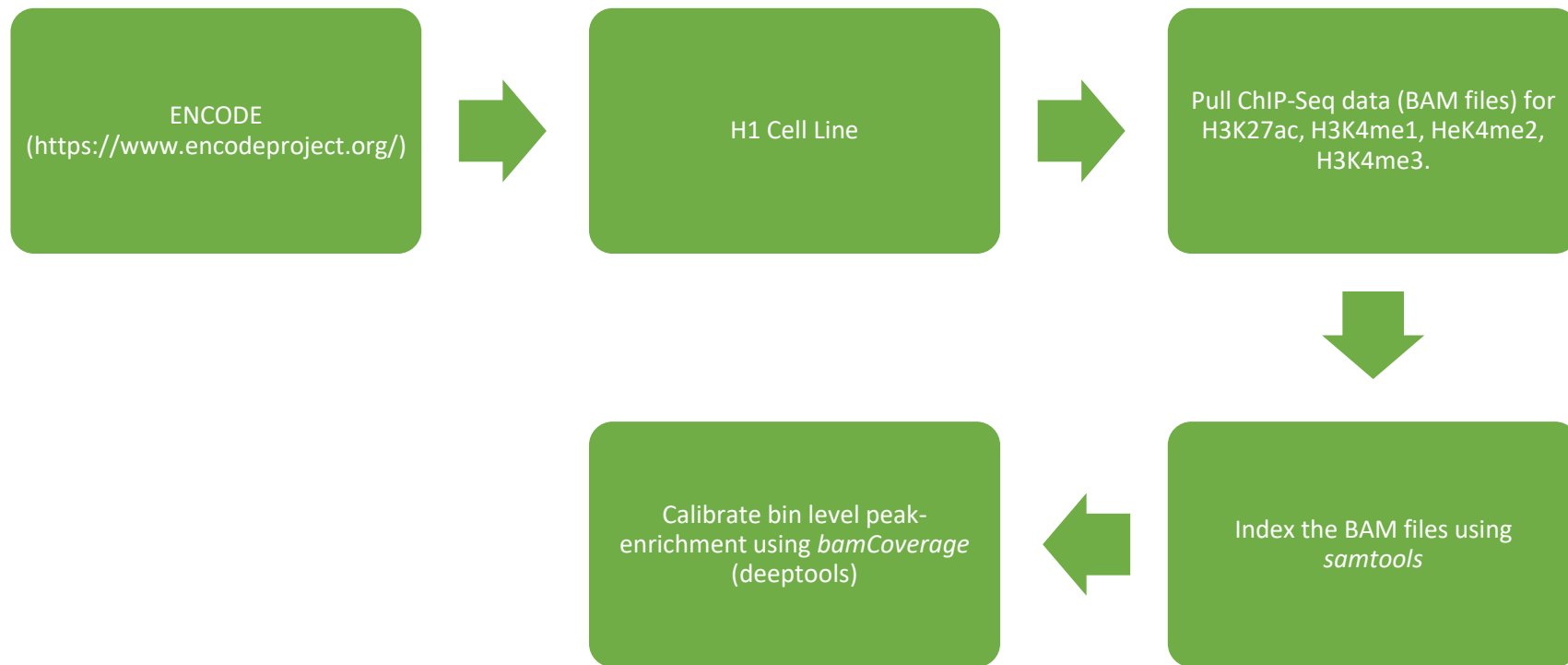
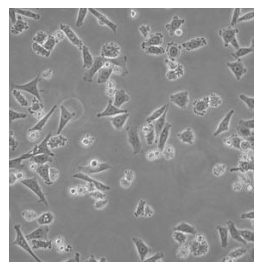


# Dataset Preparation for Enhancer Prediction via Deep Learning

Shaurya Jauhari, Mora Lab.

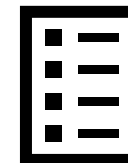
# DataSet



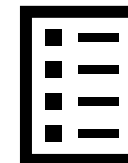


H1 Cell Line

ChIP-Seq Data  
(BAM file)  
Protein-binding  
assay



H3K27ac



H3K4me1



H3K4me2



H3K4me3

```
iMacMoraLab:H3K27ac mei$ samtools index ENCFF663SAM.bam ENCFF663SAM.bai
```

```
iMacMoraLab:H3K27ac mei$ pip install deeptools
```

```
iMacMoraLab:H3K27ac mei$ bamCoverage --bam ENCFF663SAM.bam \  
> -o ENCFF663SAM.bw \  
> --binSize 2000 \  
> --normalizeUsing RPKM \  
> --effectiveGenomeSize 2913022398 \  
> --outputFileFormat bedgraph
```

ENCFF340UJK.bw  
(H3K4me3)

ENCFF799BDH.bw  
(H3K4me2)

ENCFF441KOL.bw  
(H3K4me1)

ENCFF663SAM.bw  
(H3K27ac)



Galaxy

Tools

search tools

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

Galaxy is an open source, web-based platform  
Galaxy start [here](#) or consult our help resource  
and choose from thousands of tools from the

125+ ways to use Galaxy, right now



Galaxy

Tools

search tools

bedtools FlankBed create new intervals from the flanks of existing intervals

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

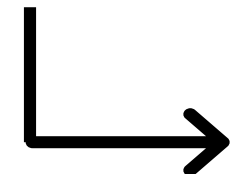
bedtools MaskFastaBed use intervals to mask sequences from a FASTA file

bedtools SubtractBed remove intervals based on overlaps

bedtools Merge BedGraph files combines coverage intervals from multiple BEDGRAPH files

bedtools BED to BAM converter

Running Your Own Galaxy  
Understanding how Galaxy works  
An in-depth tutorial



History

search datasets

BED Files Merging

5 shown, 1 hidden

2.43 GB

5: bedtools Merge BedGraph files on data 4, data 3, and data 2

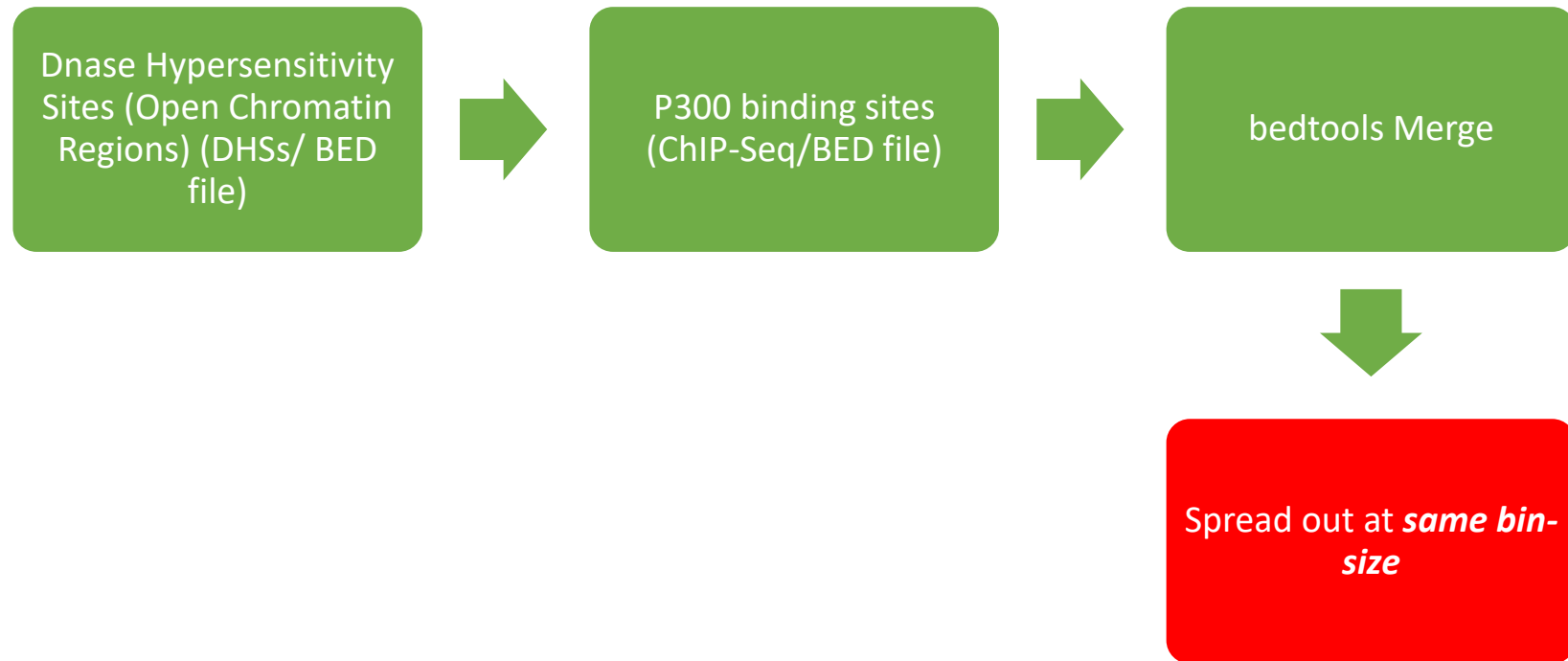
4: ENCFF340UJK.bw

3: ENCFF799BDH.bw

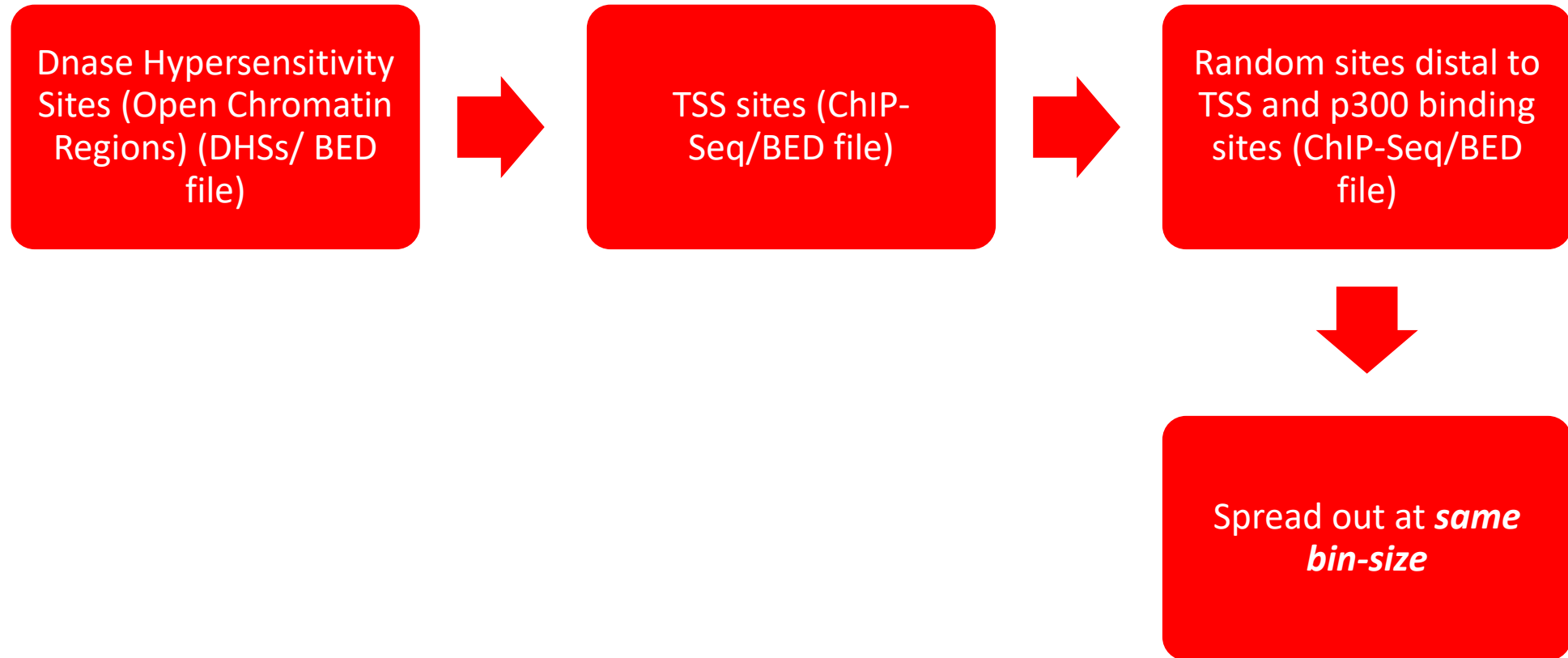
2: ENCFF441KOL.bw

1: ENCFF663SAM.bw

# Positive Training Set: Class Labels (Enhancer -> 1)



# Negative Training Set: Class Labels (Non-Enhancer -> 0)



# Web Resources

- Github Page
  - [https://github.com/shauryajauhari/Machine\\_Learning/Machine\\_Learning\\_Dep\\_Learning](https://github.com/shauryajauhari/Machine_Learning/Machine_Learning_Dep_Learning)
- Email
  - [shauryajauhari@gzhmu.edu.cn](mailto:shauryajauhari@gzhmu.edu.cn)
- R Markdown
  - <https://rpubs.com/shauryajauhari/>