Breast Cancer Blood Plasma Markers

Code

BioMed Innovation Lab

Overview

This is an example showing how to use our Blood Plasma Marker Selection Pipeline to short list candidate blood plasma markers that are more likely to be over expressed for breast cancer, compared to other cancers including colorectal, lung, prostate, skin and stomach cancers.

1. Preparation

1.1 Load the pipeline functions

This will also install and load the required libraries.

Hide

```
source("bpmsp.R", local = knitr::knit_global())
```

1.2 Loading the data sets

This might take more than 20 minutes to download near 3G data depending on your download speed. The speed might also depend on where you are located. If timeout occurs, please increase the "timeout" option, remove the partially downloaded file, then retry this step.

Hide

```
bpmsp.load_blood_plasma_markers()
```

```
## TCGA/GTEx data saved in global variable bpmsp.all_plasma_protein_genes
```

Hide

```
# Setting the time out to 1 hour (3,600 seconds) in case. If it still times out, increas
e the value.
options(timeout=3600)

bpmsp.load_recomputed_tcga_gtex_data()
```

```
## TCGA/GTEx data saved in global variable bpmsp.short_listed_tcga_gtex_rsem_gene_tpm
```

Hide

```
# Note that if the above step failed due to time out, please increase the timeout.
# Also the pipeline function has logic to bypass downloading if it finds the files, so p
lease remove the partially downloaded files before trying this step again.

bpmsp.load_ihc_data()
```

```
## IHC data are loaded into the global variables bpmsp.normal_tissue_ihc and bpmsp.tumor
_tissue_ihc
```

2. Selection of candidate markers for each cancer

Here we sequentially, instead of in parallel or in a loop, apply the pipeline to select candidate markers for each type of cancer, only for the purpose of showing how the pipeline console outputs are different for each cancer.

2.1 Breast cancer

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'breast'

cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "ESM1"       "RLN1"       "WNK4"       "CALML5"     "LYZ"
##   [6] "CFB"        "CENPE"      "XDH"        "CTHRC1"     "CRIP1"
##  [11] "PBK"        "MUC1"       "DLGAP5"     "TTK"        "FBP1"
##  [16] "MLPH"       "CXCL9"      "CAMP"       "MSX2"       "ERBB3"
##  [21] "CLEC4C"     "HPN"        "TOP2A"      "LTF"        "DSP"
##  [26] "SLC52A3"    "CCL11"      "CCL17"      "IFNG"       "CDC6"
##  [31] "EPCAM"      "CENPF"      "ST14"       "HEPACAM2"   "MZB1"
##  [36] "PAFAH1B3"   "RLN2"       "SCUBE2"     "UTS2"       "LRRC15"
##  [41] "CKMT1A"     "LYPD3"      "PIP"        "CCL7"       "IL22RA2"
##  [46] "KRT80"      "MMP7"       "APOC1"      "DIAPH3"     "GTSE1"
##  [51] "MUC5B"      "NEIL3"      "AFP"        "APOF"       "PLD4"
##  [56] "CCL5"       "F7"         "RRM2"       "CLDN3"      "TNFRSF17"
##  [61] "KRT6B"      "GZMA"       "TK1"        "CA12"       "TAT"
##  [66] "SERPINB5"   "CDC20"      "AZGP1"      "KRT23"      "KRT75"
##  [71] "SERPINA9"   "CLUL1"      "SDC1"       "TCN1"       "BAMBI"
##  [76] "TFF3"       "UMOD"       "MATN3"      "LAIR2"      "AGR2"
##  [81] "MAD2L1"     "BPIFB2"     "IRX4"       "SPP1"       "TC2N"
##  [86] "SFRP4"      "PKP3"       "HIST1H2BL"  "S100P"      "SLITRK6"
##  [91] "FOXI1"      "PGLYRP2"    "DNAI1"      "KIF2C"      "MSMB"
##  [96] "CPXM1"      "LAD1"       "HSD17B2"    "KNL1"       "CDH3"
## [101] "E2F2"       "KIF11"      "COL5A2"     "DNAH5"      "CXCL17"
## [106] "CEACAM6"    "TFF1"       "SERPINA6"   "S100A14"    "TNFSF11"
## [111] "IGSF3"      "MARCKSL1"   "SYTL5"      "SERPINA11"  "MXRA5"
## [116] "WFDC2"      "MELK"       "ACP5"       "LUM"        "MNX1"
## [121] "IQGAP3"     "SPINT2"     "IRX1"       "SDS"        "MMP3"
## [126] "ITGB6"      "FAP"        "CILP2"      "SMPDL3B"    "GALNT5"
## [131] "MMP13"      "CST2"       "PRSS8"      "STC2"       "RAB27B"
## [136] "LAMP3"      "CCNB2"      "CDK1"       "PVALB"      "CXCL11"
## [141] "DUSP4"      "SLAMF7"     "SHROOM3"    "CD207"      "LMNB1"
## [146] "KCNF1"      "SKAP1"      "CHIT1"      "MUC16"      "MDK"
## [151] "LSR"        "PMCH"       "CGN"        "LRP2"       "CDH1"
## [156] "KIFC1"      "COL1A2"     "NUF2"       "CD3G"       "TFAP2B"
## [161] "HOXC12"     "MMP10"      "MMP12"      "CXCL10"     "OLR1"
## [166] "IL19"       "HOXC11"     "HIST1H1B"   "DSC2"       "SFN"
## [171] "CCL8"       "ACTBL2"     "COL10A1"    "KRT32"      "KRT81"
## [176] "CST5"       "IL12B"      "FASLG"      "KIF14"      "HIST1H4D"
## [181] "MGP"        "KRT19"      "BPIFB1"     "COMP"       "CTSV"
## [186] "PLA2G7"     "ANO1"       "CDC45"      "ADM2"       "GALNT6"
## [191] "GGT6"       "TNFSF15"    "KRT18"      "GRP"        "COL1A1"
## [196] "DCDC2"      "PKD2L1"     "EPYC"       "ADAMDEC1"   "HIST1H4J"
## [201] "CAMSAP3"    "CDCP1"      "CST1"       "PIK3C2G"    "DSG2"
## [206] "GATA3"      "F12"        "CST4"       "KRT17"      "AGR3"
## [211] "VAMP8"      "ABCA12"     "ERVMER34-1" "ENPP1"      "TACSTD2"
## [216] "SIM1"       "AURKA"      "CDSN"       "ASPN"       "IFNL1"
## [221] "C1orf116"   "NECTIN4"    "CCL18"      "SCT"        "TPD52"
## [226] "JCHAIN"     "AURKB"      "MMP9"       "POSTN"      "MMP1"
## [231] "SEZ6L2"     "CCL22"      "ZG16B"      "CCNB1"      "IL20"
## [236] "TREM2"      "CXCL13"     "SPINT1"     "TNFRSF9"    "TCL1A"
## [241] "TNN"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "ALB"      "SERPINA1" "SERPINA3" "CP"       "F2"       "FETUB"
##   [7] "C9"       "PON1"     "AZGP1"    "SERPINA7" "FCN3"     "PLTP"
##  [13] "SERPINF1" "CLEC3B"   "CFB"      "TNC"      "MASP2"    "C4A"
##  [19] "C1QTNF7"  "ORM2"     "CD55"     "APOD"     "IL1RL1"   "MIF"
##  [25] "MFGE8"    "ENHO"     "ISG15"    "MMP3"     "TNFRSF1B" "SCUBE3"
##  [31] "SEMA3F"   "MMP10"    "IL22RA2"  "TNFSF13B" "CCL2"     "CXCL11"
##  [37] "VEGFD"    "CCL4"     "UTS2"     "HAVCR1"   "MDK"      "BMP6"
##  [43] "IL1R2"    "IL32"     "SERPINF2" "CTF1"     "IL7"      "NPPC"
##  [49] "IL2"      "EDN1"     "C4B"      "C7"       "CPN2"     "CPB2"
##  [55] "BCHE"     "PZP"      "ITIH3"    "PRG4"     "HSPG2"    "POTEJ"
##  [61] "C8G"      "KRT8"     "KRT4"     "KRT7"     "KRT19"    "WDR1"
##  [67] "GRN"      "TIMP2"    "GOLM1"    "STXBP2"   "HSPD1"    "RAB27B"
##  [73] "FLNB"     "ATRNL1"   "COLEC11"  "ACTN3"    "MEPE"     "RAB1A"
##  [79] "MGP"      "CEACAM8"  "LY6G6F"   "CILP"     "NBL1"     "RNPEP"
##  [85] "SYTL4"    "MUC1"     "GCLC"     "PRKCB"    "GALNT6"   "LRRC4B"
##  [91] "TTLL12"   "FASN"     "GMFG"     "USP14"    "SEMA3A"   "S100P"
##  [97] "VMO1"     "THSD1"    "CYB5A"    "EPHX2"    "MATN2"    "MUC5B"
## [103] "CAMK1"    "POF1B"    "CXorf36"  "TRIM58"   "ENPP5"    "IGSF10"
## [109] "SSBP1"    "HGS"      "BROX"     "UFC1"     "FAM177A1" "FREM2"
## [115] "SHMT2"    "NTN1"     "MYOM2"    "TOMM70"   "STC1"     "S100B"
## [121] "NELL1"    "ADAMTSL3" "S100A14"  "SIRPB2"   "BCAT1"    "GLCE"
## [127] "SEMA3D"   "B4GALT3"  "ADGRL1"   "SLC39A10" "AGR3"     "GLUL"
## [133] "REEP6"    "GALNT15"  "ASAP1"    "TNNI1"    "AGR2"     "CGREF1"
## [139] "NDUFS2"   "HYAL1"    "SLFN14"   "TNIK"     "PLIN1"    "NCR3LG1"
## [145] "AMN"      "KAZALD1"  "IL12A"    "ADAMTS15" "KLK14"    "DRAXIN"
## [151] "IFNLR1"   "NBN"      "SPOCK1"   "EFNA4"    "ADAMTS12" "ADAMTS7"
## [157] "ADGRF3"   "ADGRG3"   "ADGRG4"   "AR"       "ARHGEF2"  "AUTS2"
## [163] "BCL11A"   "BPIFB3"   "C14orf39" "CACNA1G"  "CDC20"    "CDH12"
## [169] "CLPB"     "COL19A1"  "CSMD2"    "DNAH8"    "DONSON"   "DOT1L"
## [175] "DTNB"     "FAM9B"    "FOXD4L1"  "GATA3"    "GCAT"     "GDF5"
## [181] "GGPS1"    "HECW2"    "HOXC11"   "HPS5"     "IKBKB"    "IQGAP3"
## [187] "IRX1"     "KBTBD11"  "KCNN2"    "KCTD4"    "KIAA0355" "KIF14"
## [193] "KIF4B"    "LAMA1"    "LIMD1"    "LRRC3"    "MEA1"     "MKI67"
## [199] "MYO5B"    "NASP"     "NEUROG3"  "NKX2-8"   "PACSIN3"  "PCDH8"
## [205] "PCNA"     "PGBD2"    "PIK3C2B"  "PKD1L1"   "POLR3A"   "POTEE"
## [211] "PRAMEF12" "PRDM15"   "PROKR2"   "RAB3D"    "RHPN1"    "RNF6"
## [217] "SALL3"    "SCRT1"    "SLC4A9"   "SYTL2"    "TFAP2B"   "TOP2A"
## [223] "TPD52L1"  "TRIM21"   "TUBB8"    "UNC5D"    "URB2"     "WNK4"
## [229] "ZFP36L1"  "ZNF232"   "ZNF300"   "ZNF445"   "ZNF507"   "ZNF566"
## [235] "ZNF780A"
```

Step 3: Select the markers short listed in both approaches

Hide

```
breast_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_markers)
breast_cancer_genesInBoth
```

```
##  [1] "WNK4"    "CFB"     "MUC1"    "TOP2A"   "UTS2"    "IL22RA2" "MUC5B"
##  [8] "CDC20"   "AZGP1"   "AGR2"    "S100P"   "S100A14" "IQGAP3"  "IRX1"
## [15] "MMP3"    "RAB27B"  "CXCL11"  "MDK"     "TFAP2B"  "MMP10"   "HOXC11"
## [22] "KIF14"   "MGP"     "KRT19"   "GALNT6"  "GATA3"   "AGR3"
```

2.2 Colorectal cancer

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'colorectal'

cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "DNAH3"      "ESM1"       "KLK10"      "NOX1"       "FGFBP1"
##   [6] "WNK4"       "MEP1A"      "LYZ"        "FGFR4"      "CXCL8"
##  [11] "XDH"        "PBK"        "ACE2"       "HMGCS2"     "DLGAP5"
##  [16] "TTK"        "DPEP1"      "CXCL9"      "PI3"        "ERBB3"
##  [21] "TOP2A"      "SLC52A3"    "CCL11"      "CCL17"      "KLK12"
##  [26] "GDA"        "VIL1"       "CDC6"       "ACSL5"      "EPCAM"
##  [31] "CENPF"      "ST14"       "HEPACAM2"   "AKR7A3"     "LAMC2"
##  [36] "CYP2C18"    "CSF2"       "PLA2G12B"   "UTS2"       "LRRC15"
##  [41] "EDAR"       "SI"         "CKMT1A"     "SLC9A2"     "FGF19"
##  [46] "CCL7"       "IL22RA2"    "LRRC19"     "KRT80"      "MMP7"
##  [51] "SLC22A11"   "NOS2"       "DIAPH3"     "NOTUM"      "GTSE1"
##  [56] "MUC5B"      "NEIL3"      "KCNQ1"      "SULT1C2"    "POF1B"
##  [61] "EPHB2"      "GALNT4"     "RRM2"       "CLDN3"      "FAM3D"
##  [66] "REG4"       "TTPA"       "TK1"        "CKMT1B"     "GIF"
##  [71] "SERPINB5"   "CDC20"      "C4BPB"      "SERPINA7"   "KRT23"
##  [76] "KRT75"      "SDC1"       "IL26"       "TCN1"       "PCSK9"
##  [81] "SLC5A1"     "TFF3"       "ALDOB"      "LAIR2"      "AGR2"
##  [86] "MAD2L1"     "FABP6"      "PKP3"       "HIST1H2BL"  "S100P"
##  [91] "KIF2C"      "KRT12"      "ZG16"       "LAD1"       "CA9"
##  [96] "HSD17B2"    "KNL1"       "PRSS3"      "HMGA2"      "PIGR"
## [101] "CDH3"       "PLS1"       "E2F2"       "KIF11"      "CEACAM6"
## [106] "TFF1"       "SERPINA6"   "S100A14"    "TNFSF11"    "FABP2"
## [111] "EPHB3"      "SPINK1"     "MELK"       "MNX1"       "IQGAP3"
## [116] "CDHR5"      "SPINT2"     "EREG"       "CASP5"      "C4BPA"
## [121] "MMP3"       "SMPDL3B"    "GALNT5"     "ATOH1"      "RETNLB"
## [126] "MMP13"      "CST2"       "PRSS8"      "AGMAT"      "KLK1"
## [131] "CDH17"      "TFF2"       "MET"        "TINAG"      "GUCA2B"
## [136] "CCNB2"      "CDK1"       "CXCL11"     "OIT3"       "CXCL1"
## [141] "REG1A"      "LMNB1"      "CFTR"       "DKK4"       "DDC"
## [146] "LSR"        "PMCH"       "DEFA5"      "AREG"       "CGN"
## [151] "CDH1"       "KIFC1"      "NUF2"       "MMP10"      "CXCL3"
## [156] "AKR1C4"     "MMP12"      "CXCL10"     "GDF15"      "PF4"
## [161] "MYH14"      "FABP1"      "GPA33"      "IL1A"       "HIST1H1B"
## [166] "DSC2"       "GGH"        "SFN"        "DEFA6"      "COL10A1"
## [171] "UGT1A1"     "LGALS4"     "CXCL5"      "CDC25A"     "KIF14"
## [176] "LIPG"       "SHH"        "CCL15"      "KRT19"      "SEMG1"
## [181] "CXADR"      "KRT20"      "CTSV"       "MYO1A"      "GUCA2A"
## [186] "CDC45"      "CEACAM1"    "ADM2"       "SULT1B1"    "GALNT6"
## [191] "GGT6"       "NFE2L3"     "PPBP"       "TNFSF15"    "CCL24"
## [196] "ANXA13"     "KRT18"      "GRP"        "REG3A"      "EPYC"
## [201] "ADAMDEC1"   "CYP2B6"     "IHH"        "CLCA1"      "KIF15"
## [206] "CDCP1"      "CST1"       "CCL20"      "DSG2"       "F12"
## [211] "CST4"       "GAL"        "CPN1"       "RBP2"       "IL17A"
## [216] "AGR3"       "GZMB"       "DSG4"       "LCN2"       "SERPINA4"
## [221] "AMN"        "LY75"       "AURKA"      "IL1B"       "GALNT3"
## [226] "C1orf116"   "NECTIN4"    "CCL18"      "JCHAIN"     "AURKB"
## [231] "MMP9"       "MMP1"       "CCL22"      "ZG16B"      "CCNB1"
## [236] "FCAMR"      "CDHR2"      "CXCL13"     "SPINT1"     "ST6GALNAC1"
## [241] "TNFRSF9"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "ALB"        "F2"         "C9"         "SERPINA6"   "PON1"
##   [6] "SERPINA7"   "PLTP"       "CFB"        "C1QTNF7"    "C1QTNF12"
##  [11] "IL1RL1"     "ADGRE5"     "ENHO"       "ISG15"      "MMP3"
##  [16] "TNFRSF1B"   "TNFSF15"    "SCUBE3"     "GDF15"      "PTX3"
##  [21] "TNFSF13B"   "CCL2"       "CXCL11"     "VEGFD"      "UTS2"
##  [26] "MDK"        "BMP6"       "IL1R2"      "IL24"       "IL32"
##  [31] "SERPINF2"   "CTF1"       "IL7"        "NPPC"       "IL2"
##  [36] "EDN1"       "SERPINC1"   "C7"         "CPN2"       "CPB2"
##  [41] "BCHE"       "ITIH3"      "ACTBL2"     "POTEJ"      "C8G"
##  [46] "KRT8"       "KRT4"       "KRT19"      "MMRN1"      "PEBP4"
##  [51] "DSG2"       "GRN"        "TIMP2"      "MYH14"      "FLNC"
##  [56] "HSPD1"      "PLS1"       "B3GNT2"     "B3GNT8"     "VIL1"
##  [61] "KRT20"      "COLEC11"    "ACTN3"      "CGN"        "MEPE"
##  [66] "CEACAM5"    "FLG2"       "CEACAM8"    "DSC2"       "LY6G6F"
##  [71] "NBL1"       "UROD"       "ASS1"       "ATP1B1"     "ROCK2"
##  [76] "CDH17"      "PCK2"       "MPST"       "LRRC4B"     "AL035078.4"
##  [81] "LSR"        "CNTN6"      "SFRP1"      "GMFG"       "SEMA3A"
##  [86] "S100P"      "VMO1"       "IL1R1"      "MEP1A"      "CPD"
##  [91] "HMGCS2"     "THSD1"      "AOC1"       "CLDN5"      "EPHX2"
##  [96] "GALNT5"     "MATN2"      "MXRA8"      "PTP4A2"     "POF1B"
## [101] "DDC"        "ATP1A1"     "PLOD3"      "TRIM58"     "IGSF10"
## [106] "TNFAIP8"    "DCTPP1"     "AGMAT"      "HGS"        "S100A10"
## [111] "UFC1"       "FREM2"      "NTN1"       "CHRD"       "TOMM70"
## [116] "NELL1"      "ADAMTSL3"   "S100A14"    "SIRPB2"     "LGALS4"
## [121] "GLCE"       "AMACR"      "ADGRL1"     "PPP5C"      "MCM3"
## [126] "ASAP1"      "RHEB"       "PDE3A"      "AGR2"       "CGREF1"
## [131] "SBSPON"     "NDUFS2"     "TRIP10"     "CLDN3"      "SLFN14"
## [136] "TNIK"       "MAN2C1"     "NXPE3"      "NCR3LG1"    "EPCAM"
## [141] "AMN"        "KAZALD1"    "ADAMTS15"   "KLK14"      "DRAXIN"
## [146] "IFNLR1"     "TREM1"      "NBN"        "FAM19A5"    "TP53"
## [151] "SPOCK1"     "EFNA4"      "ADAMTS12"   "ADAMTS7"    "ADGRF3"
## [156] "ADGRG3"     "ADGRG4"     "ATP11C"     "AURKB"      "C14orf39"
## [161] "C9orf3"     "CACNA1D"    "CACNA1G"    "CDC20"      "CDH12"
## [166] "CDK1"       "CELSR3"     "COL19A1"    "DDT"        "DNAH8"
## [171] "DONSON"     "DTNB"       "EZH2"       "FAM9B"      "FOXD4L1"
## [176] "GDF5"       "GPA33"      "GRIN1"      "HCFC1"      "HPS5"
## [181] "IFITM1"     "IQGAP3"     "IRX1"       "KBTBD11"    "KCNN2"
## [186] "KCTD4"      "KIAA0355"   "KIF14"      "KIF22"      "KIF4B"
## [191] "LAD1"       "LAMC2"      "LBR"        "LIMD1"      "LMO7"
## [196] "LRRC3"      "MEA1"       "MKI67"      "NEUROG3"    "NKX2-8"
## [201] "NT5C1B"     "PCDH7"      "PCDH8"      "PCLO"       "PCNA"
## [206] "PGBD2"      "PIK3C2B"    "PKD1L1"     "PKD2L1"     "PLIN2"
## [211] "POLR3A"     "POTEE"      "PPP6R2"     "PRAMEF12"   "PRDM15"
## [216] "PROKR2"     "SALL3"      "SLC4A9"     "SPAG9"      "SULT1B1"
## [221] "TCHH"       "TEX264"     "TOP2A"      "TUBB8"      "UNC5D"
## [226] "WDR7"       "ZDHHC15"    "ZFHX2"      "ZNF197"     "ZNF232"
## [231] "ZNF300"     "ZNF445"     "ZNF507"     "ZNF566"     "ZNF780A"
```

Step 3: Select the markers short listed in both approaches

Hide

```
colorectal_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_marke
rs)
colorectal_cancer_genesInBoth
```

```
##  [1] "MEP1A"    "HMGCS2"   "TOP2A"    "VIL1"     "EPCAM"    "LAMC2"
##  [7] "UTS2"     "POF1B"    "CLDN3"    "CDC20"    "SERPINA7" "AGR2"
## [13] "S100P"    "LAD1"     "PLS1"     "SERPINA6" "S100A14"  "IQGAP3"
## [19] "MMP3"     "GALNT5"   "AGMAT"    "CDH17"    "CDK1"     "CXCL11"
## [25] "DDC"      "LSR"      "CGN"      "GDF15"    "MYH14"    "GPA33"
## [31] "DSC2"     "LGALS4"   "KIF14"    "KRT19"    "KRT20"    "SULT1B1"
## [37] "TNFSF15"  "DSG2"     "AMN"      "AURKB"
```

2.3 Prostate cancer

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'prostate'

cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "SERPINB11"  "HAO1"       "RLN1"       "OR51E2"     "NOX1"
##   [6] "WNK4"       "XDH"        "ADAM7"      "HMGCS2"     "EPHA7"
##  [11] "SHROOM2"    "C14orf39"   "FBP1"       "MLPH"       "CXCL9"
##  [16] "ERBB3"      "REEP6"      "NAAA"       "HPN"        "KRT36"
##  [21] "LTF"        "DSP"        "SLC52A3"    "CCL11"      "CCL17"
##  [26] "TDRD1"      "KLK12"      "TGM3"       "FAM3B"      "PRSS21"
##  [31] "EPCAM"      "ST14"       "HEPACAM2"   "MZB1"       "PDLIM5"
##  [36] "WFDC12"     "ARHGAP6"    "RLN2"       "CRISP3"     "SCUBE2"
##  [41] "UTS2"       "F5"         "POTEE"      "SI"         "CKMT1A"
##  [46] "SLC9A2"     "CHGA"       "SALL3"      "PIP"        "AQP3"
##  [51] "MMP7"       "HGD"        "TNFSF10"    "SLC12A3"    "ORM2"
##  [56] "APOF"       "GGT1"       "CLDN3"      "FAM3D"      "REG4"
##  [61] "CD177"      "ENPP5"      "UNC5D"      "CALCA"      "SERPINB4"
##  [66] "AZGP1"      "CALR3"      "KLK2"       "SYTL1"      "SERPINA7"
##  [71] "EYA2"       "TUSC3"      "KRT23"      "CLUL1"      "SEMG2"
##  [76] "SDC1"       "PTK7"       "GALNT7"     "SLC5A1"     "TFF3"
##  [81] "GP2"        "ATP12A"     "UMOD"       "LAIR2"      "AGR2"
##  [86] "NPY"        "BPIFB2"     "IRX4"       "SCGN"       "SFRP4"
##  [91] "PKP3"       "HIST1H2BL"  "AMACR"      "SLITRK6"    "FOXI1"
##  [96] "CNMD"       "NANS"       "MSMB"       "GCNT1"      "H2AFJ"
## [101] "INSL5"      "LAD1"       "PIGR"       "CDH3"       "DNAH5"
## [106] "CXCL17"     "CANT1"      "OR51F2"     "TRPM4"      "TFF1"
## [111] "SEC14L2"    "SDK2"       "FOLH1"      "KLK11"      "VSTM2A"
## [116] "PCDH8"      "GCG"        "HIST1H2BC"  "PLA2G2A"    "SORD"
## [121] "MARCKSL1"   "CDH26"      "MME"        "LPA"        "SERPINA11"
## [126] "DNAH8"      "KIF5C"      "MNX1"       "DHRS7"      "SPINT2"
## [131] "DSC3"       "KCNN2"      "SMPDL3B"    "ATOH1"      "SCGB1A1"
## [136] "ACPP"       "MMP13"      "CST2"       "PRSS8"      "RAB27B"
## [141] "HIST2H4A"   "CFC1"       "PODXL2"     "SMS"        "SLCO1A2"
## [146] "DPP4"       "KLK3"       "LCP1"       "LAMP3"      "SPON2"
## [151] "CRISPLD1"   "CXCL11"     "SHROOM3"    "LRRN1"      "CD207"
## [156] "CCL25"      "CHIT1"      "DDC"        "CUX2"       "MDK"
## [161] "LSR"        "CGN"        "CDH1"       "RDH11"      "MMP10"
## [166] "MMP12"      "CXCL10"     "IGFBP2"     "GDF15"      "CRYM"
## [171] "NCAPD3"     "ENTPD5"     "DSC2"       "RGS10"      "SFN"
## [176] "CXCL6"      "SPTBN2"     "COL10A1"    "CST5"       "BCAM"
## [181] "HIST3H2A"   "HIST1H4D"   "HIST1H4B"   "CCNO"       "SHH"
## [186] "VWA1"       "PEBP4"      "IL5RA"      "GSTA1"      "KRT19"
## [191] "SEMG1"      "CXADR"      "BPIFB1"     "PLA2G7"     "IQGAP2"
## [196] "HIST1H1C"   "ADM2"       "GGT6"       "IDH1"       "SLC1A5"
## [201] "TNFSF15"    "KRT18"      "POTEJ"      "GRP"        "DCDC2"
## [206] "TGM4"       "ADAMDEC1"   "CCK"        "GNMT"       "HIST1H4J"
## [211] "CAMSAP3"    "PRIM2"      "IHH"        "CST1"       "DSG2"
## [216] "CST4"       "KRT17"      "GAL"        "GOLM1"      "AGR3"
## [221] "LCN2"       "VSIG2"      "ERVMER34-1" "TACSTD2"    "GALNT3"
## [226] "KAZALD1"    "C1orf116"   "NECTIN4"    "CCL18"      "TPD52"
## [231] "SEZ6L2"     "COL9A2"     "CCL22"      "SSTR5"      "ZG16B"
## [236] "SLC30A4"    "PPP3CA"     "CXCL13"     "SPINT1"     "ABCC4"
## [241] "ST6GALNAC1"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "ALB"       "CP"        "F2"        "AZGP1"     "PROS1"     "FCN3"
##   [7] "CST3"      "CFB"       "MASP2"     "C1QTNF7"   "IL1RL1"    "AFP"
##  [13] "ISG15"     "MMP3"      "TNFRSF1B"  "MMP10"     "GDF15"     "IL22RA2"
##  [19] "PLA2G2A"   "TNFSF13B"  "ACPP"      "CCL2"      "CXCL11"    "CX3CL1"
##  [25] "UTS2"      "BMP6"      "F3"        "IL1R2"     "TNFSF10"   "IL27"
##  [31] "IL32"      "TNFSF14"   "SERPINF2"  "CTF1"      "NPPC"      "IL2"
##  [37] "EDN1"      "SERPINC1"  "C7"        "CPN2"      "CPB2"      "BCHE"
##  [43] "PZP"       "ITIH3"     "PRG4"      "ALCAM"     "POTEJ"     "C8G"
##  [49] "KRT8"      "KRT4"      "HRNR"      "DPP4"      "IGF2R"     "KRT18"
##  [55] "YWHAQ"     "B4GALT1"   "GRN"       "TIMP2"     "GOLM1"     "FUCA1"
##  [61] "HEXB"      "RAB27B"    "FSTL1"     "COLEC11"   "ACTN3"     "GNAQ"
##  [67] "RAB1A"     "FLG2"      "IDH1"      "ESD"       "DSC2"      "LTBP2"
##  [73] "TGFB1I1"   "MANBA"     "NBL1"      "RDH11"     "FTH1"      "HSD17B4"
##  [79] "PTPRB"     "SYTL4"     "UROD"      "ENTPD5"    "SPTBN2"    "NELL2"
##  [85] "PXDN"      "MSMB"      "PRKCB"     "GAA"       "NAGA"      "TTLL12"
##  [91] "LYPLA1"    "SFRP1"     "GLO1"      "SPON2"     "FASN"      "VWCE"
##  [97] "SEMA3A"    "VMO1"      "IL1R1"     "GLA"       "KIAA0319L" "HMGCS2"
## [103] "CLDN5"     "SDF4"      "EPHX2"     "CD151"     "MATN2"     "CPE"
## [109] "MXRA8"     "PTP4A2"    "POF1B"     "SORD"      "TXNRD2"    "GPC6"
## [115] "HEBP2"     "TRIM58"    "ENPP5"     "IGSF10"    "HGS"       "NANS"
## [121] "FAM177A1"  "FREM2"     "AMIGO2"    "KITLG"     "OSTM1"     "TOMM70"
## [127] "NELL1"     "MARCKSL1"  "PRTN3"     "CRYM"      "SCAMP2"    "SIRPB2"
## [133] "B4GALT3"   "PGM3"      "DCUN1D1"   "AMACR"     "ACP6"      "ADGRL1"
## [139] "DHRS7"     "UGDH"      "TOLLIP"    "ENTPD4"    "REEP6"     "GALNT15"
## [145] "PDE3A"     "AGR2"      "CGREF1"    "SBSPON"    "NDUFS2"    "VPS28"
## [151] "HYAL1"     "CLDN3"     "TOM1"      "PUDP"      "SLFN14"    "B4GALT2"
## [157] "ALDH6A1"   "C1orf116"  "NCR3LG1"   "EPCAM"     "AMN"       "KAZALD1"
## [163] "IL12A"     "NAAA"      "ADAMTS15"  "KLK14"     "DRAXIN"    "SPOCK1"
## [169] "ADCY3"     "ADGRG3"    "ADGRG4"    "AR"        "AUTS2"     "CACNA1G"
## [175] "CDH12"     "COL19A1"   "CSMD2"     "DDT"       "DNAH8"     "DONSON"
## [181] "DOT1L"     "DTNB"      "FAM83C"    "FOLH1"     "FOXD4L1"   "HS3ST5"
## [187] "IKBKB"     "IQGAP3"    "IRX1"      "KBTBD11"   "KCNN2"     "KCTD4"
## [193] "KIF22"     "KIF4B"     "KLHL8"     "KLK2"      "KLK3"      "LMO7"
## [199] "LRRC3"     "NEUROG3"   "NKX2-8"    "ODC1"      "PIK3C2B"   "PKHD1"
## [205] "POLR3A"    "POTEE"     "PPP1R9A"   "PRAMEF12"  "PRDM15"    "PROKR2"
## [211] "PTPRR"     "RAB3D"     "SALL3"     "SLC1A5"    "SLC25A22"  "SLC4A9"
## [217] "SSFA2"     "SYTL1"     "SYTL2"     "TPD52L1"   "TUBB8"     "UNC5D"
## [223] "URB2"      "ZNF232"    "ZNF300"    "ZNF445"    "ZNF502"    "ZNF507"
## [229] "ZNF512B"   "ZNF566"    "ZNF780A"
```

Step 3: Select the markers short listed in both approaches

Hide

```
prostate_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_marker
s)
prostate_cancer_genesInBoth
```

```
##  [1] "HMGCS2"   "REEP6"    "NAAA"     "EPCAM"    "UTS2"     "POTEE"
##  [7] "SALL3"    "TNFSF10"  "CLDN3"    "ENPP5"    "UNC5D"    "AZGP1"
## [13] "KLK2"     "SYTL1"    "AGR2"     "AMACR"    "NANS"     "MSMB"
## [19] "FOLH1"    "PLA2G2A"  "SORD"     "MARCKSL1" "DNAH8"    "DHRS7"
## [25] "KCNN2"    "ACPP"     "RAB27B"   "DPP4"     "KLK3"     "SPON2"
## [31] "CXCL11"   "RDH11"    "MMP10"    "GDF15"    "CRYM"     "ENTPD5"
## [37] "DSC2"     "SPTBN2"   "IDH1"     "SLC1A5"   "KRT18"    "POTEJ"
## [43] "GOLM1"    "KAZALD1"  "C1orf116"
```

2.4 Skin cancer

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'skin'

cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "ST3GAL6"   "ESM1"      "MELTF"     "WNK4"      "RGS20"     "CSPG4"
##   [7] "LYZ"       "MMP8"      "LSAMP"     "CENPE"     "CTHRC1"    "PBK"
##  [13] "ECM1"      "FKBP10"    "PLOD3"     "DLGAP5"    "EXTL1"     "SHROOM2"
##  [19] "TTK"       "MLPH"      "CXCL9"     "PTPRZ1"    "ERBB3"     "CLEC4C"
##  [25] "TOP2A"     "UCN2"      "KRT36"     "SDCBP"     "LTB"       "CCL17"
##  [31] "CHEK1"     "SIT1"      "IFNG"      "CDC6"      "PLXNC1"    "CENPF"
##  [37] "MZB1"      "PAFAH1B3"  "VAT1"      "IL24"      "KIAA1549L" "HAMP"
##  [43] "CSF2"      "UTS2"      "LRRC15"    "MIA"       "ARPC1B"    "CCL7"
##  [49] "FEN1"      "PPY"       "GZMH"      "TXNDC5"    "LINGO1"    "BCHE"
##  [55] "APOC1"     "DIAPH3"    "L1CAM"     "PLAT"      "GTSE1"     "NEIL3"
##  [61] "SULT1C2"   "FCGR2A"    "CA6"       "CCL19"     "CCL5"      "MMP14"
##  [67] "PRAME"     "RRM2"      "TNFRSF17"  "RENBP"     "FCMR"      "CD48"
##  [73] "GZMA"      "TK1"       "OTOA"      "TPTE"      "CDC20"     "PIR"
##  [79] "GPNMB"     "GPR37"     "PKLR"      "TCN1"      "BAMBI"     "APOC4"
##  [85] "LZTS1"     "CD68"      "CD276"     "LAIR2"     "MAD2L1"    "SPP1"
##  [91] "TUBB3"     "HIST1H2BL" "HMCN1"     "KIF2C"     "CDK2"      "KRT12"
##  [97] "HLA-G"     "PTPRJ"     "CPXM1"     "PCDH7"     "KNL1"      "MYO5A"
## [103] "DUSP6"     "HMGA2"     "CDH3"      "E2F2"      "KIF11"     "TYRP1"
## [109] "PNPLA3"    "QPRT"      "TNFRSF19"  "FXYD5"     "CD5L"      "TNFSF11"
## [115] "IGSF3"     "CA14"      "ADGRG1"    "MARCKSL1"  "CPB2"      "KPNA2"
## [121] "MELK"      "ACP5"      "IQGAP3"    "VAX2"      "TYMS"      "MMP13"
## [127] "CST2"      "C1QA"      "KRT27"     "MET"       "LAMP3"     "CCNB2"
## [133] "SLC34A1"   "CDK1"      "CRISPLD1"  "RXRG"      "CXCL11"    "OIT3"
## [139] "CRTAC1"    "DUSP4"     "S100B"     "SLAMF7"    "ATP1A1"    "NRP2"
## [145] "FCRL3"     "LMNB1"     "CHIT1"     "IL21"      "MYO10"     "SNCA"
## [151] "TNFSF9"    "PMCH"      "CLEC6A"    "LRP2"      "CDH1"      "KIFC1"
## [157] "NUF2"      "CD3G"      "NLGN1"     "HOXC12"    "MMP10"     "MMP12"
## [163] "CD27"      "CXCL10"    "GDF15"     "CD63"      "HOXC11"    "LGALS3BP"
## [169] "HIST1H1B"  "FCRLB"     "GGH"       "CCL8"      "CCL3"      "ACTBL2"
## [175] "COL10A1"   "QPCT"      "GAPDHS"    "APOE"      "PMEL"      "WARS"
## [181] "CDC25A"    "FASLG"     "KIF14"     "ICAM1"     "MCM2"      "CTSV"
## [187] "PLA2G7"    "CDC45"     "CEACAM1"   "ADM2"      "BCAN"      "ACOT7"
## [193] "C1QC"      "CDH19"     "NCR1"      "COL1A1"    "EPYC"      "LILRB4"
## [199] "ADAMDEC1"  "S100A1"    "DKKL1"     "DLL3"      "HIST1H4J"  "LEF1"
## [205] "KIF15"     "CST1"      "IL2RA"     "NES"       "CCL20"     "HMGA1"
## [211] "CPN1"      "NME1"      "GZMB"      "C2"        "MCM5"      "AURKA"
## [217] "CHST11"    "GBA"       "E2F1"      "CABLES1"   "CCL18"     "CD70"
## [223] "JCHAIN"    "AURKB"     "MMP9"      "POSTN"     "MAGEC1"    "SERPINE2"
## [229] "CALU"      "MMP1"      "SEZ6L2"    "IFI30"     "CCL22"     "CCNB1"
## [235] "C1QB"      "FCAMR"     "VGF"       "TREM2"     "CXCL13"    "TNFRSF9"
## [241] "TCL1A"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "ALB"       "C4BPB"     "SERPINA6"  "PON1"      "HPR"       "PROS1"
##   [7] "PLTP"      "SERPINF1"  "CLEC3B"    "TFRC"      "CFB"       "TNC"
##  [13] "C1QTNF7"   "HP"        "SDC1"      "IGFBP7"    "ISG15"     "MMP3"
##  [19] "THBD"      "RNASE3"    "SCUBE3"    "MMP10"     "TNFSF13B"  "CXCL11"
##  [25] "CCL4"      "CX3CL1"    "MDK"       "BMP6"      "F3"        "IL1R2"
##  [31] "S100A9"    "IL24"      "IL32"      "IL7"       "NPPC"      "EDN1"
##  [37] "CPN2"      "CPB2"      "BCHE"      "KRT6A"     "KRT5"      "FLNA"
##  [43] "CD44"      "KRT75"     "KRT14"     "HSPG2"     "POTEJ"     "KRT13"
##  [49] "KRT79"     "KRT17"     "KRT76"     "WDR1"      "SLC3A2"    "DSG1"
##  [55] "GRN"       "TIMP2"     "CALML5"    "SFN"       "RAC1"      "SNX2"
##  [61] "FLNB"      "HIST2H4B"  "ACTN3"     "MANSC1"    "FABP5"     "SRC"
##  [67] "RAB1A"     "DSC1"      "DBN1"      "S100A7"    "STC2"      "NRP2"
##  [73] "SNED1"     "HPRT1"     "LY6G6F"    "TNN"       "S100A11"   "NBL1"
##  [79] "LYPD3"     "CDON"      "CD276"     "KDR"       "PLEC"      "TMEM40"
##  [85] "FSTL3"     "CHRDL2"    "FSCN1"     "PKP1"      "DSC3"      "DPP3"
##  [91] "PLD4"      "HSD17B10"  "VMO1"      "NDST1"     "IL1R1"     "CTNNB1"
##  [97] "IQGAP1"    "TMX1"      "SERPINH1"  "SERPINB5"  "FGFR2"     "CALML3"
## [103] "MXRA8"     "PCK1"      "S100A10"   "ITGA3"     "TNFRSF12A" "PITPNB"
## [109] "S100A2"    "CHRD"      "PRTN3"     "ADAMTSL3"  "S100A14"   "PITPNA"
## [115] "SIRPB2"    "S100A16"   "BCAT1"     "CLCA2"     "HNRNPR"    "KPNA2"
## [121] "MCM3"      "ATP2C1"    "REEP6"     "CGREF1"    "OAS3"      "PPME1"
## [127] "LAMB3"     "DDX21"     "SLC25A24"  "NXPE3"     "SWAP70"    "AMN"
## [133] "KAZALD1"   "ADAMTS15"  "DRAXIN"    "IFNLR1"    "TREM1"     "ITGB6"
## [139] "EFNA4"     "ADAMTS7"   "ADGRG3"    "ADGRG4"    "ALK"       "AURKB"
## [145] "BACH1"     "BCL11A"    "BCL11B"    "C14orf39"  "CACNA1G"   "CALCRL"
## [151] "CCNB1"     "CDC20"     "CDC7"      "CDK1"      "CDK4"      "CENPE"
## [157] "CLPB"      "CNMD"      "COL19A1"   "CSMD2"     "DDX47"     "DNAH8"
## [163] "DONSON"    "EFS"       "EZH2"      "FAM206A"   "FAM9B"     "FOSL1"
## [169] "GRM1"      "HIPK1"     "HMGB3"     "HOXC11"    "IQGAP3"    "ITPR3"
## [175] "KCNN2"     "KIF14"     "KIF4B"     "KLF7"      "KRT24"     "KRT6B"
## [181] "KRT6C"     "LAMC2"     "LIG1"      "LMO7"      "LRRC3"     "MCM7"
## [187] "MKI67"     "MX2"       "NASP"      "NEUROG3"   "OSBPL3"    "PARD3B"
## [193] "PCDHB7"    "PCNA"      "PHF20"     "PKD1L1"    "PKD2L1"    "PKP3"
## [199] "POLR3A"    "POTEE"     "PRAMEF12"  "PRDM15"    "PREX2"     "PROKR2"
## [205] "PRPF40A"   "RIOK2"     "RNF6"      "RXRG"      "SHANK2"    "SHOC2"
## [211] "SIAH2"     "SLC1A4"    "SLC4A9"    "SPAG9"     "SYNPO2L"   "TFAP2B"
## [217] "TOP2A"     "TOX4"      "TP63"      "TTK"       "TUBB8"     "TUSC3"
## [223] "URB2"      "WNK1"      "WRNIP1"    "ZFP36L1"   "ZNF184"    "ZNF232"
## [229] "ZNF300"    "ZNF333"    "ZNF445"    "ZNF507"    "ZNF512B"   "ZNF566"
## [235] "ZNF780A"
```

Step 3: Select the markers short listed in both approaches

Hide

```
skin_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_markers)
skin_cancer_genesInBoth
```

```
##  [1] "CENPE"  "TTK"    "TOP2A"  "IL24"   "BCHE"   "CDC20"  "CD276"  "CPB2"
##  [9] "KPNA2"  "IQGAP3" "CDK1"   "RXRG"   "CXCL11" "NRP2"   "MMP10"  "HOXC11"
## [17] "KIF14"  "AURKB"  "CCNB1"
```

2.5 Stomach cancer

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'stomach'

cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "ESM1"       "NOX1"       "MEP1A"      "LYZ"        "GC"
##   [6] "CXCL8"      "CENPE"      "XDH"        "PBK"        "MUC1"
##  [11] "HMGCS2"     "DLGAP5"     "TTK"        "FBP1"       "CXCL9"
##  [16] "AKR1B10"    "PI3"        "ERBB3"      "HIST1H4L"   "TOP2A"
##  [21] "SLC52A3"    "CCL11"      "CCL17"      "IFNG"       "GDA"
##  [26] "FAM3B"      "VIL1"       "CDC6"       "PRSS21"     "EPCAM"
##  [31] "CENPF"      "ST14"       "HEPACAM2"   "MZB1"       "AKR7A3"
##  [36] "LAMC2"      "CYP2C18"    "CSF2"       "UTS2"       "LRRC15"
##  [41] "HIST1H2BI"  "EDAR"       "POTEE"      "SI"         "CKMT1A"
##  [46] "SLC9A2"     "FGF19"      "CCL7"       "IL22RA2"    "LRRC19"
##  [51] "GAST"       "MMP7"       "HGD"        "DIAPH3"     "NOTUM"
##  [56] "GTSE1"      "MUC5B"      "NEIL3"      "SULT1C2"    "POF1B"
##  [61] "EPHB2"      "GALNT4"     "HIST1H4A"   "RRM2"       "CLDN3"
##  [66] "FAM3D"      "TNFRSF17"   "REG4"       "TK1"        "CKMT1B"
##  [71] "GIF"        "SERPINB5"   "CDC20"      "C4BPB"      "SDC1"
##  [76] "IL26"       "TCN1"       "PCSK9"      "SLC5A1"     "TFF3"
##  [81] "LAIR2"      "AGR2"       "IL17C"      "PKP3"       "HIST1H2BL"
##  [86] "S100P"      "KIF2C"      "MSMB"       "HABP2"      "LAD1"
##  [91] "CA9"        "HSD17B2"    "KNL1"       "PRSS3"      "HMGA2"
##  [96] "PIGR"       "CDH3"       "PLS1"       "E2F2"       "KIF11"
## [101] "CXCL17"     "CEACAM6"    "TFF1"       "S100A14"    "TNFSF11"
## [106] "FABP2"      "MXRA5"      "SPINK1"     "MELK"       "MNX1"
## [111] "IQGAP3"     "CDHR5"      "CASP5"      "MMP3"       "GKN1"
## [116] "SMPDL3B"    "GALNT5"     "ATOH1"      "RETNLB"     "VNN1"
## [121] "ENPP7"      "MMP13"      "CST2"       "CYP2C19"    "PRSS8"
## [126] "GREM1"      "RAB27B"     "AGMAT"      "C6orf58"    "CDH17"
## [131] "TFF2"       "LAMP3"      "TINAG"      "GUCA2B"     "CCNB2"
## [136] "CDK1"       "CXCL11"     "OIT3"       "SLAMF7"     "CXCL1"
## [141] "REG1A"      "CCL25"      "IL21"       "DDC"        "MDK"
## [146] "LSR"        "PMCH"       "CGN"        "CLEC6A"     "CDH1"
## [151] "KIFC1"      "NUF2"       "CD3G"       "MMP10"      "CXCL3"
## [156] "AKR1C4"     "MMP12"      "CXCL10"     "GDF15"      "FABP1"
## [161] "GPA33"      "IL19"       "HOXC11"     "IL1A"       "HIST1H1B"
## [166] "SFN"        "CXCL6"      "ACTBL2"     "COL10A1"    "UGT1A1"
## [171] "LGALS4"     "CXCL5"      "FASLG"      "KIF14"      "HIST1H4B"
## [176] "SHH"        "CCL15"      "GSTA1"      "KRT19"      "SEMG1"
## [181] "CXADR"      "BPIFB1"     "KRT20"      "CTSV"       "PLA2G7"
## [186] "HIST1H4C"   "MYO1A"      "CDC45"      "CEACAM1"    "ADM2"
## [191] "GALNT6"     "GGT6"       "TNFSF15"    "CCL24"      "ANXA13"
## [196] "KRT18"      "COL1A1"     "REG3A"      "EPYC"       "ADAMDEC1"
## [201] "CYP2B6"     "IHH"        "CDCP1"      "CST1"       "PIK3C2G"
## [206] "IL2RA"      "CCL20"      "DSG2"       "CST4"       "RBP2"
## [211] "IL17A"      "AGR3"       "GZMB"       "LCN2"       "POU2AF1"
## [216] "SERPINA4"   "TACSTD2"    "AMN"        "FCRL5"      "AURKA"
## [221] "IL1B"       "GALNT3"     "ALPP"       "IFNL1"      "C1orf116"
## [226] "NECTIN4"    "CCL18"      "CD70"       "JCHAIN"     "AURKB"
## [231] "MMP9"       "MMP1"       "CCL22"      "ZG16B"      "FCAMR"
## [236] "CDHR2"      "CXCL13"     "SPINT1"     "ST6GALNAC1" "TNFRSF9"
## [241] "TCL1A"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "CP"         "F2"         "C9"         "PON1"       "SERPINA7"
##   [6] "LRG1"       "PROS1"      "PLTP"       "LYZ"        "CFB"
##  [11] "C1QTNF7"    "C1QTNF12"   "CD55"       "IL1RL1"     "MFGE8"
##  [16] "CCL14"      "ISG15"      "TNFRSF1B"   "SCUBE3"     "MMP10"
##  [21] "TNFSF13B"   "CCL2"       "CXCL11"     "VEGFD"      "UTS2"
##  [26] "HAVCR1"     "BMP6"       "IL32"       "IL36B"      "SERPINF2"
##  [31] "NPPC"       "PGLYRP2"    "C7"         "CPN2"       "CPB2"
##  [36] "BCHE"       "ITIH3"      "PIGR"       "POTEJ"      "KRT8"
##  [41] "KRT4"       "CFL1"       "TPM3"       "KRT7"       "KRT19"
##  [46] "TPM2"       "WDR1"       "CDHR5"      "KRT18"      "GRN"
##  [51] "TIMP2"      "GOLM1"      "FUCA2"      "WARS"       "HSPD1"
##  [56] "PLS1"       "VIL1"       "FLNB"       "FAP"        "KRT20"
##  [61] "COLEC11"    "ACTN3"      "GNAQ"       "CGN"        "MEPE"
##  [66] "CEACAM5"    "RAB1A"      "FLG2"       "CTSS"       "CEACAM8"
##  [71] "INF2"       "DSC2"       "LY6G6F"     "TNN"        "ASL"
##  [76] "NBL1"       "SPINK1"     "CHRDL2"     "UROD"       "GALNT7"
##  [81] "MUC1"       "ASS1"       "ROCK2"      "CNTN6"      "TSKU"
##  [86] "CKMT1A"     "CKMT1B"     "COL4A2"     "GMFG"       "PRKACB"
##  [91] "ATP5MF"     "DCN"        "VWCE"       "ST6GALNAC1" "SEMA3A"
##  [96] "S100P"      "VMO1"       "IL1R1"      "MEP1A"      "CPD"
## [101] "RANBP1"     "IQGAP1"     "HSPH1"      "HMGCS2"     "CLDN5"
## [106] "SDF4"       "ATP5PB"     "GALNT5"     "MATN2"      "MXRA8"
## [111] "POF1B"      "TRIM58"     "IGSF10"     "DCTPP1"     "SSBP1"
## [116] "CD86"       "GLDN"       "FREM2"      "NTN1"       "CHRD"
## [121] "TOMM70"     "NELL1"      "MARCKSL1"   "S100A14"    "SIRPB2"
## [126] "LGALS4"     "VSIG2"      "GLCE"       "DCHS1"      "AMACR"
## [131] "ADGRL1"     "PPP5C"      "MGST3"      "FEN1"       "KPNA2"
## [136] "AGR3"       "MCM3"       "VPS13A"     "GP2"        "REEP6"
## [141] "ASAP1"      "AGR2"       "CGREF1"     "ENTPD6"     "LAMB3"
## [146] "B4GALT2"    "TNIK"       "MAN2C1"     "SLC25A24"   "C1orf116"
## [151] "AMN"        "KAZALD1"    "ADAMTS15"   "CLEC4A"     "DRAXIN"
## [156] "IFNLR1"     "TREM1"      "CDH3"       "TP53"       "EFNA4"
## [161] "ADGRF3"     "ADGRG3"     "ADGRG4"     "ASIC2"      "AURKB"
## [166] "BBC3"       "C14orf39"   "CACNA1D"    "CACNA1G"    "CCNB1"
## [171] "CDAN1"      "CDK1"       "CELSR3"     "CLPB"       "COL19A1"
## [176] "CSMD2"      "DDT"        "DNAH8"      "DONSON"     "DOT1L"
## [181] "DTNB"       "EZH2"       "FAM9B"      "FER1L6"     "FOXD4L1"
## [186] "GDF5"       "HPS5"       "IQGAP3"     "IRX1"       "ITPR3"
## [191] "KBTBD11"    "KCNN2"      "KCTD4"      "KIAA0355"   "KIF14"
## [196] "KIF22"      "KIF4B"      "LAMA1"      "LAMC2"      "LMO7"
## [201] "LRRC3"      "MAD2L1"     "MKI67"      "MYO5B"      "MYT1"
## [206] "NEUROG3"    "NKX2-8"     "OSBPL3"     "PCDH8"      "PIK3C2B"
## [211] "PKD1L1"     "PKD2L1"     "PKP3"       "POLR3A"     "POTEE"
## [216] "PRAMEF12"   "PROKR2"     "SCRT1"      "SHOC2"      "SLC4A9"
## [221] "SLC5A5"     "SRP54"      "TCHH"       "TOP2A"      "WNK2"
## [226] "ZDHHC15"    "ZFHX2"      "ZNF232"     "ZNF300"     "ZNF445"
## [231] "ZNF507"     "ZNF566"     "ZNF780A"    "ZNFX1"
```

Step 3: Select the markers short listed in both approaches

Hide

```
stomach_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_markers)
stomach_cancer_genesInBoth
```

```
##  [1] "MEP1A"      "LYZ"        "MUC1"       "HMGCS2"     "TOP2A"
##  [6] "VIL1"       "LAMC2"      "UTS2"       "POTEE"      "CKMT1A"
## [11] "POF1B"      "CKMT1B"     "AGR2"       "PKP3"       "S100P"
## [16] "PIGR"       "CDH3"       "PLS1"       "S100A14"    "SPINK1"
## [21] "IQGAP3"     "CDHR5"      "GALNT5"     "CDK1"       "CXCL11"
## [26] "CGN"        "MMP10"      "LGALS4"     "KIF14"      "KRT19"
## [31] "KRT20"      "KRT18"      "AGR3"       "AMN"        "C1orf116"
## [36] "AURKB"      "ST6GALNAC1"
```

2.6 Lung cancer

2.6.1 Subtype: LUAD

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'lung'
cancer_subtype_code = 'luad'
cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name, cancer_subtyp
e_code)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "DNAH3"      "ESM1"       "CD1C"       "FGFBP1"     "LYZ"
##   [6] "SLPI"       "CXCL8"      "APOH"       "BPIFA1"     "XDH"
##  [11] "CTHRC1"     "PBK"        "MUC1"       "TREM1"      "DLGAP5"
##  [16] "TTK"        "FBP1"       "SCGB3A1"    "MLPH"       "CXCL9"
##  [21] "FCRL2"      "CLEC4C"     "HPN"        "TOP2A"      "FGL1"
##  [26] "CCL11"      "RLN3"       "LTB"        "CCL17"      "SIT1"
##  [31] "KLK12"      "IFNG"       "FAM3B"      "VIL1"       "SFTPD"
##  [36] "PRSS21"     "EPCAM"      "CENPF"      "ST14"       "RETN"
##  [41] "MZB1"       "LAMC2"      "RLN2"       "CYP2C18"    "CSF2"
##  [46] "PLA2G12B"   "UTS2"       "LRRC15"     "LRG1"       "AMY1C"
##  [51] "AQP3"       "CCL7"       "IL22RA2"    "GZMH"       "CCL1"
##  [56] "KRT80"      "MMP7"       "APOC1"      "NOTUM"      "CTSS"
##  [61] "ORM2"       "MUC5B"      "NEIL3"      "SULT1C2"    "AGRP"
##  [66] "CCL19"      "CCL5"       "PRAME"      "RRM2"       "CLDN3"
##  [71] "TNFRSF17"   "GZMA"       "TK1"        "NKX2-8"     "SPP2"
##  [76] "CD300LF"    "CDC20"      "C4BPB"      "CNGB1"      "SDC1"
##  [81] "TCN1"       "TFF3"       "LAIR2"      "AGR2"       "BPIFB2"
##  [86] "FABP6"      "SPP1"       "TUBB3"      "KRT39"      "PKP3"
##  [91] "AMY1B"      "S100P"      "MSMB"       "HABP2"      "LAD1"
##  [96] "CA9"        "HSD17B2"    "IGFBPL1"    "B3GNT7"     "PIGR"
## [101] "CDH3"       "PLS1"       "LRRN4"      "DNAH5"      "CXCL17"
## [106] "CEACAM6"    "NAPSA"      "TFF1"       "S100A14"    "TNFSF11"
## [111] "SCIN"       "CPB2"       "MXRA5"      "SPINK1"     "WFDC2"
## [116] "MELK"       "ACP5"       "LUM"        "MNX1"       "IQGAP3"
## [121] "SPINT2"     "C4BPA"      "ITGB6"      "IL37"       "SEC14L4"
## [126] "CILP2"      "DERL3"      "SMPDL3B"    "GALNT5"     "SCGB1A1"
## [131] "MMP13"      "CST2"       "PRSS8"      "RAB27B"     "DPP4"
## [136] "MET"        "LAMP3"      "CCNB2"      "SCGB3A2"    "CXCL11"
## [141] "SLAMF7"     "CD207"      "HLA-DRA"    "CP"         "CHIT1"
## [146] "TFPI2"      "MUC16"      "IL21"       "TNFRSF13B"  "MDK"
## [151] "LSR"        "PMCH"       "CGN"        "CLEC6A"     "LRP2"
## [156] "CDH1"       "KIFC1"      "CD3G"       "C16orf89"   "MMP10"
## [161] "MMP12"      "CD27"       "CXCL10"     "GDF15"      "CCL13"
## [166] "SCEL"       "OLR1"       "ORM1"       "SFN"        "CCL8"
## [171] "ACTBL2"     "COL10A1"    "CST5"       "CXCL5"      "IL12B"
## [176] "FASLG"      "KIF14"      "SHH"        "CCL15"      "ICAM1"
## [181] "BMP3"       "GSTA1"      "KRT19"      "CXADR"      "BPIFB1"
## [186] "PLA2G7"     "PNOC"       "ADM2"       "FLRT3"      "GALNT6"
## [191] "GGT6"       "TNFSF15"    "HMGB3"      "CCL24"      "KRT18"
## [196] "FGG"        "GRP"        "PKD2L1"     "EPYC"       "ADAMDEC1"
## [201] "FGB"        "MARCO"      "CDCP1"      "CST1"       "IL2RA"
## [206] "CCL20"      "DSG2"       "CST4"       "AGR3"       "VAMP8"
## [211] "GZMB"       "ABCA12"     "LCN2"       "C2"         "POU2AF1"
## [216] "TACSTD2"    "FCRL5"      "PON3"       "GALNT3"     "ALPP"
## [221] "HAVCR1"     "C8B"        "C1orf116"   "AMY1A"      "NECTIN4"
## [226] "CCL18"      "JCHAIN"     "AURKB"      "MMP9"       "MMP1"
## [231] "IFI30"      "CCL22"      "SFTPB"      "FCAMR"      "TREM2"
## [236] "FGA"        "CXCL13"     "SPINT1"     "ST6GALNAC1" "TNFRSF9"
## [241] "TCL1A"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "ALB"       "SERPINA1"  "ORM1"      "CP"        "SERPINA6"  "PON1"
##   [7] "SERPINA7"  "C4BPA"     "LRG1"      "PLTP"      "CLEC3B"    "CFB"
##  [13] "TNC"       "C1QTNF7"   "ORM2"      "CD55"      "IGFBP7"    "ADGRE5"
##  [19] "MFGE8"     "AFP"       "CCL14"     "ISG15"     "MMP3"      "TNFRSF1B"
##  [25] "SCUBE3"    "MMP10"     "CTGF"      "TNFSF13B"  "VEGFA"     "CX3CL1"
##  [31] "BMP6"      "F3"        "IL1R2"     "S100A9"    "IL24"      "IL32"
##  [37] "SERPINF2"  "IL7"       "NPPC"      "EDN1"      "PGLYRP2"   "FGB"
##  [43] "C7"        "CPN2"      "CPB2"      "BCHE"      "PRG4"      "PIGR"
##  [49] "HSPG2"     "POTEJ"     "TPM4"      "KRT8"      "KRT17"     "TPM3"
##  [55] "KRT7"      "KRT19"     "TPM2"      "WDR1"      "KRT18"     "GRN"
##  [61] "TIMP2"     "SFTPB"     "WARS"      "HSPD1"     "SPTA1"     "FLNB"
##  [67] "COLEC11"   "ACTN3"     "MEPE"      "CEACAM5"   "RAB1A"     "FKBP1A"
##  [73] "CEACAM8"   "SNED1"     "DSC2"      "LY6G6F"    "TNN"       "NBL1"
##  [79] "SLC25A5"   "SLC25A6"   "CHRDL2"    "GALNT7"    "MUC1"      "VDAC2"
##  [85] "BSG"       "LSR"       "CNTN6"     "CKMT1A"    "CKMT1B"    "GLO1"
##  [91] "VWCE"      "SEMA3A"    "LGMN"      "VMO1"      "IL1R1"     "MEP1A"
##  [97] "CPD"       "RANBP1"    "IQGAP1"    "KIAA0319L" "SERPINH1"  "SERPINB5"
## [103] "GALNT5"    "MXRA8"     "NAPSA"     "ATP1A1"    "PLOD3"     "IGSF10"
## [109] "DCTPP1"    "GBA"       "GLDN"      "TNFRSF12A" "FREM2"     "PITPNB"
## [115] "ARF6"      "MYOM2"     "CHRD"      "TOMM70"    "NELL1"     "PRTN3"
## [121] "FKBP4"     "ADAMTSL3"  "S100A14"   "PITPNA"    "SIRPB2"    "BCAT1"
## [127] "ADGRL1"    "SLC39A10"  "FEN1"      "VPS13A"    "ATP2C1"    "REEP6"
## [133] "VARS"      "AGR2"      "CGREF1"    "ENTPD6"    "TRIP10"    "CLDN3"
## [139] "LAMB3"     "MTSS1"     "TNIK"      "MAN2C1"    "SWAP70"    "EPCAM"
## [145] "KAZALD1"   "IL12A"     "ADAMTS15"  "KLK14"     "DRAXIN"    "IFNLR1"
## [151] "TREM1"     "ITGB6"     "CDH3"      "TP53"      "SPOCK1"    "EFNA4"
## [157] "ACOT2"     "ADGRG3"    "ADGRG4"    "AQP3"      "AURKB"     "C14orf39"
## [163] "CACNA1G"   "CCNB1"     "CDAN1"     "CDK4"      "CHEK2"     "CLPB"
## [169] "COL19A1"   "CSMD2"     "DDT"       "DERL1"     "DNAH8"     "DONSON"
## [175] "DOT1L"     "ERGIC2"    "EZH2"      "FAM9B"     "FOXD4L1"   "GDF5"
## [181] "HMGB3"     "HOXC11"    "HPS5"      "HS3ST5"    "IQGAP3"    "IRX1"
## [187] "ITPR3"     "KCNN2"     "KIF14"     "KIF4B"     "KRT24"     "LAMC2"
## [193] "LIG1"      "LMO7"      "LRRC3"     "LSG1"      "MEA1"      "MKI67"
## [199] "NASP"      "NEUROG3"   "NKX2-8"    "OSBPL3"    "PCDH8"     "PCNA"
## [205] "PKD1L1"    "PKD2L1"    "PKP3"      "POLR3A"    "POTEE"     "PRAMEF12"
## [211] "PROKR2"    "PRPF40A"   "RNF6"      "SCEL"      "SLC4A9"    "SPAG9"
## [217] "SRP54"     "TOP2A"     "TP63"      "TRIM21"    "TUBB8"     "UBE2G2"
## [223] "WDR7"      "ZDHHC15"   "ZFHX3"     "ZNF300"    "ZNF445"    "ZNF507"
## [229] "ZNF566"    "ZNF780A"   "ZNFX1"
```

Step 3: Select the markers short listed in both approaches

Hide

```
luad_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_markers)
luad_cancer_genesInBoth
```

```
##  [1] "MUC1"    "TREM1"   "TOP2A"   "EPCAM"   "LAMC2"   "LRG1"    "AQP3"
##  [8] "ORM2"    "CLDN3"   "NKX2-8"  "AGR2"    "PKP3"    "PIGR"    "CDH3"
## [15] "NAPSA"   "S100A14" "CPB2"    "IQGAP3"  "C4BPA"   "ITGB6"   "GALNT5"
## [22] "CP"      "LSR"     "MMP10"   "SCEL"    "ORM1"    "KIF14"   "KRT19"
## [29] "HMGB3"   "KRT18"   "PKD2L1"  "FGB"     "AURKB"   "SFTPB"
```

2.6.2 Subtype: LUSC

Step 1: Short list markers by gene expression level

Hide

```
cancer_name = 'lung'
cancer_subtype_code = 'lusc'
cancer_case_study_barcodes <- bpmsp.get_cancer_study_barcodes(cancer_name, cancer_subtyp
e_code)
cancer_tcga_gtex_data <- bpmsp.get_tcga_gtex_data_by_barcodes(cancer_case_study_barcode
s)
tpm_shortlisted_markers <- bpmsp.short_list_markers_by_expression_level_difference(cance
r_tcga_gtex_data)
tpm_shortlisted_markers
```

```
##   [1] "DNAH3"      "SERPINB11"  "SERPINB12"  "ESM1"       "KLK8"
##   [6] "XCL1"       "FGFBP1"     "LYZ"        "SLPI"       "CXCL8"
##  [11] "BPIFA1"     "KRT6A"      "XDH"        "CTHRC1"     "ALDH3A1"
##  [16] "PBK"        "BMP7"       "TREM1"      "DLGAP5"     "TTK"
##  [21] "SCGB3A1"    "TP63"       "CXCL9"      "AKR1B10"    "PI3"
##  [26] "FCRL2"      "CLEC4C"     "CLCA2"      "TOP2A"      "DSP"
##  [31] "CCL11"      "DSG3"       "CCL17"      "KLK12"      "IFNG"
##  [36] "CDC6"       "SFTPD"      "PRSS21"     "IVL"        "EPCAM"
##  [41] "CENPF"      "ST14"       "MZB1"       "IL1RN"      "WFDC12"
##  [46] "LAMC2"      "CYP2C18"    "CSF2"       "UTS2"       "LRRC15"
##  [51] "KRT13"      "EDAR"       "CKMT1A"     "LGALS7"     "LYPD3"
##  [56] "FAM83C"     "AMY1C"      "FGF19"      "CCL7"       "IL22RA2"
##  [61] "GAST"       "KRT80"      "MMP7"       "S100A7"     "DIAPH3"
##  [66] "KRT14"      "NOTUM"      "SOST"       "GTSE1"      "MUC5B"
##  [71] "KRT3"       "NEIL3"      "POF1B"      "PRAME"      "RRM2"
##  [76] "CLDN3"      "TNFRSF17"   "CSTA"       "KRT6B"      "GZMA"
##  [81] "NECTIN1"    "S100A2"     "TK1"        "CKMT1B"     "NKX2-8"
##  [86] "SERPINB4"   "SERPINB5"   "CDC20"      "KRT75"      "SERPINA9"
##  [91] "SDC1"       "PCSK9"      "ATP12A"     "LAIR2"      "AGR2"
##  [96] "MAD2L1"     "IRX4"       "FABP6"      "SPP1"       "ITGA2"
## [101] "KRT16"      "PKP3"       "HIST1H2BL"  "SLITRK6"    "CALML3"
## [106] "KIF2C"      "KYNU"       "MSMB"       "LAD1"       "CA9"
## [111] "HSD17B2"    "KNL1"       "HMGA2"      "PIGR"       "CDH3"
## [116] "KRT34"      "E2F2"       "KIF11"      "SERPINB2"   "CXCL17"
## [121] "CEACAM6"    "NAPSA"      "S100A14"    "TNFSF11"    "IGSF3"
## [126] "MXRA5"      "WFDC2"      "MELK"       "IQGAP3"     "SPINT2"
## [131] "KRT9"       "C4BPA"      "KRT6C"      "ITGB6"      "DSC3"
## [136] "SMPDL3B"    "GALNT5"     "IL36B"      "SCGB1A1"    "MMP13"
## [141] "CST2"       "PRSS8"      "HEPHL1"     "SLCO1A2"    "LAMP3"
## [146] "CCNB2"      "SCGB3A2"    "CDK1"       "CXCL11"     "SLAMF7"
## [151] "LGALS7B"    "CXCL1"      "CD207"      "CHIT1"      "MUC16"
## [156] "NTS"        "KRT31"      "KRT5"       "LY6D"       "ULBP2"
## [161] "LSR"        "PMCH"       "CLEC6A"     "CDH1"       "KIFC1"
## [166] "NUF2"       "CD3G"       "MMP10"      "MMP12"      "CXCL10"
## [171] "KRT33A"     "CNTN5"      "IL19"       "CYP4F11"    "HOXC11"
## [176] "IL1A"       "HIST1H1B"   "DSC2"       "SFN"        "CXCL6"
## [181] "COL10A1"    "UGT1A1"     "FETUB"      "CXCL5"      "KIF14"
## [186] "HIST1H4D"   "TMEM40"     "SERPINB13"  "GSTA1"      "KRT19"
## [191] "CXADR"      "BPIFB1"     "MCM2"       "CTSV"       "PNOC"
## [196] "CDC45"      "ADM2"       "GGT6"       "IL36RN"     "CCL24"
## [201] "KRT18"      "GRP"        "EPYC"       "ADAMDEC1"   "CDCP1"
## [206] "CST1"       "IL2RA"      "CCL20"      "DSG2"       "CST4"
## [211] "KRT17"      "AGR3"       "GZMB"       "ABCA12"     "LCN2"
## [216] "POU2AF1"    "ERVMER34-1" "PLAU"       "TACSTD2"    "FCRL5"
## [221] "AURKA"      "GALNT3"     "ALPP"       "C1orf116"   "NECTIN4"
## [226] "CCL18"      "SPRR3"      "JCHAIN"     "AURKB"      "MMP9"
## [231] "MMP1"       "CCL22"      "CCNB1"      "SFTPB"      "TP73"
## [236] "TREM2"      "CXCL13"     "SPINT1"     "TNFRSF9"    "FAT2"
## [241] "TCL1A"
```

Step 2: Short list markers by IHC scores

Hide

```
ihc_shortlisted_markers <- bpmsp.short_list_markers_by_IHC_score_z_scores(cancer_name)
ihc_shortlisted_markers
```

```
##   [1] "ALB"       "SERPINA1"  "ORM1"      "CP"        "SERPINA6"  "PON1"
##   [7] "SERPINA7"  "C4BPA"     "LRG1"      "PLTP"      "CLEC3B"    "CFB"
##  [13] "TNC"       "C1QTNF7"   "ORM2"      "CD55"      "IGFBP7"    "ADGRE5"
##  [19] "MFGE8"     "AFP"       "CCL14"     "ISG15"     "MMP3"      "TNFRSF1B"
##  [25] "SCUBE3"    "MMP10"     "CTGF"      "TNFSF13B"  "VEGFA"     "CX3CL1"
##  [31] "BMP6"      "F3"        "IL1R2"     "S100A9"    "IL24"      "IL32"
##  [37] "SERPINF2"  "IL7"       "NPPC"      "EDN1"      "PGLYRP2"   "FGB"
##  [43] "C7"        "CPN2"      "CPB2"      "BCHE"      "PRG4"      "PIGR"
##  [49] "HSPG2"     "POTEJ"     "TPM4"      "KRT8"      "KRT17"     "TPM3"
##  [55] "KRT7"      "KRT19"     "TPM2"      "WDR1"      "KRT18"     "GRN"
##  [61] "TIMP2"     "SFTPB"     "WARS"      "HSPD1"     "SPTA1"     "FLNB"
##  [67] "COLEC11"   "ACTN3"     "MEPE"      "CEACAM5"   "RAB1A"     "FKBP1A"
##  [73] "CEACAM8"   "SNED1"     "DSC2"      "LY6G6F"    "TNN"       "NBL1"
##  [79] "SLC25A5"   "SLC25A6"   "CHRDL2"    "GALNT7"    "MUC1"      "VDAC2"
##  [85] "BSG"       "LSR"       "CNTN6"     "CKMT1A"    "CKMT1B"    "GLO1"
##  [91] "VWCE"      "SEMA3A"    "LGMN"      "VMO1"      "IL1R1"     "MEP1A"
##  [97] "CPD"       "RANBP1"    "IQGAP1"    "KIAA0319L" "SERPINH1"  "SERPINB5"
## [103] "GALNT5"    "MXRA8"     "NAPSA"     "ATP1A1"    "PLOD3"     "IGSF10"
## [109] "DCTPP1"    "GBA"       "GLDN"      "TNFRSF12A" "FREM2"     "PITPNB"
## [115] "ARF6"      "MYOM2"     "CHRD"      "TOMM70"    "NELL1"     "PRTN3"
## [121] "FKBP4"     "ADAMTSL3"  "S100A14"   "PITPNA"    "SIRPB2"    "BCAT1"
## [127] "ADGRL1"    "SLC39A10"  "FEN1"      "VPS13A"    "ATP2C1"    "REEP6"
## [133] "VARS"      "AGR2"      "CGREF1"    "ENTPD6"    "TRIP10"    "CLDN3"
## [139] "LAMB3"     "MTSS1"     "TNIK"      "MAN2C1"    "SWAP70"    "EPCAM"
## [145] "KAZALD1"   "IL12A"     "ADAMTS15"  "KLK14"     "DRAXIN"    "IFNLR1"
## [151] "TREM1"     "ITGB6"     "CDH3"      "TP53"      "SPOCK1"    "EFNA4"
## [157] "ACOT2"     "ADGRG3"    "ADGRG4"    "AQP3"      "AURKB"     "C14orf39"
## [163] "CACNA1G"   "CCNB1"     "CDAN1"     "CDK4"      "CHEK2"     "CLPB"
## [169] "COL19A1"   "CSMD2"     "DDT"       "DERL1"     "DNAH8"     "DONSON"
## [175] "DOT1L"     "ERGIC2"    "EZH2"      "FAM9B"     "FOXD4L1"   "GDF5"
## [181] "HMGB3"     "HOXC11"    "HPS5"      "HS3ST5"    "IQGAP3"    "IRX1"
## [187] "ITPR3"     "KCNN2"     "KIF14"     "KIF4B"     "KRT24"     "LAMC2"
## [193] "LIG1"      "LMO7"      "LRRC3"     "LSG1"      "MEA1"      "MKI67"
## [199] "NASP"      "NEUROG3"   "NKX2-8"    "OSBPL3"    "PCDH8"     "PCNA"
## [205] "PKD1L1"    "PKD2L1"    "PKP3"      "POLR3A"    "POTEE"     "PRAMEF12"
## [211] "PROKR2"    "PRPF40A"   "RNF6"      "SCEL"      "SLC4A9"    "SPAG9"
## [217] "SRP54"     "TOP2A"     "TP63"      "TRIM21"    "TUBB8"     "UBE2G2"
## [223] "WDR7"      "ZDHHC15"   "ZFHX3"     "ZNF300"    "ZNF445"    "ZNF507"
## [229] "ZNF566"    "ZNF780A"   "ZNFX1"
```

Step 3: Select the markers short listed in both approaches

Hide

```
lusc_cancer_genesInBoth = intersect(tpm_shortlisted_markers, ihc_shortlisted_markers)
lusc_cancer_genesInBoth
```

```
##  [1] "TREM1"    "TP63"     "TOP2A"    "EPCAM"    "LAMC2"    "CKMT1A"
##  [7] "CLDN3"    "CKMT1B"   "NKX2-8"   "SERPINB5" "AGR2"     "PKP3"
## [13] "PIGR"     "CDH3"     "NAPSA"    "S100A14"  "IQGAP3"   "C4BPA"
## [19] "ITGB6"    "GALNT5"   "LSR"      "MMP10"    "HOXC11"   "DSC2"
## [25] "KIF14"    "KRT19"    "KRT18"    "KRT17"    "AURKB"    "CCNB1"
## [31] "SFTPB"
```

3. Selection of candidate markers that are more specific to breast cancer

Here is the venn diagram showing how those shortlists overlap.

Hide

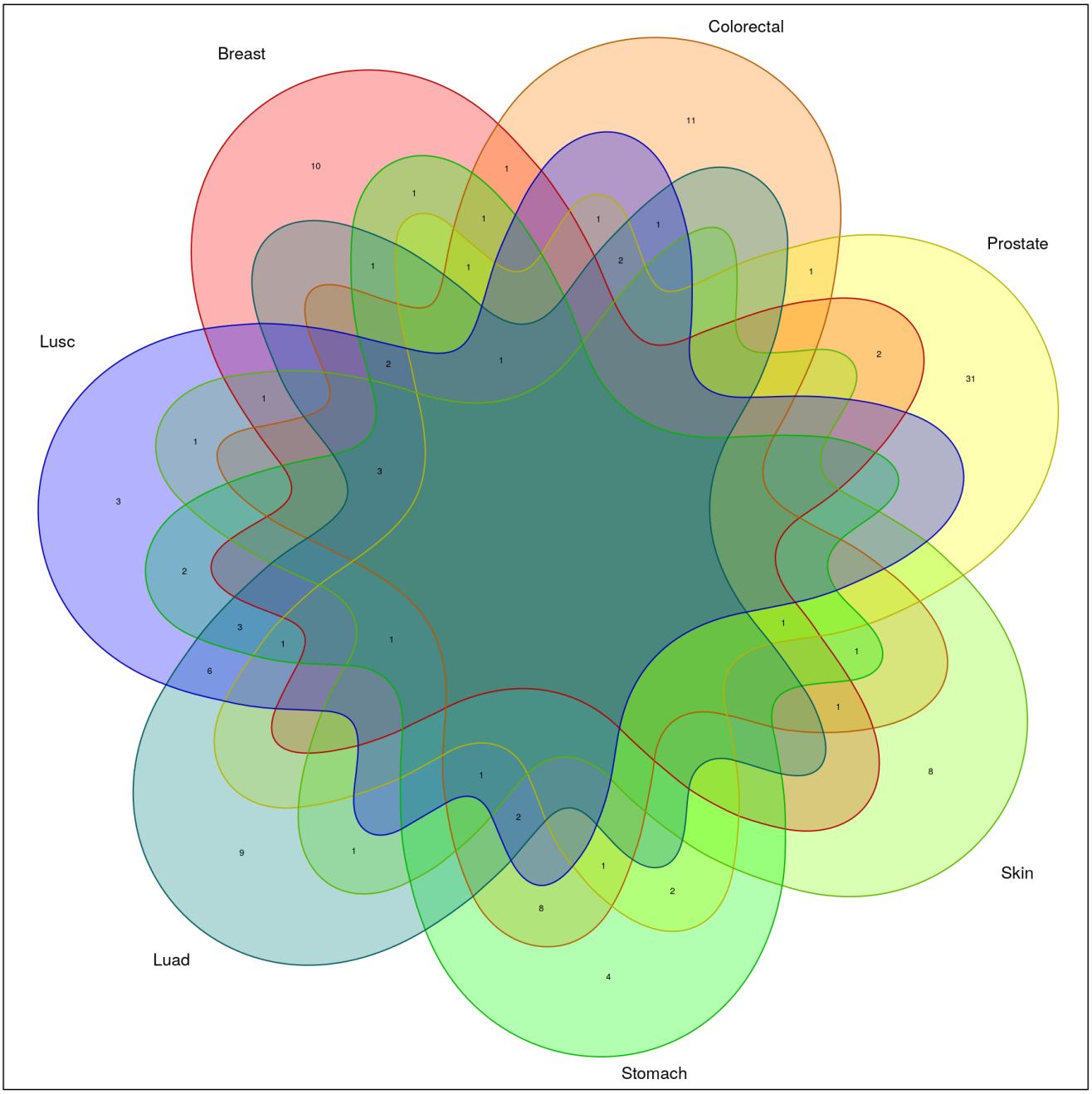
```
if(!require('venn')){
  install.packages('venn')
  library('venn')
}
```

```
## Loading required package: venn
```

Hide

```
plasma_protein_sets = list(
  Breast = breast_cancer_genesInBoth,
  Colorectal = colorectal_cancer_genesInBoth,
  Prostate = prostate_cancer_genesInBoth,
  Skin = skin_cancer_genesInBoth,
  Stomach = stomach_cancer_genesInBoth,
  Luad = luad_cancer_genesInBoth,
  Lusc = lusc_cancer_genesInBoth
)

venn(plasma_protein_sets, ilabels = TRUE, zcolor = "style")
```



The list of candidate markers that are shortlisted only for breast cancer (i.e., not shortlisted by the other cancers in this example) is:

Hide

```
Reduce(setdiff, plasma_protein_sets)
```

```
##  [1] "WNK4"    "CFB"     "IL22RA2" "MUC5B"   "IRX1"    "MDK"     "TFAP2B"
##  [8] "MGP"     "GALNT6"  "GATA3"
```