Select genes to run with Btrecase from psoriasis data

Elena

Mon Nov 26 15:39:09 2018

Contents

Sources 1
DEG in psoriatic vs healthy skin 2
DEG proximal to GWAS hits in healthy or psoriatic skin 2
Genes to follow up 3

```
## Report built by rule report from
## /home/ev250/Bayesian_inf/trecase/Scripts/stan_eff/real_data/psoriasis/run_Btrecase/Snakefile

## Load R packages and functions:
library(data.table)
library(xlsx)
library(biomaRt)
library(rmarkdown)
source("/home/ev250/Bayesian_inf/trecase/Scripts/stan_eff/real_data/psoriasis/run_Btrecase/Functions/au
```

Sources

- We have GWAS SNPs associated with psoriasis and differentially expressed genes between psoriatic and normal skin.
- We want to select genes that are likely regulated by SNPs in cis. ## GWAS SNPs For GWAS hits exported from Nat Commun. 2017; 8: 15382

```
## For each SNP we searched for genes within 500KB
## Main characteristics

## open file
gwas <- fread(snakemake@input[['gwas']])
setkey(gwas,CHROM,POS, gene_dist)
```

- Number of unique SNPs

```
length(unique(gwas$rs_id))
```

```
## [1] 53
```

- Number of unique genes

```
length(unique(gwas$gene_id))
```

```
## [1] 1493
```

- Distribution of the number of genes proximal to each SNP by gene-SNP distance

```
d <- c(0, 1 %o% 10^(c(3,4,5)), 5*10^5)
f(dt=gwas,var1="rs_id", var2="gene_id", c("Number.Genes","Number.SNPs"), d)
```

```
##        Min. 1st Qu. Median      Mean 3rd Qu. Max. Number.Genes Number.SNPs
## 0_KB      1       1      1  1.151515       1    2           38          33
```

```
## 1_KB      1       1      1  1.181818       1    2           39           33
## 10_KB     1       1      2  1.923077       2    4           75           39
## 100_KB    1       3      5  7.169811      10   25          380           53
## 500_KB    5      16     24 28.339623      33   80         1493           53
```

- Same but excluding SNPs within genes

```
f(dt=gwas[!rs_id %in% gwas[gene_dist==0,rs_id],] ,var1="rs_id",
  var2="gene_id", c("Number.Genes","Number.SNPs"), d[d>0])
```

```
##        Min. 1st Qu. Median  Mean 3rd Qu. Max. Number.Genes Number.SNPs
## 1_KB     NA      NA     NA   NaN      NA   NA            0           0
## 10_KB     1       1    1.5  1.50    2.00    2            9           6
## 100_KB    1       3    4.0  4.70    6.25   12           94          20
## 500_KB    5      14   16.5 21.05   24.25   50          421          20
```

DEG in psoriatic vs healthy skin

- Based on J Invest Dermatol. 2014 Jul;134(7):1828-1838. doi: 10.1038/jid.2014.28. Epub 2014 Jan 17.
- Gene expression is reported in RPKM, low <1, 1<=med<500, high>=500

```
## open file
drg=data.table(read.xlsx2(snakemake@input[['drg']],
                          sheetIndex=1 ,
                          colClasses=snakemake@params[['colclass']]))
```

- 7238 DEG with Case Median expression >1 and p<=10^-6 (cut-off in paper)
- 2465 DEG with Case Median expression >1 & FC >1 and p<=10^-6 (cut-off in paper)
- Distribution of FC by expression levels in up-regulated genes (p<=10^-6, cut-off in paper)

```
## Select DRG
drg <- drg[RankP<=10^-6]
f2(dt=drg[ FC>=1,], var1="CaseMedian", var2="FC",
   range1=c(0, 1, 500, max(drg$CaseMedian)))
```

```
##            Min.  1st Qu.   Median       Mean   3rd Qu.        Max.
## Low    1.282502 1.952484 2.570161   3.698117  3.926429    25.82609
## Medium 1.066252 1.312954 1.529844   4.579987  2.145677  2043.24886
## High   1.146679 1.682396 3.378544 441.865224 18.821661 19913.19291
##        Number.Genes
## Low             354
## Medium         2412
## High             53
```

- DEG with high expression in psoriatic skin are mostly keratinocytes expressed genes

DEG proximal to GWAS hits in healthy or psoriatic skin

- Look at the expression leves for genes associated with GWAS hits

```
## 22 out of 38 genes with in-gene GWAS hits are DEG, most with low-medium
## expressin levels in psoriatic skin

f3(drg, gwas, var1="Gene.Symbol",var2="CaseMedian",d=d)
```

```
##         Min. 1st Qu. Median      Mean  3rd Qu.      Max. Number.Genes
```

```
## 0_KB   0.048 3.86300 7.3305  33.42223 12.57512  512.3745           22
## 1_KB   0.048 3.96900 7.3555  32.48767 12.53225  512.3745           23
## 10_KB  0.048 1.59800 6.3285  37.28374 12.61800  512.3745           33
## 100_KB 0.008 1.86775 7.3535  51.47073 21.17600 1170.9910           95
## 500_KB 0.000 1.78875 7.5435 124.93495 25.67550 5526.0835          315
```

```
summary(drg[Gene.Symbol %in% gwas[['gene_name']] , CaseMedian ])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    0.000    1.789    7.543  124.935   25.675 5526.083
```

Genes to follow up

- Genes within 100KB to GWAS hits
- Highly expressed upregulated DEG (RPKM >=500)

```
## Total genes: 422
## Up to 100 KB from gene from gwas and upregulated DRG highly expressed in psoriatic skin
unique(c(gwas[gene_dist <= 100000, gene_name], drg[CaseMedian>=500 &FC>1, as.character(Gene.Symbol)]))
```

```
##   [1] "RP11-431K24.4"  "RP11-431K24.3"  "RNU1-7P"        "RP11-431K24.2"
##   [5] "RN7SL729P"      "RNU6-991P"      "SLC45A1"        "IFNLR1"
##   [9] "RP11-10N16.2"   "AL590683.1"     "IL22RA1"        "MYOM3"
##  [13] "AL590683.2"     "RUNX3"          "RP11-84D1.1"    "RP11-84D1.2"
##  [17] "MIR4425"        "RP4-799D16.1"   "IL23R"          "AL109843.1"
##  [21] "C1orf141"       "RNU4ATAC4P"     "RP4-763G1.1"    "RNU6-586P"
##  [25] "IL12RB2"        "RP11-74K19.1"   "GIPC2"          "DNAJB4"
##  [29] "FUBP1"          "RP11-386I14.4"  "NEXN"           "RP11-386I14.2"
##  [33] "NEXN-AS1"       "LCE3A"          "LCE3B"          "LCE3C"
##  [37] "RP1-52J10.9"    "LINC00302"      "LCE3D"          "LCE2D"
##  [41] "LCE3E"          "LCE2B"          "LCE2C"          "LCE2A"
##  [45] "LCE4A"          "FASLG"          "SLC25A38P1"     "RP1-15D23.2"
##  [49] "RNU6-693P"      "SUCO"           "DENND1B"        "RP11-448G4.4"
##  [53] "FAM204BP"       "RP11-448G4.2"   "EEF1A1P32"      "IKBKE"
##  [57] "C1orf147"       "SRGAP2"         "RP11-534L20.5"  "RASSF5"
##  [61] "RP11-534L20.4"  "EIF2D"          "AC010733.4"     "AC010733.7"
##  [65] "RN7SL632P"      "REL"            "PAPOLG"         "RNU4-51P"
##  [69] "RP11-373L24.1"  "NONOP2"         "AC010733.5"     "PUS10"
##  [73] "RP11-642D6.1"   "snoU13"         "RN7SL51P"       "IFIH1"
##  [77] "GCA"            "KCNH7"          "FAP"            "PLCL2"
##  [81] "MIR3714"        "PLCL2-AS1"      "RP11-221J22.1"  "Y_RNA"
##  [85] "RP11-221J22.2"  "NFKBIZ"         "LEPREL1"        "MTAPP2"
##  [89] "TP63"           "AC108105.1"     "ERAP1"          "CTD-2260A17.1"
##  [93] "CAST"           "CTD-2260A17.3"  "CTD-2260A17.2"  "CTC-506B8.1"
##  [97] "AC020900.2"     "ERAP2"          "IL13"           "AC004041.2"
## [101] "IL4"            "RAD50"          "AC004237.1"     "KIF3A"
## [105] "CCNI2"          "SEPT8"          "TNIP1"          "ANXA6"
## [109] "GPX3"           "CTB-17P3.4"     "CCDC69"         "AC008697.1"
## [113] "AC008703.1"     "IL12B"          "RNU4ATAC2P"     "EXOC2"
## [117] "RP1-20B11.2"    "HUS1B"          "RP11-532F6.2"   "CDKAL1"
## [121] "RP3-348I23.3"   "RP3-348I23.2"   "DPCR1"          "HCG21"
## [125] "SFTA2"          "VARS2"          "GTF2H4"         "MUC21"
## [129] "RN7SL175P"      "DDR1"           "MIR4640"        "MUC22"
## [133] "DDR1-AS1"       "RN7SKP186"      "TRAF3IP2"       "C6orf3"
## [137] "TRAF3IP2-AS1"   "FYN"            "TNFAIP3"        "RP11-356I2.4"
```

```
## [141] "RP11-10J5.1"    "RP11-240M16.1"  "RP1-111C20.3"   "RP1-111C20.4"
## [145] "RP11-13P5.1"    "TAGAP"          "RP11-13P5.2"    "FNDC1"
## [149] "RSPH3"          "ELMO1"          "RNU6-565P"      "RPS10P14"
## [153] "SNORA51"        "DDX58"          "TOPORS"         "AL353671.2"
## [157] "AL353671.1"     "TOPORS-AS1"     "NDUFB6"         "AL353671.3"
## [161] "AL353671.4"     "RP11-205M20.7"  "GVQW1"          "ACO1"
## [165] "RP11-240E2.2"   "RPS15AP27"      "CHCHD4P2"       "ZNF365"
## [169] "AC067751.1"     "CAMK2G"         "RP11-574K11.8"  "RP11-574K11.5"
## [173] "NDST2"          "RP11-574K11.31" "ZSWIM8"         "ZSWIM8-AS1"
## [177] "CHCHD1"         "FUT11"          "RMRPP1"         "PLAU"
## [181] "C10orf55"       "AC022400.2"     "SEC24C"         "ZMIZ1"
## [185] "PPIF"           "ZCCHC24"        "RP11-342M3.5"   "MED6P1"
## [189] "SNORD74"        "PTEN"           "BLOC1S2"        "PKD2L1"
## [193] "CWF19L1"        "RNU6-422P"      "PHBP9"          "SNORA12"
## [197] "RP11-316M21.6"  "CHUK"           "RP11-34D15.2"   "SCD"
## [201] "AL139819.1"     "RP11-316M21.7"  "ERLIN1"         "LINC00263"
## [205] "GPR137"         "BAD"            "KCNK4"          "RP11-783K16.10"
## [209] "TEX40"          "PLCB3"          "ESRRA"          "TRMT112"
## [213] "PRDX5"          "RP11-783K16.13" "RP11-783K16.5"  "PPP1R14B"
## [217] "AP003774.1"     "FKBP2"          "VEGFB"          "RP11-783K16.14"
## [221] "DNAJC4"         "CCDC88B"        "NUDT22"         "TRPT1"
## [225] "FERMT3"         "RPS6KA4"        "STIP1"          "MIR1237"
## [229] "CFL1"           "SNX32"          "OVOL1"          "RP11-770G2.5"
## [233] "MUS81"          "OVOL1-AS1"      "EFEMP2"         "AP5B1"
## [237] "AP001266.1"     "CTSW"           "RP11-770G2.2"   "FIBP"
## [241] "CCDC85B"        "FOSL1"          "C11orf68"       "DRAP1"
## [245] "KRT8P26"        "ZC3H12C"        "RP11-344L21.1"  "RDX"
## [249] "ETS1"           "RP11-1007G5.2"  "RP11-264E20.2"  "RP11-264E20.1"
## [253] "KLRC1"          "KLRC2"          "NKG2-E"         "KLRC3"
## [257] "KLRC4-KLRK1"    "KLRC4"          "RP11-277P12.20" "KLRK1"
## [261] "EIF2S3L"        "RP11-277P12.9"  "STAT2"          "RNU7-40P"
## [265] "IL23A"          "APOF"           "PAN2"           "CNPY2"
## [269] "RP11-977G19.10" "RP11-977G19.11" "RP11-348M3.2"   "RP11-977G19.12"
## [273] "CS"             "TIMELESS"       "COQ10A"         "RP11-977G19.14"
## [277] "ANKRD52"        "BRAP"           "RP11-686G8.2"   "ATXN2"
## [281] "RP11-686G8.1"   "PCNPP1"         "U7"             "ACAD10"
## [285] "LRRC43"         "IL31"           "B3GNT4"         "DIABLO"
## [289] "RP11-512M8.5"   "MLXIP"          "VPS33A"         "CLIP1"
## [293] "LINC00332"      "RPL17P51"       "RNY3P9"         "LINC00407"
## [297] "LINC00330"      "UBAC2"          "GPR183"         "GPR18"
## [301] "MIR623"         "HMGB3P4"        "H2AFZP3"        "RN7SKP9"
## [305] "UBAC2-AS1"      "RP11-561B11.3"  "RP11-561B11.1"  "NFKBIA"
## [309] "PSMA6"          "KIAA0391"       "DNAJC8P1"       "RP11-561B11.6"
## [313] "AL121594.1"     "RPS3AP4"        "RP11-61O1.1"    "RP11-61O1.2"
## [317] "KLF13"          "UBE2CP4"        "RMI2"           "SOCS1"
## [321] "TNP2"           "PRM3"           "SNORA48"        "PRM2"
## [325] "HNRNPCP4"       "PRM1"           "RP11-396B14.2"  "MIR548H2"
## [329] "CLEC16A"        "RP11-485G7.5"   "RP11-485G7.6"   "AC009121.1"
## [333] "CTD-3088G3.8"   "TRIM65"         "RP11-552F3.10"  "MRPL38"
## [337] "RP11-552F3.12"  "RP11-552F3.9"   "FBF1"           "TRIM47"
## [341] "WBP2"           "ACOX1"          "RP11-552F3.13"  "UNC13D"
## [345] "UNK"            "RP11-552F3.4"   "TEN1"           "TEN1-CDK3"
## [349] "PTPN2"          "RP11-773H22.1"  "RP11-773H22.2"  "RP11-973H7.1"
## [353] "SEH1L"          "FUT2"           "MAMSTR"         "RASIP1"
```

```
## [357] "RN7SL345P"      "SEC1P"          "NTN5"           "IZUMO1"
## [361] "FUT1"           "FGF21"          "CA11"           "DBP"
## [365] "SPHK2"          "AC022154.7"     "RPL18"          "FAM83E"
## [369] "RNU6-317P"      "BCAT2"          "SPACA4"         "AQP3"
## [373] "B2M"            "CALML3"         "CALML5"         "CDSN"
## [377] "CNFN"           "CRABP2"         "CSTA"           "CSTB"
## [381] "DEFB4A"         "ENO1"           "FABP5"          "GAPDH"
## [385] "GJB2"           "GSTP1"          "HLA-B"          "HLA-E"
## [389] "IFITM3"         "IFI27"          "JUP"            "KRTDAP"
## [393] "KRT5"           "KRT6A"          "KRT6B"          "KRT6C"
## [397] "KRT14"          "KRT16"          "KRT17"          "LYPD3"
## [401] "LY6D"           "PFN1"           "PI3"            "SBSN"
## [405] "SERPINB3"       "SERPINB4"       "SFN"            "SPRR1A"
## [409] "SPRR1B"         "SPRR2A"         "SPRR2B"         "SPRR2D"
## [413] "SPRR2E"         "SPRR2F"         "SPRR2G"         "S100A2"
## [417] "S100A7"         "S100A8"         "S100A9"         "S100A11"
## [421] "S100A14"        "TUBA1C"
```