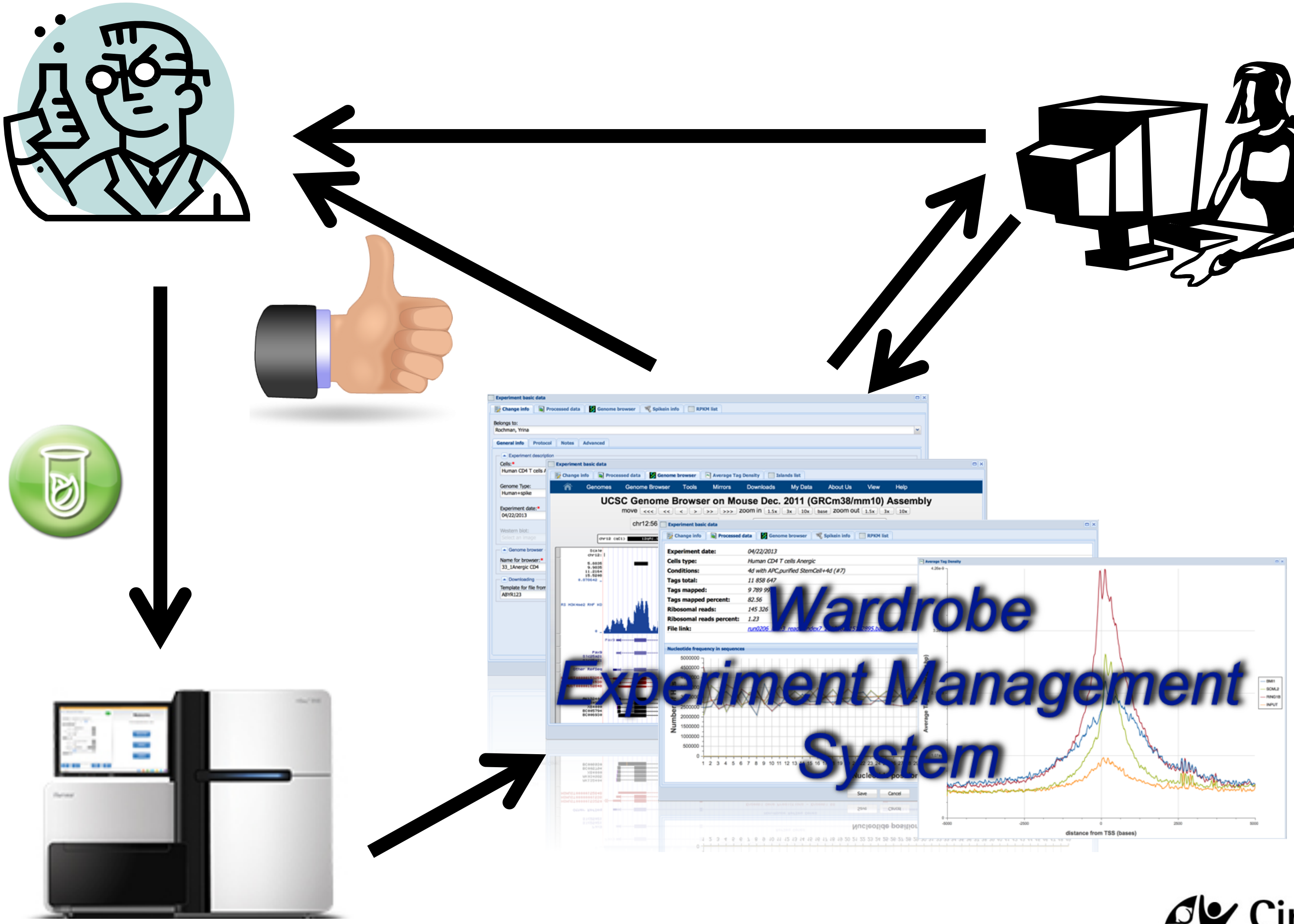
Andrey Kartashov and Artem Barski

www.biowardrobe.com

# NGS Data Analysis: Problem



# NGS Data Analysis: Solution



**Wardrobe Experiment Management System**

Experiment date: 04/22/2013
Cells type: Human CD4 T cells Anergic
Conditions: 4d with APC,purified StemCell+4d (#7)
Tags total: 11 858 647
Tags mapped: 9 789 99
Tags mapped percent: 82.56
Ribosomal reads: 145 326
Ribosomal reads percent: 1.23
File link: run0206_[illegible]

Cincinnati Children's

# Current State

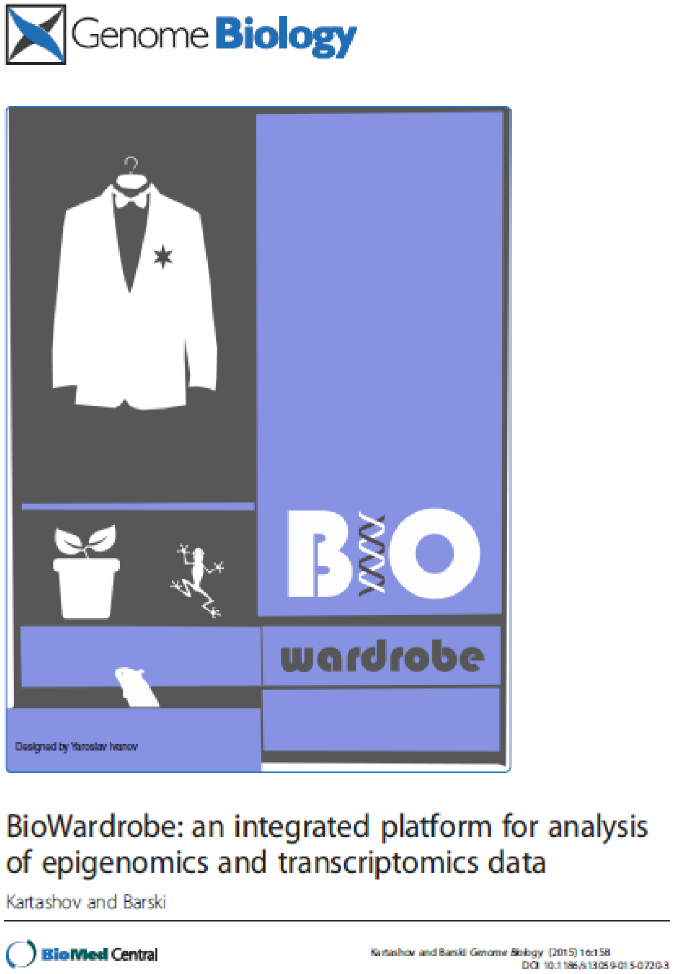
- >30 laboratories
- >80 user accounts
- 2365 libraries
  - ~650 from databases
  - 1699 ChIP-Seq/DNAse-Seq, etc
  - 666 RNA-Seq
- ~10Tb of user data
- 10 publications
- Human, Mouse, Rat, Fly, Frog

Cincinnati Children's

# Publication and Web site

Biowardrobe.com

demo.biowardrobe.com



Genome Biology

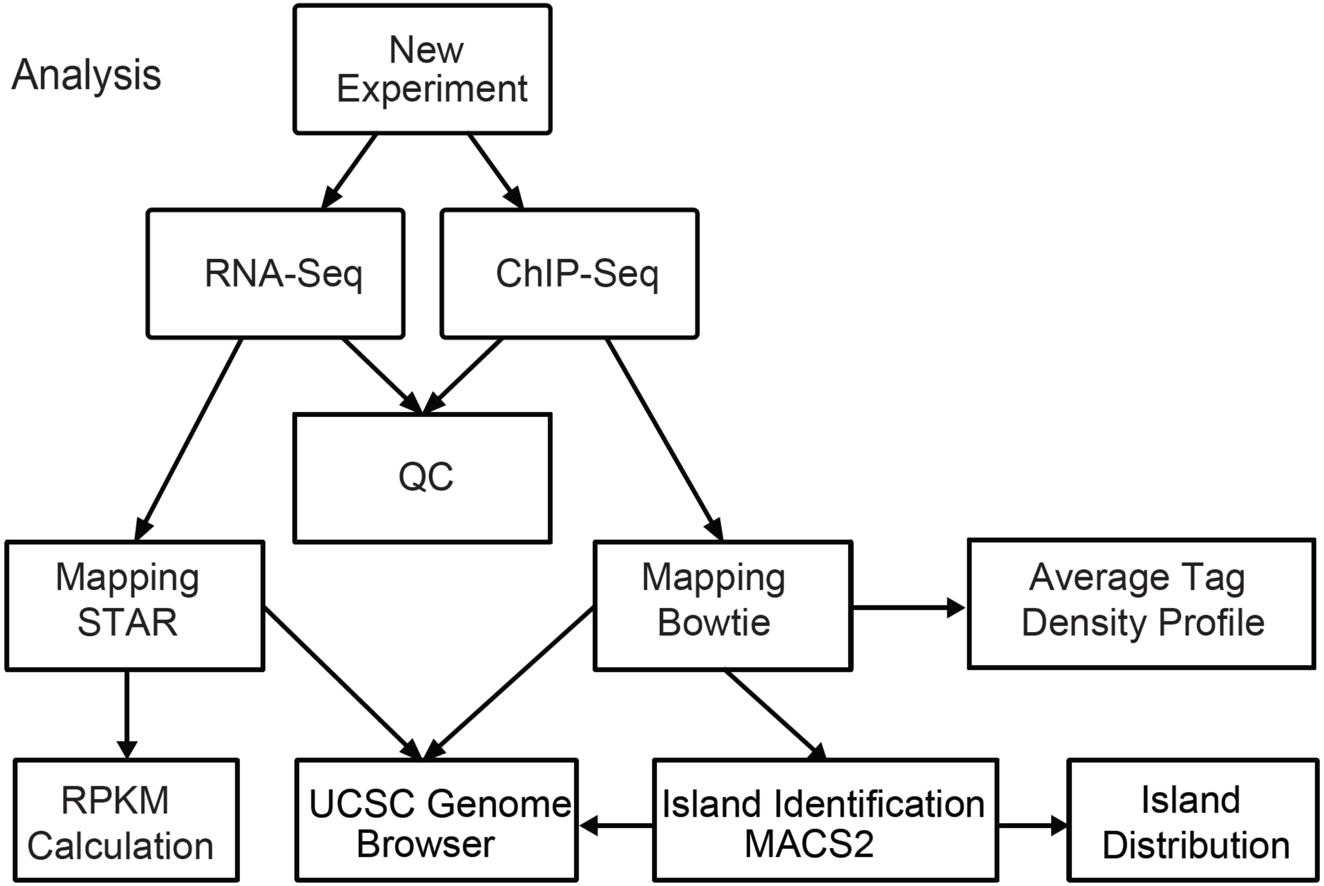
BioWardrobe: an integrated platform for analysis of epigenomics and transcriptomics data

Kartashov and Barski

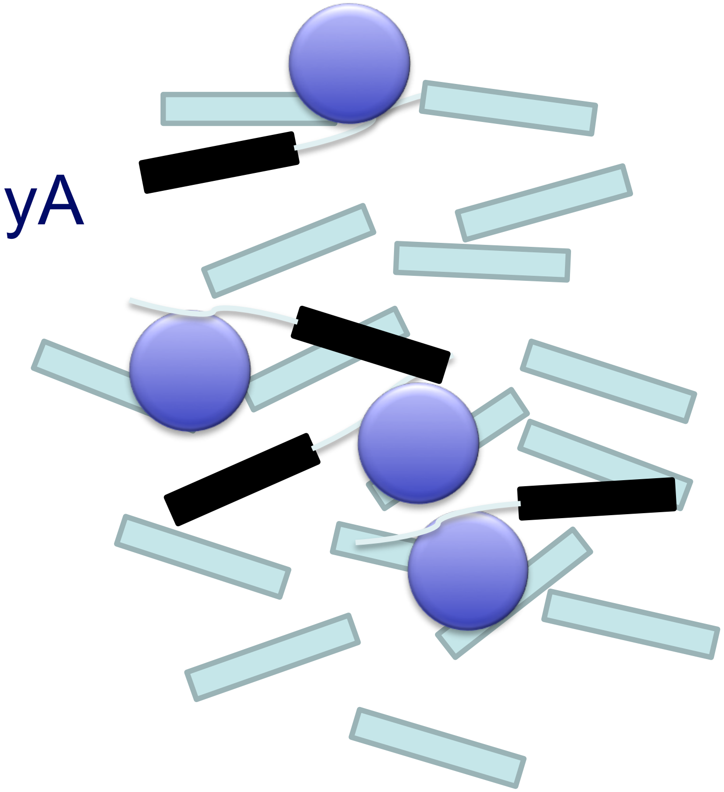
BioMed Central
Kartashov and Barski Genome Biology (2015) 16:158
DOI 10.1186/s13059-015-0720-3

# Basic Analysis

Basic Analysis



# RNA-Seq: Selection/Depletion



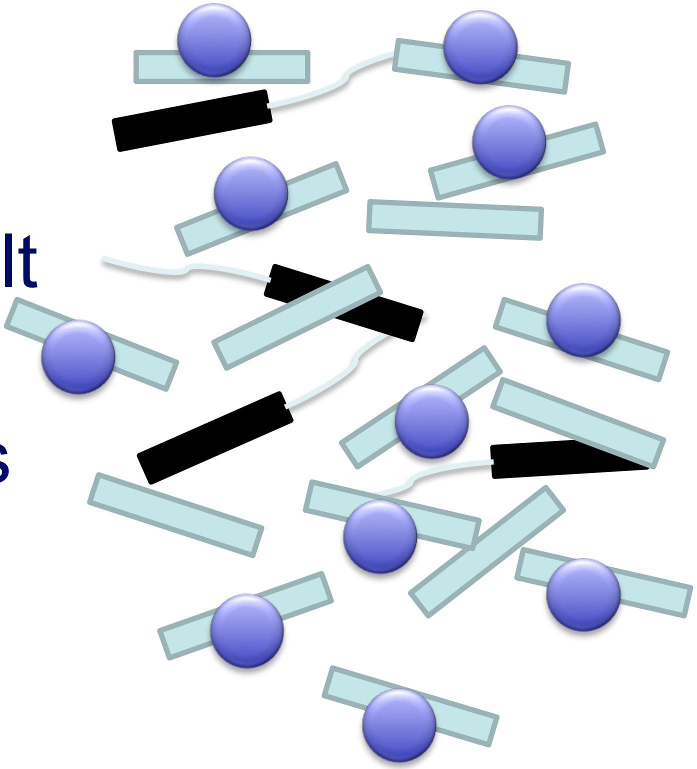
## Poly-A selection

- mRNAs (usually) have polyA tails, rRNAs and tRNAs don’t
- Cheap and easy, but requires intact RNA
- Doesn’t work on prokaryotes…
- Illumina TruSeq
- ~ 0-2% of output is rRNA

# Selection/Depletion

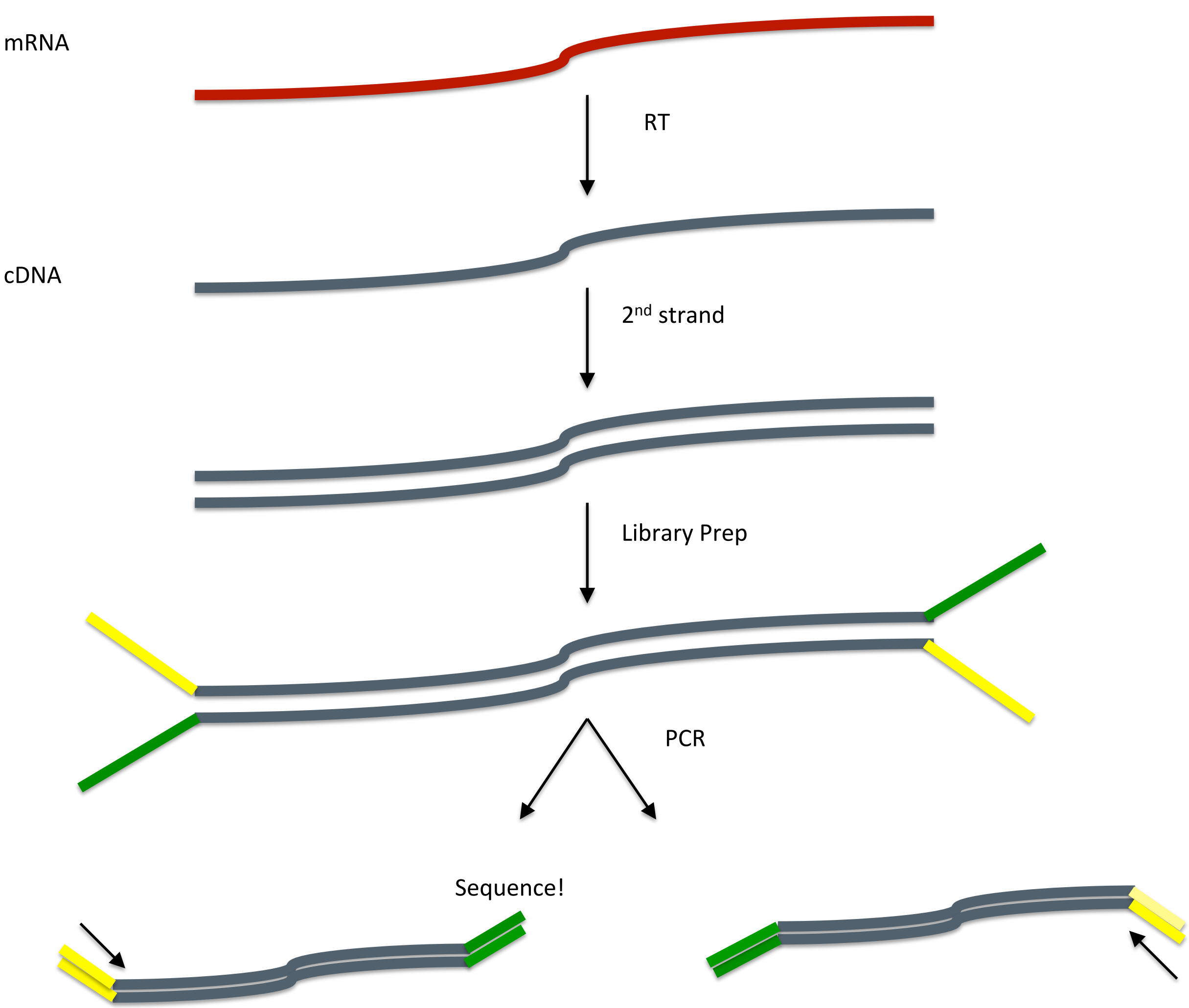
rRNA Depletion (eg RiboMinus)

- Seems to be more difficult
- Uses LNAs to capture rRNA-specific sequences
- Still dependent on intact RNA
- ~ 10-50% of output is rRNA



Cincinnati Children's

# RNA-Seq Strand Specificity



mRNA

RT

cDNA

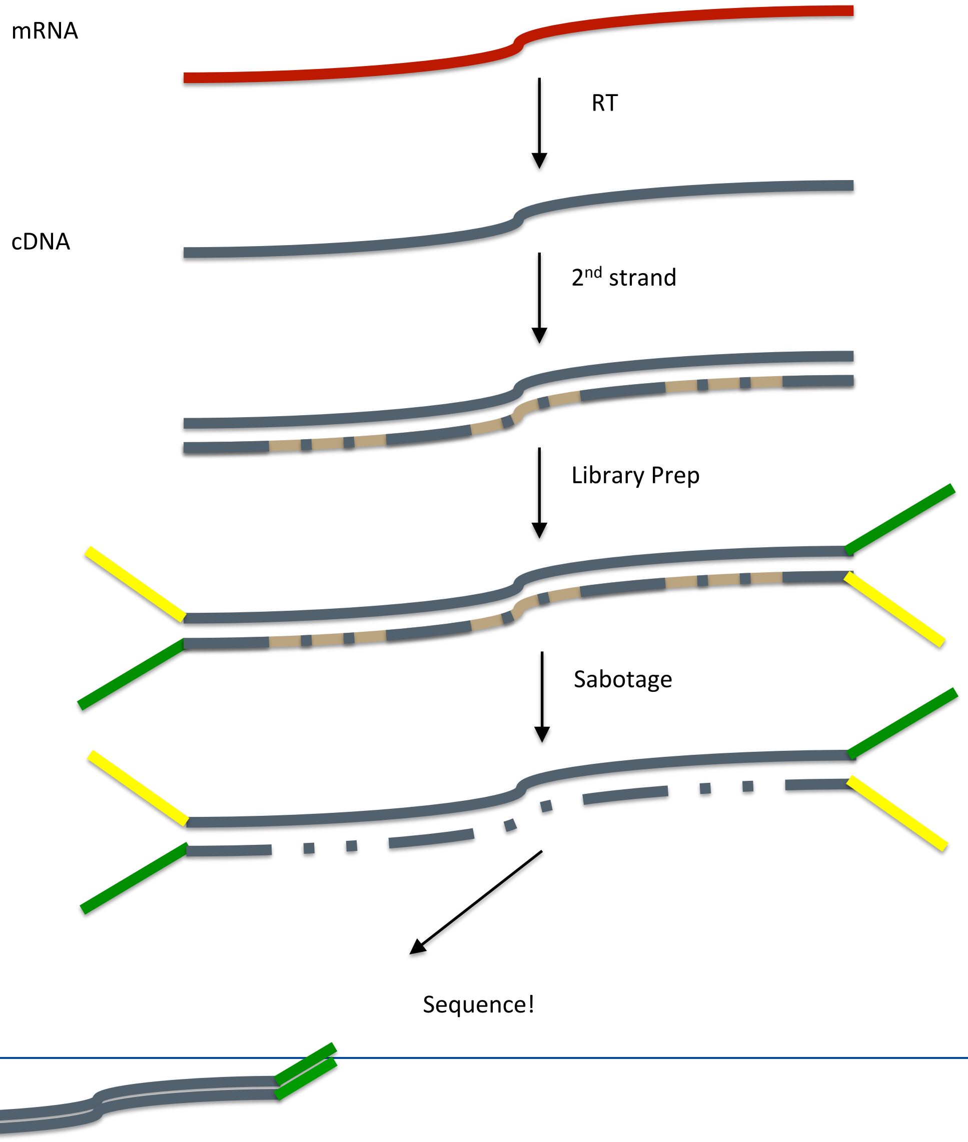
2$^{nd}$ strand

Library Prep

PCR

Sequence!

# Mapping dUTP RNA-Seq libraries



mRNA

RT

cDNA

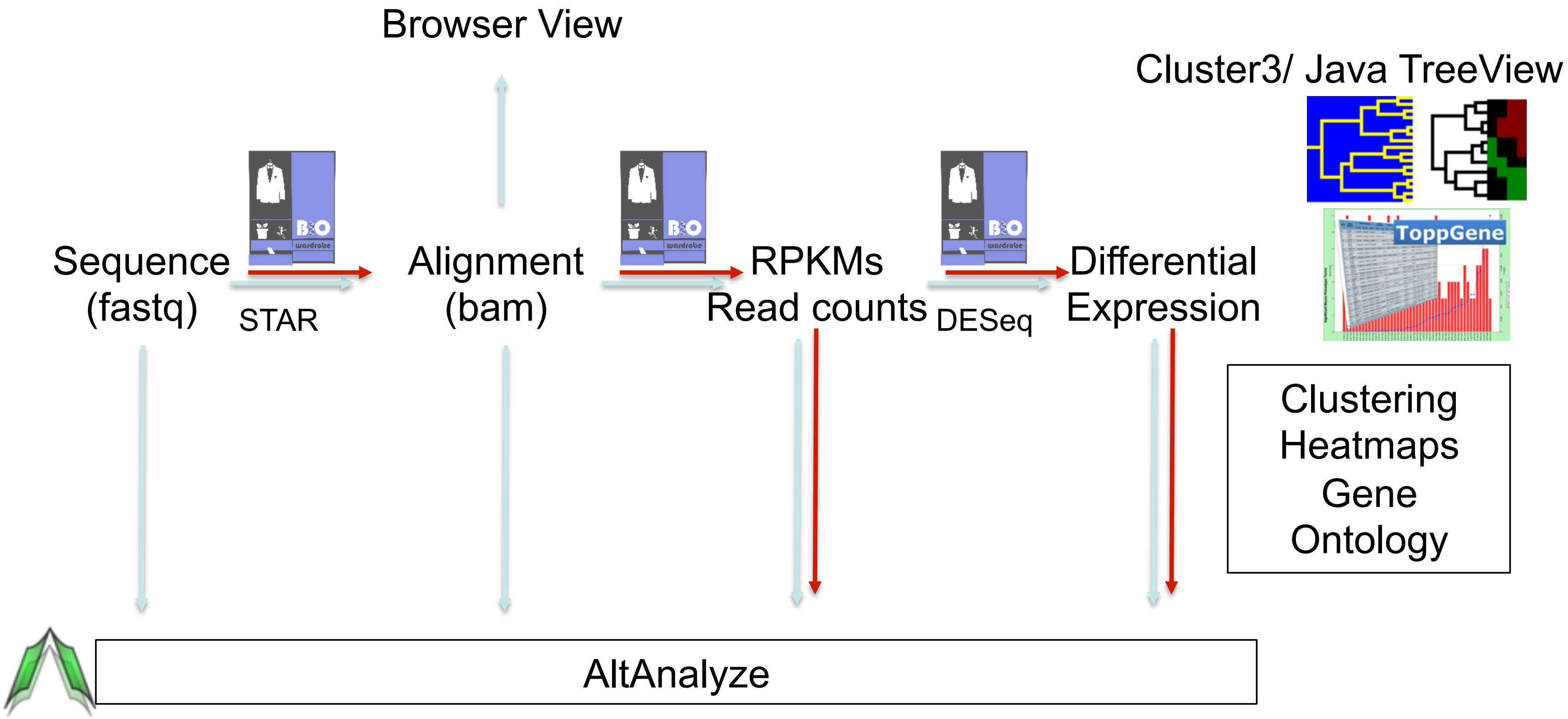
2nd strand

Library Prep

Sabotage

Sequence!

# RNA-Seq



Browser View

Cluster3/ Java TreeView

Sequence (fastq)

STAR

Alignment (bam)

RPKMs
Read counts

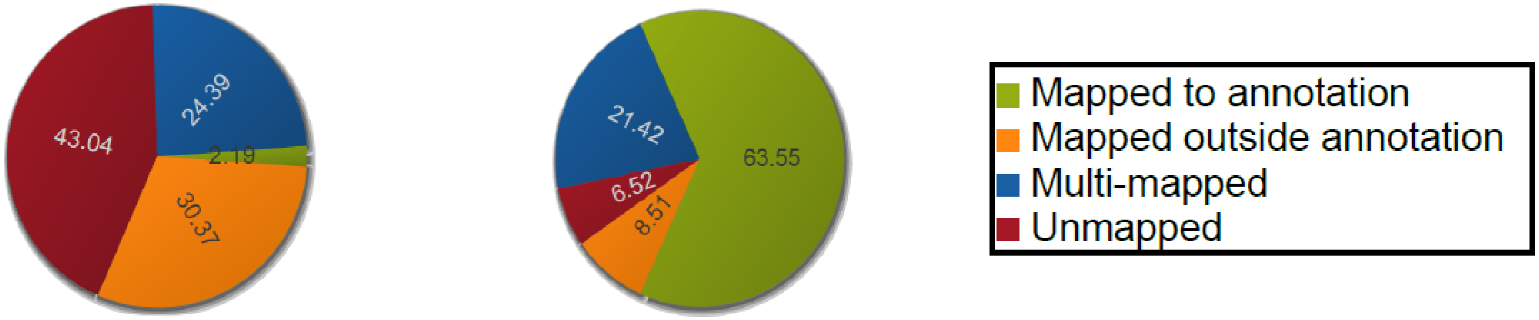
DESeq

Differential Expression

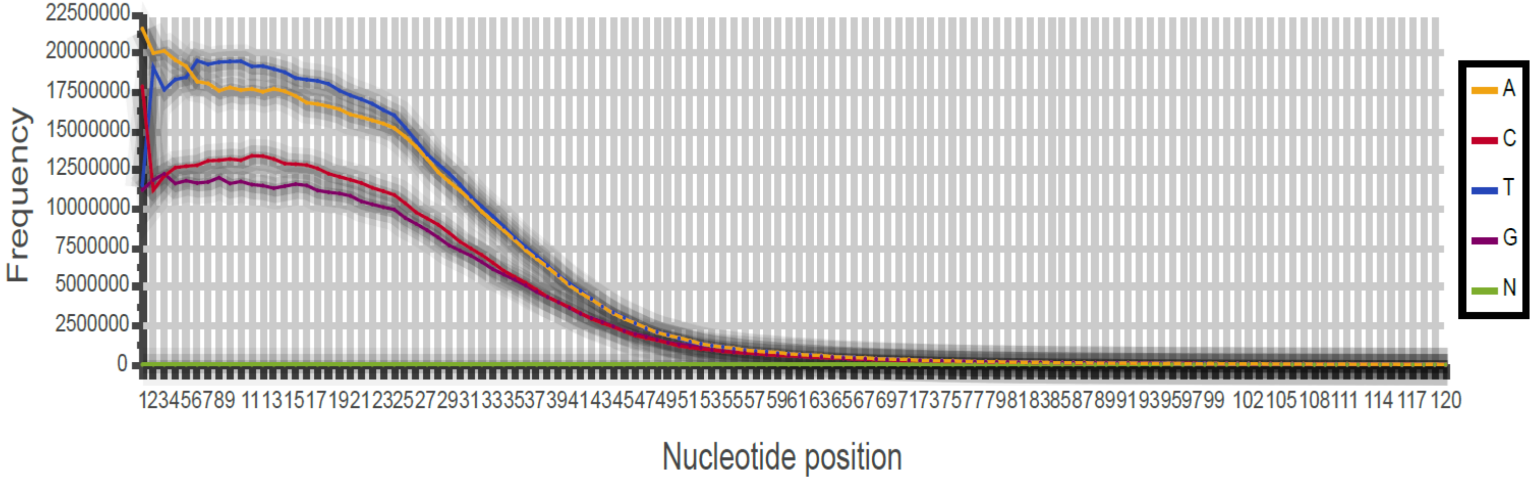
Clustering
Heatmaps
Gene
Ontology

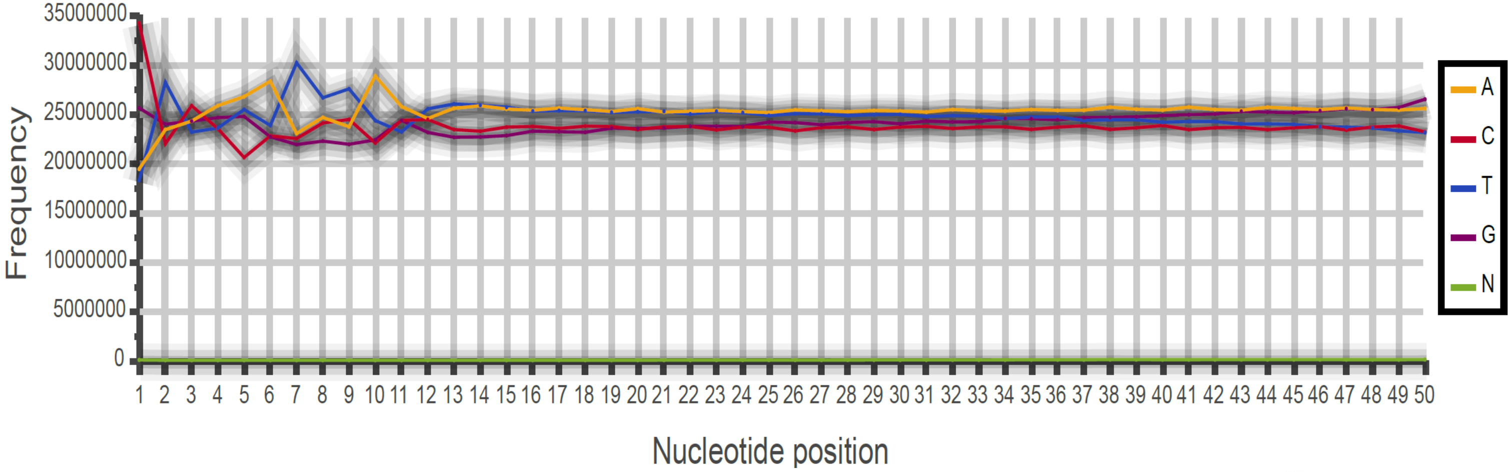
AltAnalyze

# Quality controls: Mapping



# Quality controls: Base Frequency

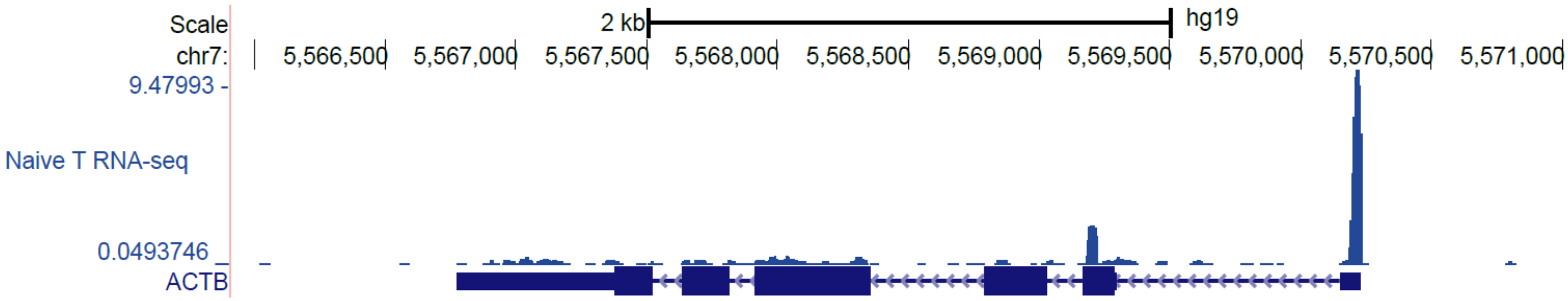


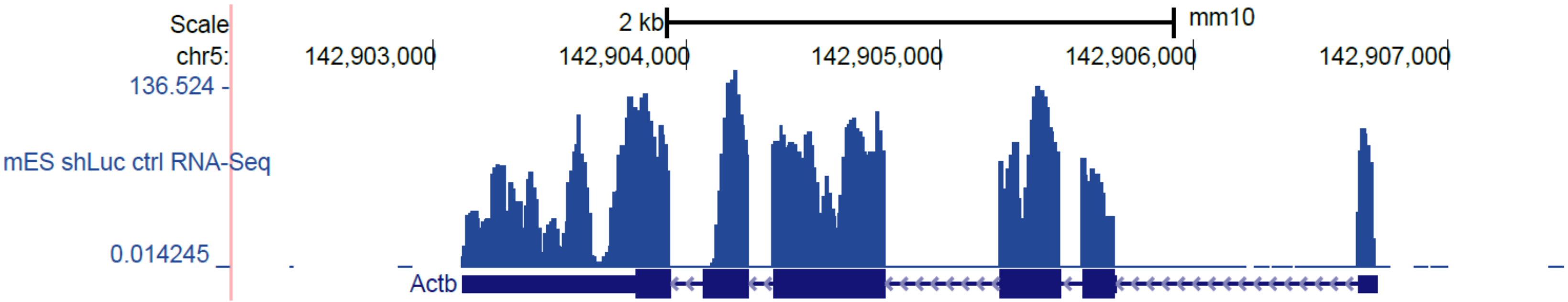


# UCSC genome browser
# RNA-Seq Coverage

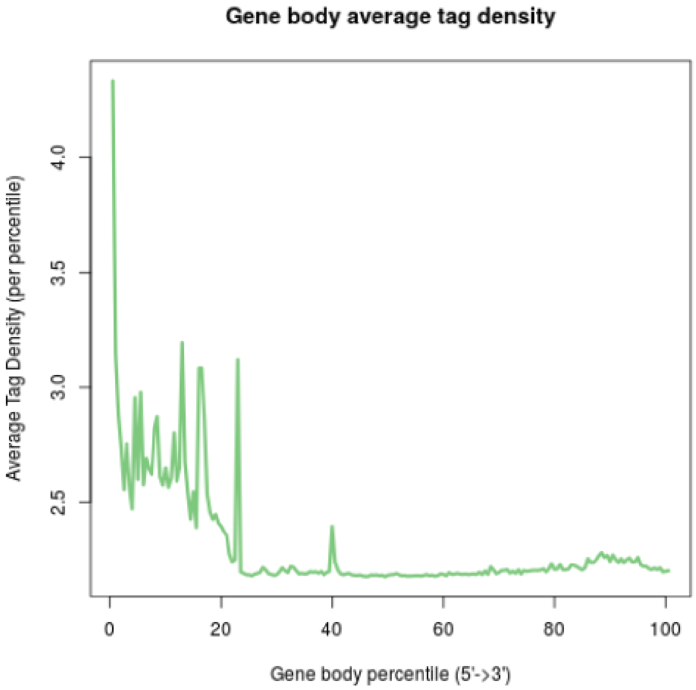
RNA: read length

# Quality controls: Browser View

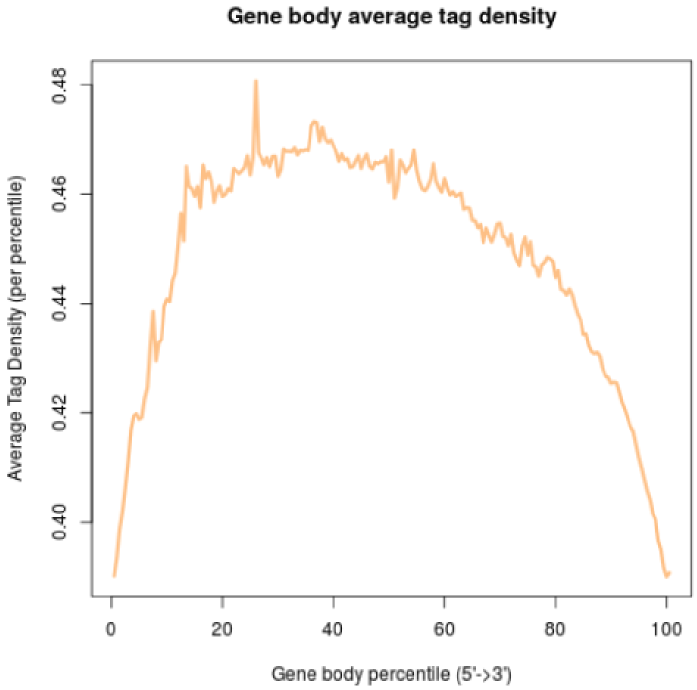




# Quality controls: average read density



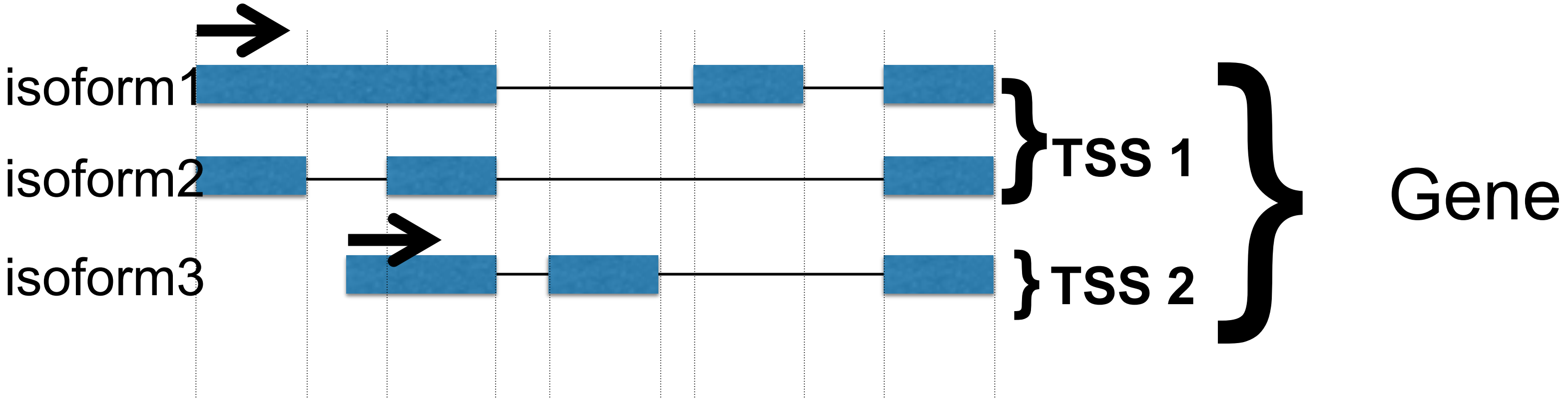
h



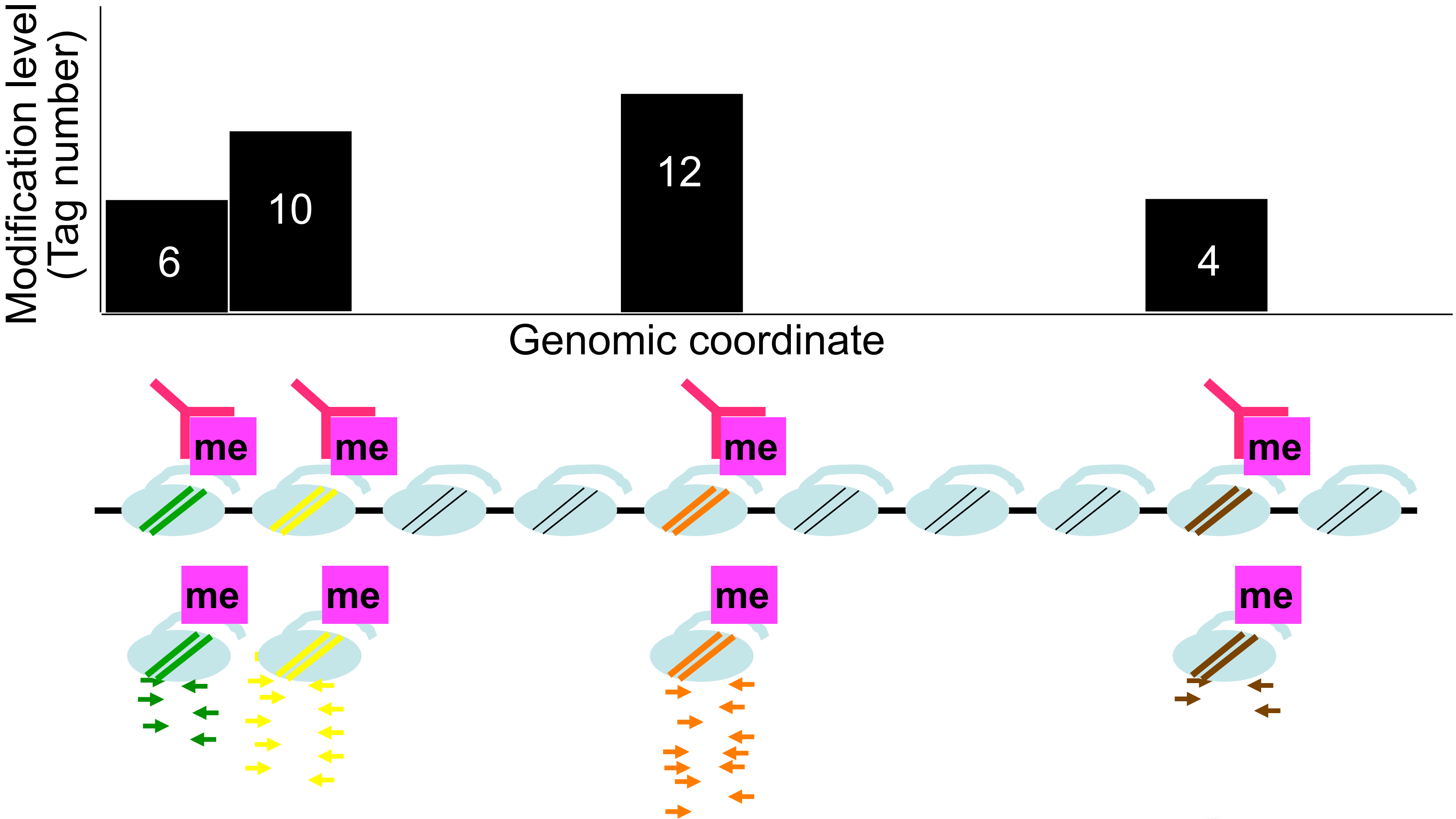
# Different RPKM grouping

With respect to annotation, RPKMs can be grouped by

- Isoforms
- Transcription Start Site, sum up all isoform's RPKMs for common TSS
- Gene name, sum up all isoform's RPKMs for the same gene name



# *Chromatin Immunoprecipitation-Sequencing (ChIP-Seq)*



Barski et al. (2007) Mikkelsen et al. (2007)

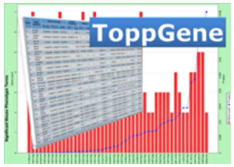
Robertson et al. (2007) Johnson et al. (2007)

Cincinnati Children's

# ChIP-Seq

Browser View

Genes



Sequence (fastq)

BowTie

Alignment (bam)

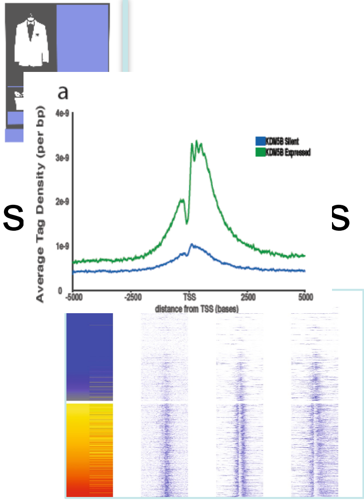
MACS

Islands

MAnorm

Differential Binding

Average Profiles

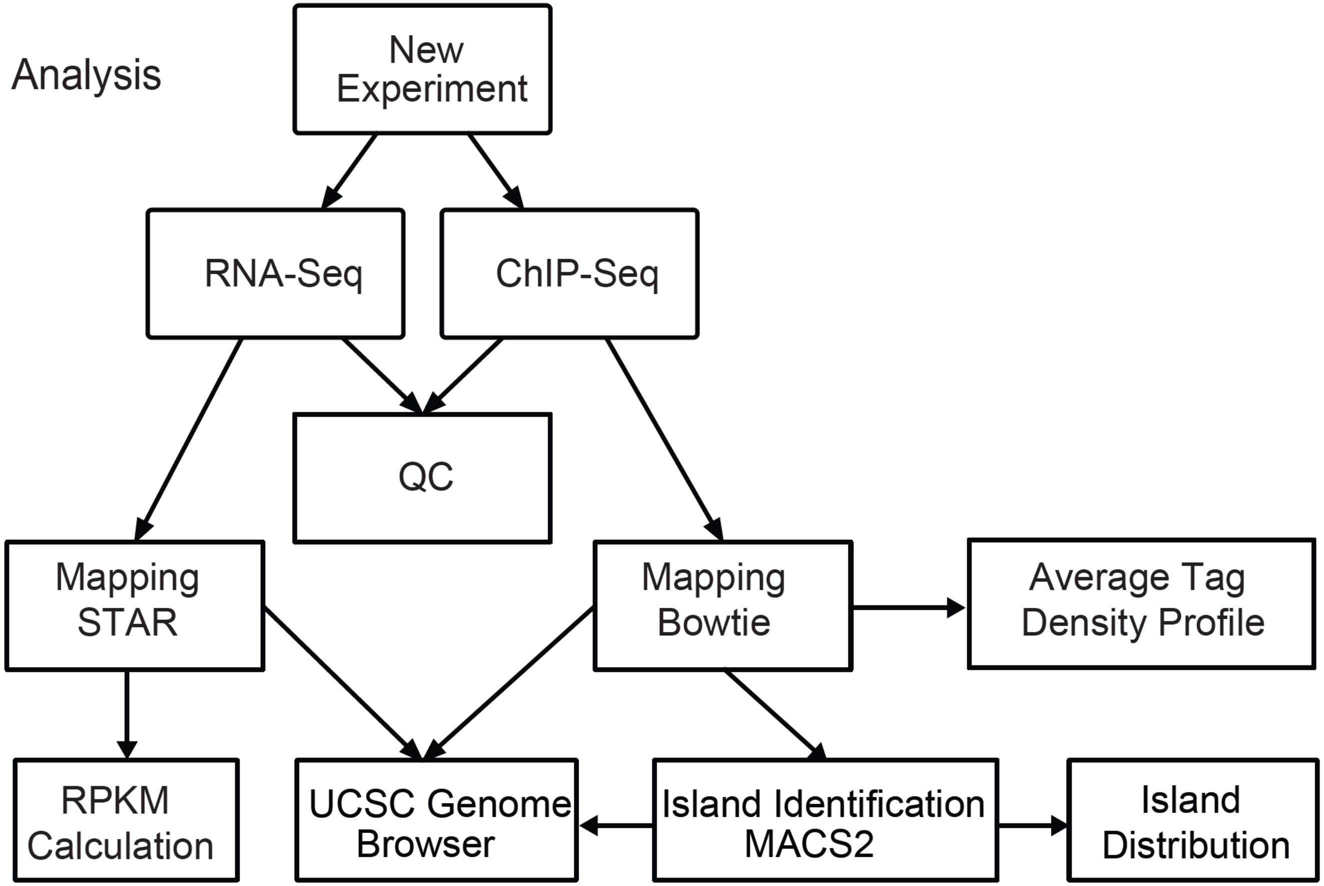


s

MEME-ChIP

# Basic Analysis

Basic Analysis


New Experiment

RNA-Seq

ChIP-Seq

QC

Mapping STAR

Mapping Bowtie

Average Tag Density Profile

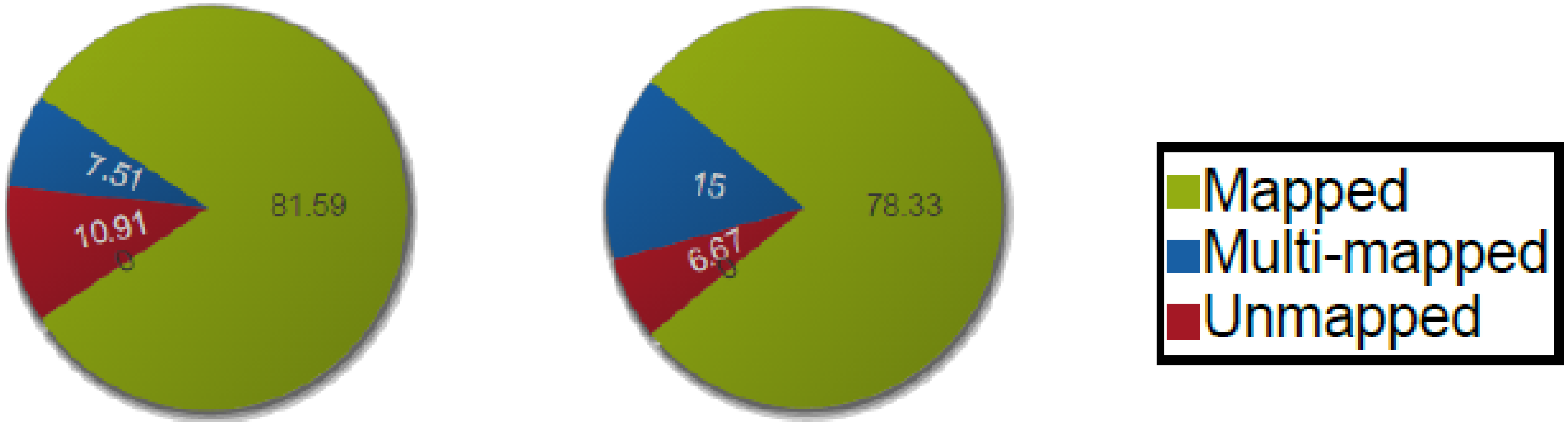
RPKM Calculation

UCSC Genome Browser

Island Identification MACS2

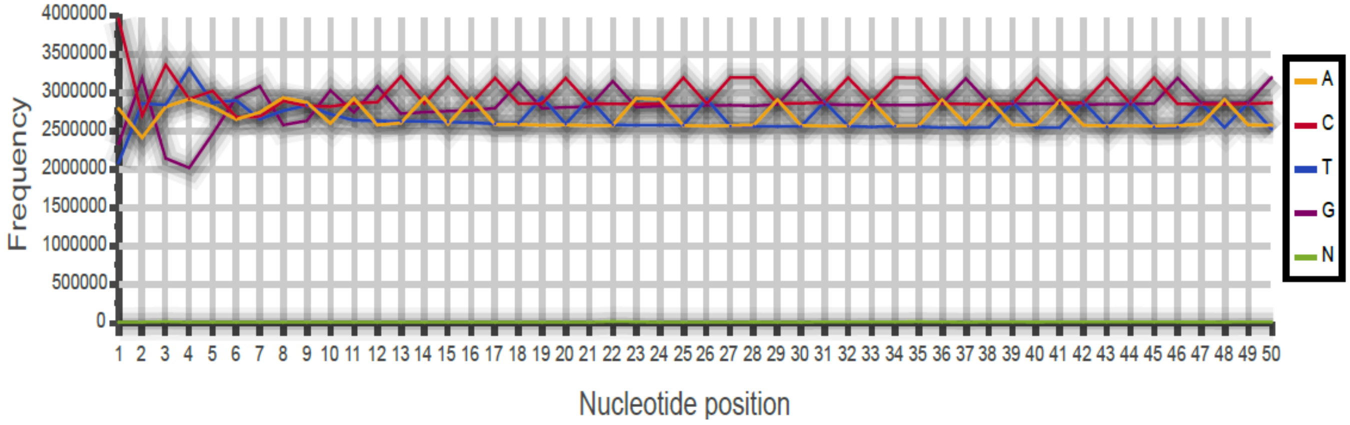
Island Distribution

# Quality controls: Mapping

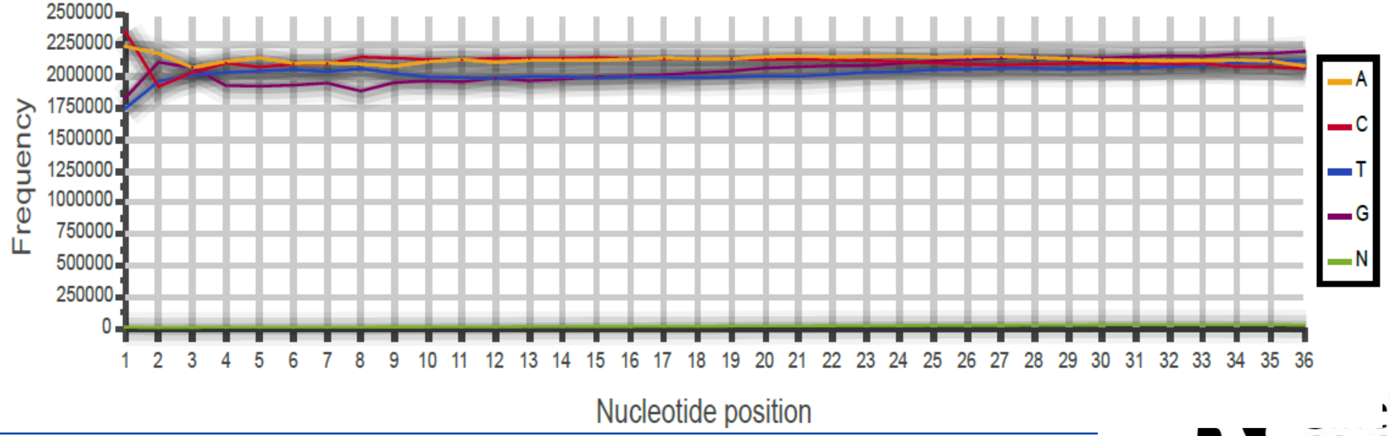


# Quality controls: Base Frequency

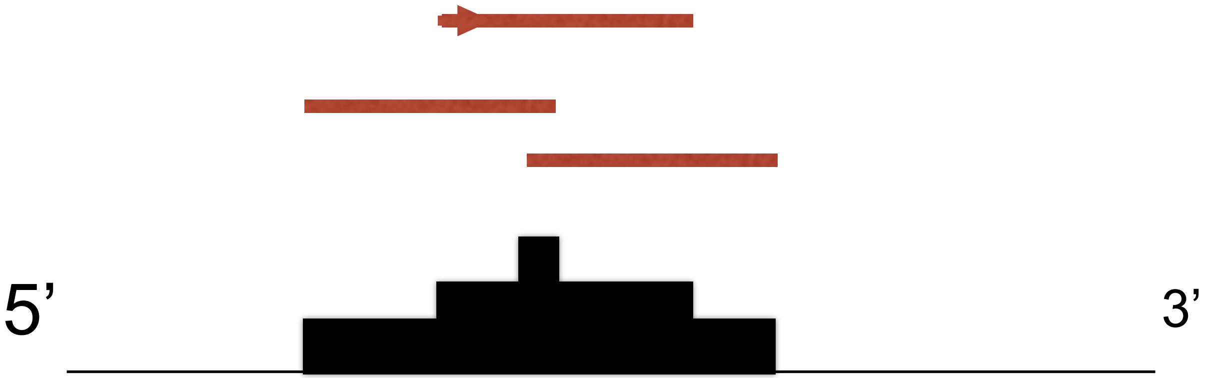
c



d

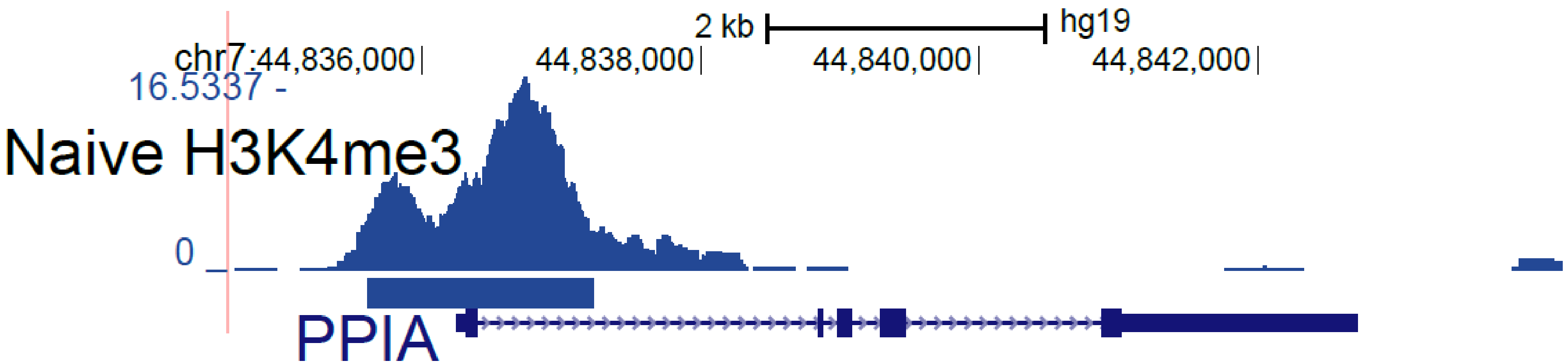


# UCSC genome browser
# Coverage for ChIP-Seq

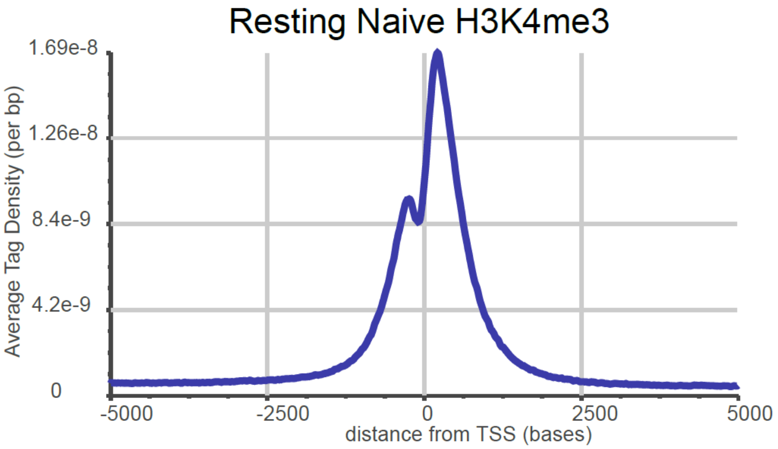


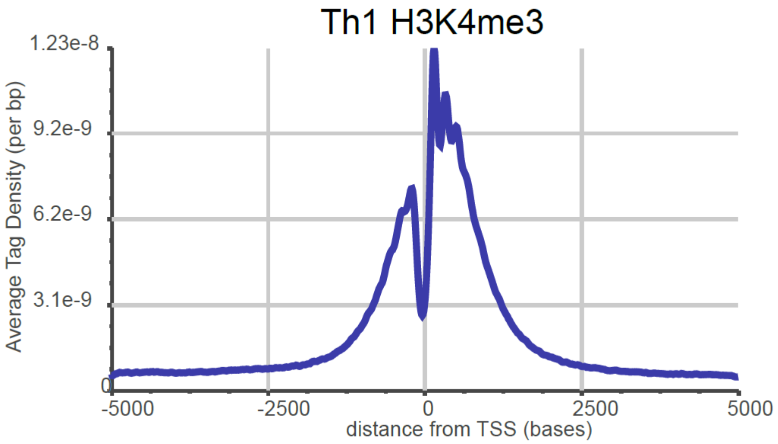
DNA: Estimated fragment size for single reads, original fragment size for pair-end reads

# Quality controls: Browser View

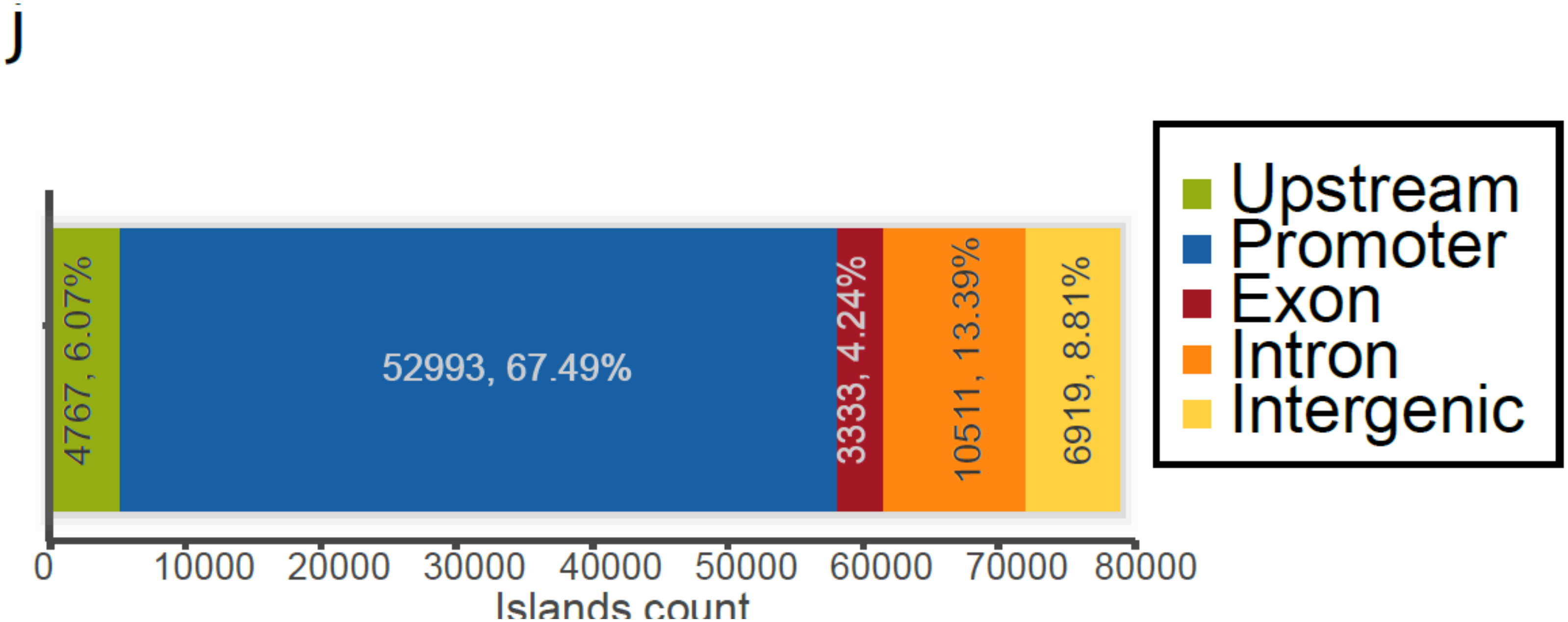


# Quality controls: average read density


Resting Naive H3K4me3


Th1 H3K4me3

# Island Distribution



# Subscriptions

| | New libraries | Old library storage | Manual analysis | Cost/year |
|---|---|---|---|---|
| **Mini** | 25 | 50 | | $800 |
| **Midi** | 50 | 100 | | $1500 |
| **Maxi** | 100 | 200 | 2hrs | $2750 |
| **Unlimited** | Unlimited | Unlimited | 5hrs | $4000 |

# Facility Services and fees

- Initial analysis and 1 year storage: $35
- Storage after the 1st year: $10/year ($0.82/month)
- Manual help (e.g. submission of data to NCBI): $100/hr
- First 3 libraries: free
- Fees will be reduced/waived for collaborators who have funded grants with us

# Advanced Analysis