**Output for interval step:**

**INFO [2024-08-01 19:36:29] Loading /mapabality/wgEncodeCrgMapabilityAlign100mer.bigWig...**

**INFO [2024-08-01 19:36:42] Loading PureCN 2.11.0...**

**INFO [2024-08-01 19:36:42] Processing /hg19/IDT.Exomev2.probes.hg19.bed...**

**INFO [2024-08-01 19:36:42] Loading /Reptiming/wgEncodeUwRepliSeqK562WaveSignalRep1.bigWig...**

**WARN [2024-08-01 19:36:43] Found 319004 overlapping intervals, starting at line 2.**

**WARN [2024-08-01 19:36:45] Target intervals were not sorted.**

**INFO [2024-08-01 19:36:48] Averaging reptiming into bins of size 100000...**

**INFO [2024-08-01 19:36:51] Splitting 265448 large targets to an average width of 400.**

**INFO [2024-08-01 19:37:01] Tiling off-target regions to an average width of 200000.**

**WARN [2024-08-01 19:37:20] 248 intervals without mappability score (248 on-target).**

**INFO [2024-08-01 19:37:20] Removing 10182 intervals with low mappability score (<0.60).**

**INFO [2024-08-01 19:37:24] Calculating GC-content...**

**WARN [2024-08-01 19:38:07] Cannot find all contig lengths while exporting interval file.**

**Warning message:**

**In .merge\_two\_Seqinfo\_objects(x, y) :**

**Each of the 2 combined objects has sequence levels not in the other:**

**- in 'x': chr1\_gl000192\_random, chr4\_ctg9\_hap1, chr6\_apd\_hap1, chr6\_cox\_hap2, chr6\_dbb\_hap3, chr6\_mann\_hap4, chr6\_mcf\_hap5, chr6\_qbl\_hap6, chr6\_ssto\_hap7, chr17\_ctg5\_hap1, chrUn\_gl000225**

**- in 'y': chrM**

**Make sure to always combine/compare objects based on the same reference**

**genome (use suppressWarnings() to suppress this warning).**

**Loading required package: TxDb.Hsapiens.UCSC.hg19.knownGene**

**Loading required package: GenomicFeatures**

**Loading required package: AnnotationDbi**

**Loading required package: org.Hs.eg.db**

**WARN [2024-08-01 19:38:22] Attempted adding gene symbols to intervals. Heuristics have been used to pick symbols for overlapping genes.**

**WARN [2024-08-01 19:38:23] Cannot find all contig lengths while exporting interval file.**

**Output for coverage and GC of tumor bam sample:**

Warning messages for 4603:

1: The dot-dot notation (`..level..`) was deprecated in ggplot2 3.4.0.

ℹ Please use `after\_stat(level)` instead.

ℹ The deprecated feature was likely used in the PureCN package.

Please report the issue to the authors.

2: `stat\_contour()`: Zero contours were generated

3: In min(x) : no non-missing arguments to min; returning Inf

4: In max(x) : no non-missing arguments to max; returning -Inf

5: `stat\_contour()`: Zero contours were generated

6: In min(x) : no non-missing arguments to min; returning Inf

7: In max(x) : no non-missing arguments to max; returning -Inf

8: Removed 986 rows containing non-finite outside the scale range

(`stat\_density2d()`).

**OUTPUT for CREATING Normal DB:**

Error in vglm.fitter(x = x, y = y, w = w, offset = offset, Xm2 = Xm2, :

could not obtain valid initial values. Try using 'etastart', 'coefstart' or 'mustart', else family-specific arguments such as 'imethod'.

In addition: There were 50 or more warnings (use warnings() to see the first 50)

WARN [2024-07-12 14:15:46] Could not fit beta binomial dist for hs37d5:25864993 (alt 86,23,62, ref 0,0,0, fa 0.705,0.697,0.873).

Error in vglm.fitter(x = x, y = y, w = w, offset = offset, Xm2 = Xm2, :

could not obtain valid initial values. Try using 'etastart', 'coefstart' or 'mustart', else family-specific arguments such as 'imethod'.

**OUT put for run PureCN in tumor sample:**

**Loading required package: DNAcopy**

**Loading required package: VariantAnnotation**

**Loading required package: BiocGenerics**

**Attaching package: ‘BiocGenerics’**

**The following objects are masked from ‘package:stats’:**

**IQR, mad, sd, var, xtabs**

**The following objects are masked from ‘package:base’:**

**anyDuplicated, aperm, append, as.data.frame, basename, cbind,**

**colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,**

**get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,**

**match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,**

**Position, rank, rbind, Reduce, rownames, sapply, setdiff, table,**

**tapply, union, unique, unsplit, which.max, which.min**

**Loading required package: MatrixGenerics**

**Loading required package: matrixStats**

**Attaching package: ‘MatrixGenerics’**

**The following objects are masked from ‘package:matrixStats’:**

**colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,**

**colCounts, colCummaxs, colCummins, colCumprods, colCumsums,**

**colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,**

**colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,**

**colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,**

**colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,**

**colWeightedMeans, colWeightedMedians, colWeightedSds,**

**colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,**

**rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,**

**rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,**

**rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,**

**rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,**

**rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,**

**rowWeightedMads, rowWeightedMeans, rowWeightedMedians,**

**rowWeightedSds, rowWeightedVars**

**Loading required package: GenomeInfoDb**

**Loading required package: S4Vectors**

**Loading required package: stats4**

**Attaching package: ‘S4Vectors’**

**The following object is masked from ‘package:utils’:**

**findMatches**

**The following objects are masked from ‘package:base’:**

**expand.grid, I, unname**

**Loading required package: IRanges**

**Loading required package: GenomicRanges**

**Loading required package: SummarizedExperiment**

**Loading required package: Biobase**

**Welcome to Bioconductor**

**Vignettes contain introductory material; view with**

**'browseVignettes()'. To cite Bioconductor, see**

**'citation("Biobase")', and for packages 'citation("pkgname")'.**

**Attaching package: ‘Biobase’**

**The following object is masked from ‘package:MatrixGenerics’:**

**rowMedians**

**The following objects are masked from ‘package:matrixStats’:**

**anyMissing, rowMedians**

**Loading required package: Rsamtools**

**Loading required package: Biostrings**

**Loading required package: XVector**

**Attaching package: ‘Biostrings’**

**The following object is masked from ‘package:base’:**

**strsplit**

**Attaching package: ‘VariantAnnotation’**

**The following object is masked from ‘package:base’:**

**tabulate**

**INFO [2024-08-05 20:41:51] Loading PureCN 2.11.0...**

**INFO [2024-08-05 20:42:02] Mean coverages: chrX: 211.10, chrY: 2.14, chr1-22: 224.13.**

**INFO [2024-08-05 20:42:02] Sample sex: F**

**WARN [2024-08-05 20:42:03] Recommended to provide --fun-segmentation PSCBS.**

**INFO [2024-08-05 20:42:03] ------------------------------------------------------------**

**INFO [2024-08-05 20:42:03] PureCN 2.11.0**

**INFO [2024-08-05 20:42:03] ------------------------------------------------------------**

**INFO [2024-08-05 20:42:03] Arguments: -tumor.coverage.file /Output/4603-C/4603\_tumor\_coverage\_loess\_hg19.txt.gz -log.ratio -seg.file -vcf.file /Tumour\_Samples/4603\_TNh\_lifted\_over.vcf.gz -genome hg19 -sex ? -args.setPriorVcf 6 -args.setMappingBiasVcf NULL -args.filterIntervals 100,0.05 -args.segmentation 0.005,NULL, -sampleid 4603\_C -min.ploidy 1.4 -max.ploidy 6 -max.non.clonal 0.2 -max.homozygous.loss 0.05,1e+07 -log.ratio.calibration 0.1 -model.homozygous FALSE -error 0.001 -interval.file /Intervals/baits\_c\_hg19\_intervals.txt -min.logr.sdev 0.15 -max.segments 300 -plot.cnv TRUE -vcf.field.prefix PureCN. -cosmic.vcf.file -DB.info.flag DB -POPAF.info.field POP\_AF -Cosmic.CNT.info.field Cosmic.CNT -model beta -post.optimize FALSE -BPPARAM -log.file / /Output/4603\_C.log -normal.coverage.file <data> -normalDB <data> -args.filterVcf <data> -fun.segmentation <data> -test.num.copy <data> -test.purity <data> -speedup.heuristics <data>**

**INFO [2024-08-05 20:42:03] Loading coverage files...**

**INFO [2024-08-05 20:42:09] Mean target coverages: 227X (tumor) 219X (normal).**

**INFO [2024-08-05 20:42:11] Mean coverages: chrX: 211.10, chrY: 2.14, chr1-22: 224.13.**

**INFO [2024-08-05 20:42:11] Mean coverages: chrX: 236.02, chrY: 1.40, chr1-22: 215.70.**

**INFO [2024-08-05 20:42:36] Removing 1145 intervals with missing log.ratio.**

**INFO [2024-08-05 20:42:36] Removing 18 low/high GC targets.**

**INFO [2024-08-05 20:42:36] Removing 2235 intervals excluded in normalDB.**

**INFO [2024-08-05 20:42:36] Removing 4 intervals with low total coverage in normal (< 150.00 reads).**

**INFO [2024-08-05 20:42:36] normalDB provided. Setting minimum coverage for segmentation to 0.0015X.**

**INFO [2024-08-05 20:42:36] Removing 15119 low count (< 100 total reads) intervals.**

**INFO [2024-08-05 20:42:37] Removing 5 low coverage (< 0.0015X) intervals.**

**WARN [2024-08-05 20:42:37] Not enough off-target intervals. Ignoring them (254038 on-target, 6181 off-target, ratio 0.02).**

**INFO [2024-08-05 20:42:37] Removing 6181 off-target intervals.**

**INFO [2024-08-05 20:42:37] Using 254038 intervals (254038 on-target, 0 off-target).**

**INFO [2024-08-05 20:42:37] No off-target intervals. If this is hybrid-capture data, consider adding them.**

**INFO [2024-08-05 20:42:38] AT/GC dropout: 1.09 (tumor), 1.00 (normal), 1.09 (coverage log-ratio).**

**WARN [2024-08-05 20:42:38] High GC-bias in normalized tumor vs normal log2 ratio.**

**INFO [2024-08-05 20:42:38] Loading VCF...**

**INFO [2024-08-05 20:42:44] Found 301711 variants in VCF file.**

**INFO [2024-08-05 20:42:46] 145611 (48.3%) variants annotated as likely germline (DB INFO flag).**

**WARN [2024-08-05 20:42:56] Found 19 variants with missing allelic fraction starting with chr1:216689978\_T/TACACACACACACACAC. Removing them.**

**WARN [2024-08-05 20:42:57] Did not find base quality scores, will use global error rate of 0.0010 instead.**

**INFO [2024-08-05 20:42:58] 4603 is tumor in VCF file.**

**INFO [2024-08-05 20:43:01] 745 homozygous and 776 heterozygous variants on chrX.**

**INFO [2024-08-05 20:43:01] Sex from VCF: F (Fisher's p-value: < 0.0001, odds-ratio: 1.29).**

**INFO [2024-08-05 20:43:02] Removing 0 low quality variants with non-offset BQ < 25.**

**INFO [2024-08-05 20:43:02] Global base quality score of 29**

**INFO [2024-08-05 20:43:02] Minimum number of supporting reads ranges from 2 to 24, depending on coverage and BQS.**

**INFO [2024-08-05 20:43:46] Initial testing for significant sample cross-contamination: maybe**

**INFO [2024-08-05 20:43:48] Removing 213889 variants with AF < 0.030 or AF >= 0.970 or insufficient supporting reads or depth < 15.**

**INFO [2024-08-05 20:43:48] Total size of targeted genomic region: 40.62Mb (61.27Mb with 50bp padding).**

**INFO [2024-08-05 20:43:49] 15.0% of targets contain variants.**

**INFO [2024-08-05 20:43:49] Removing 26097 variants outside intervals.**

**INFO [2024-08-05 20:43:49] Setting somatic prior probabilities for likely germline hits to 0.000500 or to 0.500000 otherwise.**

**INFO [2024-08-05 20:43:49]** **VCF does not contain somatic status. For best results, consider providing mapping.bias.file when matched normals are not available.**

**INFO [2024-08-05 20:43:49] Excluding 29117 novel or poor quality variants from segmentation.**

**INFO [2024-08-05 20:43:50] Sample sex: F**

**INFO [2024-08-05 20:43:50] Segmenting data...**

**INFO [2024-08-05 20:43:50] Interval weights found, will use weighted CBS.**

**INFO [2024-08-05 20:43:50] Loading pre-computed boundaries for DNAcopy...**

**INFO [2024-08-05 20:43:50] Setting undo.SD parameter to 0.750000.**

**INFO [2024-08-05 20:45:40] Setting undo.SD parameter to 1.125000.**

**Setting multi-figure configuration**

**INFO [2024-08-05 20:48:24] Setting prune.hclust.h parameter to 0.200000.**

**INFO [2024-08-05 20:48:41] Found 3566 segments with median size of 0.07Mb.**

**INFO [2024-08-05 20:48:41] Removing 415 variants outside segments.**

**INFO [2024-08-05 20:48:41] Using 61291 variants.**

**INFO [2024-08-05 20:48:47] Mean standard deviation of log-ratios: 0.28 (MAPD: 0.20)**

**INFO [2024-08-05 20:48:47] 2D-grid search of purity and ploidy...**

**INFO [2024-08-05 20:59:10] Local optima: 0.31/2.4, 0.31/2, 0.82/4, 0.92/5, 0.78/2**

**INFO [2024-08-05 20:59:10] Testing local optimum 1/5 at purity 0.31 and total ploidy 2.40...**

**INFO [2024-08-05 21:07:18] Testing local optimum 2/5 at purity 0.31 and total ploidy 2.00...**

**INFO [2024-08-05 21:15:27] Testing local optimum 3/5 at purity 0.82 and total ploidy 4.00...**

**INFO [2024-08-05 21:22:59] Testing local optimum 4/5 at purity 0.92 and total ploidy 5.00...**

**INFO [2024-08-05 21:29:18] Testing local optimum 5/5 at purity 0.78 and total ploidy 2.00...**

**INFO [2024-08-05 21:36:08] Skipping 1 solutions that converged to the same optima.**

**INFO [2024-08-05 21:36:08] Fitting variants with beta model for local optimum 1/5...**

**INFO [2024-08-05 21:36:10] Fitting variants for purity 0.36, tumor ploidy 4.46 and contamination 0.01.**

**INFO [2024-08-05 21:49:16] Optimized purity: 0.36**

**INFO [2024-08-05 21:49:16] Fitting variants with beta model for local optimum 3/5...**

**INFO [2024-08-05 21:49:18] Fitting variants for purity 0.85, tumor ploidy 4.54 and contamination 0.01.**

**INFO [2024-08-05 22:02:19] Optimized purity: 0.85**

**INFO [2024-08-05 22:02:19] Fitting variants with beta model for local optimum 4/5...**

**INFO [2024-08-05 22:02:23] Fitting variants for purity 0.95, tumor ploidy 5.50 and contamination 0.01.**

**INFO [2024-08-05 22:15:32] Optimized purity: 0.95**

**INFO [2024-08-05 22:15:32] Fitting variants with beta model for local optimum 5/5...**

**INFO [2024-08-05 22:15:34] Fitting variants for purity 0.59, tumor ploidy 2.61 and contamination 0.01.**

**INFO [2024-08-05 22:28:40] Optimized purity: 0.59**

**INFO [2024-08-05 22:28:40] Initial guess of contamination rate: 0.047**

**INFO [2024-08-05 22:28:40] Optimizing contamination rate of optimum 1/4...**

**INFO [2024-08-05 22:41:49] Optimized contamination rate: 0.050**

**INFO [2024-08-05 22:41:49] Initial guess of contamination rate: 0.041**

**INFO [2024-08-05 22:41:49] Optimizing contamination rate of optimum 2/4...**

**INFO [2024-08-05 22:54:58] Optimized contamination rate: 0.041**

**INFO [2024-08-05 22:54:58] Optimized contamination rate: 0.000**

**INFO [2024-08-05 22:54:58] Initial guess of contamination rate: 0.037**

**INFO [2024-08-05 22:54:58] Optimizing contamination rate of optimum 4/4...**

**INFO [2024-08-05 23:08:05] Optimized contamination rate: 0.047**

**INFO [2024-08-05 23:08:05] Done.**

**INFO [2024-08-05 23:08:05] ------------------------------------------------------------**

**INFO [2024-08-05 23:08:18] Generating output files...**

**Warning messages:**

**1: In par(par.mar) : argument 1 does not name a graphical parameter**

**2: In par(par.mar) : argument 1 does not name a graphical parameter**

**3: In par(par.mar) : argument 1 does not name a graphical parameter**

**4: In par(par.mar) : argument 1 does not name a graphical parameter**

**INFO [2024-08-05 23:09:00] Tumor/normal noise ratio: 14.047**

**WARN [2024-08-05 23:09:00****] Extensive noise in tumor compared to normals.**