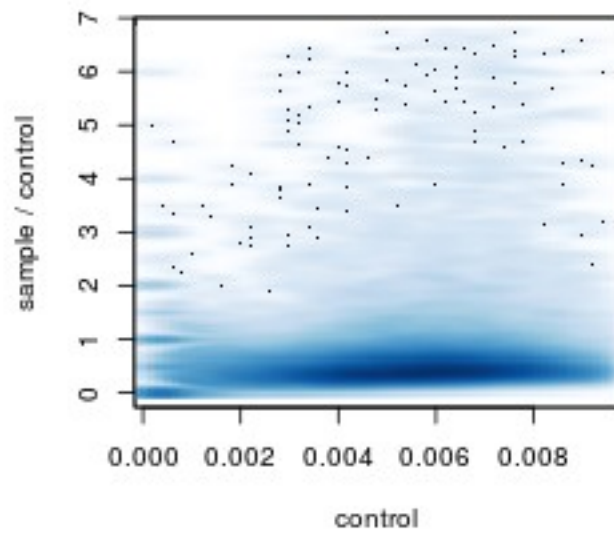
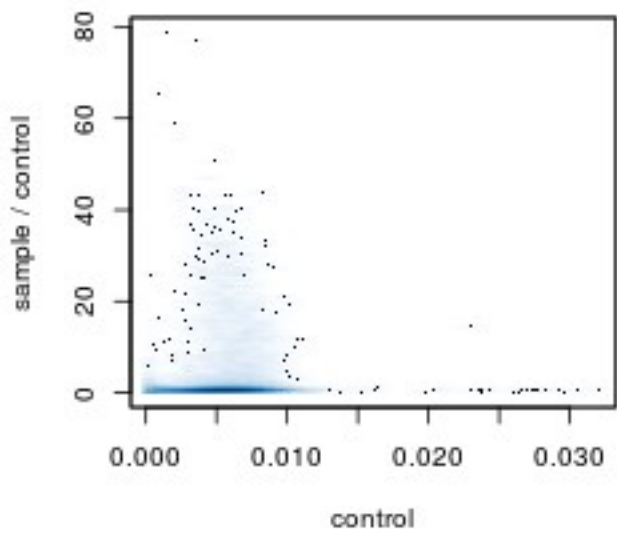
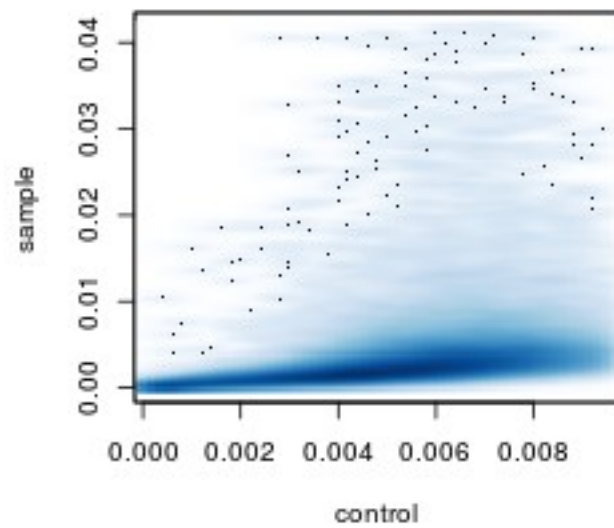
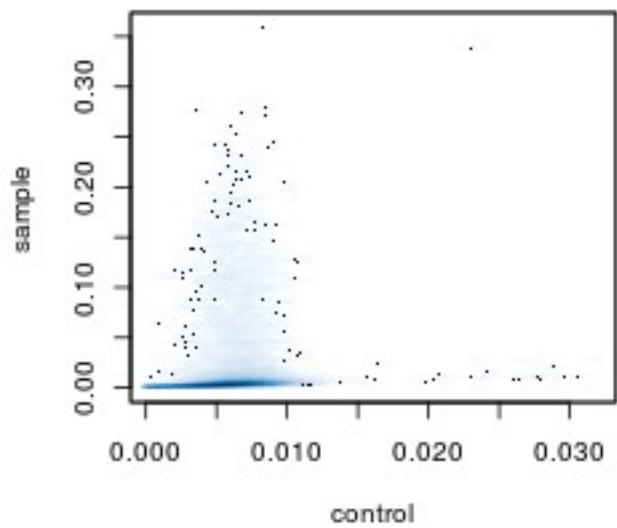
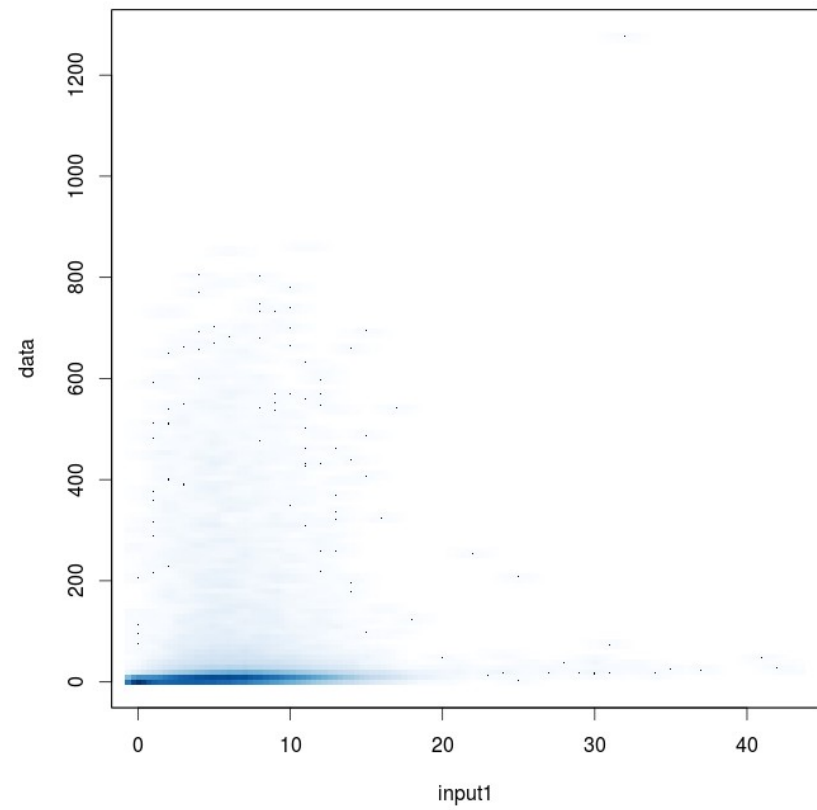
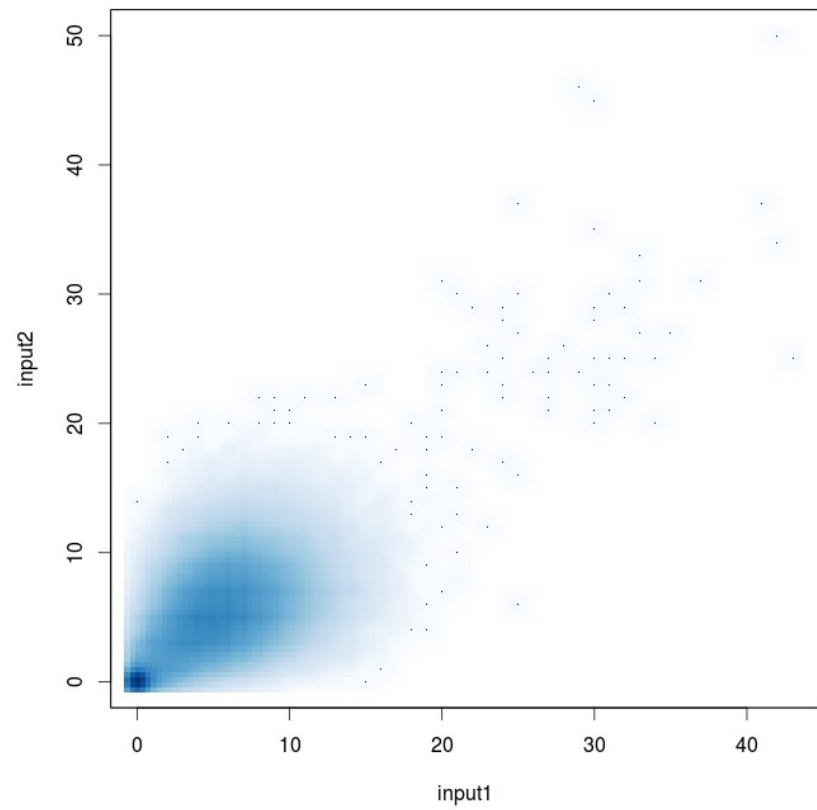


# ChIP-Seq analysis ideas

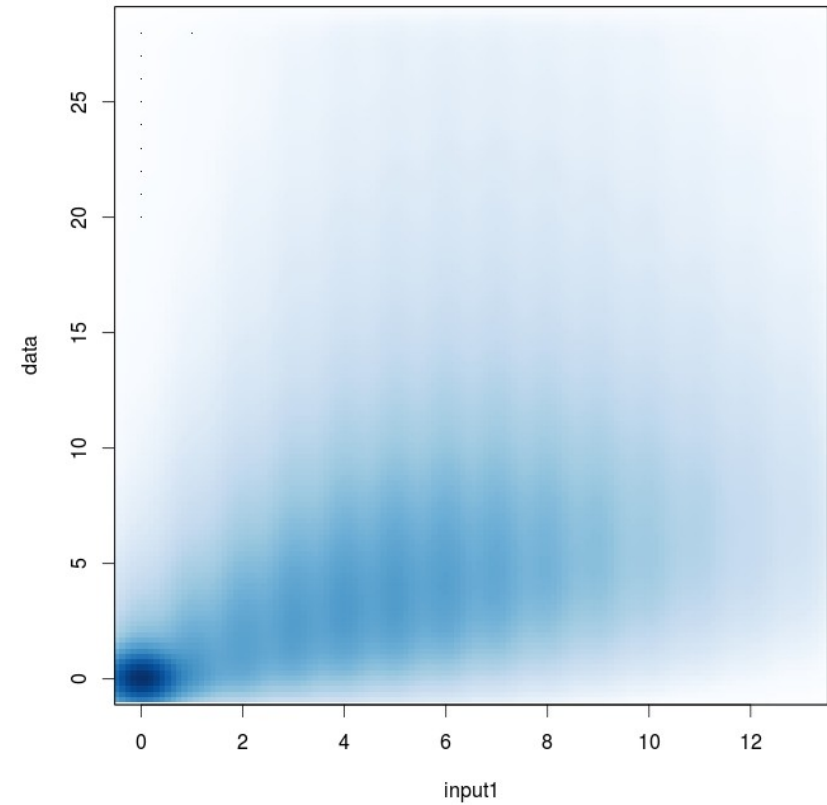
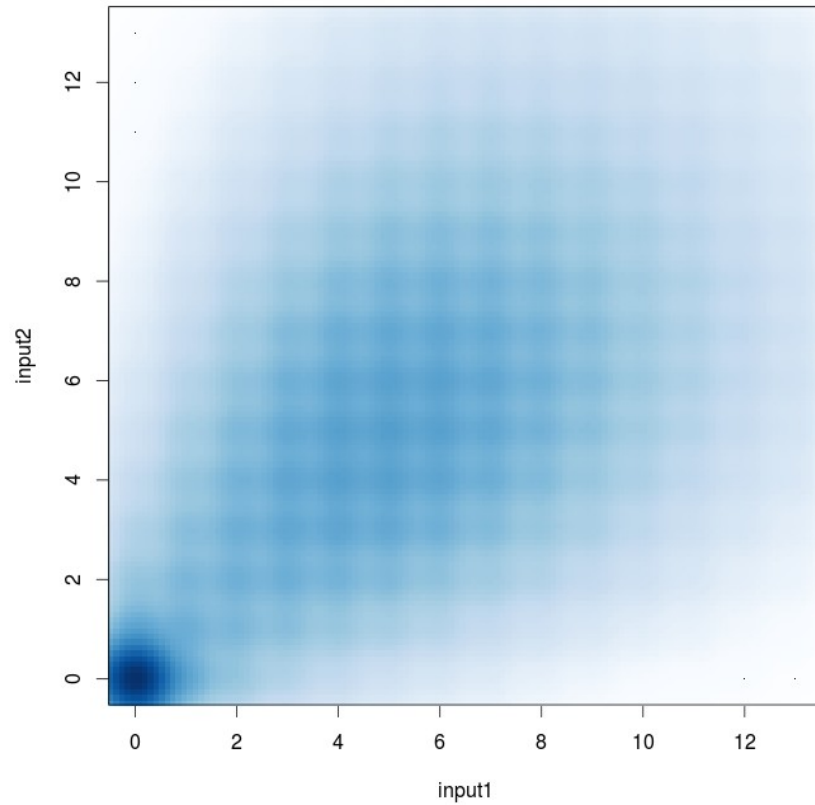
# ChIP-Seq depends of background

- Scatter plot of bincount
- Binsize 5000 nucleotides





# Let's look closer



# Yet another enrichment model

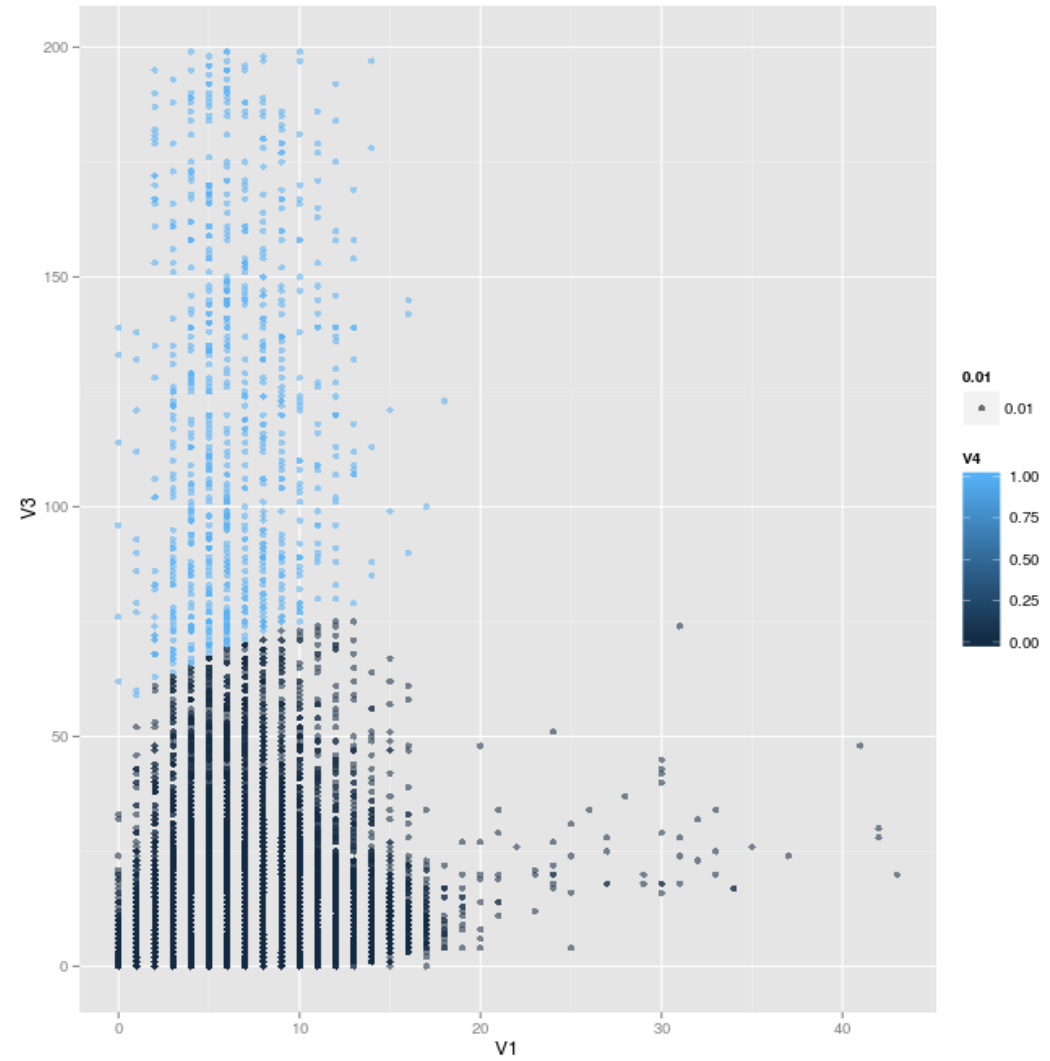
- Let's look at two dimensional distributions:
  - (input, input)
  - (input, data)
- Assume (Input, Data) is a mixture of two distributions:
  - Scaled (input, input)
  - Uniform

# Yet another enrichment model

- Approximate distribution (input, input) using mixture of Poisson distributions
- Scale (input, input) mixture to fit data

# Result

- TODO





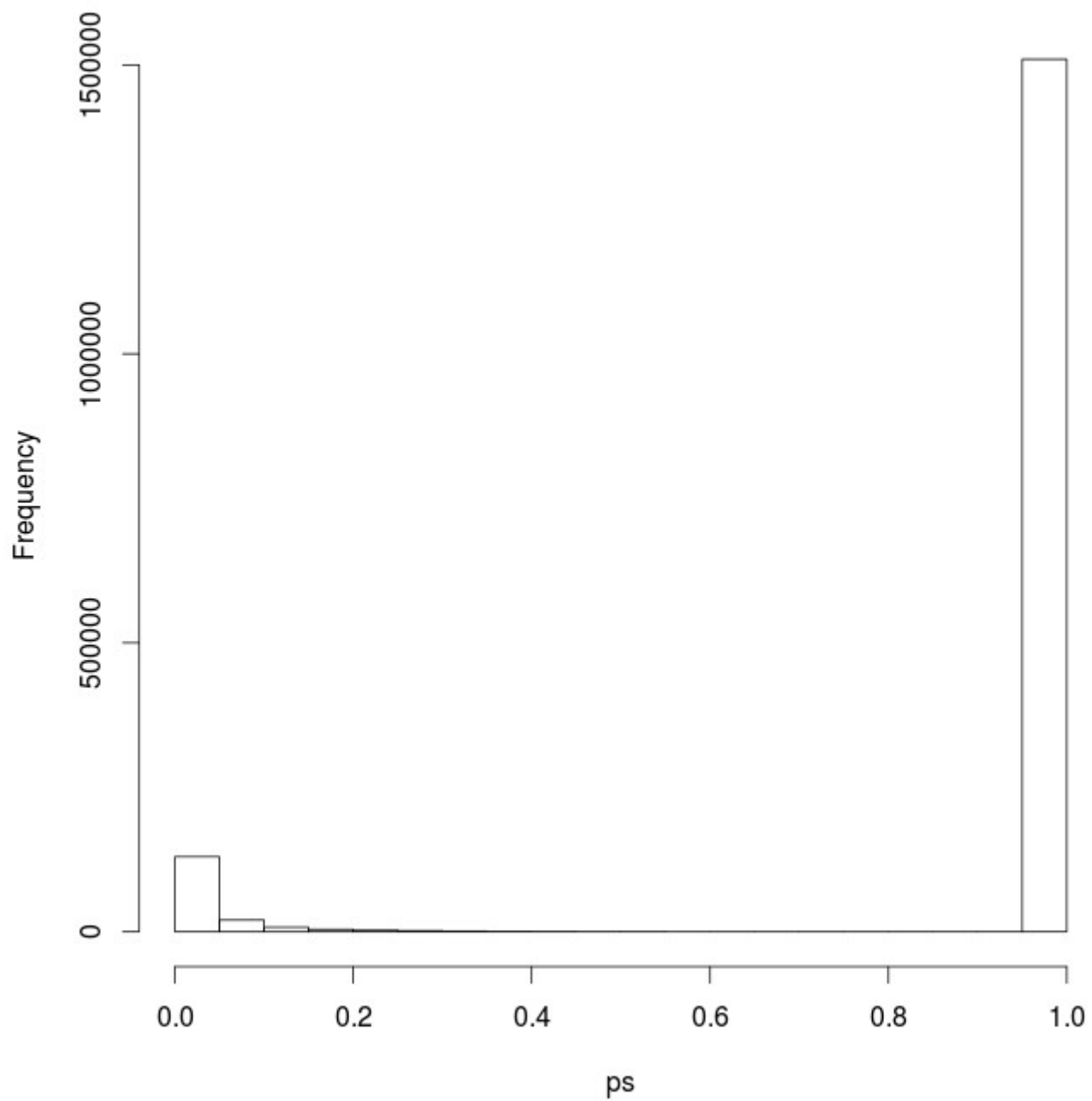
# Small scale dependence

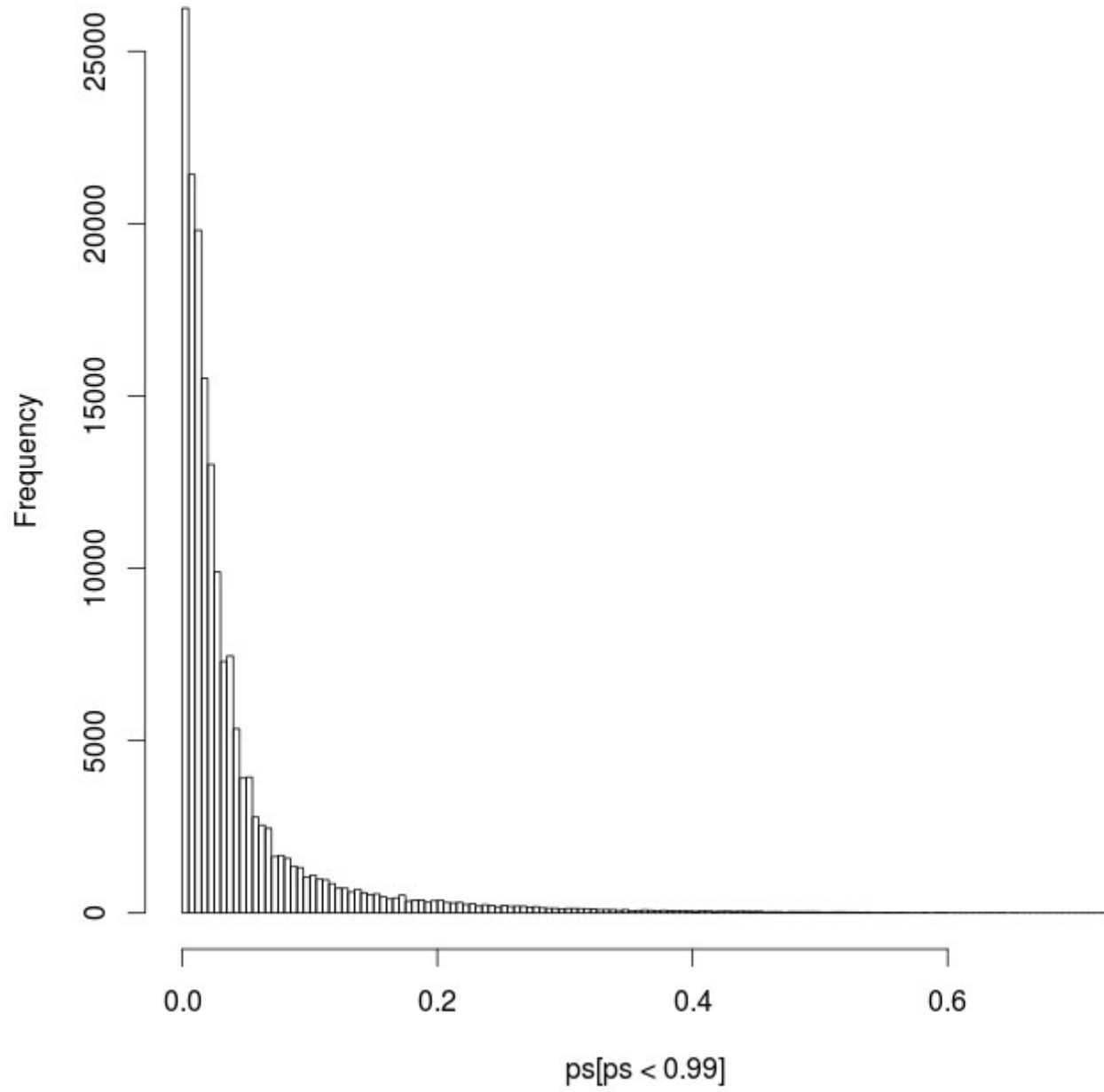
For every bin, divide nucleotides to 4 groups:

- First chip-seq have tag.
- Second chip-seq have tag.
- Both
- None

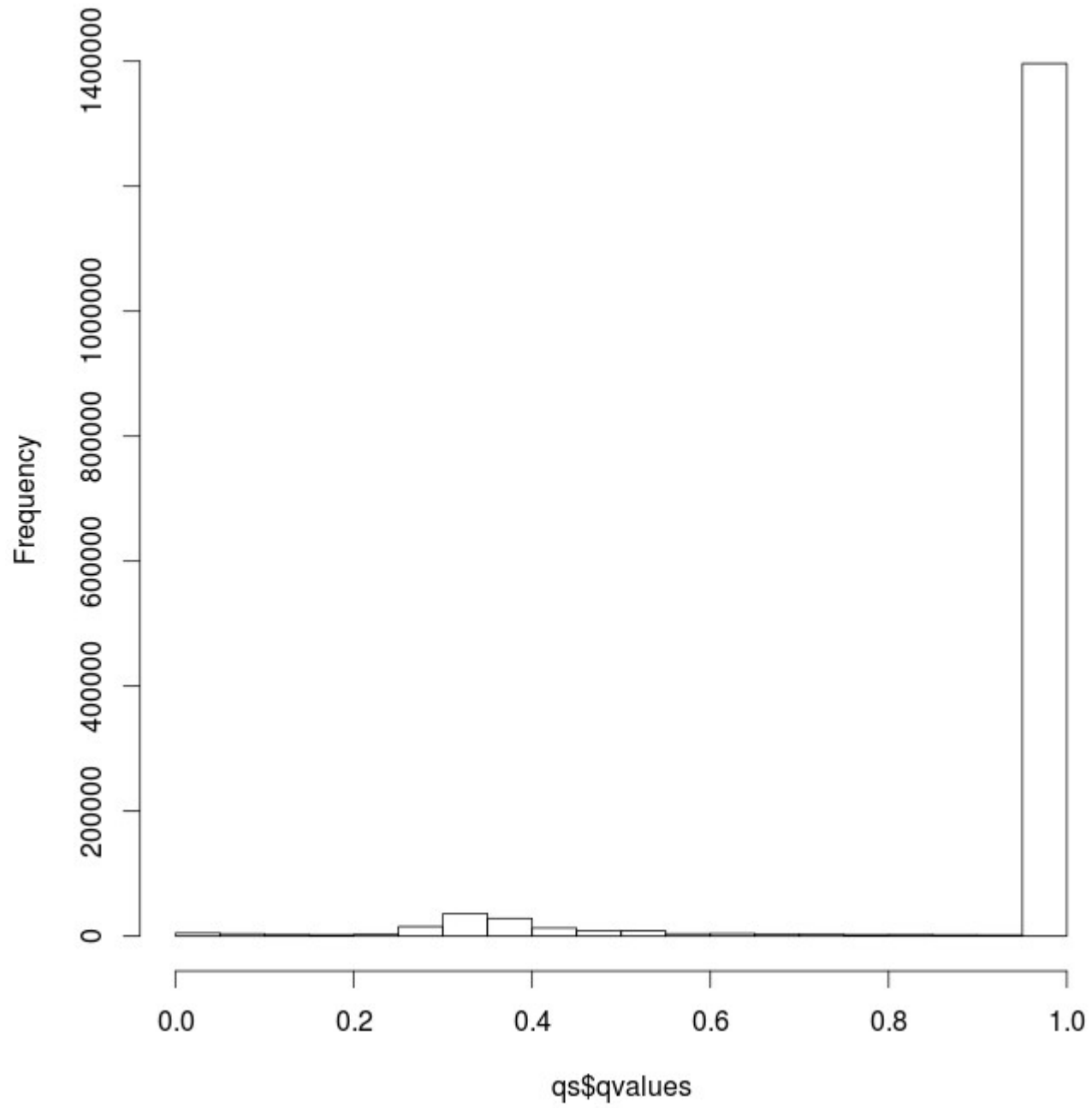
Make Fisher test for every bin

Histogram of ps

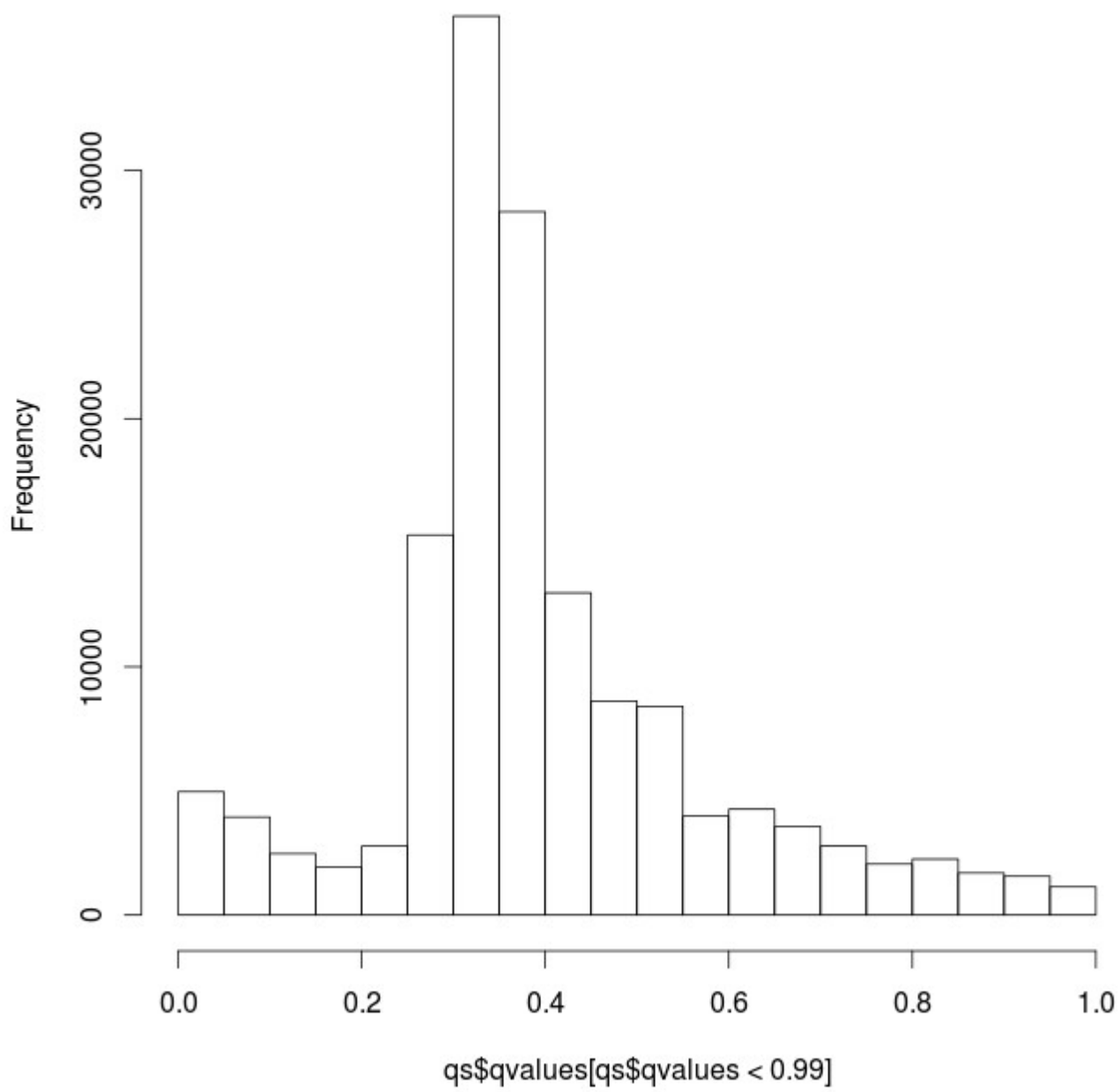




Histogram of qs\$qvalues



**Histogram of qs\$qvalues[qs\$qvalues < 0.99]**



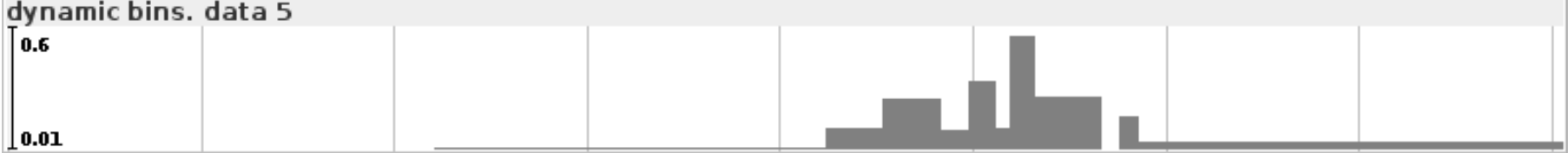
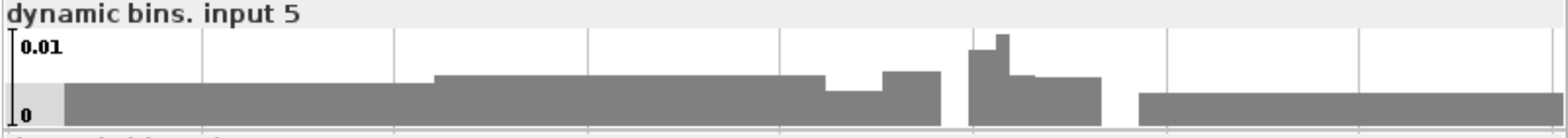
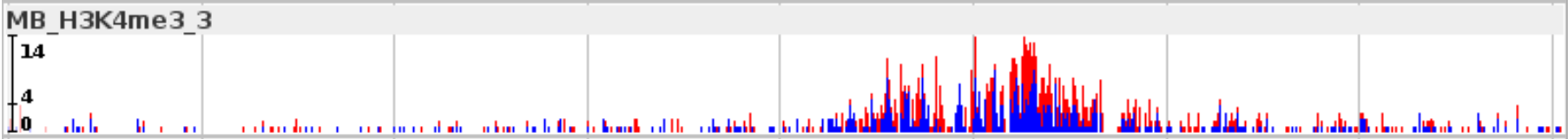
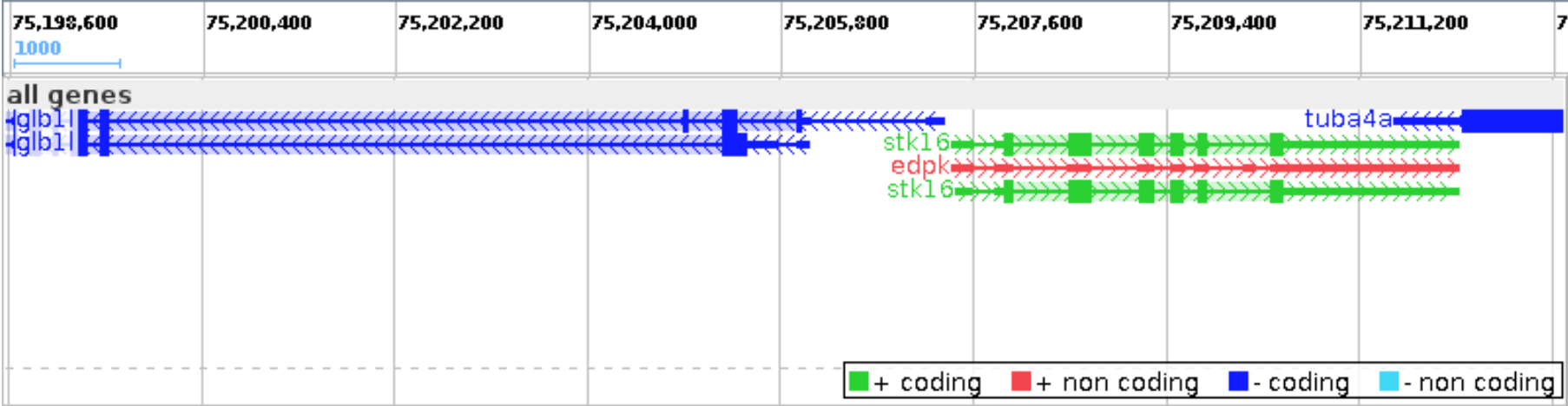
# Dynamic bins

- Bin can cover enrichment partially
- Assume bin has uniform coverage
- Coverage of nucleotide has Bernoulli distribution
- maximise

$$\sum_{t \in \text{tracks}} \sum_i c_{t,i} \log(p_i) + nc_{t,i} \log(1 - p_i) - \text{binsNumber} \cdot \text{penalty}$$

# Coverage in non uniform

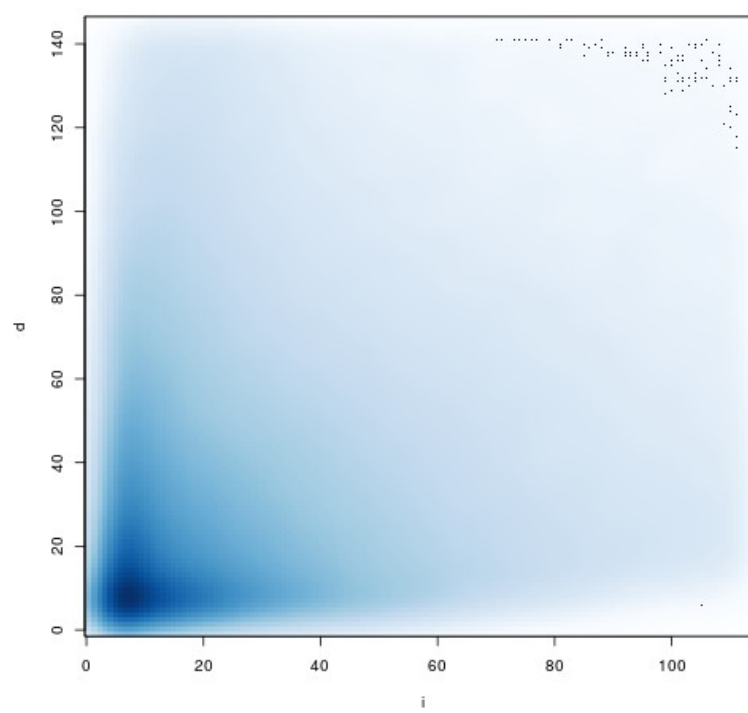
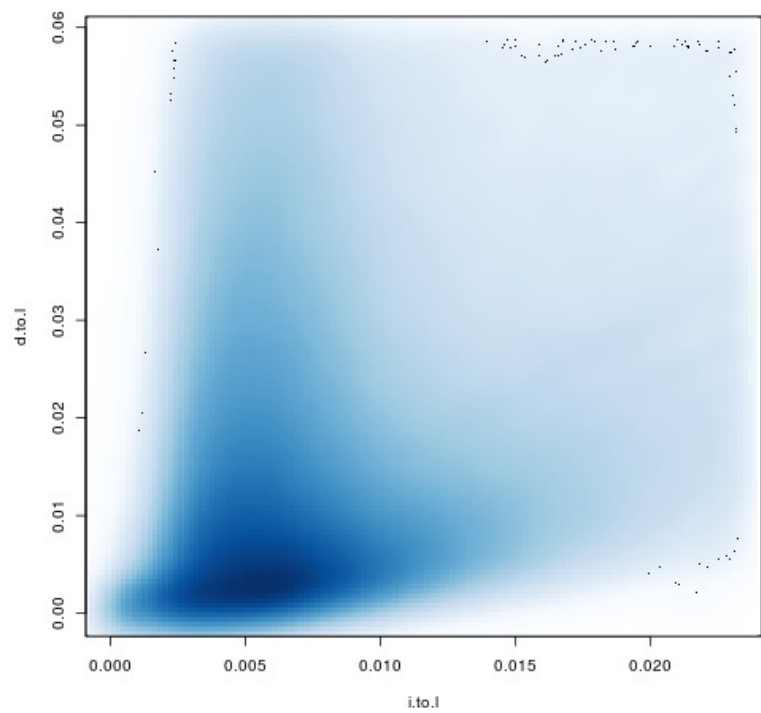
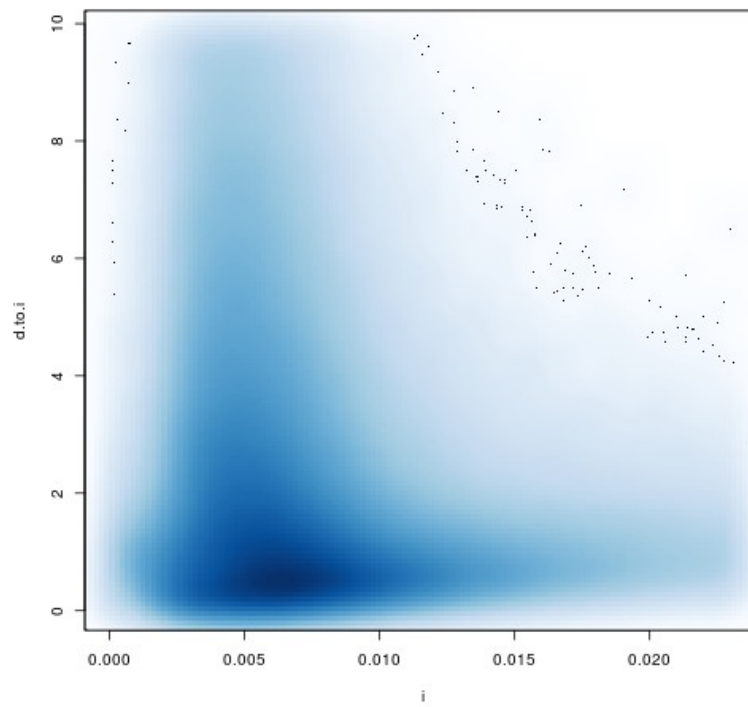
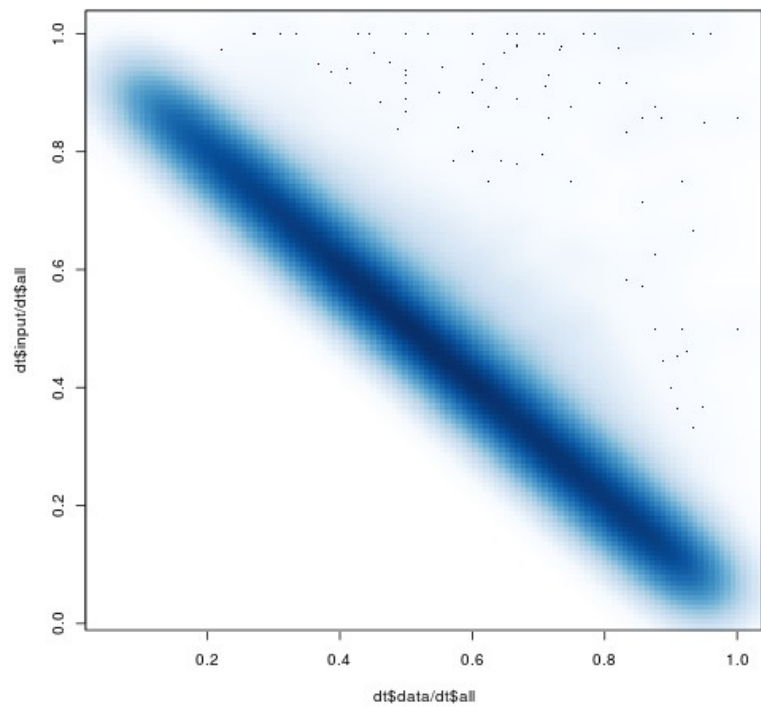
- Remove nucleotides completely without coverage
- Each bin characterised by two parameters,  $P(\text{input} \mid \text{input or data})$ ,  $P(\text{data} \mid \text{input or data})$



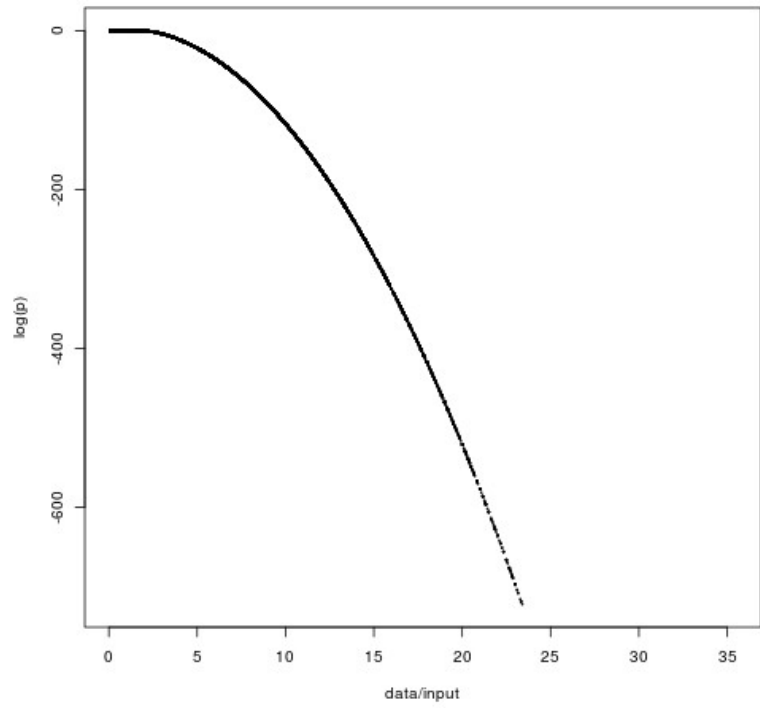
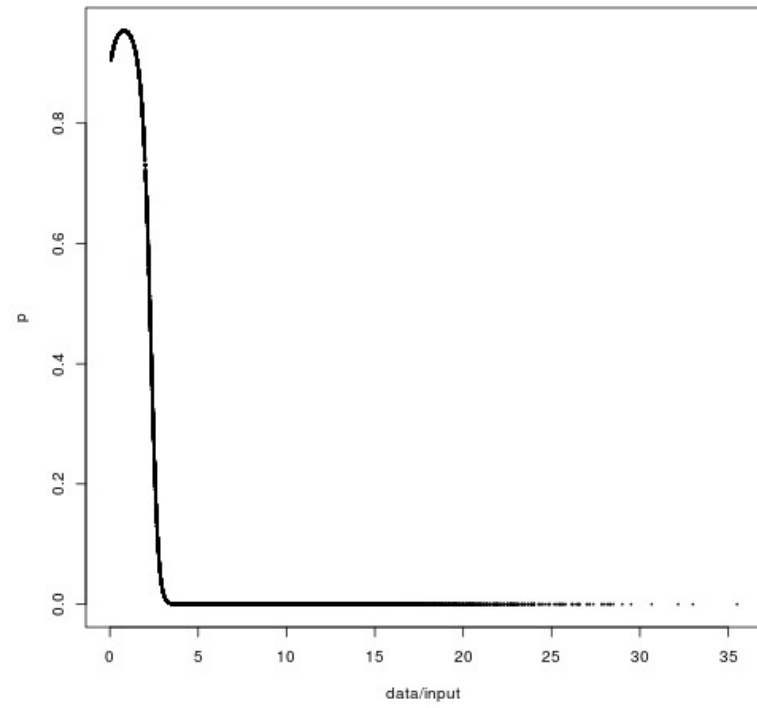
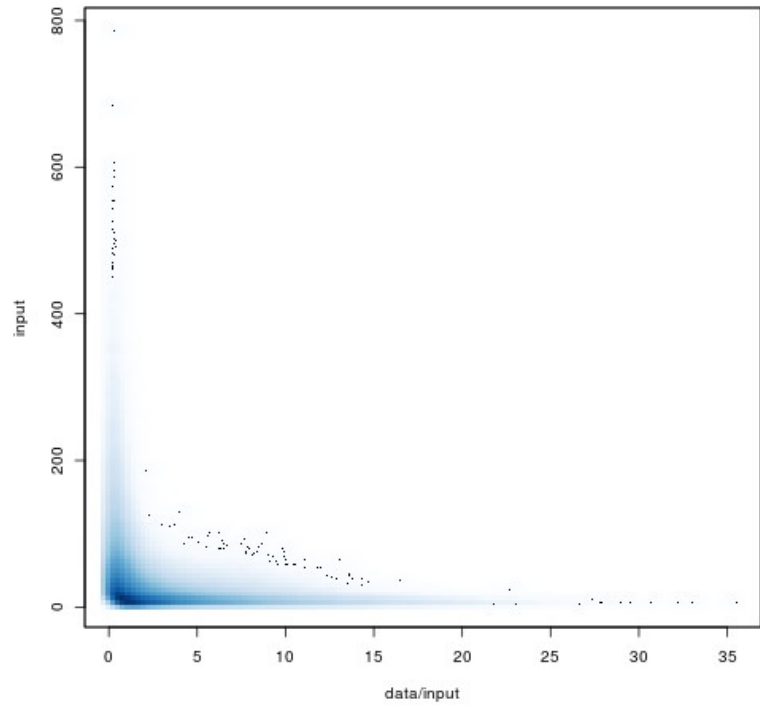
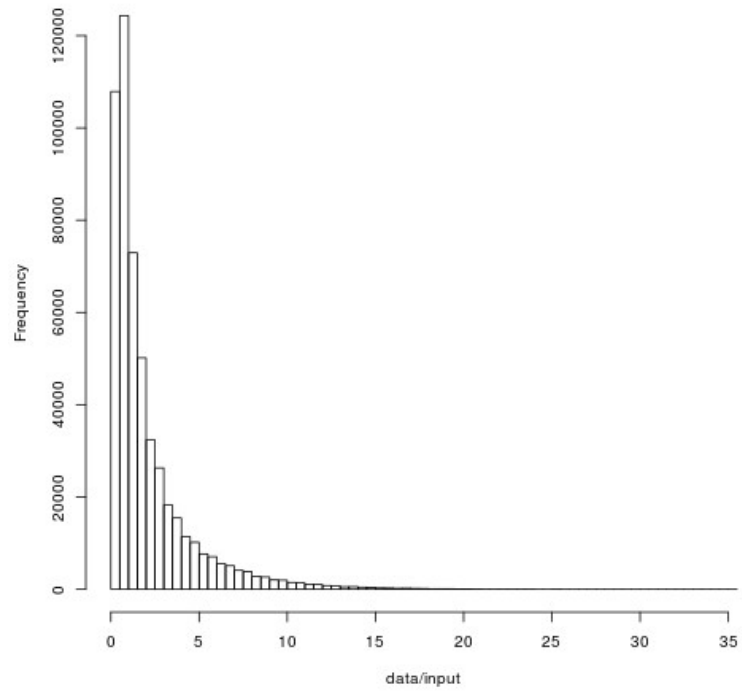


# Enrichment

- Calculate for every bin data divided to input
- Fit mixture of two normal distributions



Histogram of dt\data/dt\$input



# Combine these ideas

- TO DO