**Pipeline: creating count-matrices from BAM files; needed for FacetsY segmentation**

Whole exome sequencing data for 9 organoid samples were made available @ /warm/results/share/Schultz/cheny1. snp-pileup function with parameters -v -A -q15 -Q20 -r10 -g was run to create count matrices on each individual sample. dbsnp\_137.hg19\_\_RmDupsClean\_\_plusPseudo50\_\_DROP\_SORT.vcf.gz served as the reference file for accessing heterozygous SNP positions with the human hg19 reference genome. Count-matrices were analyzed using the FacetsY algorithm.

**Specific statements on individual samples**

**A ASC1\_6669E**

Facets-QC metrics failed for ASC1\_6669E. However, the overall segmentation pattern when using default parameters (i.e. cval 100, snp.nbdh = 250) looks reasonable good. The tumor purity of 98% was estimated (which makes sense, as one may assume that we were exclusively propagating cancer cells in vivo) and the ploidy was estimated around 2.1 (i.e. no traces of genome doubling within the organoid, in vivo model). Sequencing reads for the more informative Yp arm are sparse. The highest sequencing reads density is observed at the Yq 11.21 band, within which no particular gene is located, however should still provide us with a stable prediction. Based on FacetsY, this particular organoid sample has an intactY-chromosome, though with low confidence. The signals are there and allude to an intact Y-chromosome, however large parts (esp., Yq covering most coding genes) are missing (sequencing coverage), and hence my confidence is decreased.

*Segmentation profile*

Chart, scatter chart

Description automatically generated

Gaussian mixture model: building two clusters, wherein most input marker positions fall within the red curve; meaning that the population median CnLR centers around 0.06, which indicates an intact Y-chromosome.

![Chart, line chart

Description automatically generated]()

**B BM61\_6669B**

Pretty much the same pattern as in the organoid sample above. The purity is high (0.95), and the sample appears diploid (2.17). Sequencing reads derived from the Y-chromosome again cluster around the Yq 11.21. cytoband, which makes the whole representation of the Y-chromosome difficult. However, given these restricted circumstances, we do have clear signals of a Y-chromosome loss!

*Gaussian mixture model: The dashed line centers around -0.83 which is a clear indication of Y-chromosome loss; though with a restricted genomic representation (as most signals come from Yq 11.21.)*

![Chart, line chart

Description automatically generated]() ![Chart, histogram

Description automatically generated]()

**C ST111\_6669**

Again, restricted view on genome (due to sequencing read distribution);

Given these circumstances I consider this organoid sample as Y-chromosome lost (although with low confidence).

**D ST121\_6669**

No indication of Y-chromosome loss!

**E ST154\_6669**

This sample has an intact Y-chromosome. Again, restricted view on the whole chromosome. However, this zoom-in portend even to a Y-chromosome gain; meaning that there could potentially be two copies.

**F ST171\_6669**

This sample has a Y-chromosome loss (high confidence);

Looking into naïve algorithm (log2(Tumor.Count/Normal.Count) [without gc-correction])

**F ST43\_6669**

This sample as an intact Y-chromosome. It is highly likely that the Y-chromosome in this particular sample got duplicated (2 copies; whereas the sample itself is NOT WGD)

**G ST60\_6669**

Clear indication of Y-chromosome loss (high confidence)

**H ST88\_6669**

Low confidence about this sample. Can either be intact or lost. The signals tend towards loss; however not entirely clear.

**Summary**

|  |  |  |
| --- | --- | --- |
| Sample | Call | Comment |
| ASC1\_6669E | Intact | - |
| BM61\_6669B | Loss | restricted resolution |
| ST111\_6669 | Loss | low confidence |
| ST121\_6669 | Intact | - |
| ST154\_6669 | Intact | potentially two copies |
| ST171\_6669 | Loss | - |
| ST43\_6669 | Intact | potentially two copies |
| ST60\_6669 | Loss | - |
| ST88\_6669 | N/A | N/A |

Appendix



*‘normal resolution’ ‘restricted’ resolution (signals just from colored region)*