

#6.2

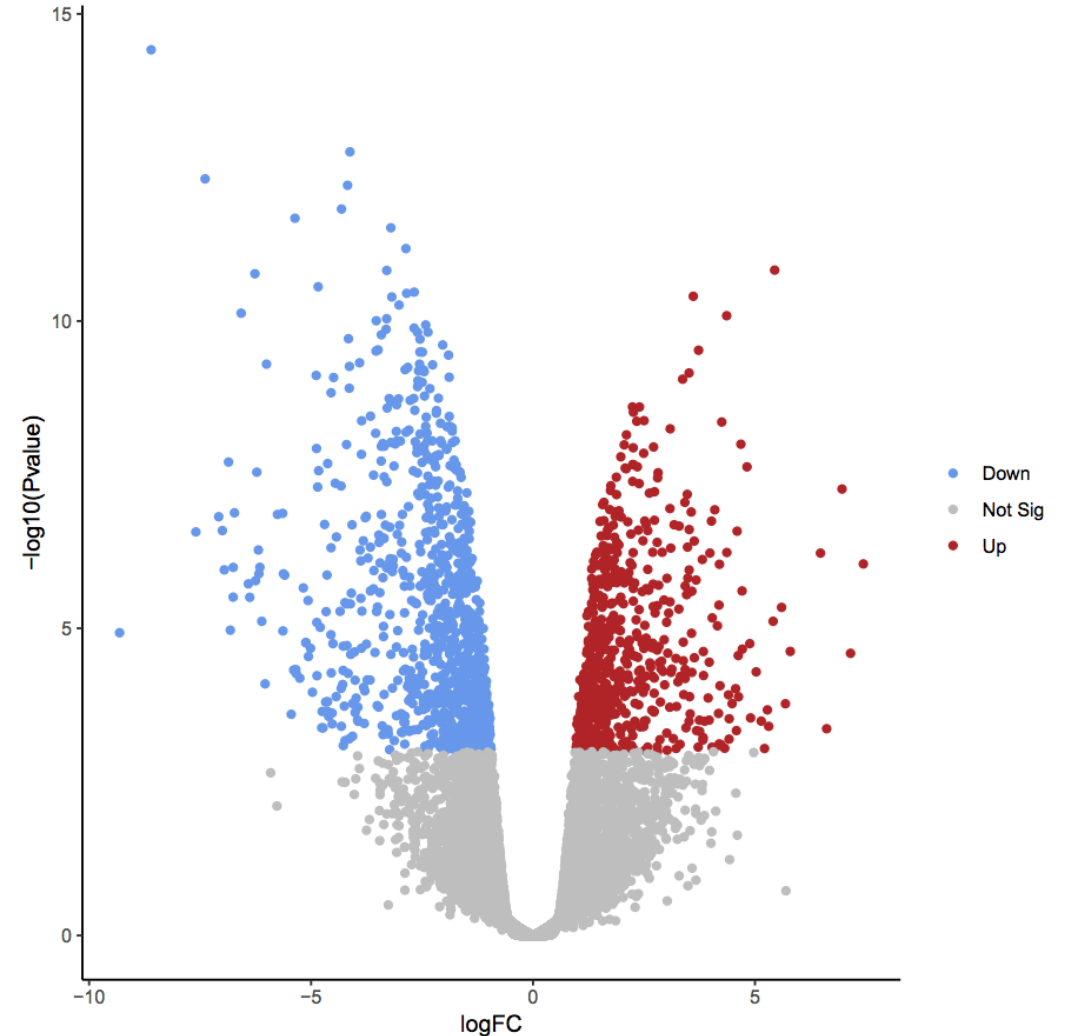
Volcano Plots

Kaplan Meier Curves

Density Curves

Volcano Plots

- Useful, convenient way to visualize magnitude and significance at the same time
- Statistical values should be platform-appropriate
 - Remember: RNA-seq is not normally distributed!



Volcano Plots

- 1. Take RNA-seq differential expression data from DESeq2
- 2. Download the **EnhancedVolcano** package from BioConductor
`BiocManager::install('EnhancedVolcano')`

```
> library(DESeq2)
> count_table <- read.table(
  "C:\\Users\\jorda\\Desktop\\projects\\bioinformatics-bootcamp\\class_6_2\\sce_mct1_03hr_counts.txt",
  sep = '\\t',
  header = TRUE,
  row.names = 1,
  check.names=F)

> sample_table <- read.table(
  text = readLines(
    "C:\\Users\\jorda\\Desktop\\projects\\bioinformatics-bootcamp\\class_6_2\\sce_mct1_03hr_metadata.txt",
    warn = FALSE),
  header = TRUE,
  sep = '\\t')

> dds <- DESeqDataSetFromMatrix(
> countData = count_table,
> colData = sample_table,
> design = ~Genotype)
> dds <- DESeq(dds)
> res <- results(dds)
> resOrdered <- res[order(res$padj),]
```

Volcano Plots

```
> library(EnhancedVolcano)
> EnhancedVolcano(
  resOrdered,
  lab = rownames(resOrdered),
  x = 'log2FoldChange',
  y = 'padj',
  title = 'mct1 deletion versus WT',
  pCutoff = 10e-10,
  FScutoff = 1)
```

Some notes:

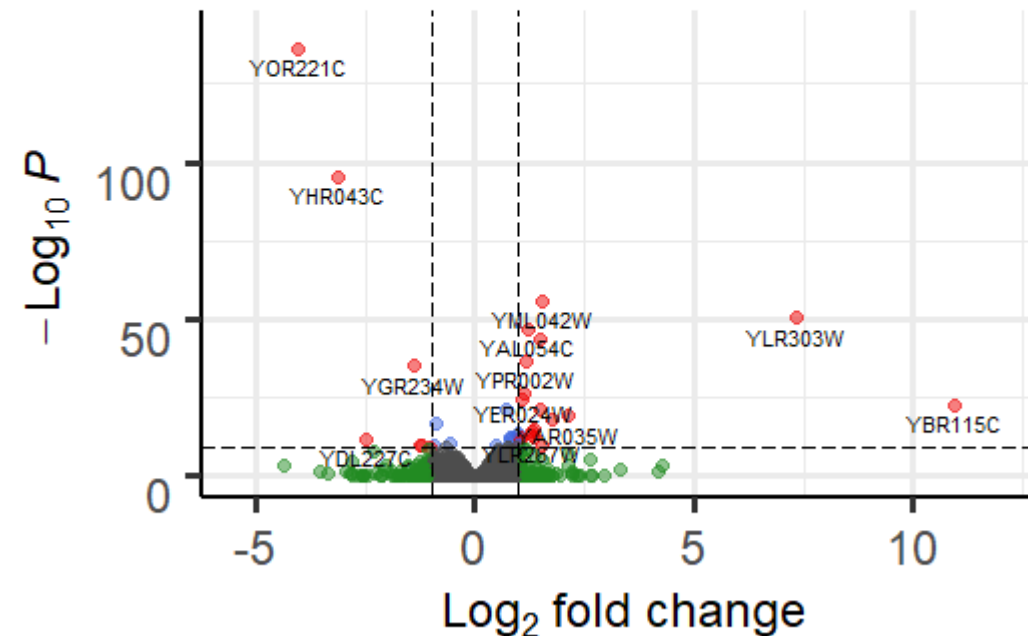
- Use gene names
- Label axes correctly (“FDR” or “p-adj”, not just “p”)
- Check the documentation for more information:

<https://github.com/kevinblighe/EnhancedVolcano>

mct1 deletion versus WT

EnhancedVolcano

● NS ● Log₂ FC ● p-value ● p – value and log




Total = 7127 variables

Kaplain-Meier Curves

- Estimate survival as a function of a condition
 - Is a drug having a significant effect on outcome?
 - Does a genotype or phenotype improve outcome?

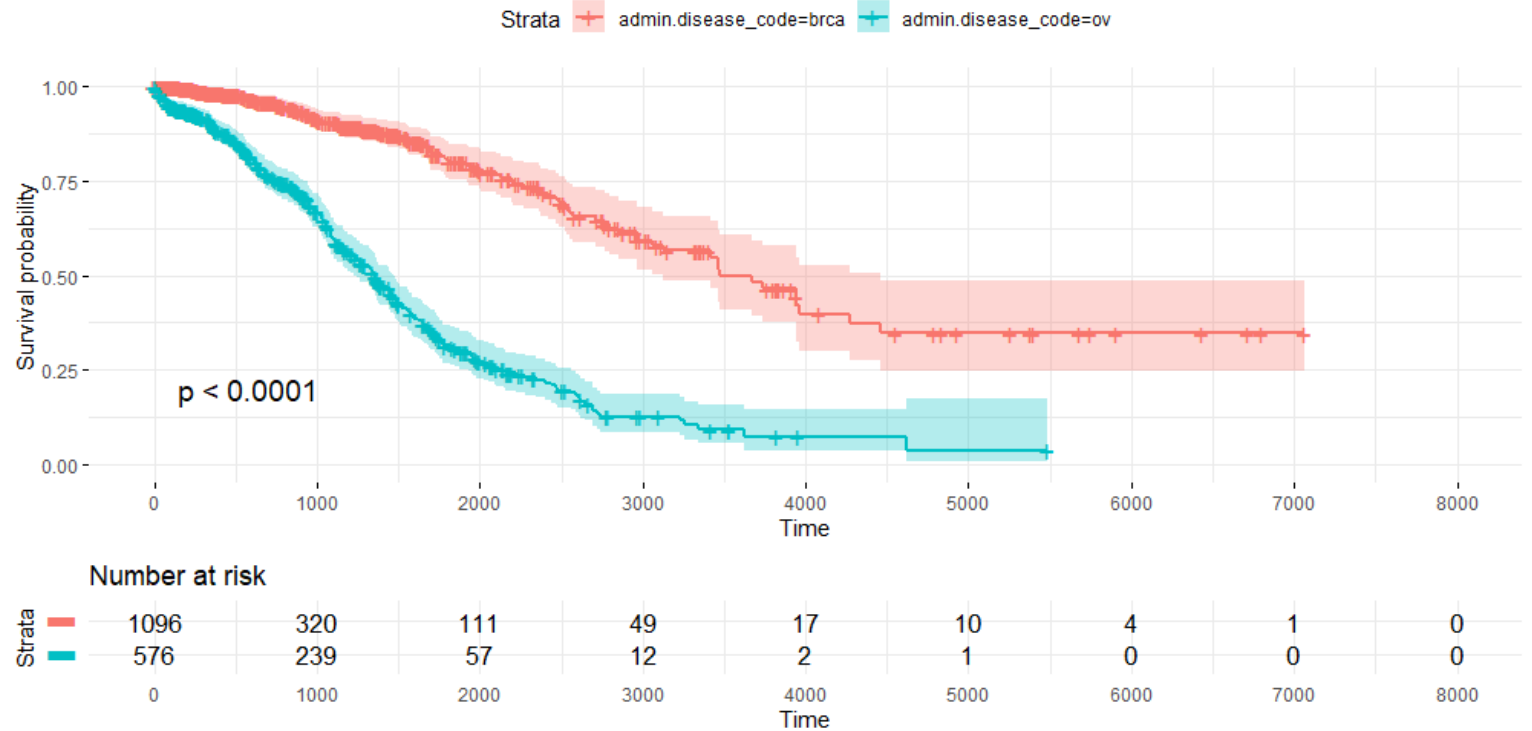
```
> install.packages('survminer')  
> BiocManager::install('RTCGA.clinical')  
  
> library(survminer)  
> library(survival)  
> library(RTCGA.clinical)
```

There are a lot of open source resources like this!



Kaplain-Meier Curves

```
> survivalTCGA(BRCA.clinical, OV.clinical, extract.cols = "admin.disease_code")
> fit <- survfit(Surv(times, patient.vital_status) ~ admin.disease_code, data = BRCAOV.survInfo)
> ggsurvplot(
  fit,
  data = BRCAOV.survInfo,
  risk.table = TRUE,
  pval = TRUE,
  conf.int = TRUE,
  break.time.by = 1000,
  ggtheme = theme_minimal(),
  risk.table.y.text.col = T,
  risk.table.y.text = FALSE)
```



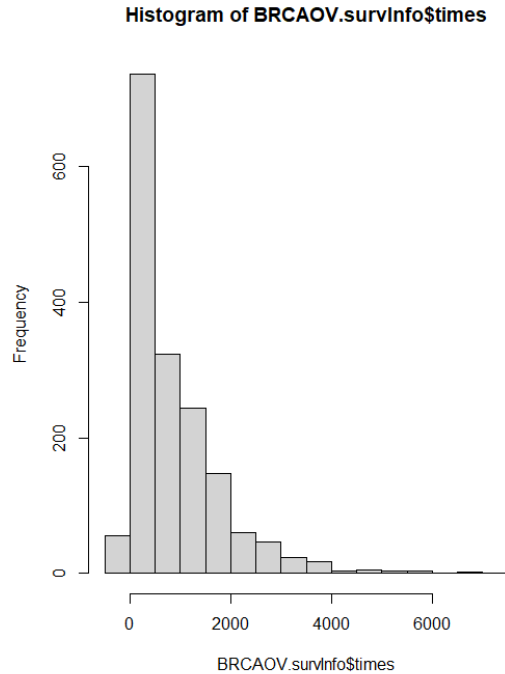
Kaplain-Meier Curves

- Breaking down the inputs
 - Metadata-based:
 - Phenotype/genotype information
 - Survival time after diagnosis
 - Cumulative probability that at time x, someone is still alive/healthy/etc.

```
> head(BRCAOV.survInfo, 30)
```

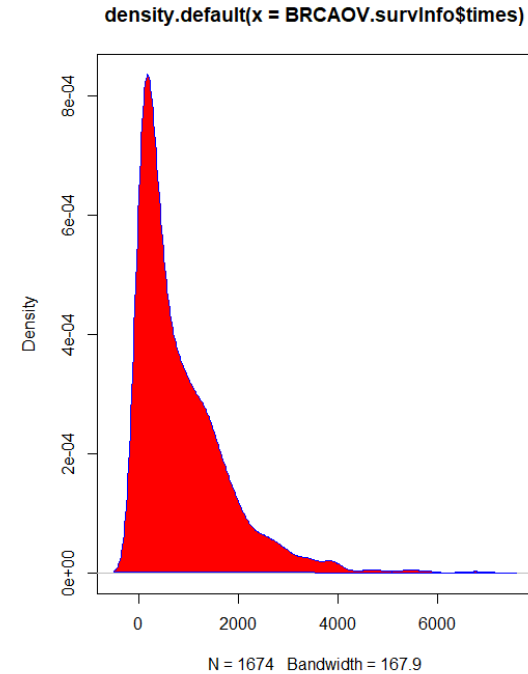
	times	bcr_patient_barcode	patient.vital_status	admin.disease_code
1	3767	TCGA-3C-AAAU	0	brca
2	3801	TCGA-3C-AALI	0	brca
3	1228	TCGA-3C-AALJ	0	brca
4	1217	TCGA-3C-AALK	0	brca
5	158	TCGA-4H-AAAK	0	brca
6	1477	TCGA-5L-AAT0	0	brca
7	1471	TCGA-5L-AAT1	0	brca
8	12	TCGA-5T-A9QA	0	brca
9	259	TCGA-A1-A0SB	0	brca
10	437	TCGA-A1-A0SD	0	brca
11	1321	TCGA-A1-A0SE	0	brca
12	1463	TCGA-A1-A0SF	0	brca
13	434	TCGA-A1-A0SG	0	brca
14	1437	TCGA-A1-A0SH	0	brca
15	635	TCGA-A1-A0SI	0	brca
16	416	TCGA-A1-A0SJ	0	brca
17	967	TCGA-A1-A0SK	1	brca
18	242	TCGA-A1-A0SM	0	brca
19	1196	TCGA-A1-A0SN	0	brca
20	852	TCGA-A1-A0SO	0	brca
21	584	TCGA-A1-A0SP	0	brca
22	554	TCGA-A1-A0SQ	0	brca
23	3153	TCGA-A2-A04N	0	brca
24	548	TCGA-A2-A04P	1	brca
25	2179	TCGA-A2-A04Q	0	brca
26	2365	TCGA-A2-A04R	0	brca
27	1950	TCGA-A2-A04T	0	brca
28	671	TCGA-A2-A04U	0	brca
29	1920	TCGA-A2-A04V	1	brca
30	1918	TCGA-A2-A04W	0	brca

Density Curve



Number at a time

```
> hist(BRCAOV.survInfo$times)
```



Proportion at a time

```
> d <- density(BRCAOV.survInfo$times)
> plot(d)
> polygon(d, col='red', border='blue')
```

Overlapping plots

```
> d1 <- density(OV.survInfo$times)
> d2 <- density(BRCA.survInfo$times)

> plot(d1)
> polygon(d1, col='red')

> par(new=TRUE)

> plot(d2)
> polygon(d2, col='blue')
```

