

```

library(DESeq2);load("CancerAndParacancer.RData")

## Loading required package: S4Vectors

## Loading required package: stats4

## 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, saveRDS, setdiff,
##     table, tapply, union, unique, unsplit, which.max, which.min

## 
## 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: GenomeInfoDb

## Loading required package: SummarizedExperiment

## 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: 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
```

```

expr <- ExprMat; pdat <- pDataMat; feat <- FeatureMat
same_samples <- intersect(colnames(expr), rownames(pdat))
expr <- expr[, same_samples, drop = F]
pdat <- pdat[same_samples, , drop = F]
expr <- expr[rowSums(expr) >= 954, , drop = F]
expr <- expr[!duplicated(rownames(expr)), , drop = F]
feat <- feat[!duplicated(rownames(feat)), , drop = F]
new_col <- intersect(c("gene_symbol", "symbol", "chromosome", "description", "gene_biotype"), colnames(feat))
feat <- feat[, new_col, drop = F]
feat <- feat[rownames(expr), , drop = F]
colnames(pdat)[1] <- "condition"
pdat$condition <- factor(pdat$condition)
pdat$condition <- relevel(pdat$condition, ref = "paracancerous")
dds <- DESeqDataSetFromMatrix(countData = expr, colData = pdat, design = ~ condition)
dds <- DESeq(dds); res <- results(dds)
```

```

## estimating size factors
```

```

## estimating dispersions
```

```

## gene-wise dispersion estimates
```

```

## mean-dispersion relationship
```

```

## final dispersion estimates
```

```

## fitting model and testing
```

```

## -- replacing outliers and refitting for 1371 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)
```

```

## estimating dispersions
```

```

## fitting model and testing
```

```

res_ordered <- res[order(res$padj, na.last = NA), ]
des_order <- as.data.frame(res_ordered)
normal_des <- as.data.frame(counts(dds, normalized = T))
des_order$gene <- rownames(des_order)
top_genes <- rownames(des_order)[1:15]
top_genes <- top_genes[top_genes %in% rownames(FeatureMat)]
cutoff <- min(rowSums(expr)[top_genes]); cat("cutoff is:", cutoff, "\n")
```

```

## cutoff is: 9296
```

```

top <- cbind(GeneID = top_genes, FeatureMat[top_genes, , drop = F],
  des_order[top_genes, c("log2FoldChange", "pvalue", "padj", "stat", "lfcSE", "baseMean"), drop = F])
up_order <- order(des_order$log2FoldChange, decreasing = T)
up_genes <- rownames(des_order)[up_order][1:10]
up_genes <- up_genes[up_genes %in% rownames(FeatureMat)]
cutoff <- min(rowSums(expr)[up_genes]); cat("cutoff is:", cutoff, "\n")
```

```

## cutoff is: 1043
```

```
top_up <- cbind(GeneID = up_genes, FeatureMat[up_genes, , drop = F],
  des_order[up_genes, c("log2FoldChange", "pvalue", "padj", "stat", "lfcSE", "baseMean"), drop = F])
down_order <- order(des_order$log2FoldChange, decreasing = F)
down_genes <- rownames(des_order)[down_order][1:10]
down_genes <- down_genes[down_genes %in% rownames(FeatureMat)]
cutoff <- min(rowSums(expr)[down_genes]);cat("cutoff is:", cutoff, "\n")
```

```
## cutoff is: 954
```

```
top_down <- cbind(GeneID = down_genes, FeatureMat[down_genes, , drop = F],
  des_order[down_genes, c("log2FoldChange", "pvalue", "padj", "stat", "lfcSE", "baseMean"), drop = F])
summary(top);knitr::kable(top)
```

```
##      GeneID          GeneID        Symbol       Description
##  Length:15      Min.   : 1589  Length:15      Length:15
##  Class :character 1st Qu.: 4898  Class :character  Class :character
##  Mode  :character Median : 9340  Mode  :character  Mode  :character
##                           Mean   : 79939
##                           3rd Qu.:182952
##                           Max.  :286133
##      Synonyms          GeneType      EnsemblGeneID      Status
##  Length:15      Length:15      Length:15      Length:15
##  Class :character Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##      ChrAcc          ChrStart      ChrStop       Orientation
##  Length:15      Length:15      Length:15      Length:15
##  Class :character Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##      Length     GOFunctionID      GOProcessID      GOComponentID
##  Min.   : 947  Length:15      Length:15      Length:15
##  1st Qu.: 2186  Class :character  Class :character  Class :character
##  Median : 2429  Mode  :character  Mode  :character  Mode  :character
##  Mean   : 4340
##  3rd Qu.: 5370
##  Max.   :13899
##      GOFunction      GOProcess      GOComponent      log2FoldChange
##  Length:15      Length:15      Length:15      Min.   :-7.681
##  Class :character Class :character  Class :character  1st Qu.:-4.934
##  Mode  :character  Mode  :character  Mode  :character  Median :-4.777
##                           Mean   :-4.722
##                           3rd Qu.:-4.088
##                           Max.   :-3.357
##      pvalue          padj          stat          lfcSE
##  Min.   :0.000e+00  Min.   :0.000e+00  Min.   :-19.64  Min.   :0.2385
##  1st Qu.:0.000e+00  1st Qu.:0.000e+00  1st Qu.:-15.38  1st Qu.:0.2853
##  Median :2.520e-45  Median :6.770e-42  Median :-14.13  Median :0.3125
##  Mean   :4.419e-42  Mean   :6.613e-39  Mean   :-14.87  Mean   :0.3156
##  3rd Qu.:6.347e-43  3rd Qu.:1.164e-39  3rd Qu.:-13.75  3rd Qu.:0.3529
##  Max.   :3.319e-41  Max.   :4.766e-38  Max.   :-13.44  Max.   :0.3911
##      baseMean
##  Min.   :136.3
##  1st Qu.:217.3
##  Median :435.4
##  Mean   :1081.2
##  3rd Qu.:1428.7
##  Max.   :4452.2
```

GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID	Status	ChrAcc
221476	221476	221476	PI16	peptidase inhibitor 16	CD364 CRISP9 MSMBBP PSPBP	protein-coding	ENSG00000164530	active NC_000
1675	1675	1675	CFD	complement factor D	ADIPSIN ADN DF PFD	protein-coding	ENSG00000197766	active NC_000
286133	286133	286133	SCARA5	scavenger receptor class A member 5	NET33 Tesr	protein-coding	ENSG00000168079	active NC_000

GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID	Status	ChrAcc
7123	7123	7123 CLEC3B	C-type lectin domain family 3 member B	MCDR4 TN TNA	protein-coding	ENSG00000163815	active	NC_000
9340	9340	9340 GLP2R	glucagon like peptide 2 receptor		protein-coding	ENSG0000065325	active	NC_000
11170	11170	11170 FAM107A	family with sequence similarity 107 member A	DRR1 TU3A	protein-coding	ENSG00000168309	active	NC_000
219348	219348	219348 PLAC9	placenta associated 9		protein-coding	ENSG00000189129	active	NC_000
146556	146556	146556 C16orf89	chromosome 16 open reading frame 89		protein-coding	ENSG00000153446	active	NC_000
221091	221091	221091 LRRN4CL	LRRN4 C-terminal like		protein-coding	ENSG00000177363	active	NC_000
1589	1589	1589 CYP21A2	cytochrome P450 family 21 subfamily A member 2	CA21H CAH1 CPS1 CYP21 CYP21B P450c21B	protein-coding	ENSG00000231852	active	NC_000
2674	2674	2674 GFRA1	GDNF family receptor alpha 1	GDNFR GDNFR-alpha-1 GDNFRA GFR-ALPHA-1 GFRalpha-1 RET1L RETL1 RHD4 TRNR1	protein-coding	ENSG00000151892	active	NC_000
1590	1590	1590 CYP21A1P	cytochrome P450 family 21 subfamily A member 1, pseudogene	CYP21A CYP21P P450c21A	pseudo	ENSG00000290788	active	NC_000
7146	7146	7146 TNXA	tenascin XA (pseudogene)	D6S103E HXBL TNX XA	pseudo		active	NC_000

GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID	Status	ChrAcc
55022	55022	55022 PID1	phosphotyrosine interaction domain containing 1	HMFN2073 NYGGF4 P-CLI1 PCLI1	protein-coding	ENSG00000153823	active	NC_000
7148	7148	7148 TNXB	tenascin XB	EDS3 EDSCLL EDSCLL1 HXBL TENX TN-X TNX TNXB1 TNXB2 TNXB5 VUR8 XB XBS	protein-coding	ENSG00000168477	active	NC_000

```
summary(top_up);knitr::kable(top_up)
```

```

##   GeneID      GeneID      Symbol      Description
##   Length:10    Min. : 1472    Length:10    Length:10
##   Class :character  1st Qu.: 19677    Class :character  Class :character
##   Mode :character   Median : 109906    Mode :character  Mode :character
##               Mean : 41924921
##               3rd Qu.:101929231
##               Max. :109729169
##   Synonyms      GeneType      EnsemblGeneID      Status
##   Length:10    Length:10    Length:10    Length:10
##   Class :character  Class :character  Class :character  Class :character
##   Mode :character   Mode :character  Mode :character  Mode :character
##
##
##
##   ChrAcc      ChrStart      ChrStop      Orientation
##   Length:10    Length:10    Length:10    Length:10
##   Class :character  Class :character  Class :character  Class :character
##   Mode :character   Mode :character  Mode :character  Mode :character
##
##
##
##   Length      GOFunctionID      GOProcessID      GOComponentID
##   Min. : 679    Length:10    Length:10    Length:10
##   1st Qu.:1236    Class :character  Class :character  Class :character
##   Median :1519    Mode :character   Mode :character  Mode :character
##   Mean : 1804
##   3rd Qu.:2160
##   Max. : 4158
##   GOFunction      GOProcess      GOComponent      log2FoldChange
##   Length:10    Length:10    Length:10    Min. :5.448
##   Class :character  Class :character  Class :character  1st Qu.:5.788
##   Mode :character   Mode :character   Mode :character  Median :5.931
##               Mean :6.141
##               3rd Qu.:6.124
##               Max. :7.555
##   pvalue      padj      stat      lfcSE
##   Min. : 0.000e+00    Min. :0.000e+00    Min. : 3.806    Min. : 0.6162
##   1st Qu.: 0.000e+00    1st Qu.:0.000e+00    1st Qu.:5.354    1st Qu.: 0.7731
##   Median : 9.300e-10    Median :1.230e-08    Median : 6.512    Median : 0.9580
##   Mean : 1.763e-05    Mean :6.626e-05    Mean : 6.558    Mean : 1.0348
##   3rd Qu.:1.525e-07    3rd Qu.:1.219e-06    3rd Qu.: 7.230    3rd Qu.: 1.1248
##   Max. : 1.410e-04    Max. :5.105e-04    Max. : 9.996    Max. : 1.9849
##   baseMean
##   Min. : 10.84
##   1st Qu.: 21.86
##   Median : 33.03
##   Mean : 41.44
##   3rd Qu.: 45.57
##   Max. :132.55

```

	GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID	Status	ChrAcc
51350	51350	51350	KRT76	keratin 76	HUMCYT2A KRT2B KRT2P	protein-coding	ENSG00000185069	active	NC_00001:
9119	9119	9119	KRT75	keratin 75	CK-75 K6HF K75 KB18 PFB hK6hf	protein-coding	ENSG00000170454	active	NC_00001:
1472	1472	1472	CST4	cystatin S		protein-coding	ENSG00000101441	active	NC_000021:
101929412	101929412	101929412	LINC02212	long intergenic non-protein coding RNA 2212		ncRNA	ENSG00000249396	active	NC_00000:
101928687	101928687	101928687	LNCAROD	lncRNA activating regulator of DKK1	A-ROD LINC01468 lnc-MBL2-4	ncRNA	ENSG00000231131	active	NC_000010:

GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID	Status	ChrAcc
6704	6704	6704 SPRR2E	small proline rich protein 2E		protein-coding	ENSG00000203785	active	NC_00000:
56033	56033	56033 BARX1	BARX homeobox 1		protein-coding	ENSG00000131668	active	NC_00000:
163778	163778	163778 SPRR4	small proline rich protein 4		protein-coding	ENSG00000184148	active	NC_00000:
105373485	105373485	105373485 LOC105373485	uncharacterized LOC105373485		ncRNA		active	NC_00000:
109729169	109729169	109729169 LINC02154	long intergenic non-protein coding RNA 2154		ncRNA	ENSG00000235385	active	NC_00002:

```
summary(top_down);knitr::kable(top_down)
```

```

##   GeneID          GeneID          Symbol          Description
## Length:10      Min. :    70  Length:10      Length:10
## Class :character 1st Qu.:  4770  Class :character  Class :character
## Mode  :character Median :  7276  Mode  :character  Mode  :character
##               Mean : 10283197
##               3rd Qu.: 187198
##               Max. :101927499
##   Synonyms        GeneType        EnsemblGeneID      Status
## Length:10      Length:10      Length:10      Length:10
## Class :character Class :character  Class :character  Class :character
## Mode  :character Mode  :character  Mode  :character  Mode  :character
##
##
##
##   ChrAcc          ChrStart        ChrStop        Orientation
## Length:10      Length:10      Length:10      Length:10
## Class :character Class :character  Class :character  Class :character
## Mode  :character Mode  :character  Mode  :character  Mode  :character
##
##
##
##   Length       GOFunctionID      GOProcessID      GOComponentID
## Min. : 382.0  Length:10      Length:10      Length:10
## 1st Qu.: 767.5 Class :character  Class :character  Class :character
## Median :1741.0 Mode  :character  Mode  :character  Mode  :character
## Mean   :2201.6
## 3rd Qu.:2241.2
## Max.   :6769.0
##   GOFunction      GOProcess      GOComponent      log2FoldChange
## Length:10      Length:10      Length:10      Min.   :-7.681
## Class :character Class :character  Class :character  1st Qu.:-6.584
## Mode  :character Mode  :character  Mode  :character  Median :-6.087
##               Mean   :-6.317
##               3rd Qu.:-5.924
##               Max.   :-5.751
##   pvalue          padj          stat          lfcSE
## Min.   :0.000e+00  Min.   :0.000e+00  Min.   :-19.638  Min.   :0.3872
## 1st Qu.:0.000e+00  1st Qu.:0.000e+00  1st Qu.:-12.127  1st Qu.:0.5233
## Median :0.000e+00  Median :0.000e+00  Median :-11.327  Median :0.5599
## Mean   :5.015e-17  Mean   :2.762e-15  Mean   :-11.969  Mean   :0.5514
## 3rd Qu.:1.000e-22  3rd Qu.:1.900e-20  3rd Qu.:-9.968  3rd Qu.:0.6075
## Max.   :5.015e-16  Max.   :2.762e-14  Max.   :-8.111  Max.   :0.7293
##   baseMean
## Min.   : 19.44
## 1st Qu.: 145.78
## Median : 242.39
## Mean   : 4194.07
## 3rd Qu.: 1797.90
## Max.   :34495.23

```

	GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID
221476	221476	221476	PI16	peptidase inhibitor 16	CD364 CRISP9 MSMBBP PSPBP	protein-coding	ENSG00000164530

GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID
4653	4653	4653 MYOC	myocilin	GLC1A GPOA JOAG JOAG1 TIGR	protein-coding	ENSG00000034971
572558	572558	572558 PGM5-AS1	PGM5 antisense RNA 1	FAM233A	ncRNA	ENSG00000224958
5212	5212	5212 VIT	vitrin	VIT1	protein-coding	ENSG00000205221
84366	84366	84366 PRAC1	PRAC1 small nuclear protein	C17orf92 PRAC	protein-coding	ENSG00000159182
70	70	70 ACTC1	actin alpha cardiac muscle 1	ACTC ASD5 CMD1R CMH11 LVNC4	protein-coding	ENSG00000159251

GeneID	GeneID	Symbol	Description	Synonyms	GeneType	EnsemblGeneID
9340	9340	9340 GLP2R	glucagon like peptide 2 receptor		protein-coding	ENSG00000065325
101927499	101927499	101927499 LOC101927499	uncharacterized LOC101927499		ncRNA	
1674	1674	1674 DES	desmin	CDCD3 CSM1 CSM2 LGMD1D LGMD1E LGMD2R	protein-coding	ENSG00000175084
5121	5121	5121 PCP4	Purkinje cell protein 4	PEP-19	protein-coding	ENSG00000183036