



Salzburg

2024





Salzburg

2024

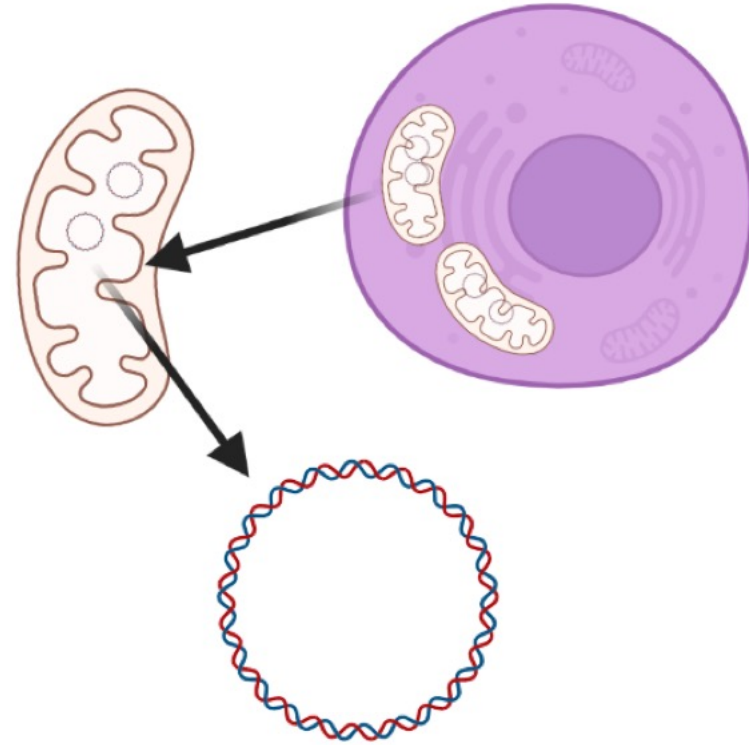
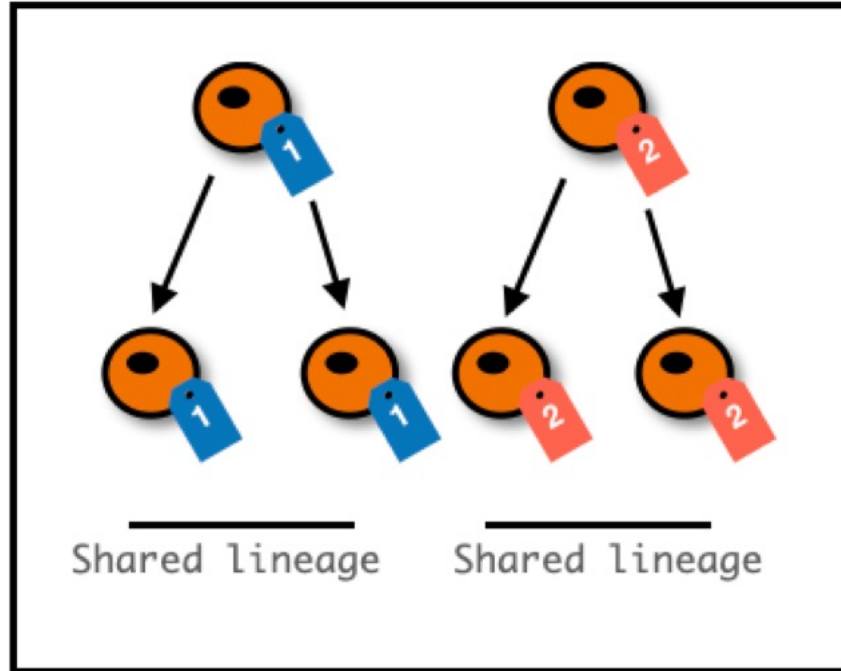
# ScMitoMut: Single Cell Lineage Informative Mitochondrial Mutation Calling Tool

Wenjie SUN



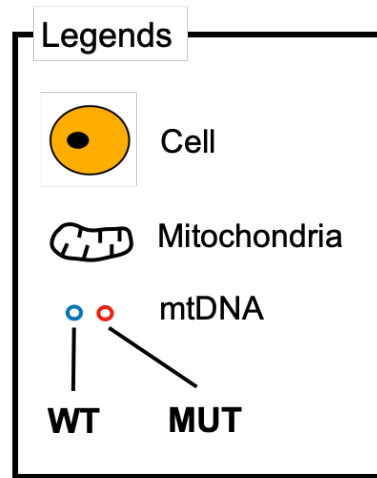
No conflicts of interest to disclose.

# Using mtDNA somatic mutation to follow the lineage

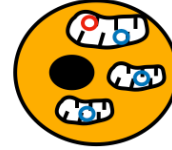


**Challenge: Calling lineage informative mtDNA in single cell sequencing data**

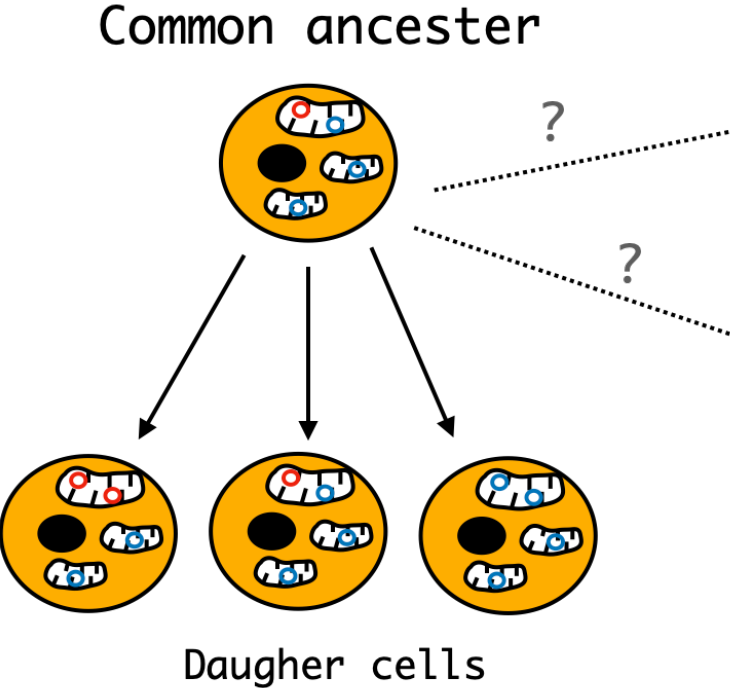
# Fitting single cell mtDNA mutation with beta-binomial distribution



Common ancestor

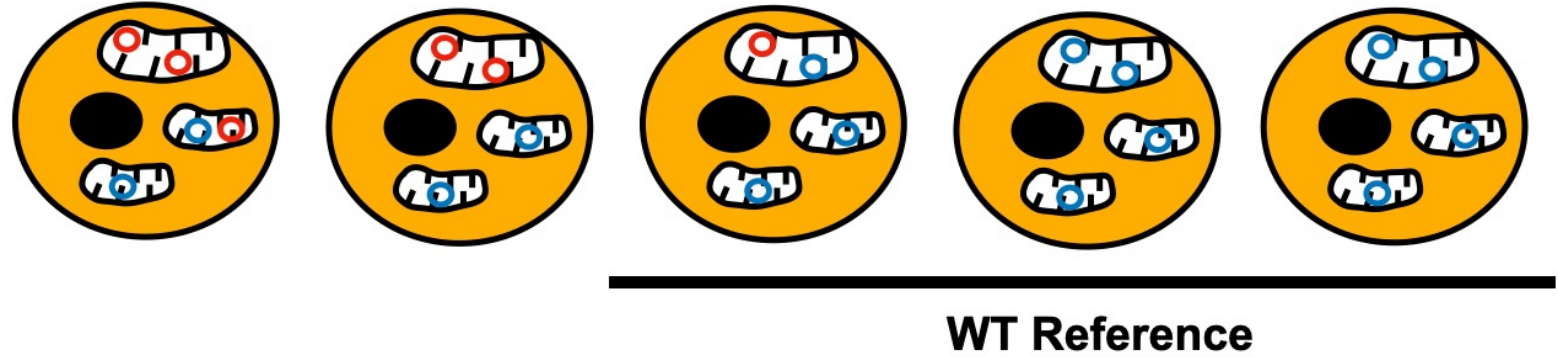


# Calling mutation with statistical test



# Preselect WT reference using **binomial-mixture model**

Step1: Preselect WT cells



# Quick Prototyping with R

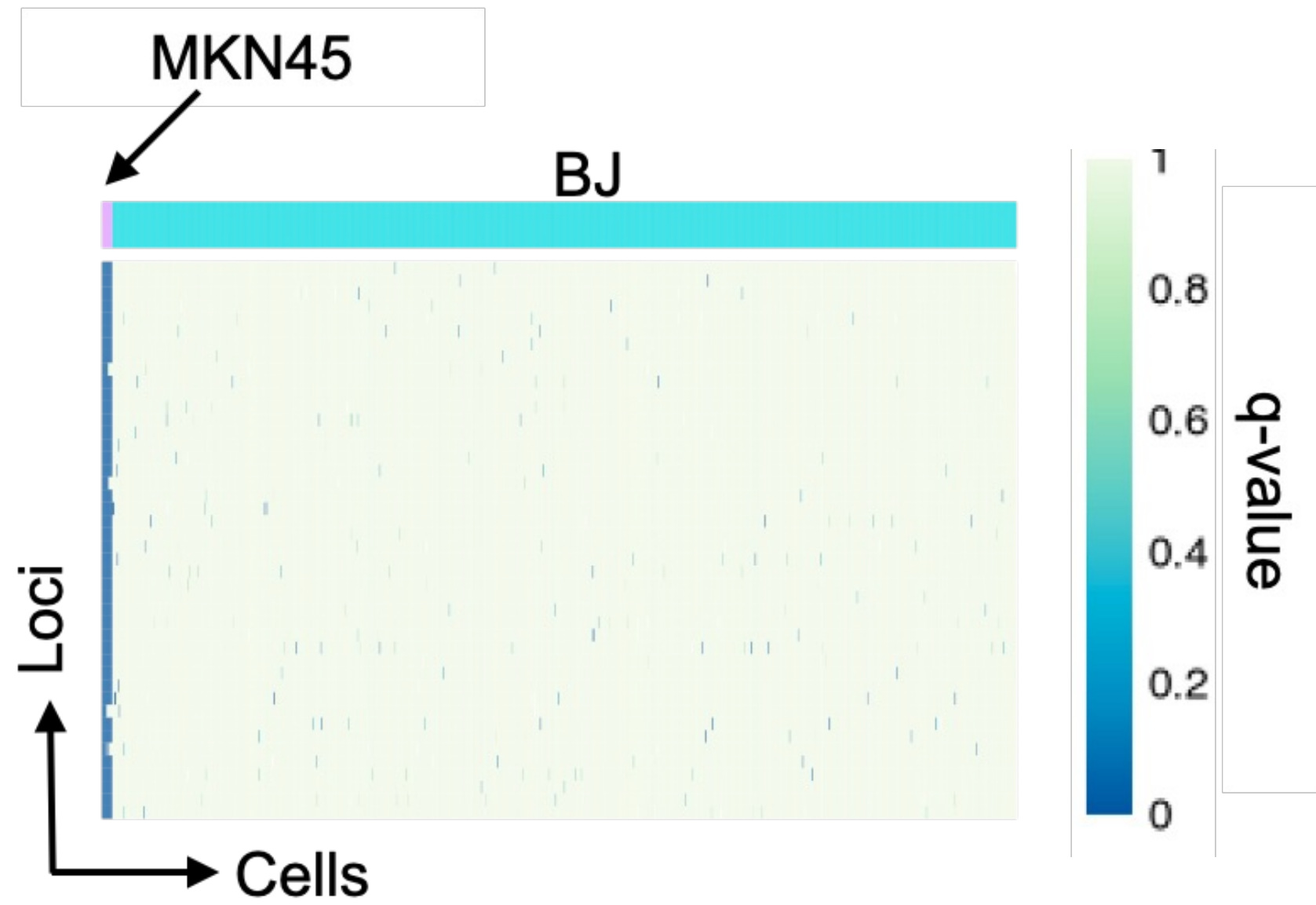
**{data.table} -> handling data**

**{mixtolls} -> fitting binomial-mixture model**

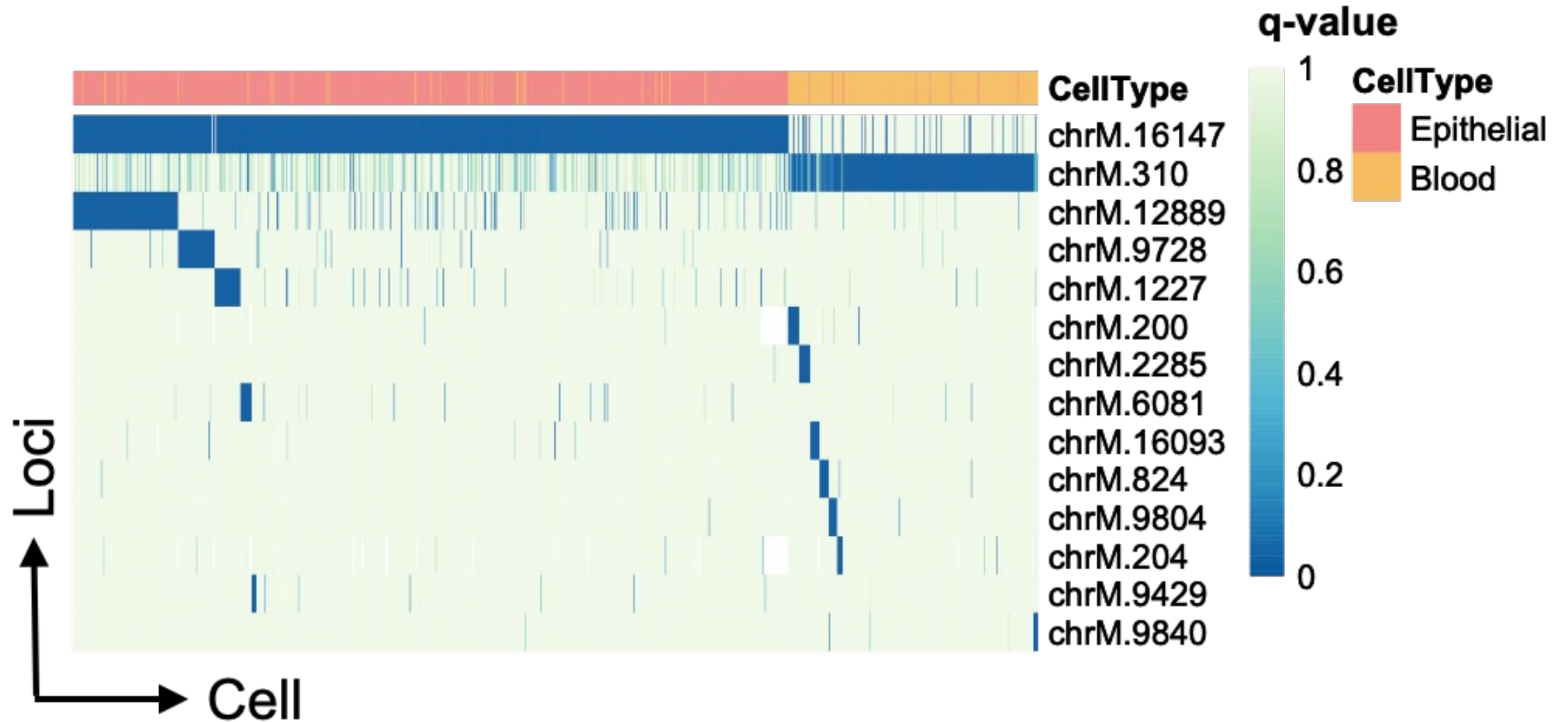
**{VGAM} -> fitting beta-binomial model**



mtDNA somatic mutations seperate two cell lines



# mtDNA somatic mutations separate two tissue types



# Memory requirement

10,000 cells

18,000 bp

4 base: A, T, C, G

Memory needs to keep **input data**:

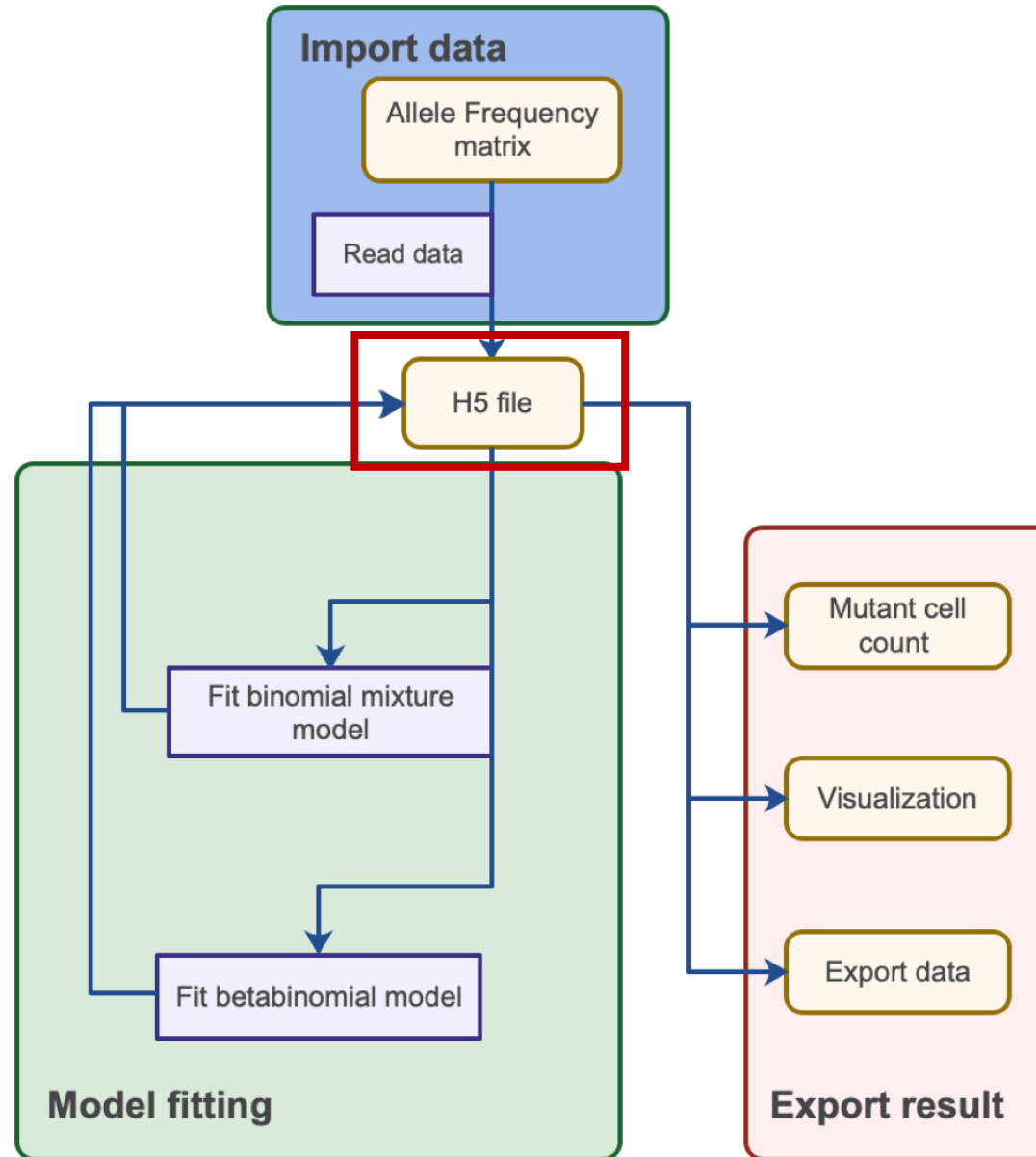
$10,000 * 18,000 * 4 * 8 \text{ bytes} = \mathbf{5.36GB}$

**X 4 (Intermediate results)**

R eats a lot of memory?



# Using H5 file to optimizing the memory usage



# Model fitting CPU time

18,000 bp

**1 sec per locus**  
**18,000 sec (5 hours)**

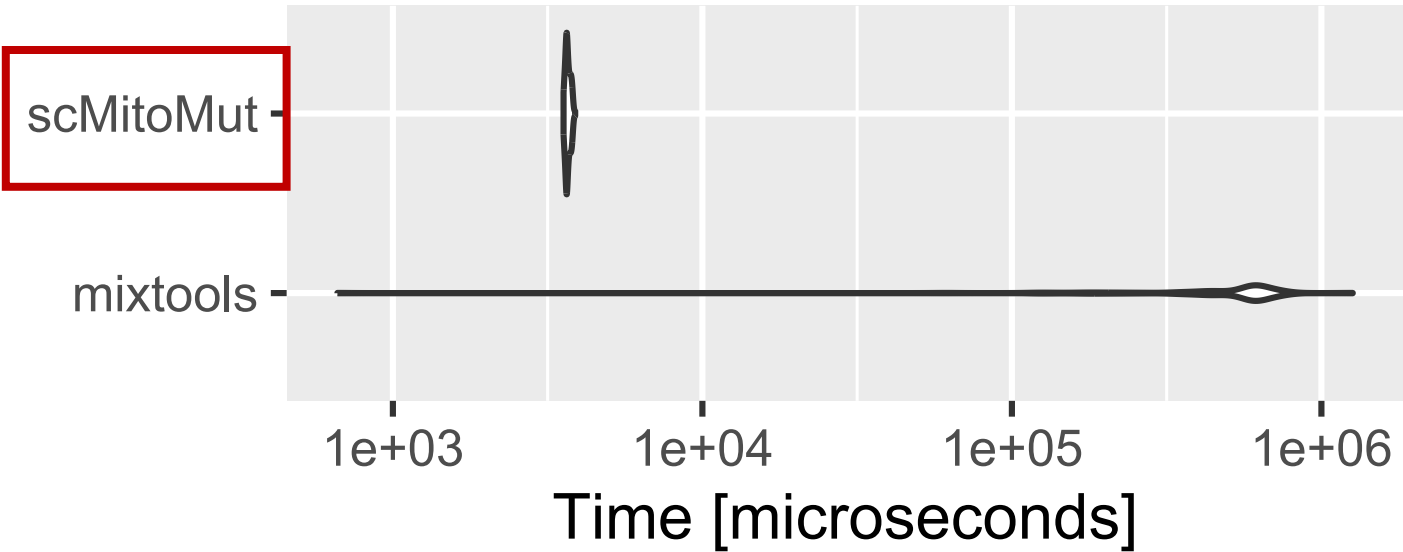
R too slow?



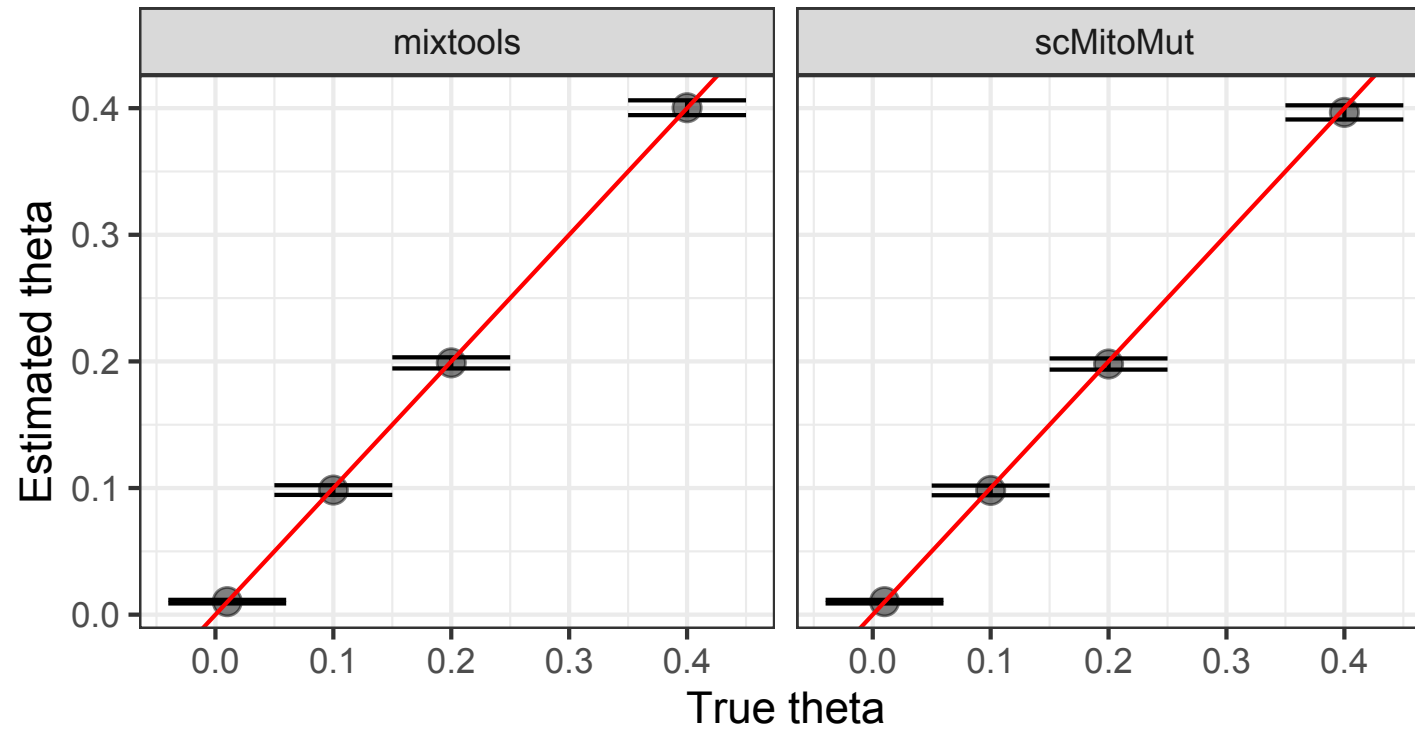




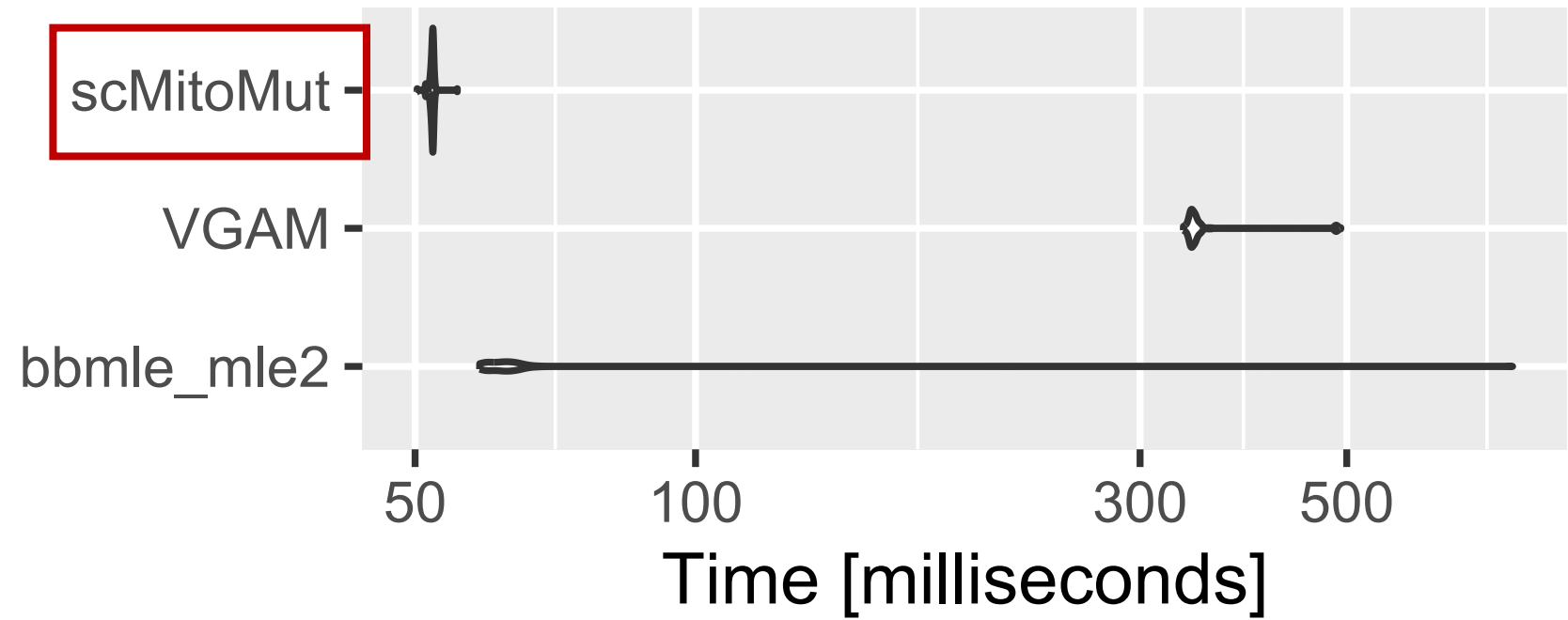
# Improvded binomial-mixture model fitting speed



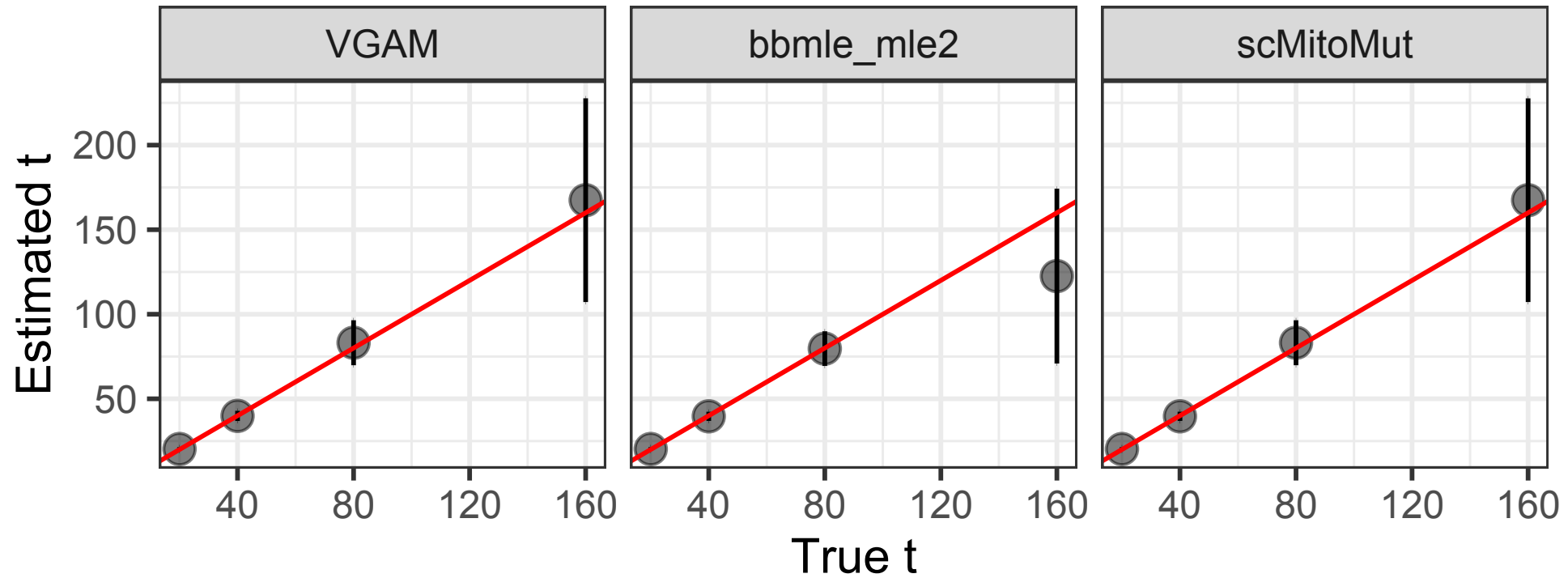
# There is no difference in model fitting results



# Improvded beta-binomial model fitting time



# Fitting results: scMitoMut eq VGAM better than bbml





**Complex model can be simple;  
“optimizing” is possible.**