

recount (DER analyses)

Kai Kammers and Shannon Ellis

July 22, 2016

Contents

Download study data	2
Obtain count data	3
DE analysis	4
Annotate regions with genomic information	7
Gene level analysis	9
QC data	9
DE analysis : gene-level	11
Compare DER findings to gene-level analysis	13
Compare DERs to known genomic annotation	14
Compare DERs to known genomic annotation	18
Reproducibility	19

Here is an example of how to download and analyze a `RangedSummarizedExperiment` object. While differential expression analysis of RNA-Seq data is often done at the gene and/or exon level, there is utility in looking for differential expression using an annotation-independent approach. Here, without the aid of any annotation, expressed regions are summarized, filtered, and differential expression analyses are carried out. These results are compared to those from annotation-dependent analyses carried out on the same data to assess whether additional information can be garnered from this approach. Data herein come from triple negative and HER2-positive breast cancer samples from SRA study id SRP032798.

We first load the required packages.

```
## load libraries
library('recount')
library('SummarizedExperiment')
library('limma')
library('edgeR')
library('qvalue')
library('topGO')
library('matrixStats')
library('RSkittleBrewer')
library('derfinder')
```

```

library('derfinderPlot')
library('BiocParallel')
library('GenomicRanges')
library('bumphunter')
library('downloader')
library('TxDb.Hsapiens.UCSC.hg38.knownGene')

## set colors
trop <- RSkittleBrewer('tropical')[c(1, 2)]

```

Download study data

Expressed region data are downloaded for each chromosome and summarized into a count matrix (`covMat`) where each region is a row and each sample a column.

```

chrs <- paste0('chr', c(1:22, 'X', 'Y'))
bp <- SerialParam() ## Change if you have access to more cores

## Download region level data
if(!file.exists('regions_SRP032789.Rdata')) {
  regions_list <- bplapply(chrs, function(chr) {
    regs <- expressed_regions('SRP032789', chr, cutoff = 5L,
      maxClusterGap = 3000L, verbose = FALSE)
    return(regs)
  }, BPPARAM = bp)
  names(regions_list) <- chrs
  regions <- unlist(GRangesList(regions_list))

  ## Save the regions
  save(regions, regions_list, file = 'regions_SRP032789.Rdata')
} else {
  load('regions_SRP032789.Rdata')
}

## Compute coverage matrix for study SRP032789
if(!file.exists('covMat_SRP032789.Rdata')) {
  covMat <- bplapply(chrs, function(chr) {
    coverageMatrix <- coverage_matrix('SRP032789', chr,
      regions_list[[chr]], verboseLoad = FALSE)
    return(coverageMatrix)
  }, BPPARAM = bp)
  covMat <- do.call(rbind, covMat)

  ## Round the coverage matrix to integers
  covMat <- round(covMat, 0)
  save(covMat, file = 'covMat_SRP032789.Rdata')
} else {
  load('covMat_SRP032789.Rdata')
}

```

Obtain count data

After summarizing data across expressed regions, data are filtered to only include samples of interest (TNBC and HER2+ tumor samples, (`covMat_filt`) and regions with greater than 5 mean read counts across samples (`counts`).

```
## Download phenotype data from
## http://trace.ncbi.nlm.nih.gov/Traces/study/?acc=SRP032789
pheno <- read.table('SraRunTable_SRP032789.txt', sep = '\t',
  header=TRUE,
  stringsAsFactors = FALSE)

## Check ordering of samples
pheno <- pheno[pheno$Run_s %in% colnames(covMat), ]
identical(pheno$Run_s, colnames(covMat))

## [1] FALSE

## Obtain correct order for pheno data
pheno <- pheno[match(colnames(covMat), pheno$Run_s), ]
identical(pheno$Run_s, colnames(covMat))

## [1] TRUE

head(cbind(pheno$Run_s, colnames(covMat)))

##      [,1]      [,2]
## [1,] "SRR1027171" "SRR1027171"
## [2,] "SRR1027173" "SRR1027173"
## [3,] "SRR1027174" "SRR1027174"
## [4,] "SRR1027175" "SRR1027175"
## [5,] "SRR1027176" "SRR1027176"
## [6,] "SRR1027177" "SRR1027177"

## Find tumor type information
group <- pheno$tumor_type_s
table(group)

## group
## HER2 Positive Breast Tumor      Non-TNBC Breast Tumor
##                               5                  6
## Normal Breast Organoids        TNBC Breast Tumor
##                               3                  6

## Subset data to HER2 and TNBC type
covMat_filt <- covMat[, group %in% c('HER2 Positive Breast Tumor', 'TNBC Breast Tumor')]
group <- group[group %in% c('HER2 Positive Breast Tumor', 'TNBC Breast Tumor')]
dim(covMat_filt)

## [1] 328481      11
```

```

rownames(covMat_filt) <- rownames(regions)

## Filter count matrix
counts <- covMat_filt
filter <- apply(counts, 1, function(x) mean(x) > 5)
counts <- counts[filter, ]
dim(counts)

## [1] 291247      11

## Obtain chromosome and position information for regions included after filtering
regions_counts <- regions[filter, ]

```

DE analysis

Differential expression analysis is carried out at the expressed-region level using `voom` and `limma`. Results for each region are plotted using a volcano plot to compare the effect size of the differential expression [as measured by the $\log_2(fold - change)$ in expression] and its significance [$-\log_{10}(p - value)$].

```

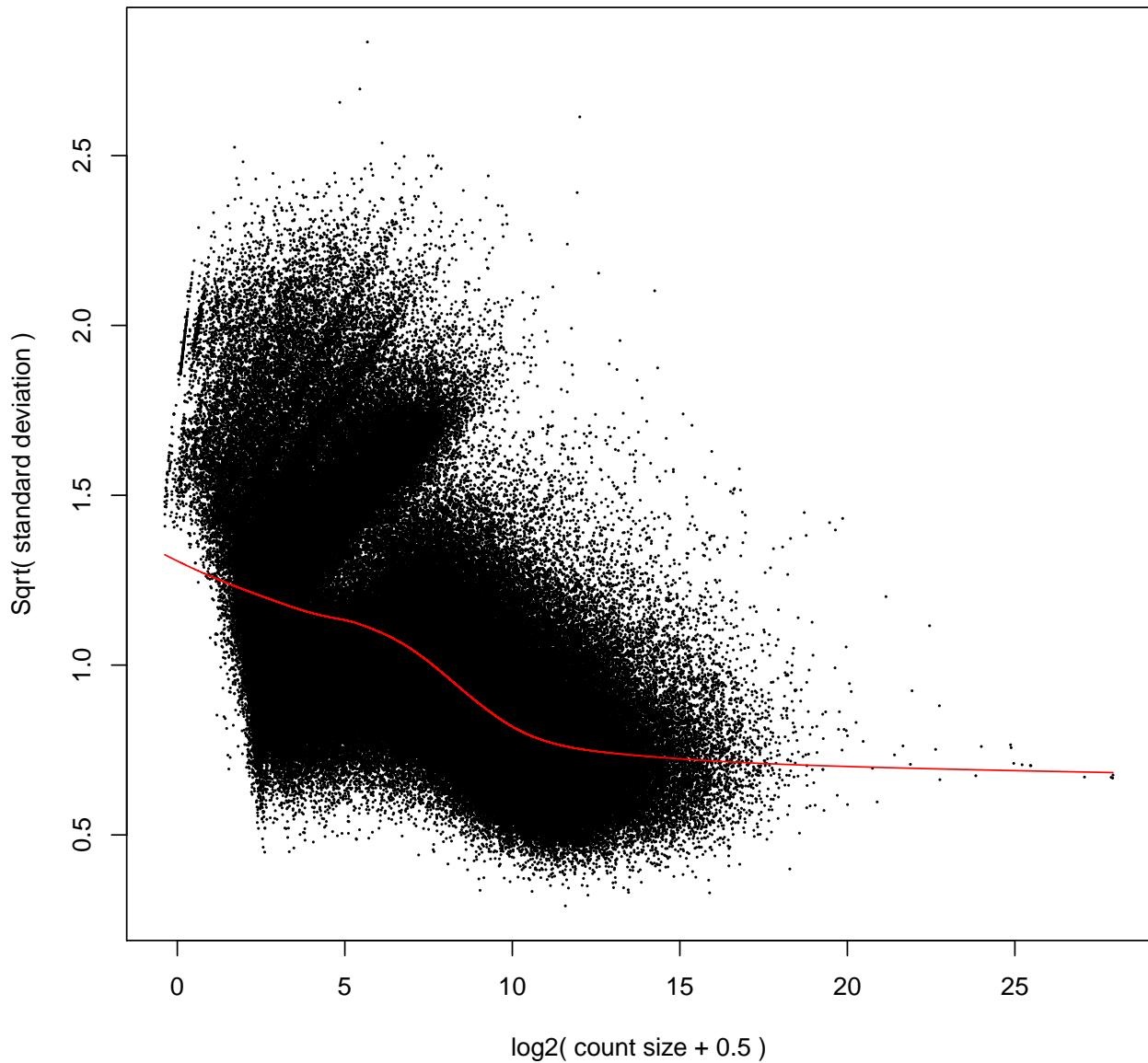
design <- model.matrix(~ group)
design

## (Intercept) groupTNBC Breast Tumor
## 1          1             1
## 2          1             1
## 3          1             1
## 4          1             1
## 5          1             1
## 6          1             0
## 7          1             0
## 8          1             0
## 9          1             0
## 10         1             0
## 11         1             1
## attr(,"assign")
## [1] 0 1
## attr(,"contrasts")
## attr(,"contrasts")$group
## [1] "contr.treatment"

dge <- DGEList(counts = counts)
dge <- calcNormFactors(dge)
v <- voom(dge, design, plot = TRUE)

```

voom: Mean–variance trend



```
fit <- lmFit(v, design)
fit <- eBayes(fit)
log2FC <- fit$coefficients[, 2]
p.mod <- fit$p.value[, 2]
q.mod <- qvalue(p.mod)$q
res.regions <- data.frame(log2FC, p.mod, q.mod)
sum(res.regions$q.mod < 0.05)

## [1] 35809

## Add differential expression information to GRanges object containing information
## about where the expressed regions are located in the genome (regions_counts)
regions_counts$log2FC <- res.regions$log2FC
regions_counts$p.mod <- res.regions$p.mod
```

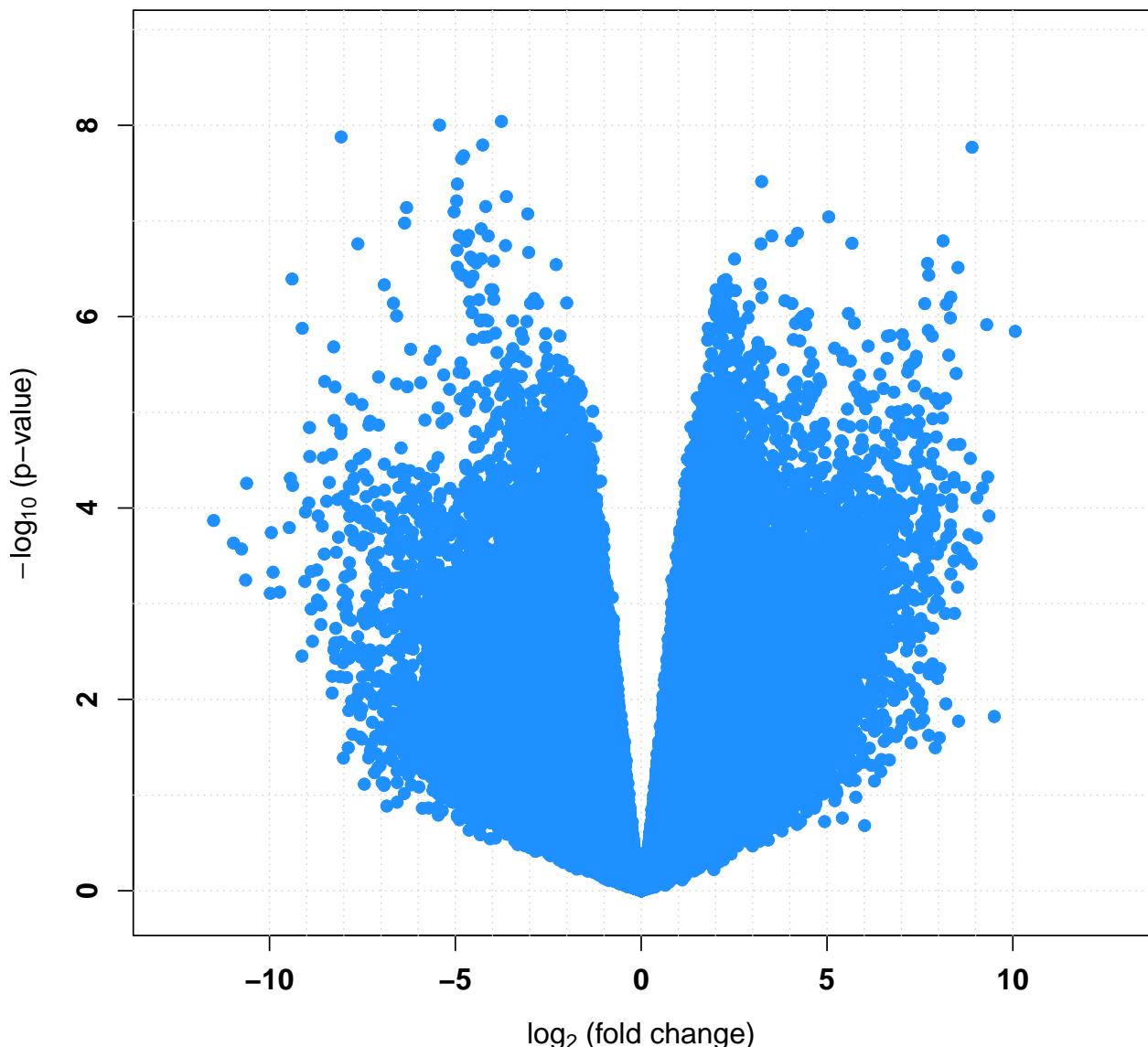
```

regions_counts$q.mod <- res.regions$q.mod

## Volcano plot
par(font.lab = 2, cex.lab = 1.2, font.axis = 2, cex.axis = 1.2)
rx2 <- c(-1, 1) * 1.1 * max(abs(log2FC))
ry2 <- c(-0.1, max(-log10(p.mod))) * 1.1
plot(log2FC, -log10(p.mod),
     pch = 19, xlim = rx2, ylim = ry2, col = trop[2],
     xlab = bquote(paste(log[2], ' (fold change)')),
     ylab = bquote(paste(-log[10], ' (p-value)')))
abline(v = seq(-10, 10, 1), col = 'lightgray', lty = 'dotted')
abline(h = seq(0, 23, 1), col = 'lightgray', lty = 'dotted')
points(log2FC, -log10(p.mod), pch = 19, col = trop[2])
title('Volcano plot: TNBC vs. HER2+ in SRP032789 (er level)')

```

Volcano plot: TNBC vs. HER2+ in SRP032789 (er level)



Annotate regions with genomic information

After detecting regions differentially expressed between TNBC and HER2+ samples, we can annotate where in the genome these regions are found and compare them back to the results from the gene- and exon-level differential expression analyses to determine what new information is gained by utilizing expressed region level data.

```
## Obtain significant (q<0.05) regions
regions_counts_sig <- regions_counts[regions_counts$q.mod < 0.05, ]

## Identify the top regions by highest total coverage
# top <- regions_counts_sig[order(regions_counts_sig$area, decreasing = TRUE)[1:100], ]

## Annotate significant regions
genes <- annotateTranscripts(TxDb.Hsapiens.UCSC.hg38.knownGene)

## No annotationPackage supplied. Trying org.Hs.eg.db.

## Getting TSS and TSE.

## Getting CSS and CSE.

## Getting exons.

## Annotating genes.

## Match detected regions with known gene annotation
## Warning: This can take some time
## (dependent upon how many regions are included for annotation)
# annotation_der <- matchGenes(regions_counts_sig, genes)

## Merge DER information with annotation information
# regions_annotated <- cbind(regions_counts_sig, annotation_der)
```

There are 35809 statistically significant regions detected ($q < 0.05$). We can use a venn diagram to visually represent where they are found in the genome.

```
## Get Annotation
if(!file.exists('genomicState.Hsapiens.BioMart.ENSEMBLMARTENSEMBL.GRCh38.p5.Rdata')) {
  ## Genomic state created by https://github.com/nellore/runs/blob/master/gtex/DER_analysis/coverageMatrix
  download('https://github.com/nellore/runs/blob/master/gtex/DER_analysis/coverageMatrix/genomicState.Rdata')
}

load('genomicState.Hsapiens.BioMart.ENSEMBLMARTENSEMBL.GRCh38.p5.Rdata')
gs_raw <- genomicState.Hsapiens.BioMart.ENSEMBLMARTENSEMBL.GRCh38.p5$fullGenome
gs <- renameSeqlevels(gs_raw, paste0('chr', seqlevels(gs_raw)))
gs_exons <- gs[gs$theRegion == "exon"]

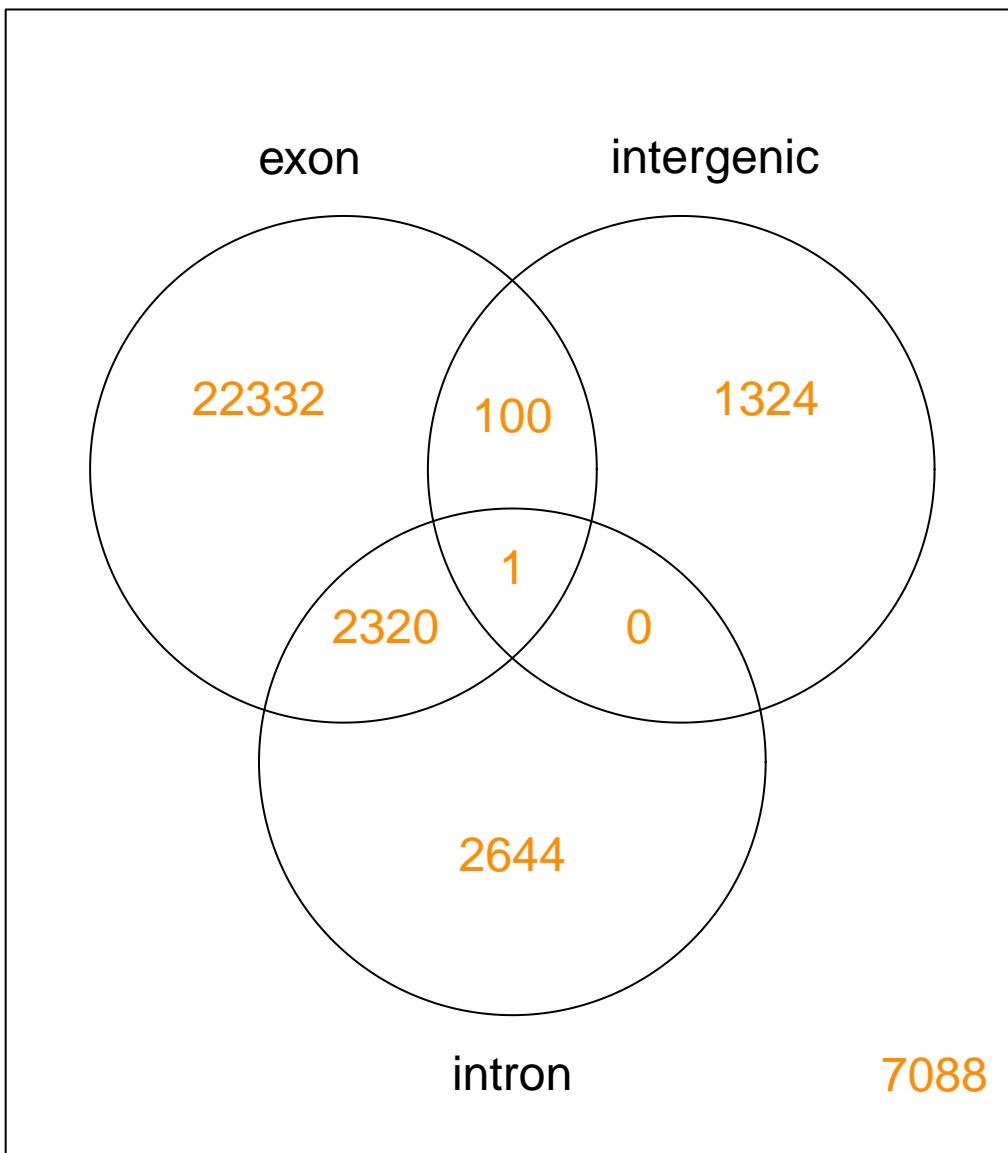
## Make venn diagram
## Venn Diagram for significant DERs (q<0.05)
annoRegs_sigDER <- annotateRegions(regions_counts_sig, gs, minoverlap = 20)
```

```

## 2016-07-25 23:07:18 annotateRegions: counting
## 2016-07-25 23:07:18 annotateRegions: annotating
#pdf(file = 'plots/VennDiagram_sigDERs.pdf')
vennRegions(annoRegs_sigDER, main = 'Significant DERs (q<0.05)', counts.col = trop[1])

```

Significant DERs (q<0.05)



```

## exon intergenic intron Counts
## 1      0        0      0    7088

```

```

## 2    0      0      1    2644
## 3    0      1      0    1324
## 4    0      1      1      0
## 5    1      0      0   22332
## 6    1      0      1   2320
## 7    1      1      0    100
## 8    1      1      1      1
## attr(,"class")
## [1] "VennCounts"

#dev.off()

```

Gene level analysis

In order to compare these differentially expressed regions with gene-level results, we must carry out differential expression in these same data at the gene level.

We first download the project of interest (SRP032798), obtaining expression data for the study of interest.

```

## Find the project of interest (SRP032789), e.g. with parts of the abstract
project_info <- abstract_search('To define the digital transcriptome of three breast cancer')

## Download the gene-level RangedSummarizedExperiment data
if(!file.exists(file.path('SRP032789', 'rse_gene.Rdata'))) {
  download_study(project_info$project)
}

## Load the data
load(file.path(project_info$project, 'rse_gene.Rdata'))
rse_gene

## class: RangedSummarizedExperiment
## dim: 23779 20
## metadata(0):
## assays(1): counts
## rownames(23779): 1 10 ... 9994 9997
## rowData names(2): gene_id bp_length
## colnames(20): SRR1027171 SRR1027173 ... SRR1027190 SRR1027172
## colData names(18): project sample ... avg_read_length bigwig_file

```

QC data

Downloaded count data are first scaled to take into account differing coverage between samples. Phenotype data (`pheno`) are obtained and ordered to match the sample order of the gene expression data (`rse_gene`). Only those samples that are HER2-positive or TNBC are included for analysis. Prior to differential gene expression analysis, count data are obtained in matrix format and then filtered to only include those genes with greater than five average normalized counts across all samples.

```

## Scale counts by taking into account the total coverage per sample
rse <- scale_counts(rse_gene)

## Download additional phenotype data from
## http://trace.ncbi.nlm.nih.gov/Traces/study/?acc=SRP032789
pheno <- read.table('SraRunTable_SRP032789.txt', sep = '\t',
  header = TRUE,
  stringsAsFactors = FALSE)

## Obtain correct order for pheno data
pheno <- pheno[match(rse$run, pheno$Run_s), ]
identical(pheno$Run_s, rse$run)

## [1] TRUE

head(cbind(pheno$Run_s, rse$run))

##      [,1]      [,2]
## [1,] "SRR1027171" "SRR1027171"
## [2,] "SRR1027173" "SRR1027173"
## [3,] "SRR1027174" "SRR1027174"
## [4,] "SRR1027175" "SRR1027175"
## [5,] "SRR1027176" "SRR1027176"
## [6,] "SRR1027177" "SRR1027177"

## Obtain grouping information
colData(rse)$group <- pheno$tumor_type_s
table(colData(rse)$group)

## 
## HER2 Positive Breast Tumor      Non-TNBC Breast Tumor
##                               5                      6
## Normal Breast Organoids       TNBC Breast Tumor
##                               3                      6

## Subset data to HER2 and TNBC types
rse <- rse[, rse$group %in% c('HER2 Positive Breast Tumor',
  'TNBC Breast Tumor')]
rse

## class: RangedSummarizedExperiment
## dim: 23779 11
## metadata(0):
## assays(1): counts
## rownames(23779): 1 10 ... 9994 9997
## rowData names(2): gene_id bp_length
## colnames(11): SRR1027171 SRR1027173 ... SRR1027187 SRR1027172
## colData names(19): project sample ... bigwig_file group

```

```

## Obtain count matrix
counts <- assays(rse)$counts

## Filter count matrix
filter <- apply(counts, 1, function(x) mean(x) > 5)
counts <- counts[filter, ]
dim(counts)

## [1] 17874      11

counts_genes <- counts

```

DE analysis : gene-level

Using limma and voom, differentially expressed genes are detected at the gene level.

```

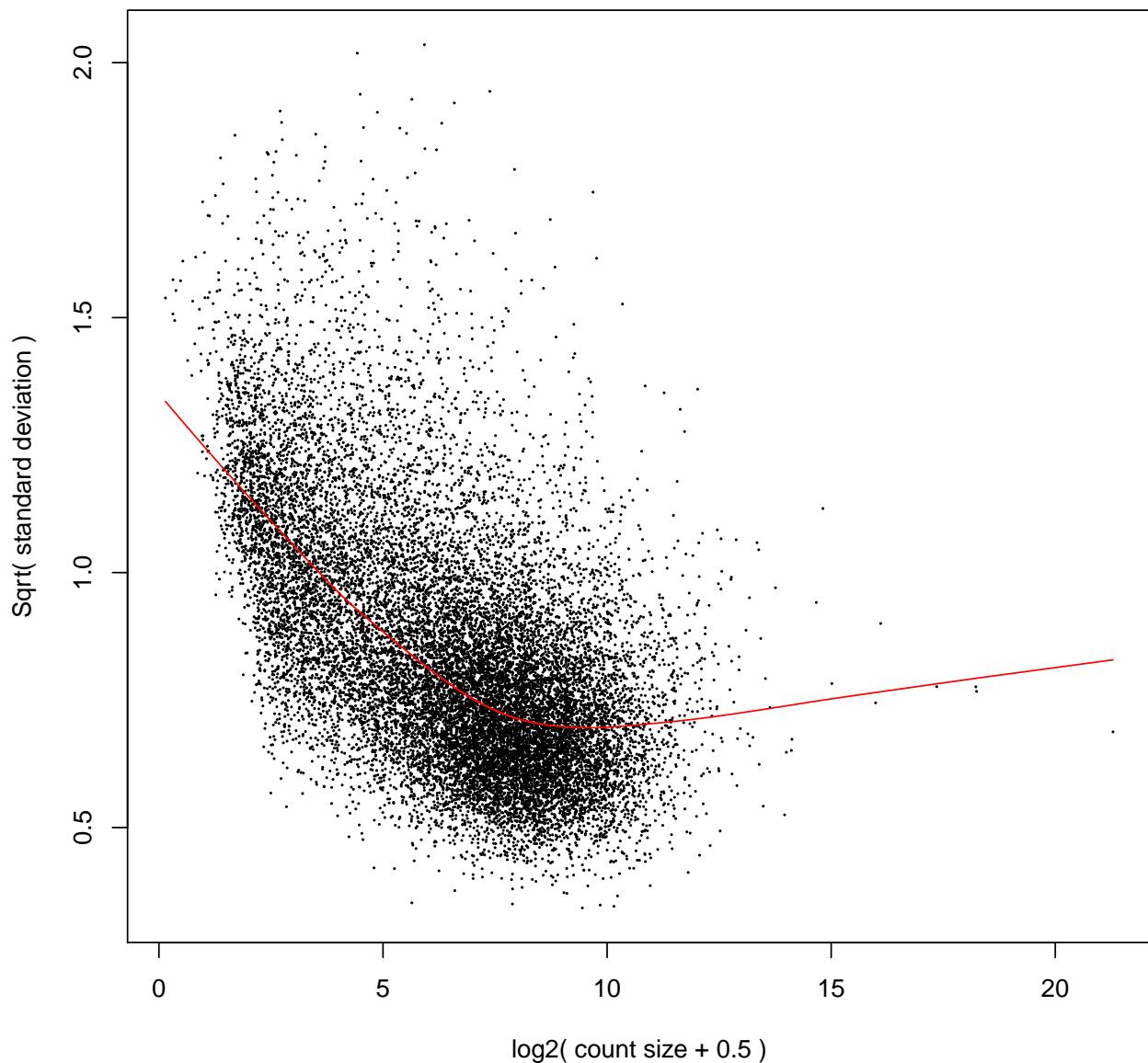
## Perform differential expression analysis with limma-voom
design <- model.matrix(~ rse$group)
design

##      (Intercept) rse$groupTNBC Breast Tumor
## 1              1                      1
## 2              1                      1
## 3              1                      1
## 4              1                      1
## 5              1                      1
## 6              1                      0
## 7              1                      0
## 8              1                      0
## 9              1                      0
## 10             1                      0
## 11             1                      1
## attr(),"assign")
## [1] 0 1
## attr(),"contrasts")
## attr(),"contrasts")$`rse$group`
## [1] "contr.treatment"

dge <- DGEList(counts = counts)
dge <- calcNormFactors(dge)
v <- voom(dge, design, plot = TRUE)

```

voom: Mean–variance trend



```
fit <- lmFit(v, design)
fit <- eBayes(fit)
log2FC <- fit$coefficients[, 2]
p.mod <- fit$p.value[, 2]
q.mod <- qvalue(p.mod)$q
res.genes <- data.frame(log2FC, p.mod, q.mod)
rownames(res.genes) <- rownames(counts)
sum(res.genes$q.mod < 0.05)
```

```
## [1] 1611
```

Compare DER findings to gene-level analysis

By avoiding using gene or exon-level annotation for quantification, regions usually excluded from analysis can be included. Here, the same data set is utilized for differential expression; however, we determine the number of significantly differentially expressed regions (DERs, $q < 0.05$) that do not overlap any gene. These regions are places within the genome that would be completely missed by the common gene-level analyses.

```
## Determine DERs not detected by gene level analysis
## Get DEGs into GRanges object format
genes_GRanges <- SummarizedExperiment::rowRanges(rse_gene$filter, [])
genes_GRanges$log2FC <- res.genes$log2FC
genes_GRanges$p.mod <- res.genes$p.mod
genes_GRanges$q.mod <- res.genes$q.mod

## For each region, determine if region overlaps with a gene included in the gene-level analysis
overlap <- countOverlaps(regions_counts_sig, genes_GRanges, type="any")
overlap <- countOverlaps(regions_counts, genes_GRanges, type="any")

regions_counts$overlap <- overlap
regions_counts_ingroup <- regions_counts[overlap>0]
regions_counts_notingroup <- regions_counts[overlap==0]

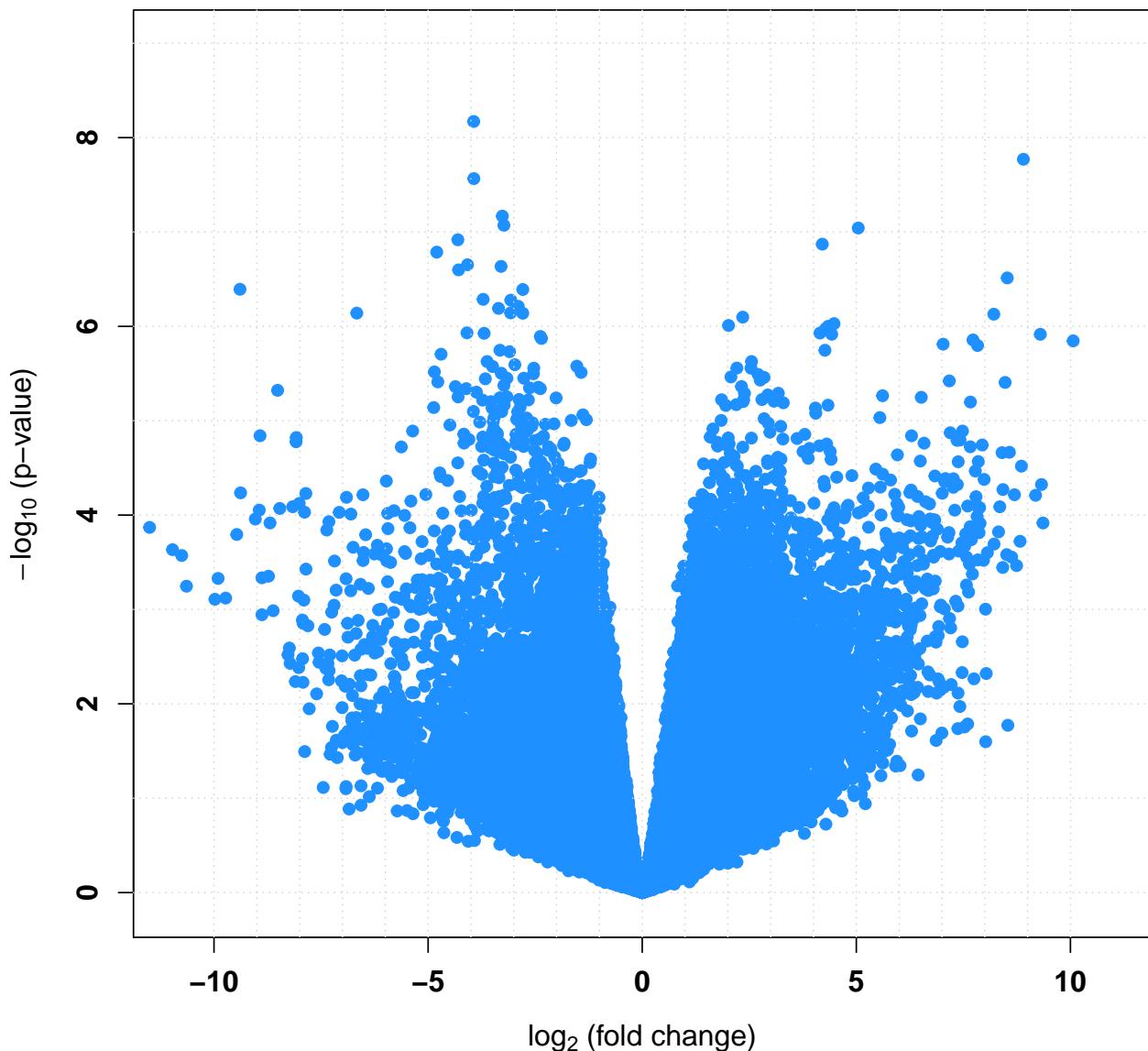
## For significant DERs ( $q < 0.05$ ), how many sites do not fall within a gene
## (and thus would not be detected by gene-level analyses

## number of DERs that do not overlap with genes in gene-level analysis
length(regions_counts_notingroup[regions_counts_notingroup$q.mod<0.05])

## [1] 4462

## Volcano plot of DERs (that do not overlap with genes)
par(font.lab = 2, cex.lab = 1.2, font.axis = 2, cex.axis = 1.2)
rx2 <- c(-11, 11)
ry2 <- c(-0.1, max(-log10(p.mod))) * 1.1
plot(log2FC, -log10(p.mod),
     pch = 19, xlim = rx2, ylim = ry2, col = trop[2],
     xlab = bquote(paste(log[2], ' (fold change)')),
     ylab = bquote(paste(-log[10], ' (p-value)')))
abline(v = seq(-10, 10, 1), col = 'lightgray', lty = 'dotted')
abline(h = seq(0, 23, 1), col = 'lightgray', lty = 'dotted')
points(regions_counts_notingroup$log2FC, -log10(regions_counts_notingroup$p.mod), pch = 19, col = trop[2])
title('Volcano plot: TNBC vs. HER2+ in SRP032789 (DERs not in genes)')
```

Volcano plot: TNBC vs. HER2+ in SRP032789 (DERs not in genes)



There are 4462 DERs that do not overlap with genes included in the gene-level analysis.

These range in size from 1 bps to 15879 bps in length, with a median length of 39 bps.

Compare DERs to known genomic annotation

Beyond an interest in those regions not detected in the gene-level analysis, we are also interested in those regions that do not overlap any annotated gene. First, we quantify where DERs are found within the genome that do not overlap any known exons. Then, we highlight single locus coverage plots for a few of the longest DERs.

```
## For each region, determine if region overlaps with a gene included
overlapANY <-countOverlaps(regions_counts,gs_exons,type="any")
```

```

regions_counts$overlapANY <- overlapANY
regions_counts_ingroupANY <- regions_counts[regions_counts$overlapANY>0]
regions_counts_notinANYgene <- regions_counts[regions_counts$overlapANY==0]

## Number of DERs that do not overlap with any genes
length(regions_counts_notinANYgene[regions_counts_notinANYgene$q.mod<0.05])

## [1] 6613

## Get top 1% longest regions outside of genes to visualize
regions = subset(regions_counts_notinANYgene, width(regions_counts_notinANYgene) > quantile(width(regions_counts_notinANYgene), 0.99))
#order so most significant are plotted first
regions <- regions[order(regions$p.mod), ]

## make region plots
tIndexes = split(1:length(rse$group), rse$group)

## Get required information for the plots
## Annotate regions
annoRegs <- annotateRegions(regions, gs, minoverlap = 1)

## 2016-07-25 23:07:28 annotateRegions: counting

## 2016-07-25 23:07:29 annotateRegions: annotating

## Load full coverage
SRP032789_urls <- subset(recount_url, project == 'SRP032789' & file_name != 'mean_SRP032789.bw')
##### Leo! change recount_url$path to recount_url$url for submission #####
files <- SRP032789_urls$url[grep('bw$', SRP032789_urls$url)]
names(files) <- gsub('.bw', '', SRP032789_urls$file_name[grep('bw$', SRP032789_urls$url)])

## Only include TNBC and HER2+ samples
files <- subset(files, names(files) %in% colnames(covMat_filt))

## Obtain fullCoverage matrix for all chromosomes
fullCov <- fullCoverage(files = files, chrs = paste0('chr', c(1:22, 'X', 'Y')), which = regions, verbose = FALSE)

## Find nearest annotation with bumphunter::matchGenes()
nearestAnnotation <- matchGenes(x = regions, subject = genes)

## .....

## Get the region coverage
geneRegionCov <- getRegionCoverage(fullCov=fullCov, regions=regions,
  targetSize = 4e+07, totalMapped=colData(rse)$reads_aligned, verbose = FALSE)
geneRegionCovMeans = lapply(geneRegionCov, function(x) {
  cat(".")
  sapply(tIndexes, function(ii) rowMeans(x[,ii]))
})

```

```

## .....  

## Plot region coverage for the longest DERs
pdf('DERs_notingene.pdf', h = 5, w = 7)
plotRegionCoverage(regions=regions,
  regionCoverage=geneRegionCovMeans,
  groupInfo= factor(unique(rse$group)), colors = trop,
  nearestAnnotation=nearestAnnotation,
  annotatedRegions=annoRegs,
  ask=FALSE, whichRegions=1:10, verbose=FALSE,
  txdb = TxDb.Hsapiens.UCSC.hg38.knownGene)
dev.off()  

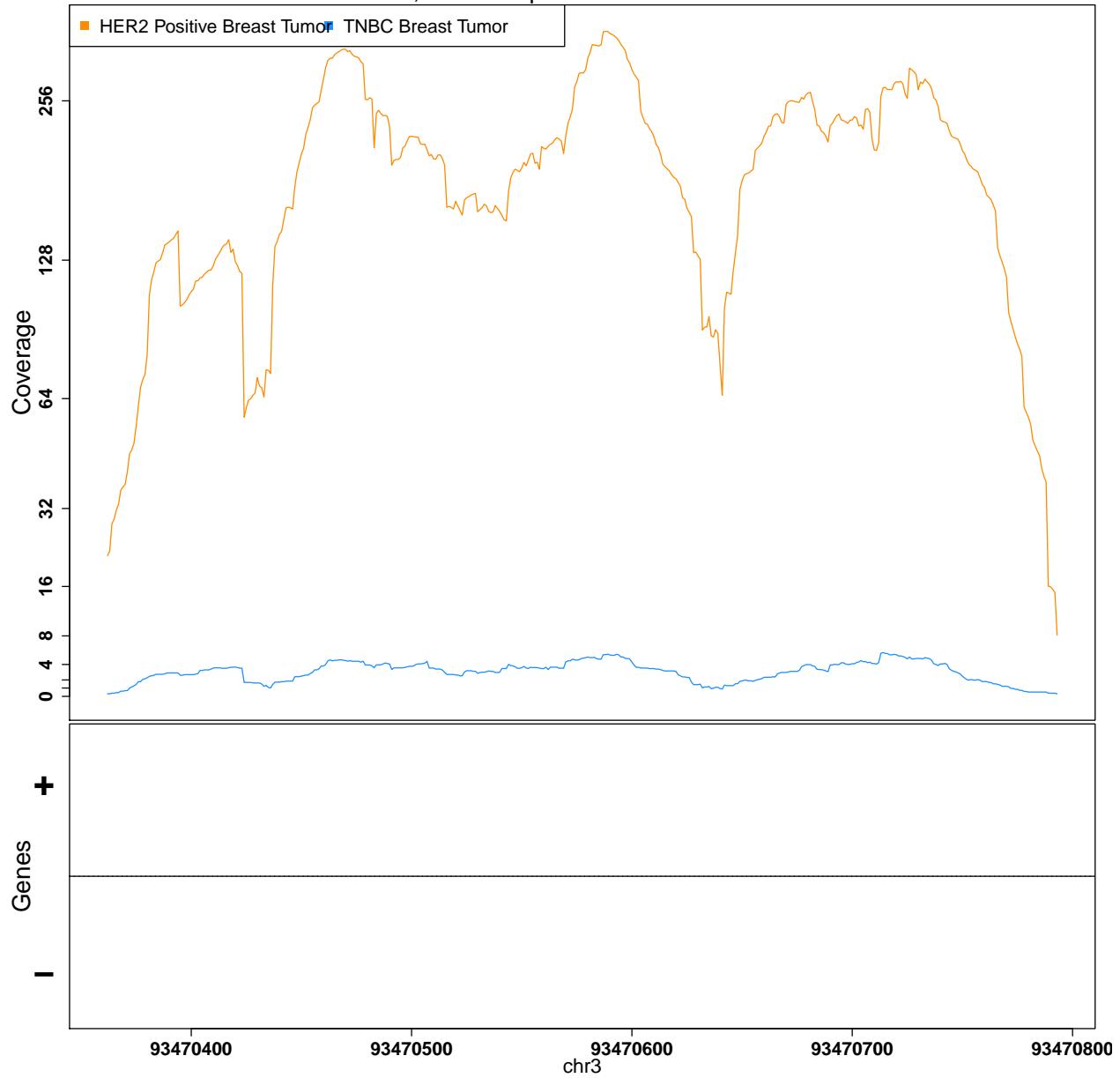
## RStudioGD
##      2  

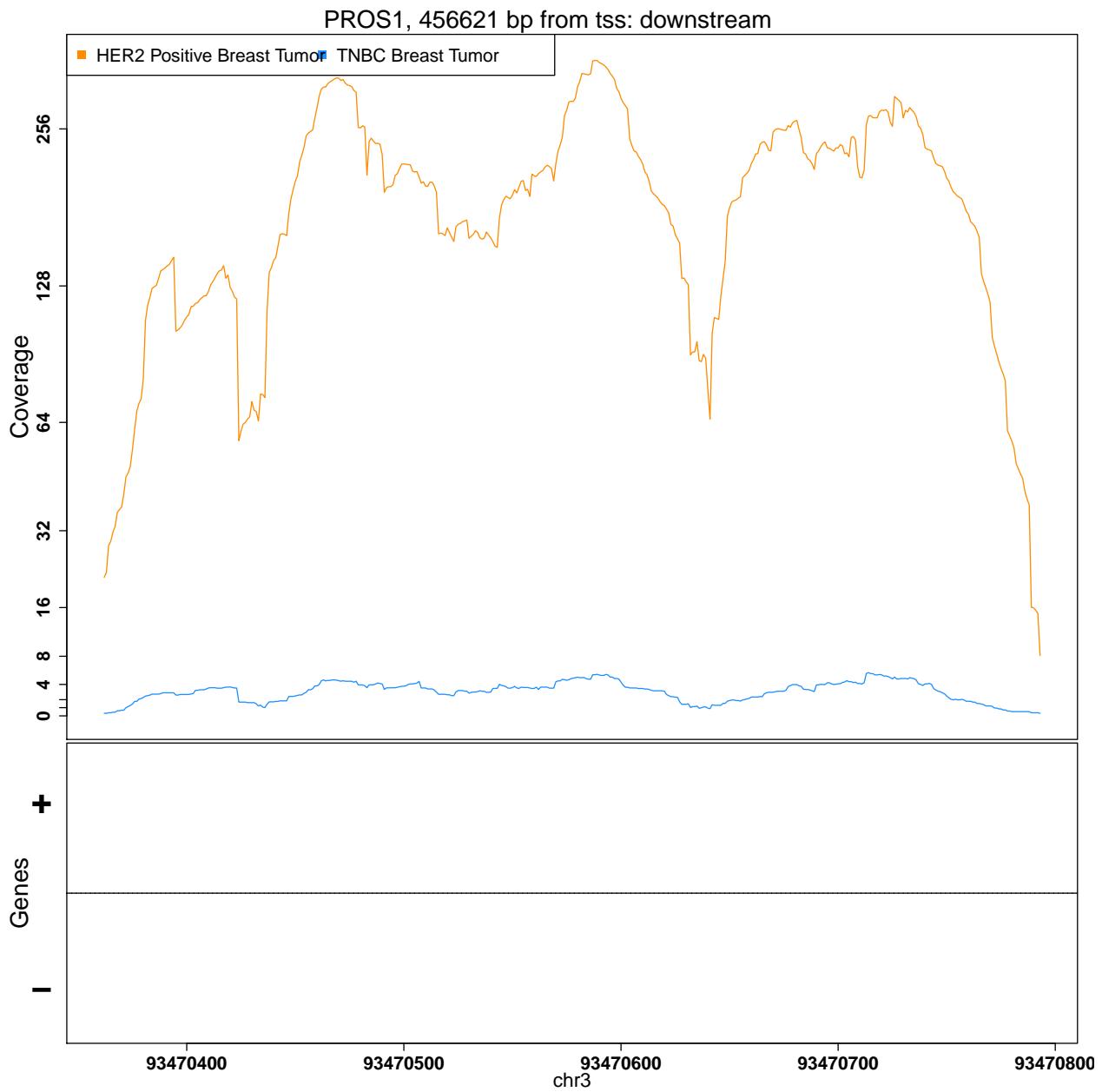
  

## Example of DER outside of any exonic region
plotRegionCoverage(regions=regions,
  regionCoverage=geneRegionCovMeans,
  groupInfo= factor(unique(rse$group)), colors = trop,
  nearestAnnotation=nearestAnnotation,
  annotatedRegions=annoRegs,
  ask=FALSE, whichRegions=6, verbose=FALSE,
  txdb = TxDb.Hsapiens.UCSC.hg38.knownGene)

```

PROS1, 456621 bp from tss: downstream





There are 6613 DERs that do not overlap with any known genes.

These range in size from 1 bps to 15879 bps in length, with a median length of 22 bps.

Compare DERs to known genomic annotation

Finally, to compare the DER analysis to gene-level results, we must assign each DER to its nearest gene, creating the output file: “AnnotatedDERs.Rdata”.

```
## Assign regions to annotate with their genomic position
regions = regions_counts

## Obtain annotation information
ensemblAnno = annotateRegions(regions,gs)
```

```

## 2016-07-25 23:13:17 annotateRegions: counting

## 2016-07-25 23:13:19 annotateRegions: annotating

countTable = ensemblAnno$countTable

## Assign overlap between regions and genomic annotation
dA = distanceToNearest(regions_counts, genes)
regions$nearestSymbol = genes$Entrez[subjectHits(dA)]
regions$distToGene = mcols(dA)$distance
mcols(regions) = cbind(mcols(regions), countTable)

## Only include regions that are within genes to compare to gene-level findings
regions_exons <- regions[regions$exon==1]
regions_exons_distToGene <- regions_exons[regions_exons$distToGene==0]

## Extract p-value and nearest gene information to be used in comparison to gene-level analysis
annotatedDERs <- regions_exons_distToGene$p.mod
names(annotatedDERs) <- regions_exons_distToGene$nearestSymbol

## Save file for use in comparison analysis
save(annotatedDERs, file="AnnotatedDERs.Rdata")

```

Reproducibility

This analysis report was made possible thanks to:

- R (Lawrence, Huber, Pagès, Aboyoun, et al., 2013)
- *BiocParallel* (R Core Team, 2016)
- *BiocStyle* (Alexa and Rahnenfuhrer, 2016)
- *derfinder* (Allaire, Cheng, Xie, McPherson, et al., 2016)
- *devtools* (Bengtsson, 2016)
- *downloader* (Boettiger, 2016)
- *edgeR* (Chang, 2015)
- *GenomicRanges* (Collado-Torres, Nellore, Frazee, Wilks, et al., 2016)
- *knitcitations* (Collado-Torres, Nellore, and Leek, 2016)
- *matrixStats* (Frazee, 2016)
- *recount* (Morgan, Obenchain, Lang, and Thompson, 2016)
- *rmarkdown* (Morgan, Obenchain, Hester, and Pagès, 2016)
- *RSkittleBrewer* (Oleś, Morgan, and Huber, 2016)
- *SummarizedExperiment* (Robinson, McCarthy, and Smyth, 2010)
- *topGO* (Law, Chen, Shi, and Smyth, 2014)
- *limma* (Wickham and Chang, 2016)

Bibliography file

- [1] A. Alexa and J. Rahnenfuhrer. topGO: Enrichment Analysis for Gene Ontology. R package version 2.25.0. 2016.
- [2] J. Allaire, J. Cheng, Y. Xie, J. McPherson, et al. rmarkdown: Dynamic Documents for R. R package version 1.0. 2016. URL: <https://CRAN.R-project.org/package=rmarkdown>.

- [3] H. Bengtsson. *matrixStats*: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.50.2. 2016. URL: <https://CRAN.R-project.org/package=matrixStats>.
- [4] C. Boettiger. *knitcitations*: Citations for ‘Knitr’ Markdown Files. R package version 1.0.7.1. 2016. URL: <https://github.com/cboettig/knitcitations>.
- [5] W. Chang. *downloader*: Download Files over HTTP and HTTPS. R package version 0.4. 2015. URL: <https://CRAN.R-project.org/package=downloader>.
- [6] L. Collado-Torres, A. Nellore, A. C. Frazee, C. Wilks, et al. “Flexible expressed region analysis for RNA-seq with derfinder”. In: *bioRxiv* (2016). DOI: 10.1101/015370. URL: <http://biorkxiv.org/content/early/2016/05/07/015370>.
- [7] L. Collado-Torres, A. Nellore and J. T. Leek. *recount*: Explore and download data from the recount project. R package version 0.99.0. 2016. URL: <https://github.com/leekgroup/recount>.
- [8] A. Frazee. *RSkittleBrewer*: Fun with R Colors. R package version 1.1. 2016. URL: <https://github.com/alyssafrazee/RSkittleBrewer>.
- [9] C. Law, Y. Chen, W. Shi and G. Smyth. “Voom: precision weights unlock linear model analysis tools for RNA-seq read counts”. In: *Genome Biology* 15 (2014), p. R29.
- [10] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, et al. “Software for Computing and Annotating Genomic Ranges”. In: *PLoS Computational Biology* 9 (8 2013). DOI: 10.1371/journal.pcbi.1003118. URL: <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- [11] M. Morgan, V. Obenchain, J. Hester and H. Pagès. *SummarizedExperiment*: SummarizedExperiment container. R package version 1.3.7. 2016.
- [12] M. Morgan, V. Obenchain, M. Lang and R. Thompson. *BiocParallel*: Bioconductor facilities for parallel evaluation. R package version 1.7.5. 2016. URL: <https://github.com/Bioconductor/BiocParallel>.
- [13] A. Oleś, M. Morgan and W. Huber. *BiocStyle*: Standard styles for vignettes and other Bioconductor documents. R package version 2.1.19. 2016. URL: <https://github.com/Bioconductor/BiocStyle>.
- [14] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2016. URL: <https://www.R-project.org/>.
- [15] M. D. Robinson, D. J. McCarthy and G. K. Smyth. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26 (2010), pp. -1.
- [16] H. Wickham and W. Chang. *devtools*: Tools to Make Developing R Packages Easier. R package version 1.12.0. 2016. URL: <https://CRAN.R-project.org/package=devtools>.

```
## Time spent creating this report:
diff(c(timestart, Sys.time()))
```

```
## Time difference of 8.627276 mins
```

```
## Date this report was generated
message(Sys.time())
```

```
## 2016-07-25 23:13:21
```

```
## Reproducibility info
options(width = 120)
devtools::session_info()
```

```
## Session info -----
```

```

## setting value
## version R version 3.3.0 (2016-05-03)
## system x86_64, darwin13.4.0
## ui      RStudio (0.99.902)
## language (EN)
## collate en_US.UTF-8
## tz      America/New_York
## date   2016-07-25

## Packages ----

## package          * version  date     source
## acepack           1.3-3.3  2013-05-03 CRAN (R 3.1.0)
## AnnotationDbi    * 1.35.4   2016-07-08 Bioconductor
## AnnotationHub    2.5.5    2016-07-17 Bioconductor
## bibtex            0.4.0    2014-12-31 CRAN (R 3.3.0)
## Biobase           * 2.33.0   2016-05-27 Bioconductor
## BiocGenerics     * 0.19.2   2016-07-08 Bioconductor
## BiocInstaller    1.23.6   2016-07-12 Bioconductor
## BiocParallel      * 1.7.5    2016-07-21 Bioconductor
## BiocStyle          * 2.1.19   2016-07-11 Bioconductor
## biomaRt            2.29.2   2016-05-30 Bioconductor
## Biostrings         2.41.4   2016-06-17 Bioconductor
## biovizBase         1.21.0   2016-05-27 Bioconductor
## bitops              1.0-6    2013-08-17 CRAN (R 3.3.0)
## BSgenome            1.41.2   2016-06-17 Bioconductor
## bumphunter         * 1.13.1   2016-07-11 Bioconductor
## chron               2.3-47   2015-06-24 CRAN (R 3.2.0)
## cluster              2.0.4    2016-04-18 CRAN (R 3.3.0)
## codetools            0.2-14   2015-07-15 CRAN (R 3.3.0)
## colorspace            1.2-6    2015-03-11 CRAN (R 3.1.3)
## data.table            1.9.6    2015-09-19 CRAN (R 3.2.0)
## DBI                  0.4-1    2016-05-08 CRAN (R 3.2.5)
## derfinder             * 1.7.9    2016-07-11 Bioconductor
## derfinderHelper       1.7.3    2016-05-27 Bioconductor
## derfinderPlot          * 1.7.6    2016-05-27 Bioconductor
## devtools              1.12.0   2016-06-24 CRAN (R 3.3.0)
## dichromat             2.0-0    2013-01-24 CRAN (R 3.0.2)
## digest                0.6.9    2016-01-08 CRAN (R 3.3.0)
## doRNG                 1.6      2014-03-07 CRAN (R 3.3.0)
## downloader             * 0.4     2015-07-09 CRAN (R 3.2.0)
## edgeR                  * 3.15.2   2016-07-08 Bioconductor
## ensemblDb              1.5.9    2016-07-08 Bioconductor
## evaluate                0.9     2016-04-29 CRAN (R 3.3.0)
## foreach                 * 1.4.3    2015-10-13 CRAN (R 3.2.0)
## foreign                 0.8-66   2015-08-19 CRAN (R 3.3.0)
## formatR                  1.4     2016-05-09 CRAN (R 3.3.0)
## Formula                 1.2-1    2015-04-07 CRAN (R 3.1.3)
## GenomeInfoDb             * 1.9.4    2016-07-14 Bioconductor
## GenomicAlignments        1.9.6    2016-07-17 Bioconductor
## GenomicFeatures          * 1.25.16   2016-07-24 Bioconductor
## GenomicFiles              1.9.11   2016-06-03 Bioconductor
## GenomicRanges             * 1.25.9    2016-06-26 Bioconductor
## GGally                  1.2.0    2016-07-01 CRAN (R 3.3.0)

```

```

##  ggbio           1.21.2   2016-05-27 Bioconductor
##  ggplot2          2.1.0    2016-03-01 CRAN (R 3.2.4)
##  GO.db            * 3.3.0    2016-05-25 Bioconductor
##  graph             1.51.0   2016-05-27 Bioconductor
##  gridExtra         2.2.1    2016-02-29 CRAN (R 3.2.4)
##  gtable            0.2.0    2016-02-26 CRAN (R 3.2.3)
##  Hmisc              3.17-4   2016-05-02 CRAN (R 3.2.5)
##  htmltools          0.3.5    2016-03-21 CRAN (R 3.3.0)
##  httpuv             1.3.3    2015-08-04 CRAN (R 3.2.0)
##  httr               1.2.1    2016-07-03 CRAN (R 3.3.0)
##  interactiveDisplayBase 1.11.3   2016-05-27 Bioconductor
##  IRanges            * 2.7.12   2016-07-24 Bioconductor
##  iterators          * 1.0.8    2015-10-13 CRAN (R 3.2.0)
##  knitrCitations     * 1.0.7.1  2016-06-06 Github (cboettig/knitcitations@76c128a)
##  knitr               1.13     2016-05-09 CRAN (R 3.3.0)
##  lattice              0.20-33  2015-07-14 CRAN (R 3.3.0)
##  latticeExtra        0.6-28   2016-02-09 CRAN (R 3.2.3)
##  limma               * 3.29.17  2016-07-24 Bioconductor
##  locfit              * 1.5-9.1  2013-04-20 CRAN (R 3.1.0)
##  lubridate            1.5.6    2016-04-06 CRAN (R 3.3.0)
##  magrittr             1.5      2014-11-22 CRAN (R 3.3.0)
##  Matrix               1.2-6    2016-05-02 CRAN (R 3.3.0)
##  matrixStats          * 0.50.2   2016-04-24 CRAN (R 3.2.5)
##  memoise              1.0.0    2016-01-29 CRAN (R 3.2.3)
##  mime                 0.5      2016-07-07 CRAN (R 3.3.0)
##  munsell              0.4.3    2016-02-13 CRAN (R 3.2.3)
##  nnet                  7.3-12   2016-02-02 CRAN (R 3.3.0)
##  org.Hs.eg.db          * 3.3.0    2016-05-25 Bioconductor
##  OrganismDbi          1.15.1   2016-05-27 Bioconductor
##  pkgmaker              0.22     2014-05-14 CRAN (R 3.3.0)
##  plyr                  1.8.4    2016-06-08 CRAN (R 3.3.0)
##  qvalue                * 2.5.2    2016-05-27 Bioconductor
##  R6                     2.1.2    2016-01-26 CRAN (R 3.2.3)
##  RBGL                  1.49.1   2016-05-27 Bioconductor
##  RColorBrewer          1.1-2    2014-12-07 CRAN (R 3.1.2)
##  Rcpp                   0.12.6   2016-07-19 CRAN (R 3.3.0)
##  RCurl                  1.95-4.8 2016-03-01 CRAN (R 3.2.4)
##  recount                * 0.99.0   2016-06-06 Github (leekgroup/recount@bfa4681)
##  RefManageR             0.10.13   2016-04-04 CRAN (R 3.3.0)
##  registry               0.3      2015-07-08 CRAN (R 3.3.0)
##  reshape                 0.8.5    2014-04-23 CRAN (R 3.1.0)
##  reshape2                1.4.1    2014-12-06 CRAN (R 3.1.2)
##  RJSONIO                 1.3-0    2014-07-28 CRAN (R 3.2.0)
##  rmarkdown                * 1.0     2016-07-08 CRAN (R 3.3.0)
##  rngtools                 1.2.4    2014-03-06 CRAN (R 3.3.0)
##  rpart                   4.1-10   2015-06-29 CRAN (R 3.3.0)
##  Rsamtools                1.25.0   2016-05-27 Bioconductor
##  RSkittleBrewer          * 1.1     2016-06-07 Github (alyssafrazee/RSkittleBrewer@230d1d0)
##  RSQLite                  1.0.0    2014-10-25 CRAN (R 3.1.2)
##  rstudioapi                0.6      2016-06-27 CRAN (R 3.3.0)
##  rtracklayer                1.33.11  2016-07-24 Bioconductor
##  S4Vectors                 * 0.11.10  2016-07-21 Bioconductor
##  scales                   0.4.0    2016-02-26 CRAN (R 3.2.3)
##  shiny                     0.13.2   2016-03-28 CRAN (R 3.2.4)

```

```
## SparseM * 1.7      2015-08-15 CRAN (R 3.2.0)
## stringi      1.1.1    2016-05-27 CRAN (R 3.3.0)
## stringr      1.0.0    2015-04-30 CRAN (R 3.3.0)
## SummarizedExperiment * 1.3.7    2016-07-08 Bioconductor
## survival     2.39-5   2016-06-26 CRAN (R 3.3.0)
## topGO        2.25.0   2016-05-27 Bioconductor
## TxDb.Hsapiens.UCSC.hg38.knownGene * 3.1.3    2016-06-20 Bioconductor
## VariantAnnotation 1.19.8   2016-07-12 Bioconductor
## withr         1.0.2    2016-06-20 CRAN (R 3.3.0)
## XML           3.98-1.4  2016-03-01 CRAN (R 3.2.4)
## xtable        1.8-2    2016-02-05 CRAN (R 3.2.3)
## XVector       0.13.7   2016-07-24 Bioconductor
## yaml          2.1.13   2014-06-12 CRAN (R 3.3.0)
## zlibbioc      1.19.0   2016-05-27 Bioconductor
```