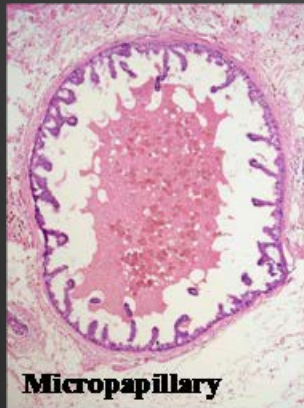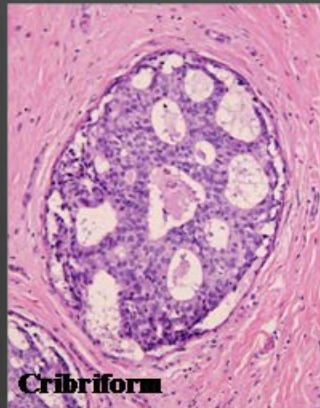# Estimating normal cell content and tumor subclone structure from matched tumor-normal NGS data

Yi Qiao, Gábor T. Marth
Boston College
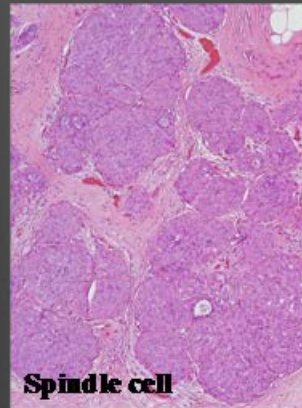
Sample Admixture
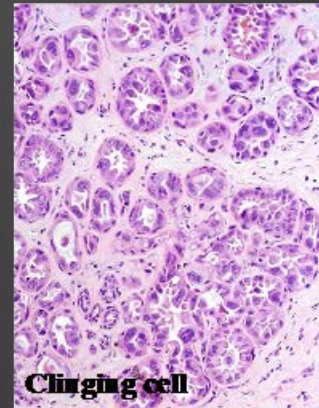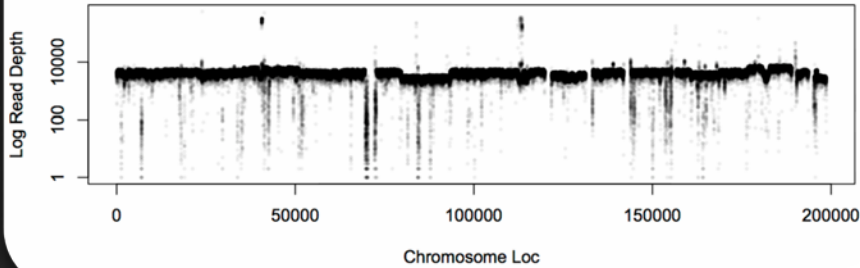

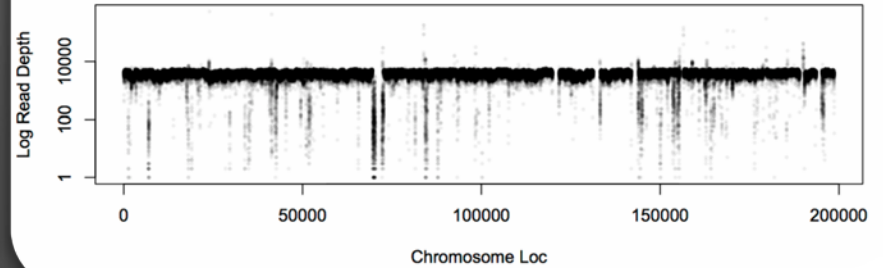Micropapillary Cribriform Spindle cell Clinging cell

Copy Number Analysis

Tumor Read Depth
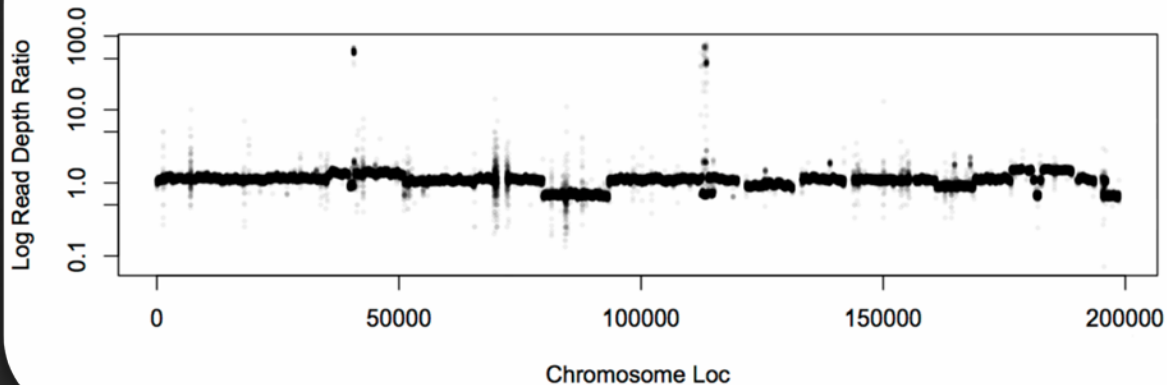
Normal Read Depth

Read Depth Ratio

Normal Contamination

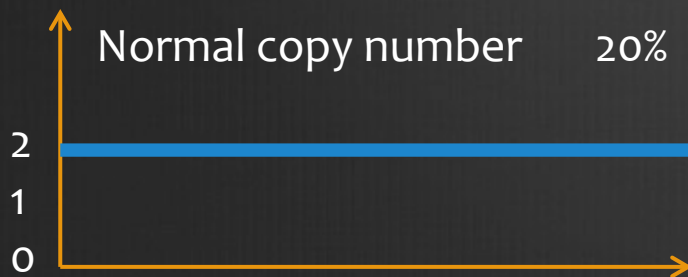Histogram of Tumor / Normal Ratio (a)

FFT de-noised histogram signal (b)

0.7

Chromosome 19 of sample TCGA-04-1371 (c)

Tumor copy number   80%
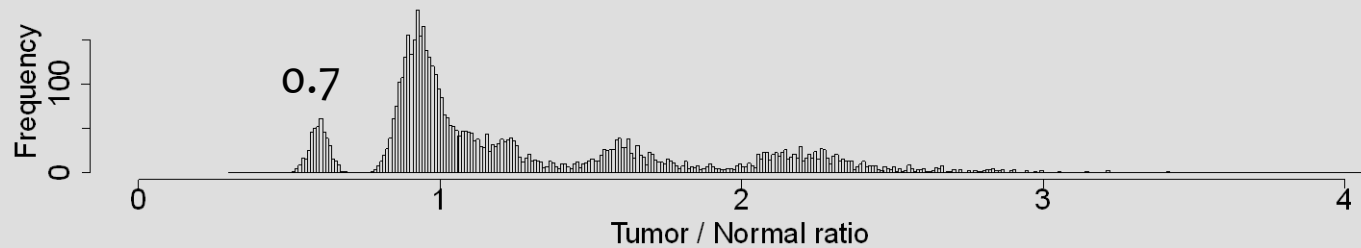
2
1
0

Normal copy number   20%

2
1
0

Observed data (ratio)

1
0.5
0

<- 0.7
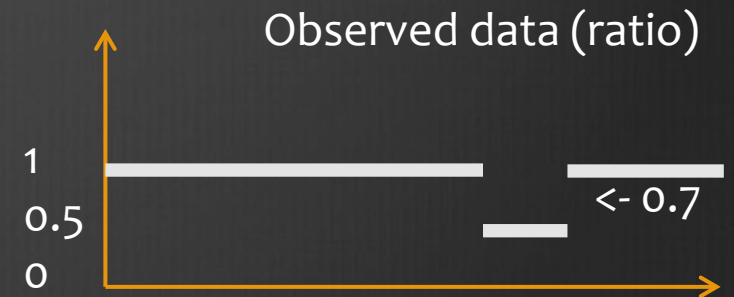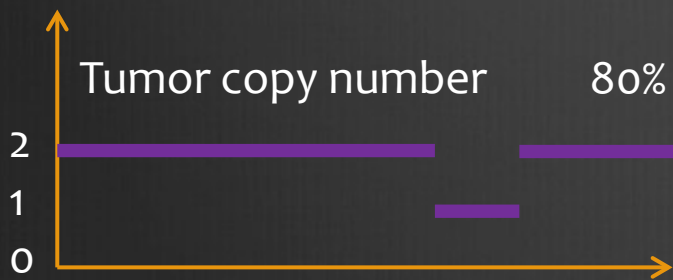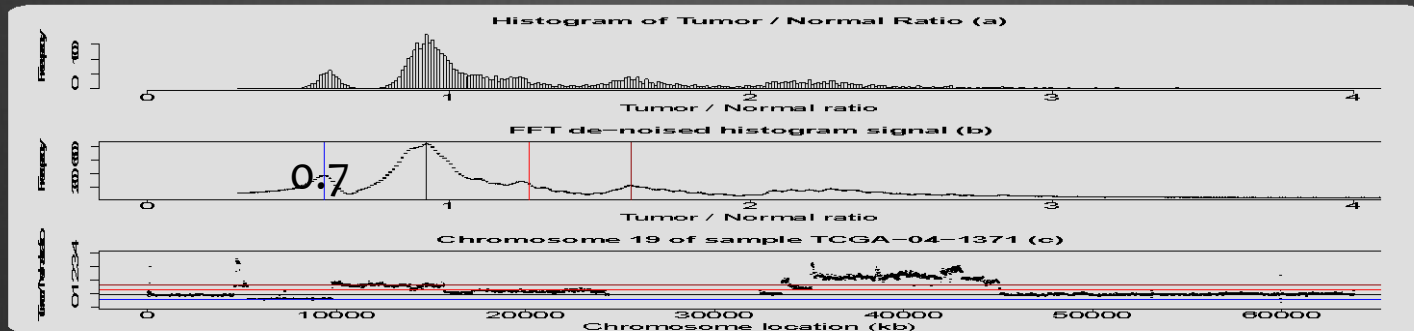
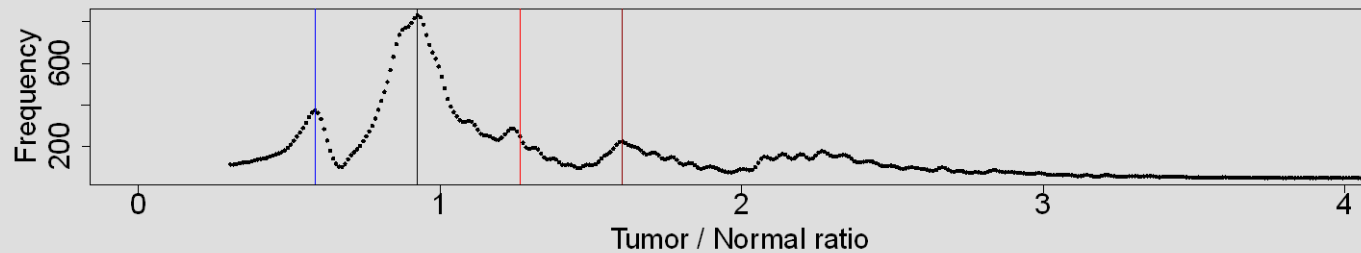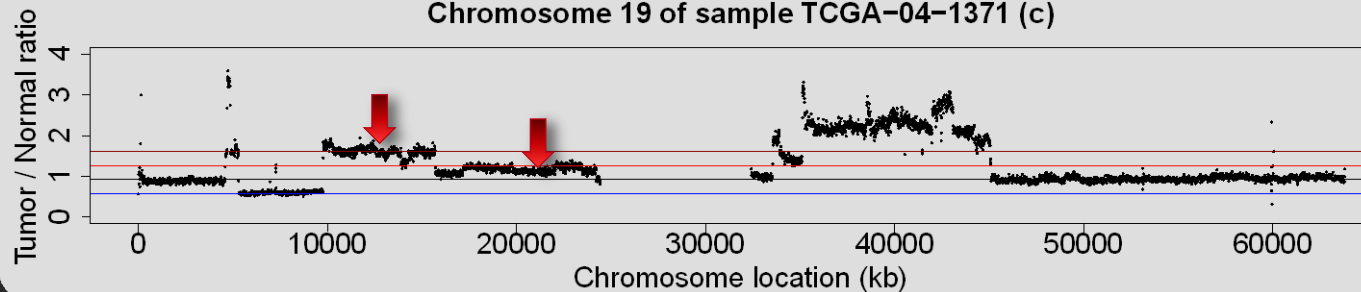$$\frac{T_{CN} * r + N_{CN} * (1-r)}{N_{CN}} = Ratio_{Observed}$$

Normal Contamination
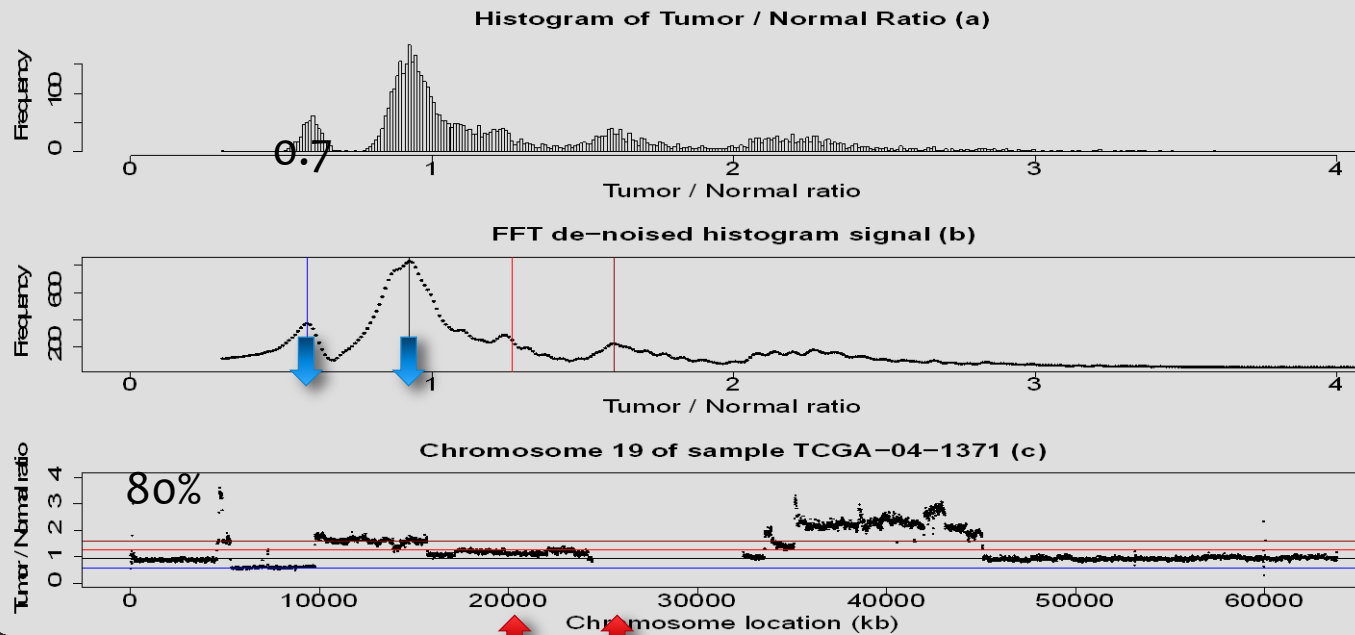
Histogram of Tumor / Normal Ratio (a)

FFT de-noised histogram signal (b)

Chromosome 19 of sample TCGA−04−1371 (c)

Normal Contamination

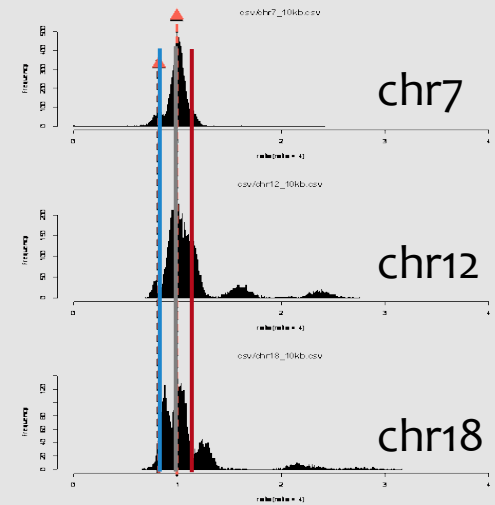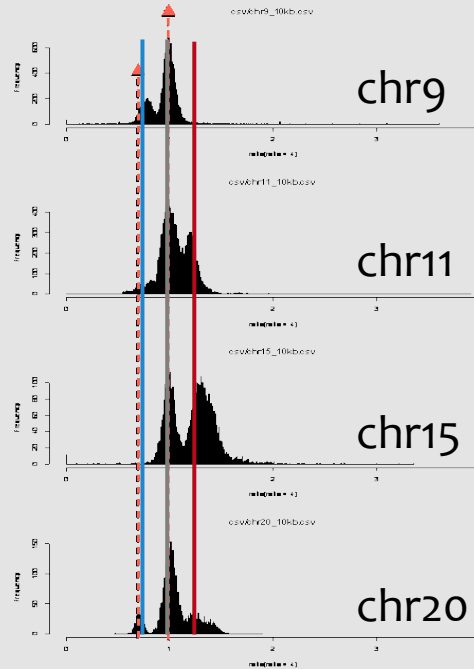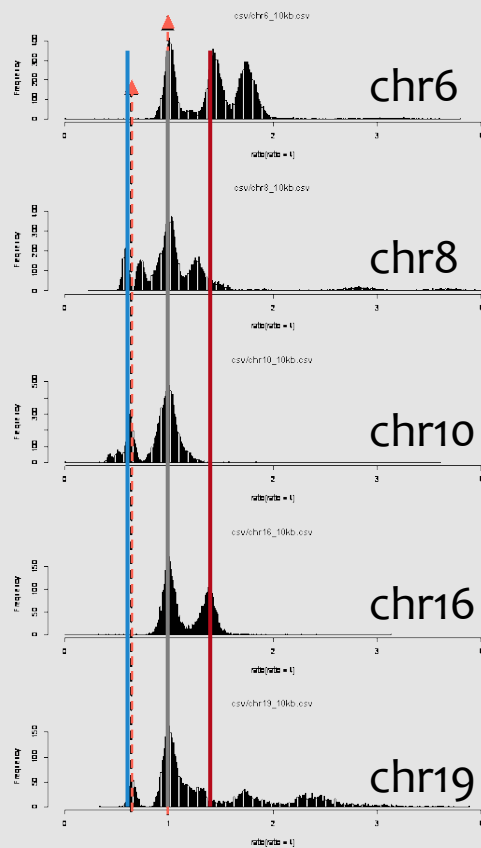Histogram of Tumor / Normal Ratio (a)

FFT de-noised histogram signal (b)

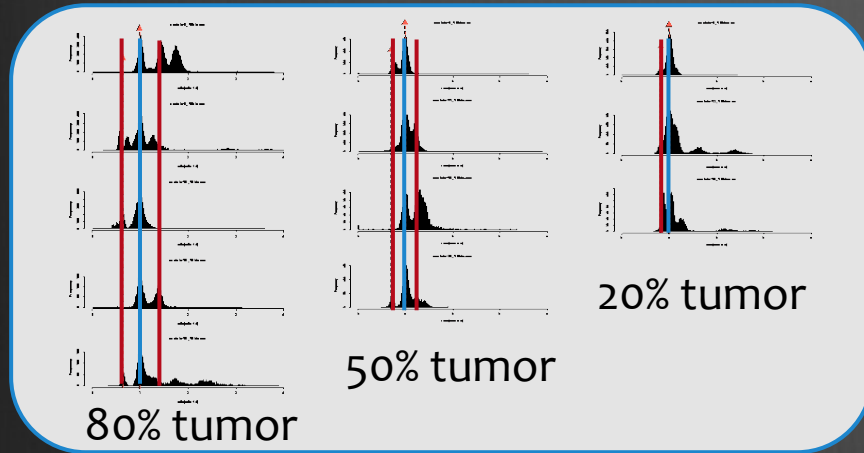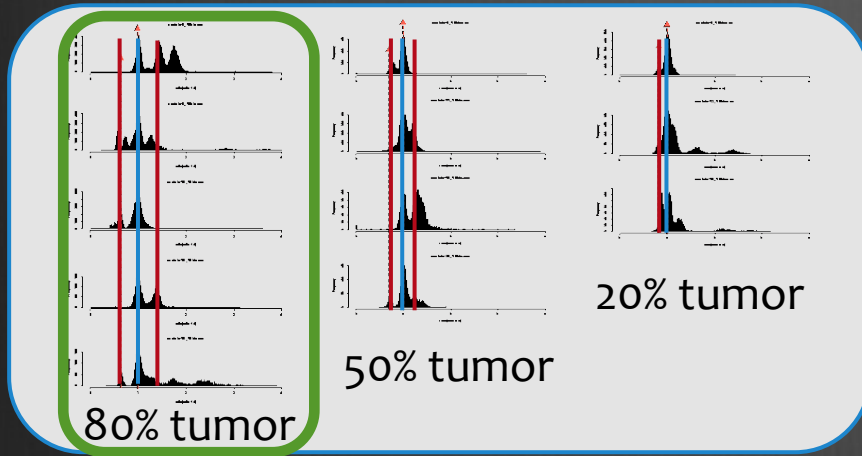Chromosome 19 of sample TCGA-04-1371 (c)

Tumor Heterogeneity

chr6

chr8

chr10

chr16

chr19

chr9

chr11

chr15

chr20

chr7

chr12

chr18

80% tumor
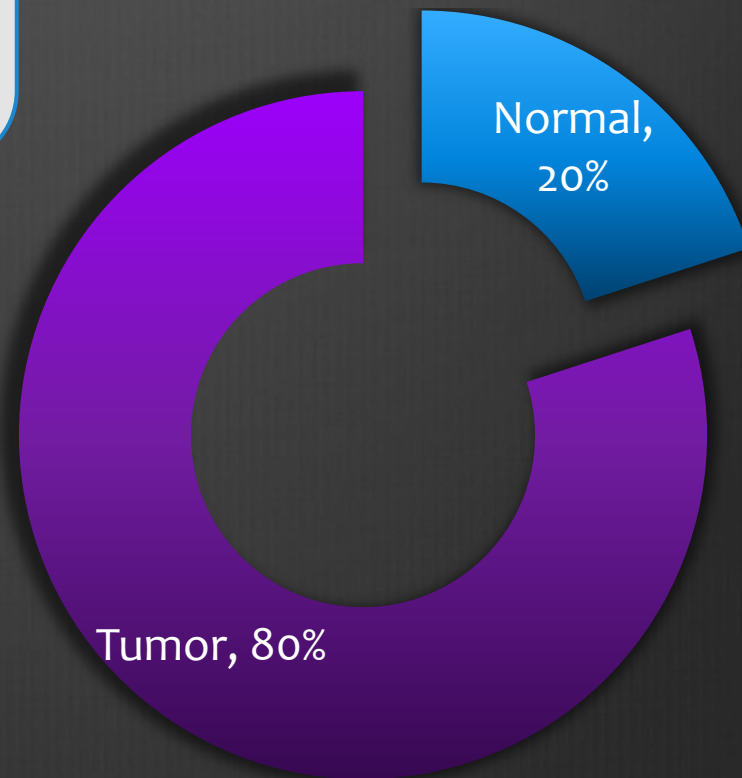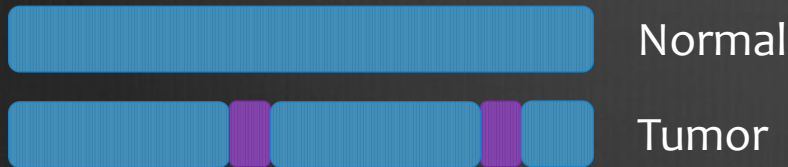
50% tumor

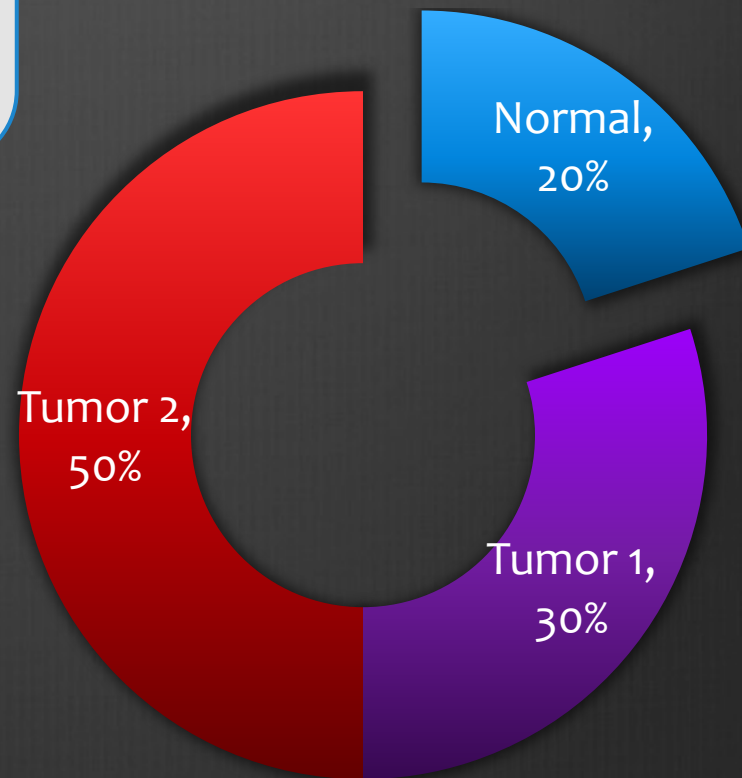20% tumor

Tumor Heterogeneity

80% tumor

50% tumor

20% tumor

Hierarchical
Subclone Structure

Tumor Heterogeneity
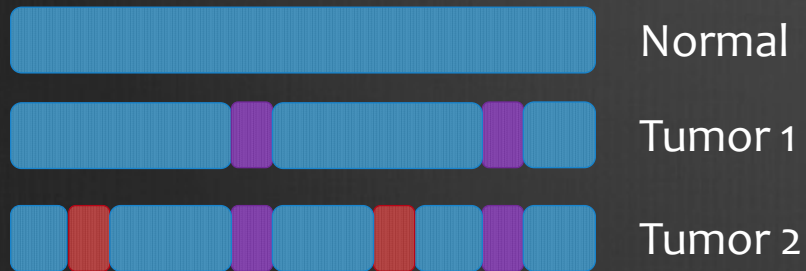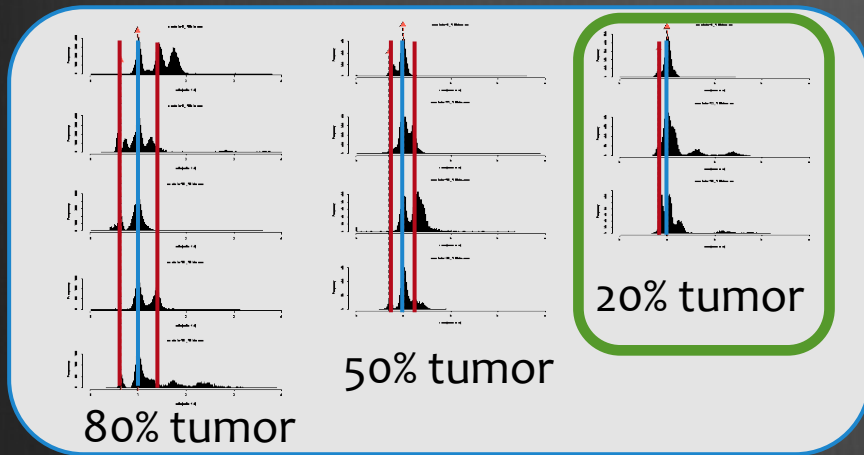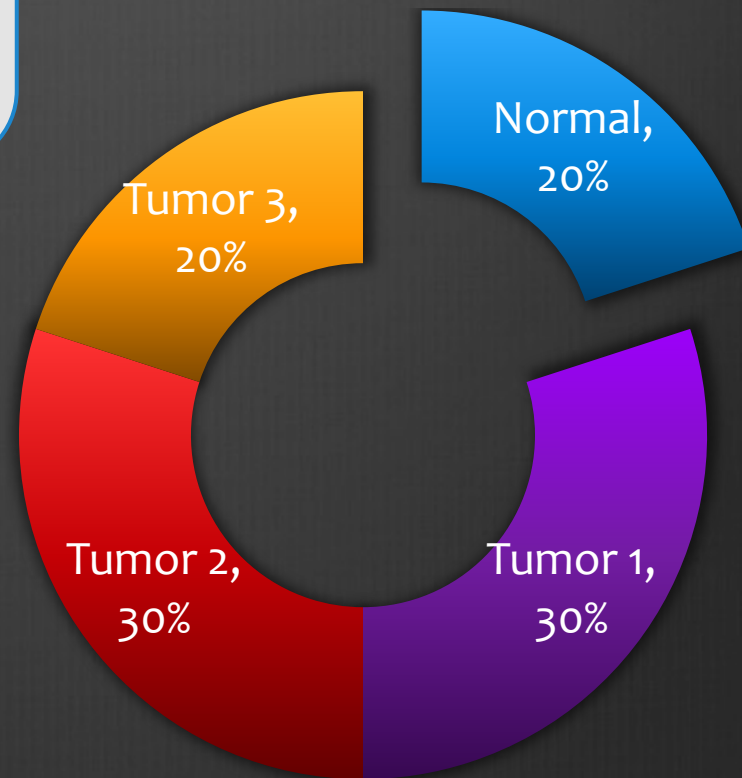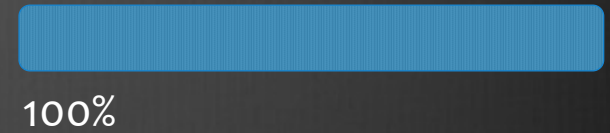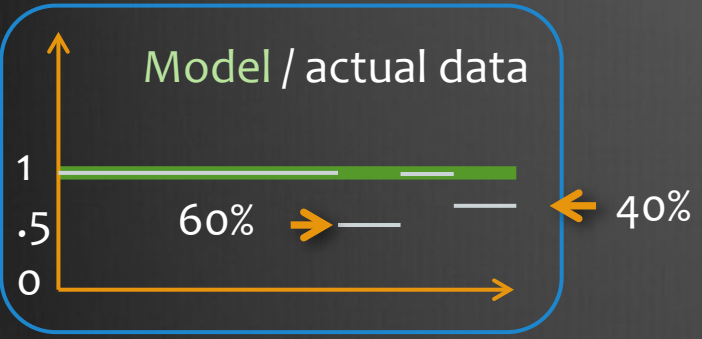
80% tumor

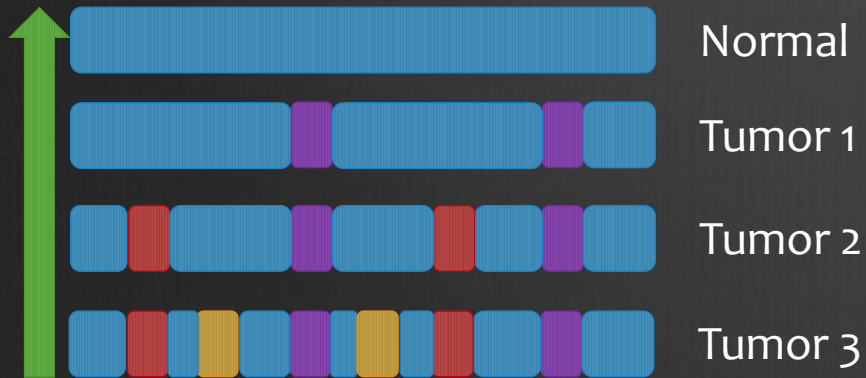50% tumor

20% tumor

Hierarchical
Subclone Structure

Normal

Tumor

Normal, 20%

Tumor, 80%

Tumor Heterogeneity

80% tumor

50% tumor

20% tumor

Hierarchical Subclone Structure

Normal

Tumor 1

Tumor 2

Normal, 20%

Tumor 1, 30%

Tumor 2, 50%

Tumor Heterogeneity
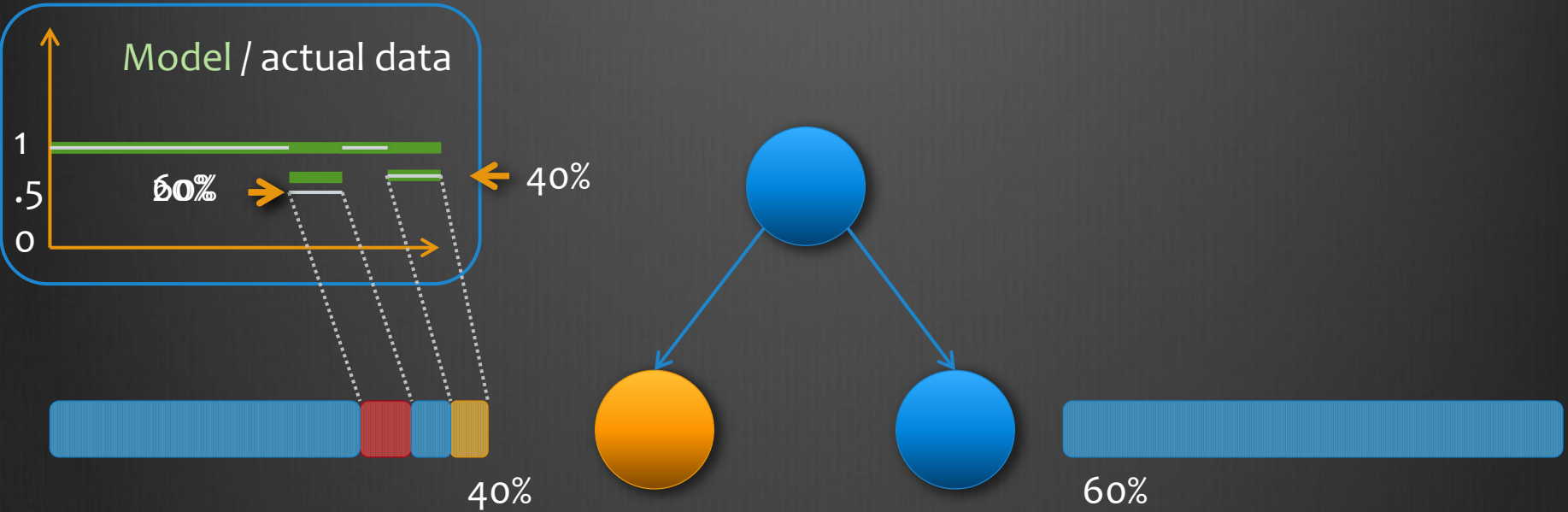

80% tumor

50% tumor

20% tumor

Hierarchical Subclone Structure

Normal

Tumor 1

Tumor 2

Tumor 3

Normal, 20%

Tumor 1, 30%

Tumor 2, 30%

Tumor 3, 20%

Uncover the structure
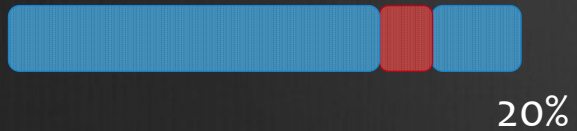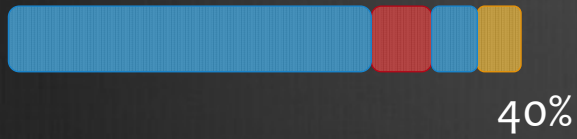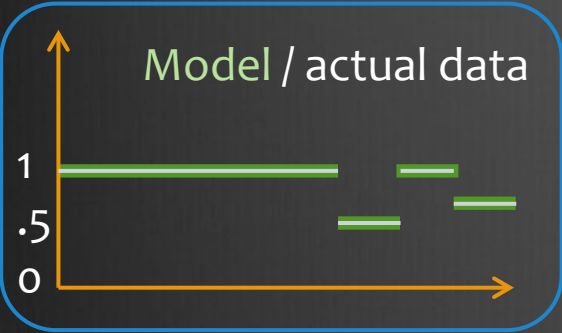

Model / actual data

1
.5
0

60% → 40%

100%

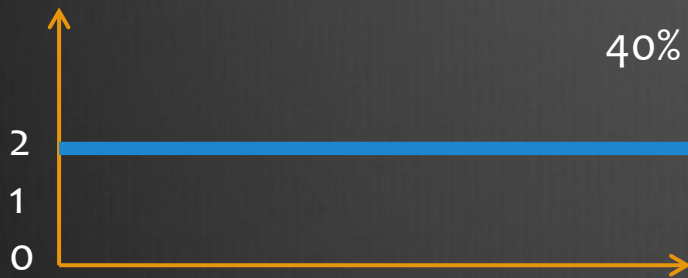# Uncover the structure

Uncover the structure
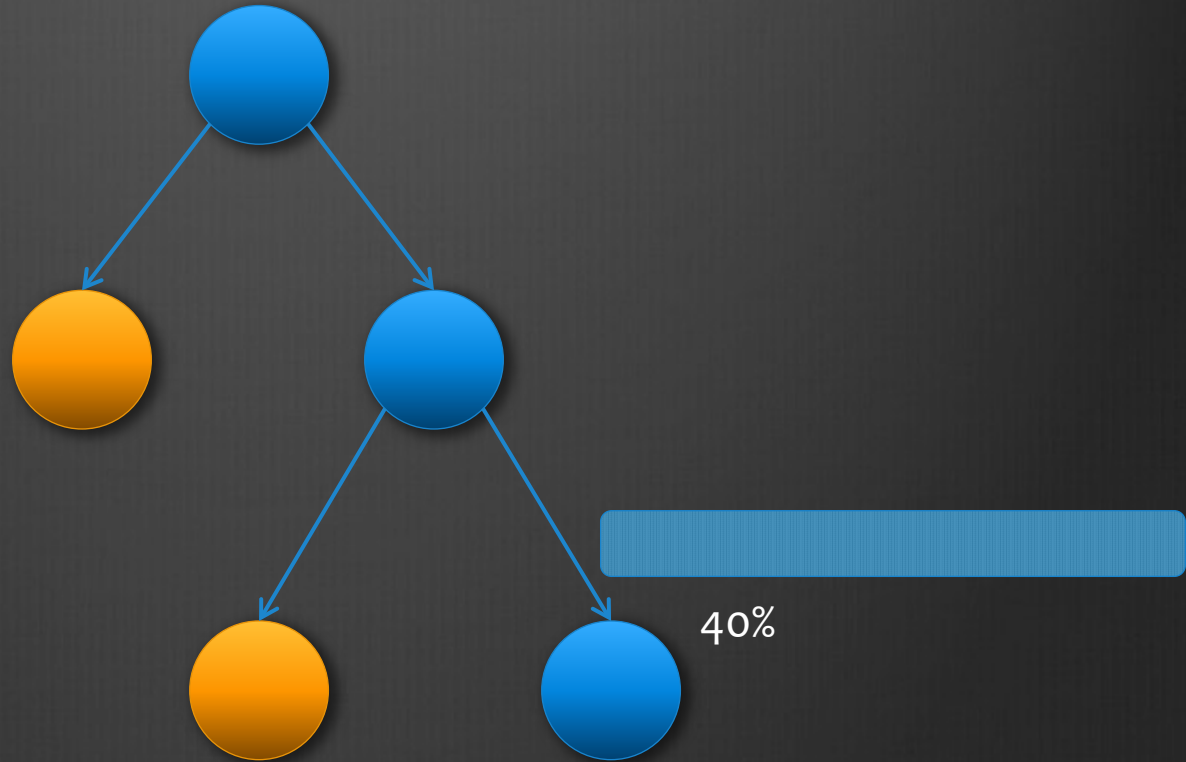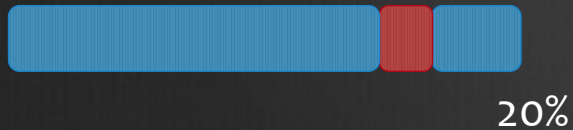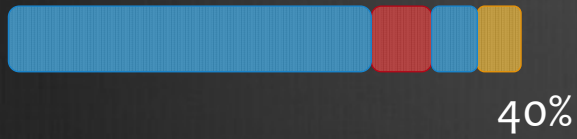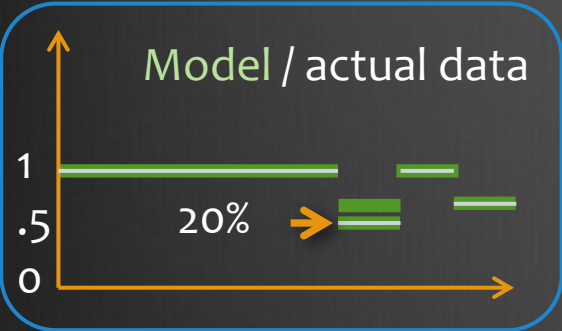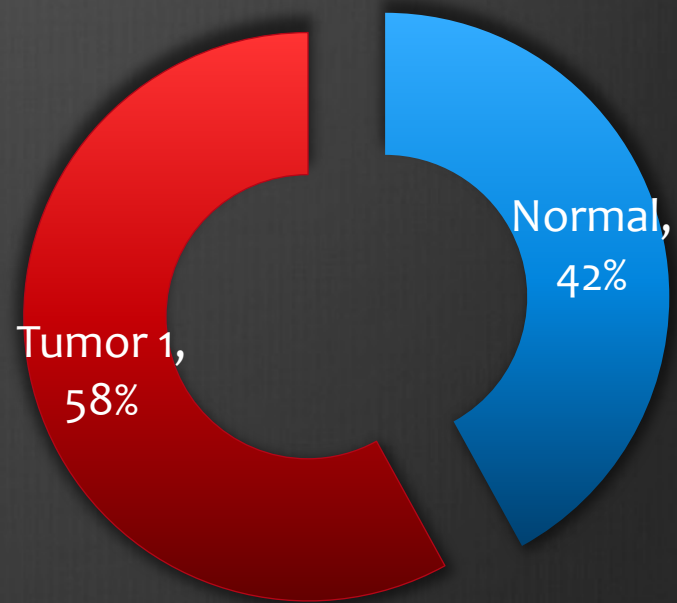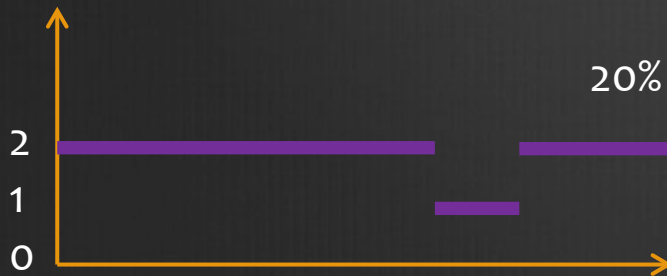
Uncover the structure
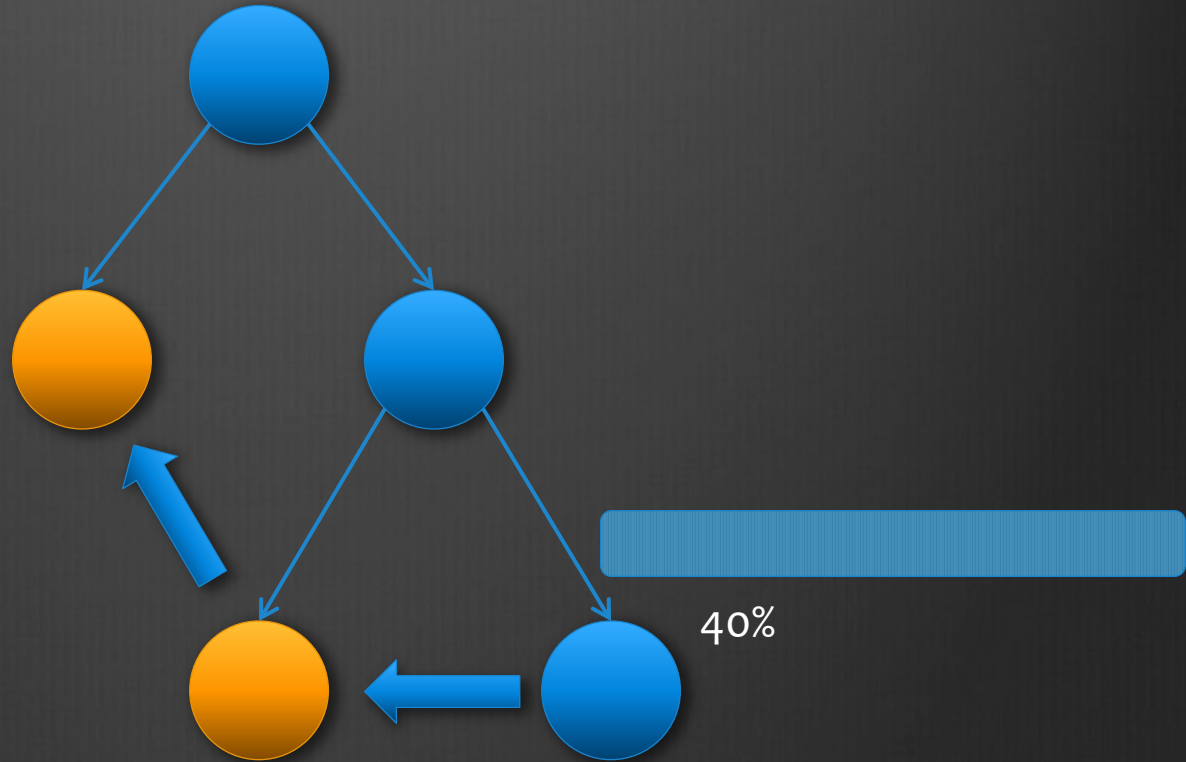
Uncover the structure
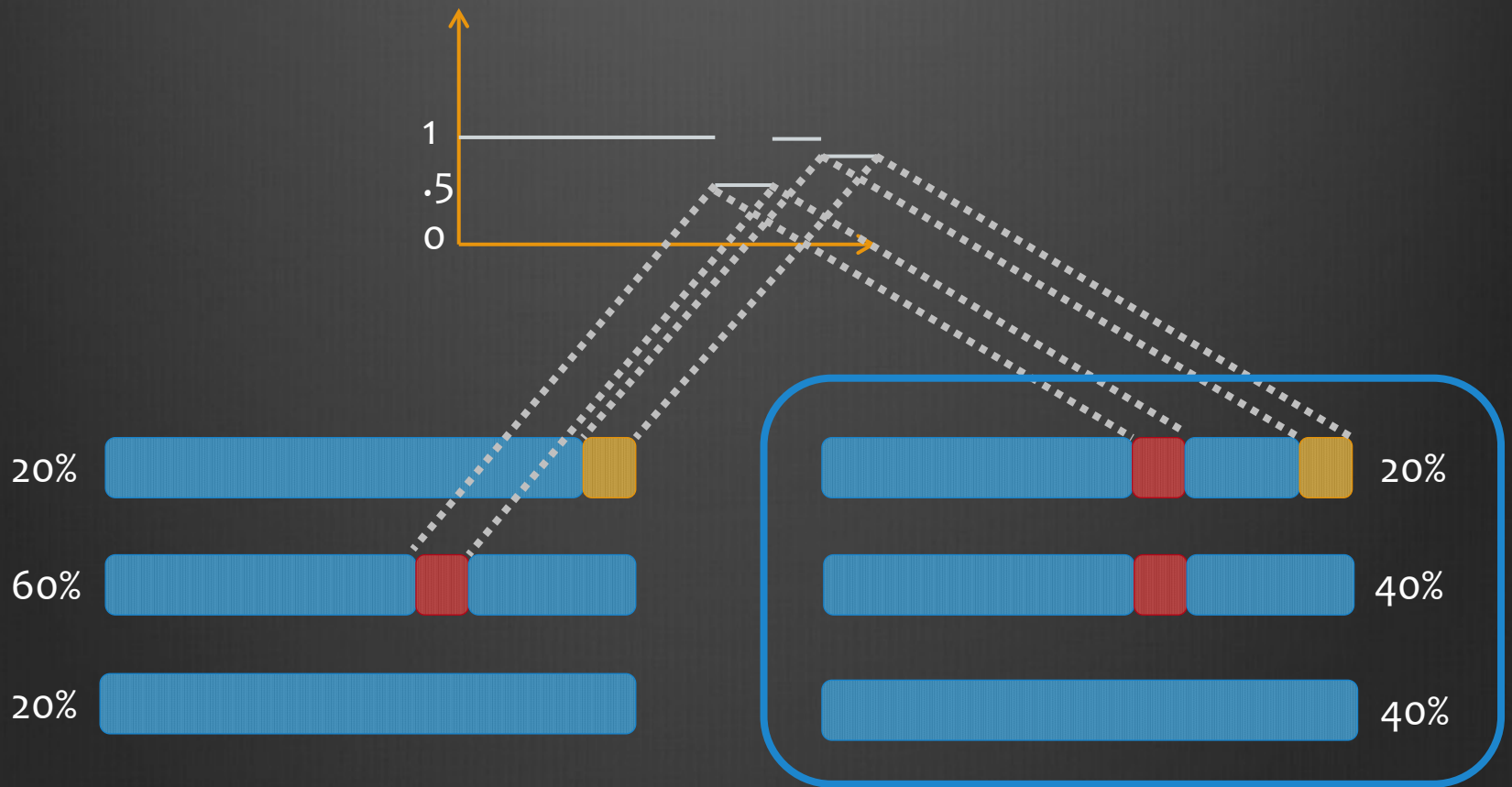
Results

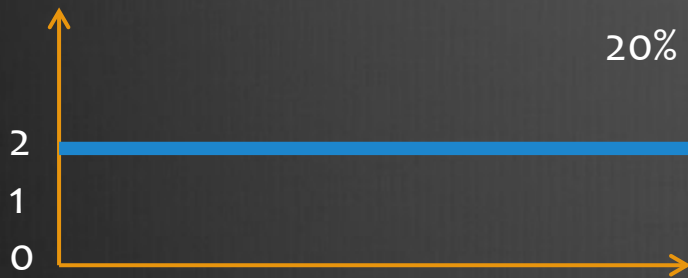# Results

Simulation #2

20%

2
1
0

60%

2
1
0

20%

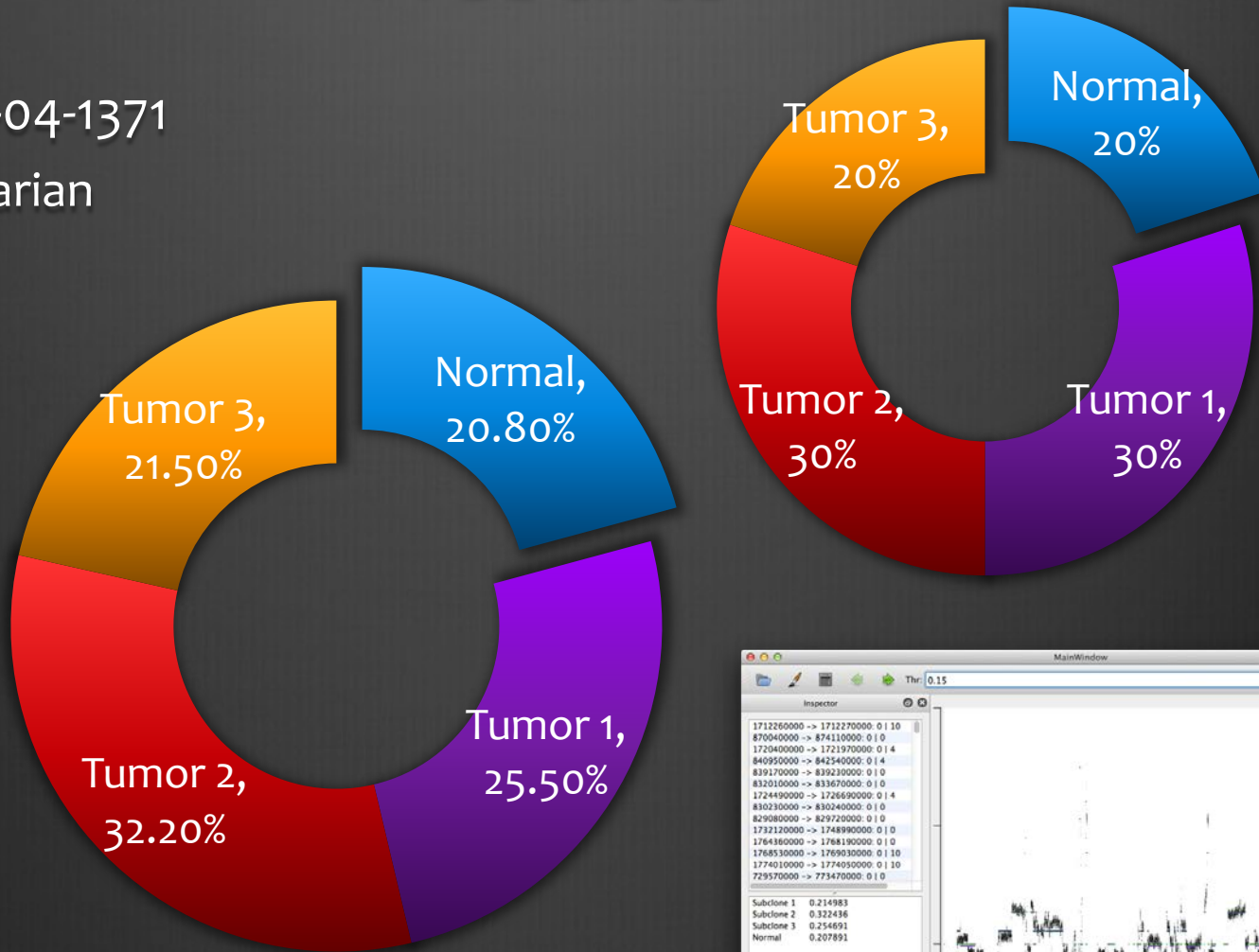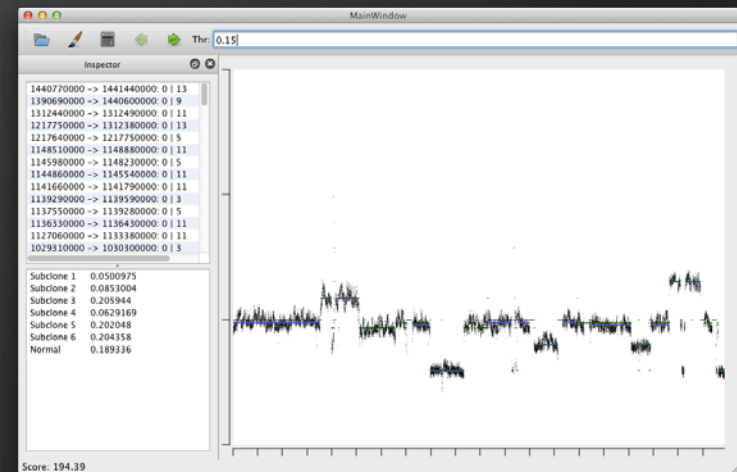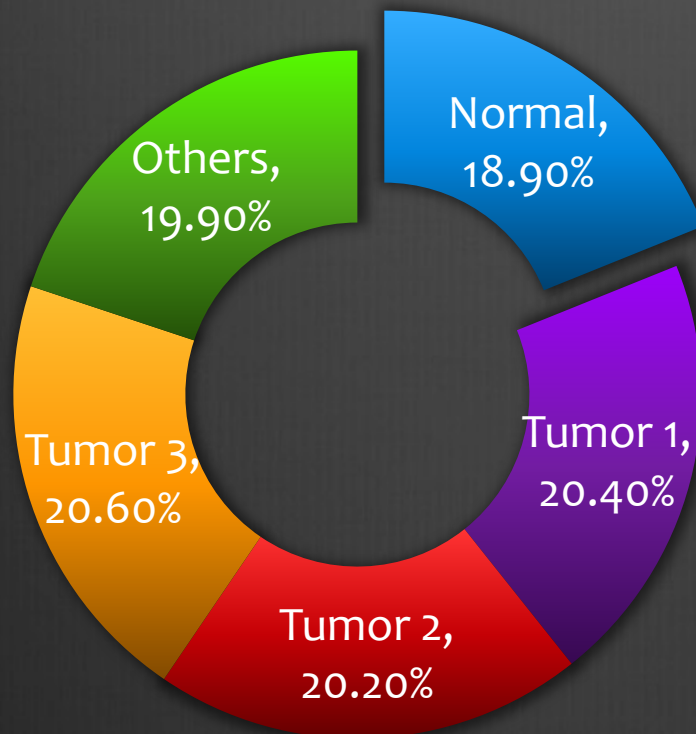2
1
0

Result:

Normal, 23%
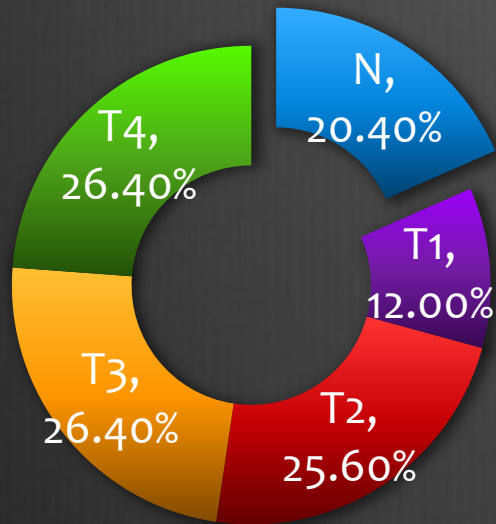
Tumor 1, 20%

Tumor 2, 57%
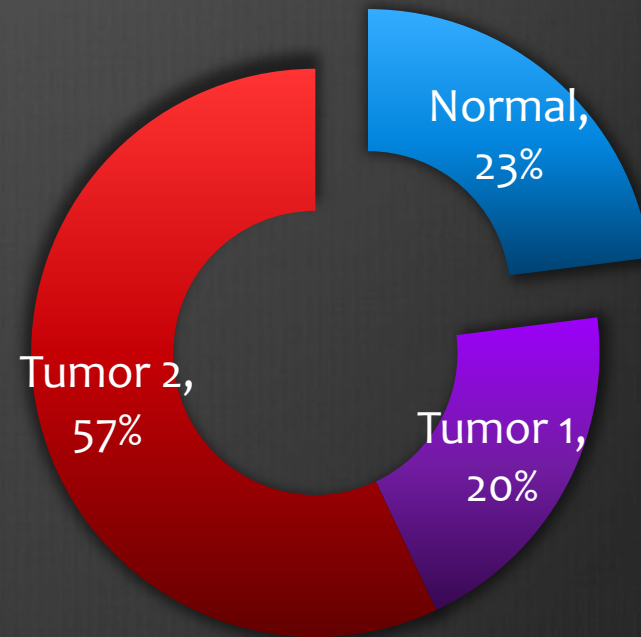
Results

- TCGA-04-1371
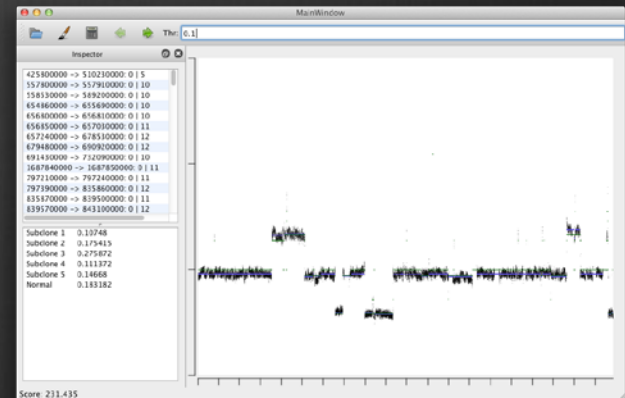 - Ovarian
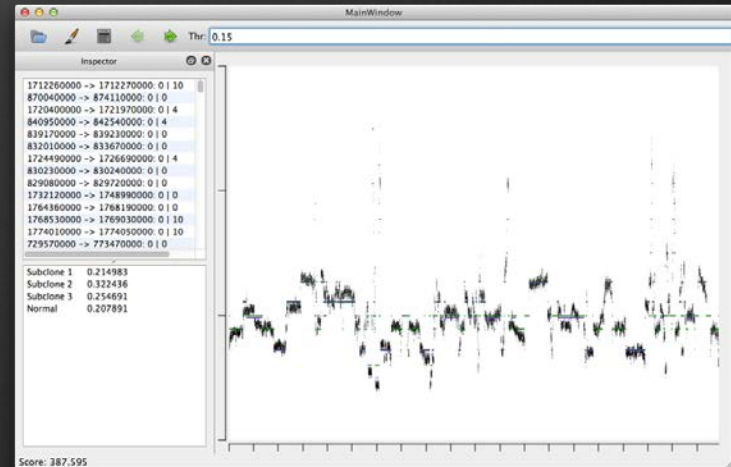
Results

- TCGA-06-0152
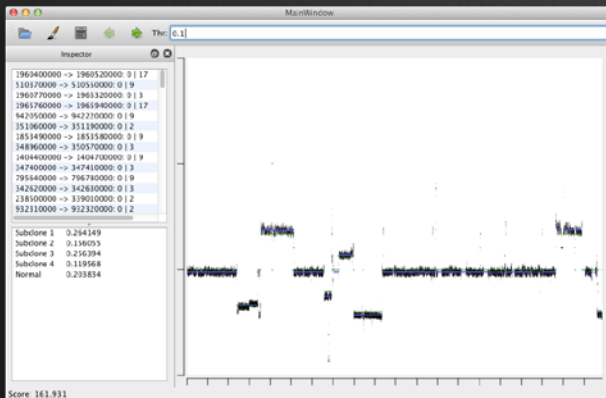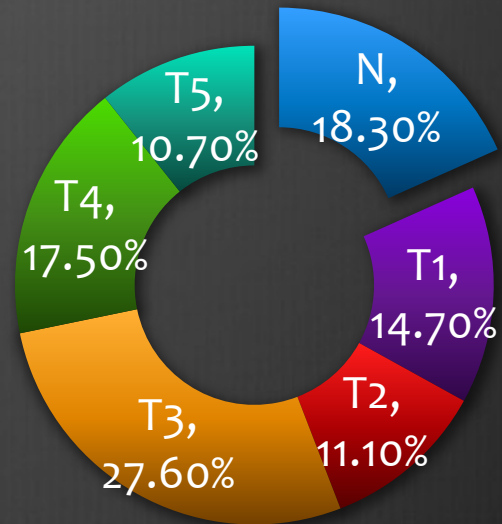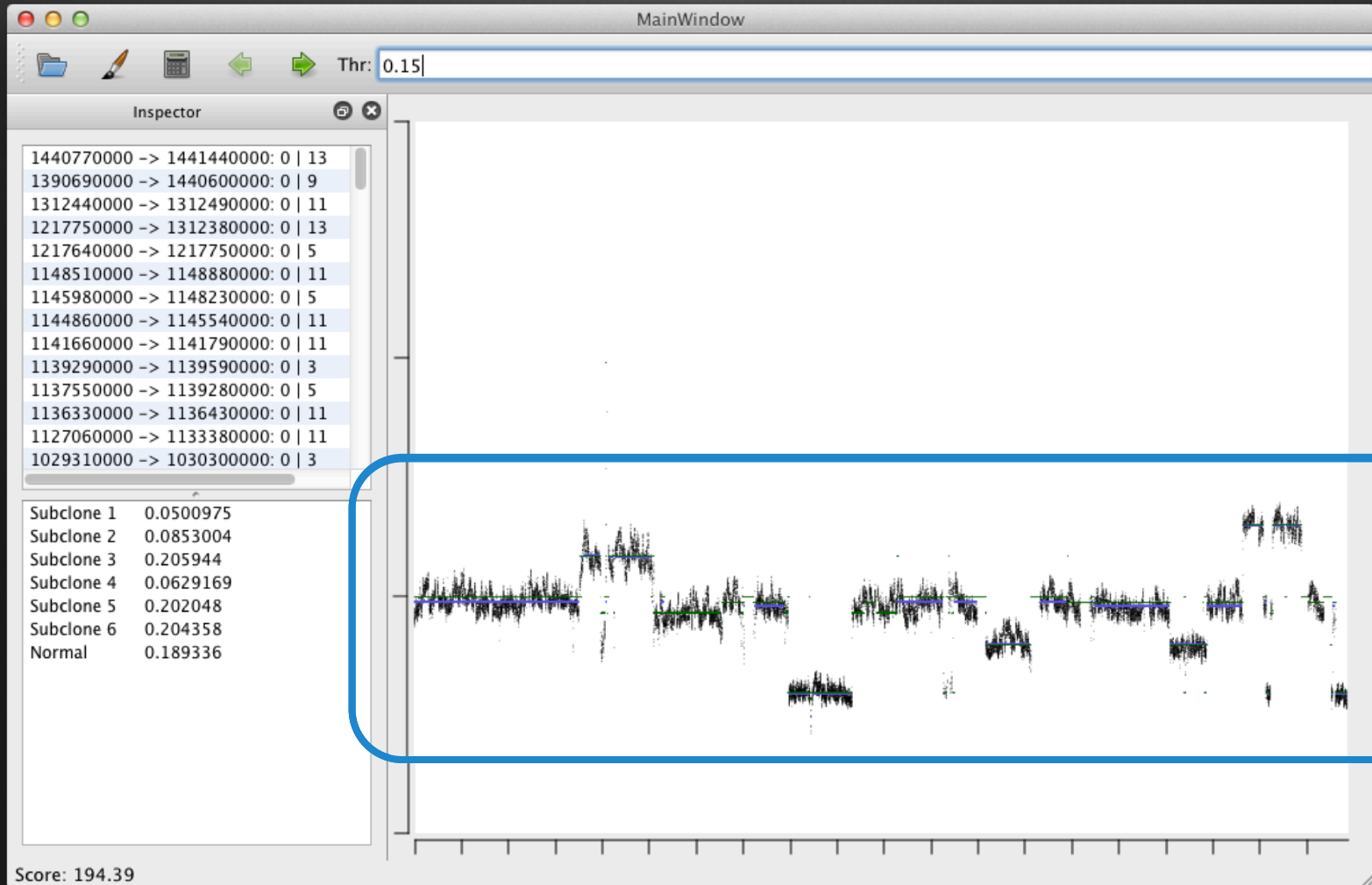 - Glioblastoma

Some more results

TCGA-06-0185 (GBM)

TCGA-06-0145 (GBM)

Interface

Discussion

- Advantages:
  - Simultaneously estimates both the normal cell admixture ratio and the tumor heterogeneity
  - Fast Decomposition of subclone structure: $O(L)$ iterations
  - Provides prior to down-stream analysis (SNP calling, etc…)

- Method independent of CNV caller.

- Structure model is biologically motivated

- Future Direction:
  - Validation (looking for collaborators)
  - Work with capture data

Acknowledgement