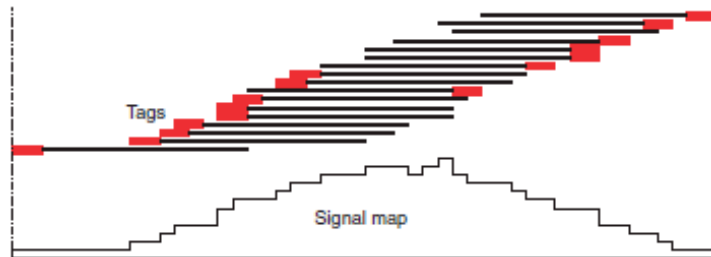


PeakSeq: [Rozowsky, 2009]

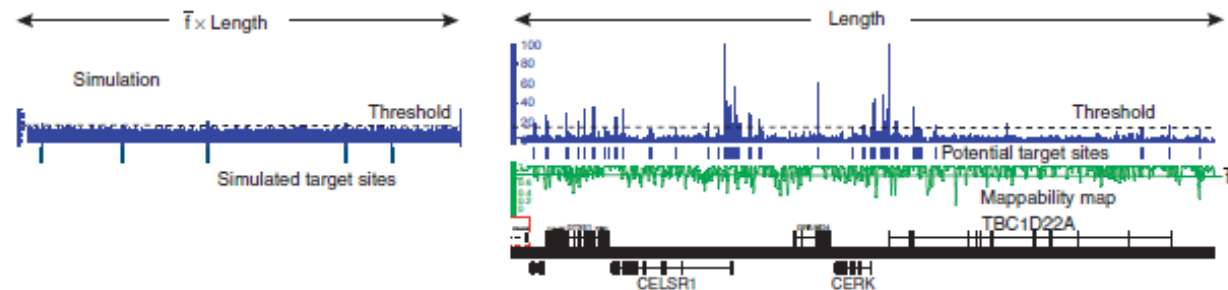
1. Constructing signal maps



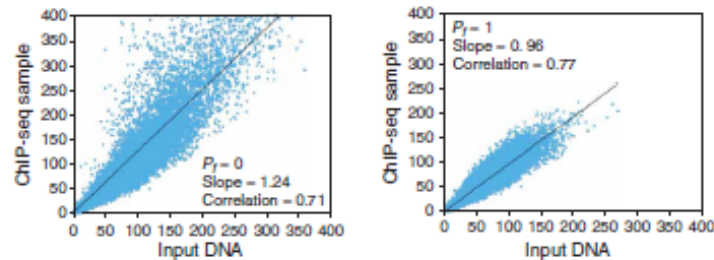
- Extend mapped tags to DNA fragment
- Map of number of DNA fragments at each nucleotide position

2. First pass: determining potential binding regions by comparison to simulation

- Simulate each segment
- Determine a threshold satisfying the desired initial false discovery rate
- Use the threshold to identify potential target sites



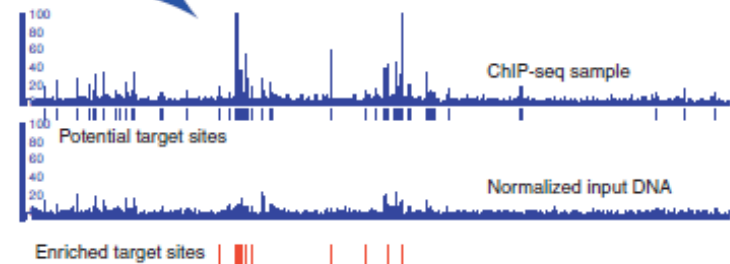
3. Normalizing control to ChIP-seq sample



- Select fraction of potential peaks to exclude (parameter P_1)
- Count tags in bins along chromosome for ChIP-seq sample and control
- Determine slope of least squares linear regression

4. Second pass: scoring enriched target regions relative to control

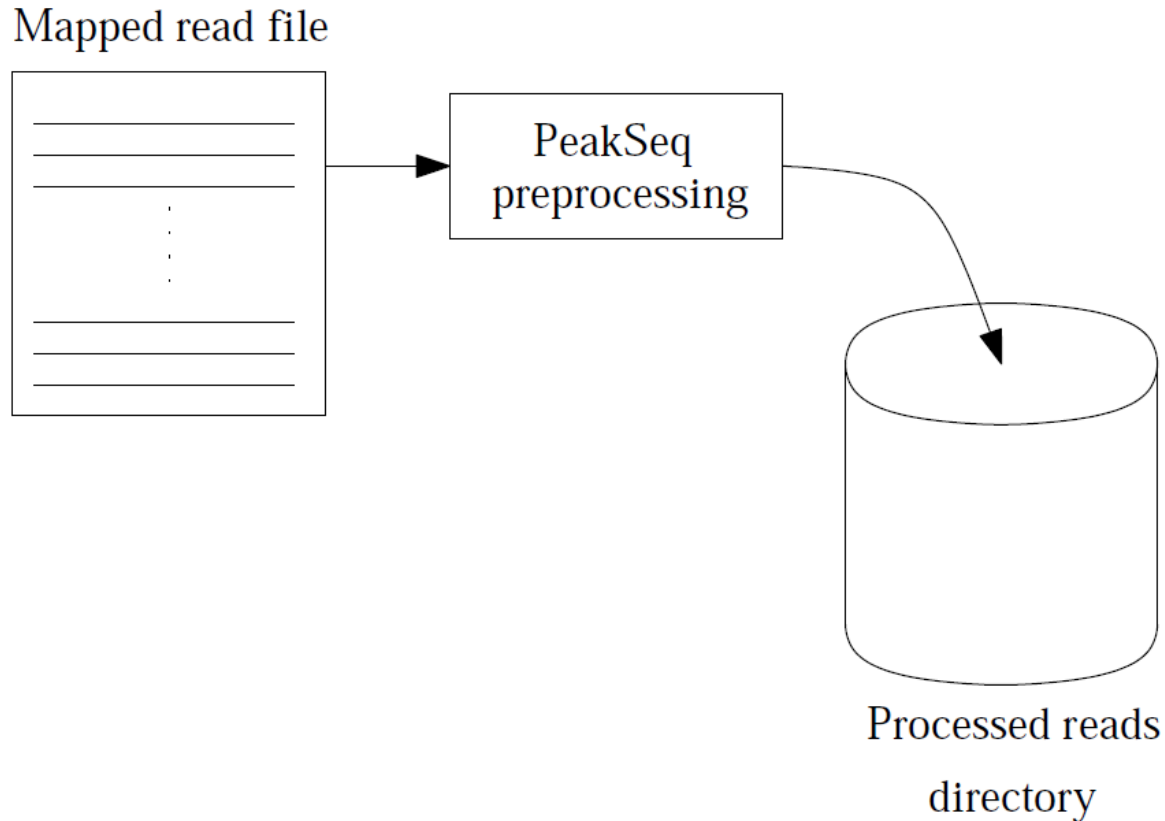
- For potential binding sites calculate the fold enrichment
- Compute a P -value from the binomial distribution
- Correct for multiple hypothesis testing and determine enriched target sites



PeakSeq: Updates

- The code is re-organized and re-written
- Supports multiple input read formats
 - SAM, ELAND, Default BOWTIE format, tagAlign
 - Supports BAM via piping SAM output from samtools
- Easier to setup than previous version
- Works faster:
 - Observed that peak calling on the Pol2 dataset on the website finishes in half the time
- There is also new code for setting up PeakSeq for peak calling on a large dataset
- Available at: http://archive.gersteinlab.org/proj/PeakSeq/Scoring_ChIPSeq/Code/C/

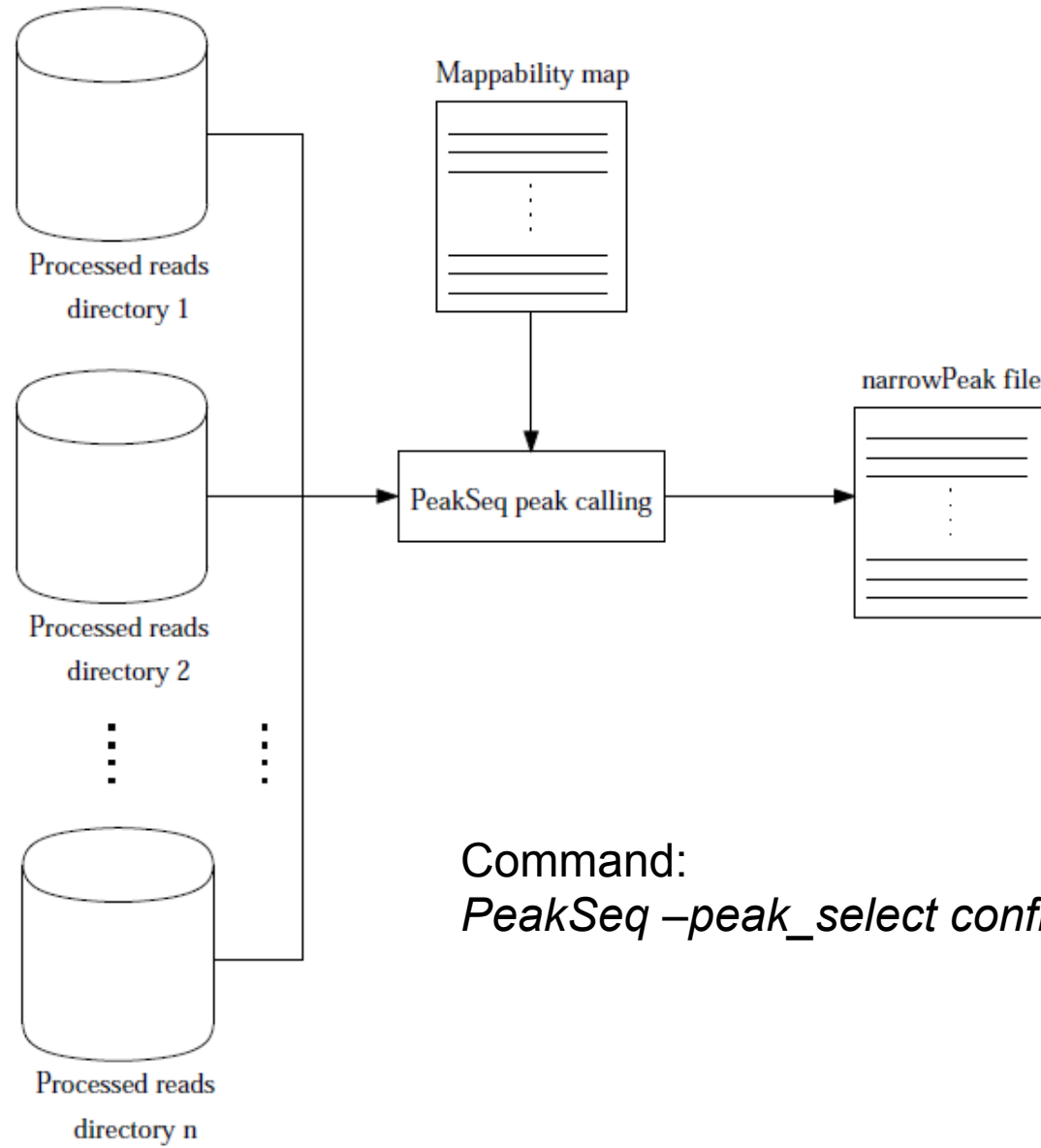
PeakSeq: Preprocessing



- Command:

```
PeakSeq -preprocess tagAlign  
wgEncodeSydhTfbsK562Pol2StdAlnRep1.tagAlign chr_ids.txt  
processed_reads_dir
```

PeakSeq: Peak calling



Command:
PeakSeq -peak_select config.dat

Configuration file (config.dat)

```
Experiment_id Sydh_GM12878_TR4_24
chromosome_list_file maternal_chr_id_list.txt
Enrichment_fragment_length 200
target_FDR 0.05
N_Simulations 50
Minimum_interpeak_distance 200
Mappability_map_file Mapability_HG_Maternal.txt
ChIP_Seq_reads_data_dirs
    wgEncodeSydhTfbsGm12878Tr4StdAlnRep1_mat_0
    wgEncodeSydhTfbsGm12878Tr4StdAlnRep2_mat_1
Input_reads_data_dirs
    wgEncodeSydhTfbsGm12878InputStdAlnRep1_mat_0
narrowPeak_output_file_path Sydh_GM12878_TR4_24.narrowPeak
Simulation_seed 434708749
Background_model Simulated
max_Qvalue 0.05
```

PeakSeq: Background models

- There is an experimental Poisson background
 - Specified by setting “Background” entry to “Poisson”
 - Sets the threshold for each window to the average expected read depth for that window

$$Threshold = \frac{n_{reads} \times l_{fragment}}{N_{mappable}}$$

- Generates a relaxed threshold and lacks the target FDR requirement