Figures

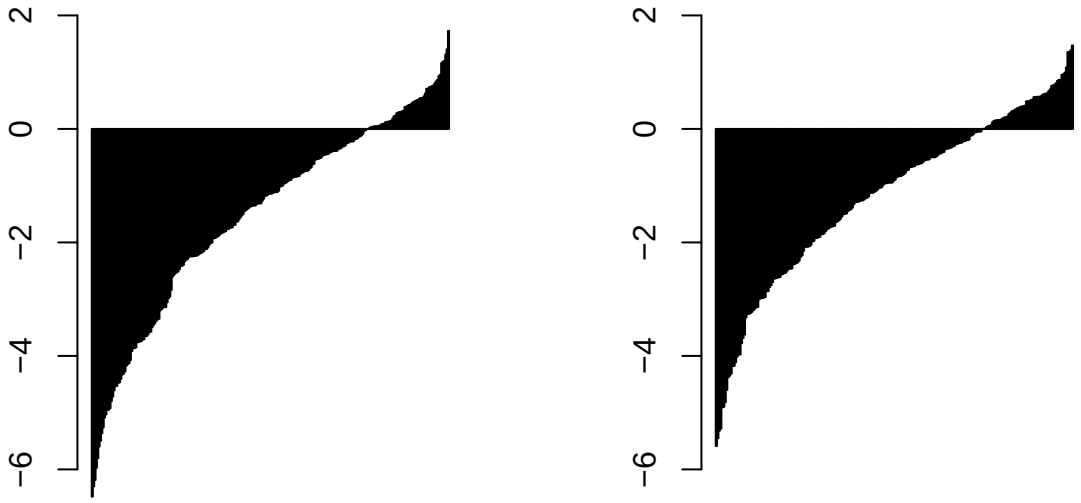
Skanderup lab

9/19/2018

Pan-cancer crosstalk analysis

Expression of ligands and receptors across tumor types

```
# pan-cancer expression of ligands
df.l = unique(df[,c('l','t','c_l','s_l')])
df.l = df.l[df.l$t!='MEDIAN',]
df.l$lfc = log2((df.l$c_l+1)/(df.l$s_l+1))
df.l = df.l[order(df.l$lfc),]
df.l.pancan = aggregate(df.l$lfc,list(df.l$l),median)
colnames(df.l.pancan) = c('l','lfc')
df.l.pancan = df.l.pancan[order(df.l.pancan$lfc),]

# pan-cancer expression of receptors
df.r = unique(df[,c('r','t','c_r','s_r')])
df.r = df.r[df.r$t!='MEDIAN',]
df.r$lfc = log2((df.r$c_r+1)/(df.r$s_r+1))
df.r = df.r[order(df.r$lfc),]
df.r.pancan = aggregate(df.r$lfc,list(df.r$r),median)
colnames(df.r.pancan) = c('r','lfc')
df.r.pancan = df.r.pancan[order(df.r.pancan$lfc),]

par(mfrow=c(1,2))
barplot(df.l.pancan$lfc,ylim=c(-7,2))
barplot(df.r.pancan$lfc,ylim=c(-7,2))
```




```
# list top-10 ligands and receptors
head(df.l.pancan,n=10)
```

```
##          l       lfc
```

```
## 29    C1QB -6.479623
## 42    CCL5 -6.297799
## 39   CCL21 -6.187251
## 30      C3 -5.971518
## 73   CXCL9 -5.802816
## 36   CCL19 -5.598068
## 71  CXCL13 -5.494986
## 75     DCN -5.361811
## 194  PTPRC -5.261131
## 154    LTB -5.085601
```

```
tail(df.l.pancan,n=10)[10:1,]
```

```
##          l       lfc
## 186 PODXL2 1.7274613
## 26    BMP7 1.4038044
## 252  VEGFA 1.3117658
## 80     DTL 1.1975560
## 162   MST1 1.1855930
## 85   EFNA3 1.1598409
## 217 SEMA4G 0.9539774
## 84   EFNA1 0.9311229
## 11     AMH 0.8674052
## 95   FGF12 0.8391557
```

```
head(df.r.pancan,n=10)
```

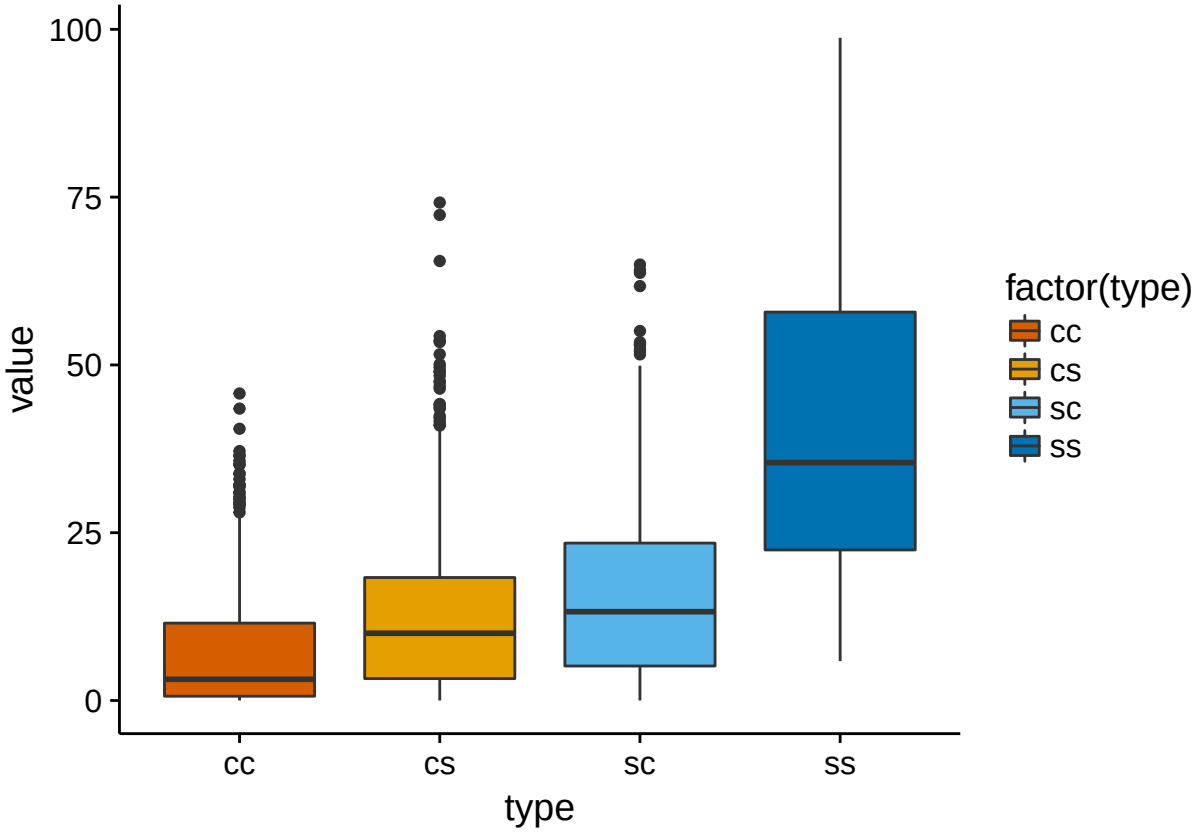
```
##          r       lfc
## 100  IL2RG -5.588500
## 121  ITGB2 -5.450845
## 34    CD74 -5.317843
## 240 TYROBP -5.283264
## 25     CD2 -4.908353
## 29     CD4 -4.895632
## 42   CSF1R -4.816277
## 27    CD27 -4.597750
## 73    FPR3 -4.377222
## 116  ITGAL -4.348399
```

```
tail(df.r.pancan,n=10)[10:1,]
```

```
##          r       lfc
## 65   ERBB3 1.4713160
## 145   LRP4 1.3864947
## 146   LRP5 1.3569002
## 50    DDR1 1.3551940
## 176 PLXNB1 1.0842515
## 196  RTN4R 0.9931948
## 64   ERBB2 0.9627734
## 85    HMMR 0.9548589
## 68   FGFR3 0.8679795
## 147   LRP6 0.8625430
```

Pan-cancer crosstalk scores

```
# cc cs sc ss nn
ct.colors = cbPalette <- c("#D55E00", "#E69F00", "#56B4E9", "#0072B2",  "#BBBBBB")

df.med = list()
for(lr2 in unique(df$lr)) {
  xx = df %>% filter(lr == lr2)
  rr = c(median(xx$cc,na.rm = T),median(xx$ss,na.rm = T),median(xx$sc,na.rm = T),median(xx$cs,na.rm = T
  df.med[[lr2]] = rr
}
df.med = data.frame(do.call(rbind,df.med))
df.med$lr = rownames(df.med)
colnames(df.med) = c('cc','ss','sc','cs','nn','lr')
df.med$r = sub('.+_',"\\1",df.med$lr)

df.med.g = gather(df.med,'type','value',-r,-lr,-nn)
p3b = ggplot((df.med.g), aes(type,value)) + geom_boxplot(aes(fill=factor(type))) + scale_fill_manual(va
p3b
```



Count LR pairs with strong autocrine cancer signaling across tumor types (median RC score > 40).

```
count(df.med %>% filter(cc > 40))
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1     3
```

```
count(df.med %>% filter(ss > 40))
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   264
```

```
count(df.med %>% filter(cs > 40))
```

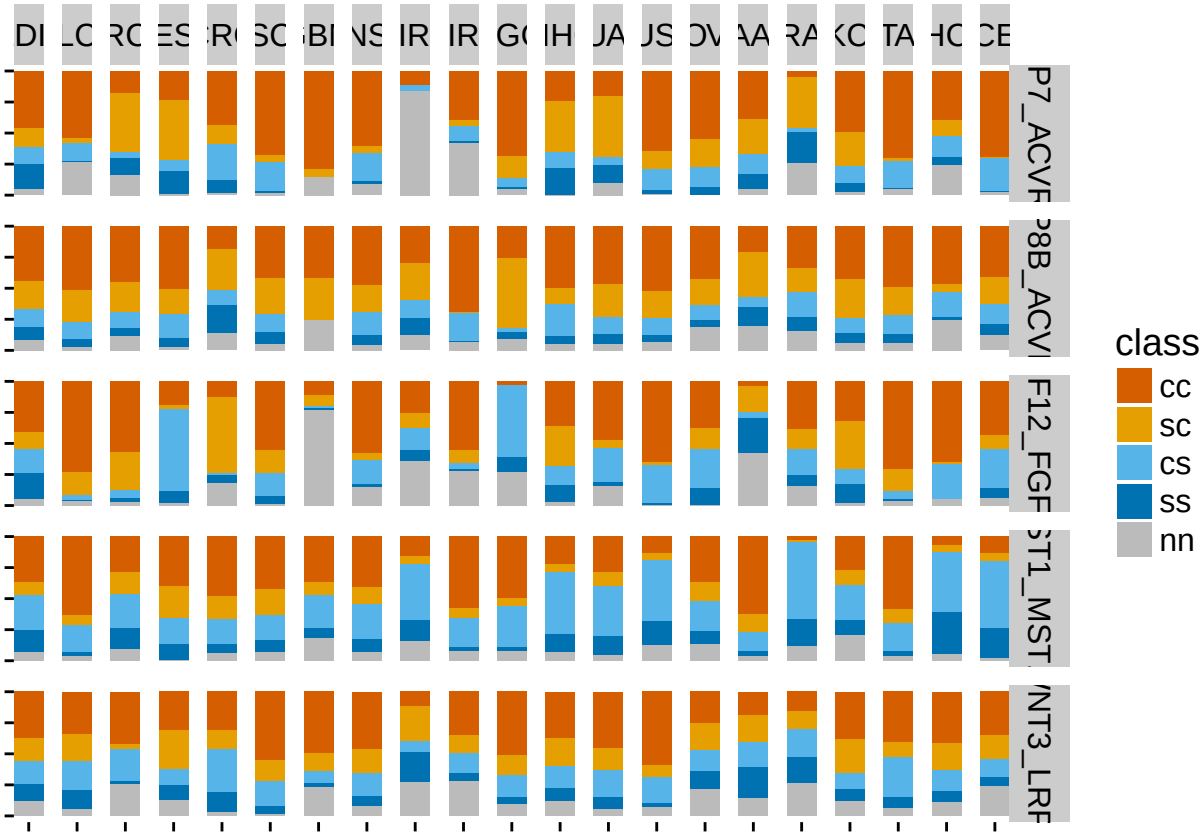
```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    26
```

```
count(df.med %>% filter(sc > 40))
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    37
```

Top-5 C->C crosstalk pairs

```
top5.lr = (df.med[order(-df.med$cc),])[1:5,'lr']
df.top5 = gather(df %>% filter(lr %in% top5.lr),'type','value', cc,sc,cs,ss,nn)
df.top5$type <- factor(df.top5$type, levels = c('cc','sc','cs','ss','nn'))
df.top5$t <- factor(df.top5$t, levels = c(c('MEDIAN'),setdiff(levels(df.top5$t),c('MEDIAN'))))
p3c = ggplot(df.top5, aes(x = "", y=value, fill = factor(type))) + geom_bar(width = 1, stat = "identity
  theme(axis.line = element_blank(), axis.text = element_blank()) + scale_fill_manual(values=ct.colors)
  labs(fill="class", x=NULL, y=NULL) + facet_grid(lr ~ t)
p3c
```

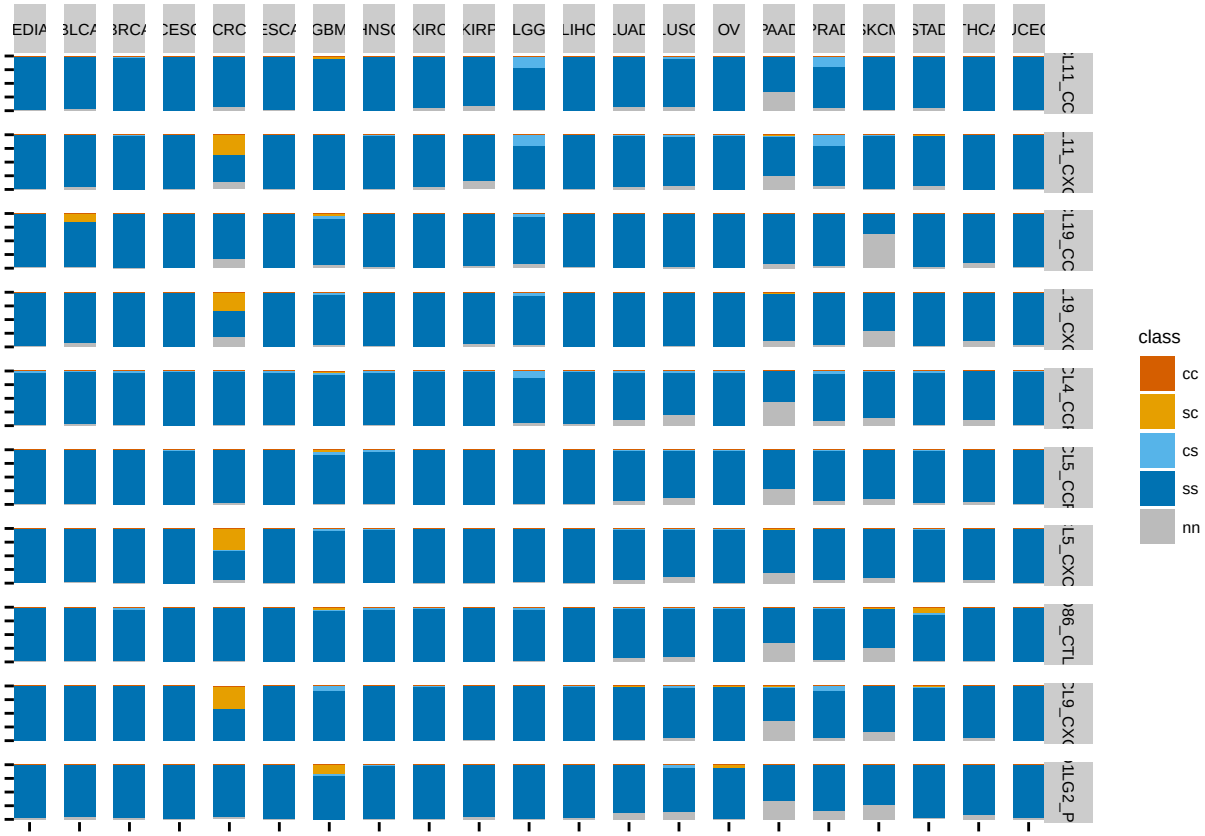


Top-5 S->C crosstalk pairs

```
top5.lr = (df.med[order(-df.med$sc),])[1:5,'lr']
df.top5 = gather(df %>% filter(lr %in% top5.lr),'type','value', cc,sc,cs,ss,nn)
df.top5$type <- factor(df.top5$type, levels = c('cc','sc','cs','ss','nn'))
df.top5$t <- factor(df.top5$t, levels = c(c('MEDIAN'),setdiff(levels(df.top5$t),c('MEDIAN'))))
p3d = ggplot(df.top5, aes(x = "", y=value, fill = factor(type))) + geom_bar(width = 1, stat = "identity
  theme(axis.line = element_blank(), axis.text = element_blank()) + scale_fill_manual(values=ct.colors)
  labs(fill="class", x=NULL, y=NULL) + facet_grid(lr ~ t)
p3d
```

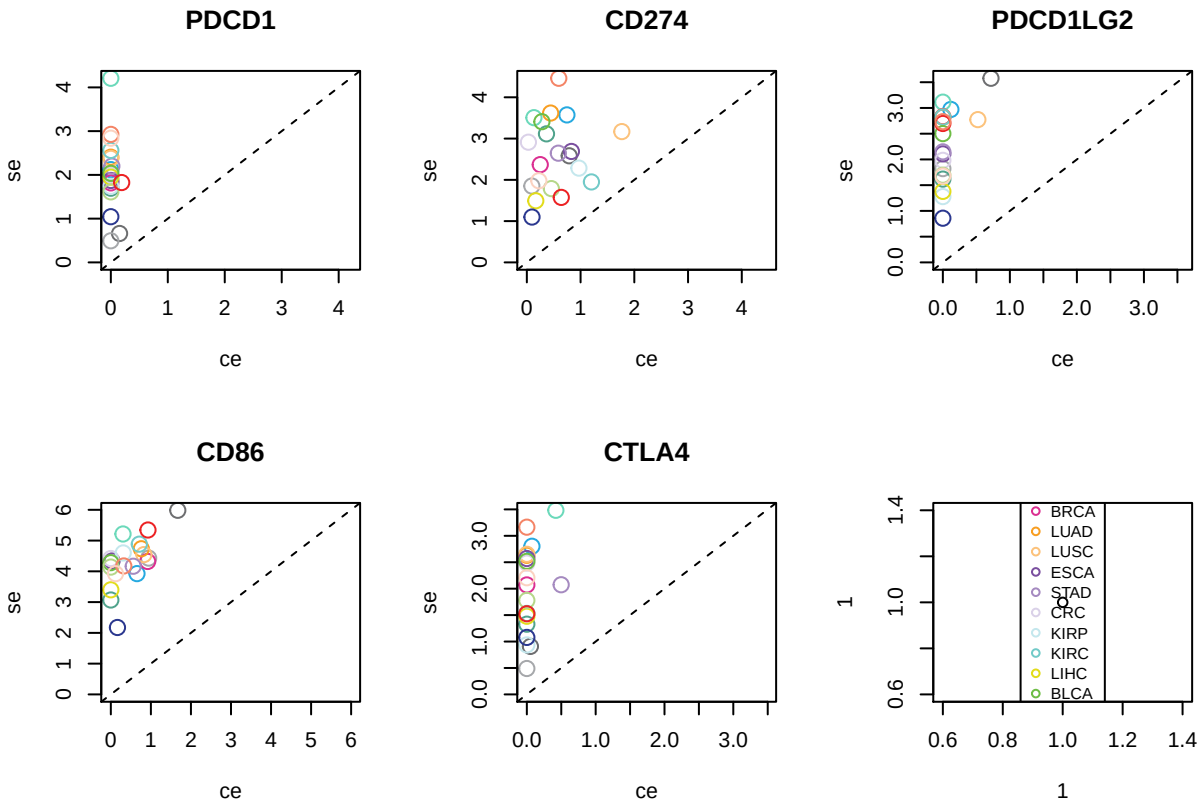
Top 10 stroma-stroma interactons

```
topss.lr = (df.med[order(-df.med$ss),])[1:10,'lr']
df.topss = gather(df %>% filter(lr %in% topss.lr),'type','value', cc,sc,cs,ss,nn)
df.topss$type <- factor(df.topss$type, levels = c('cc','sc','cs','ss','nn'))
df.topss$t <- factor(df.topss$t, levels = c(c('MEDIAN'),setdiff(levels(df.topss$t),c('MEDIAN'))))
pss = ggplot(df.topss, aes(x = "", y=value, fill = factor(type))) + geom_bar(width = 1, stat = "identity
  theme(axis.line = element_blank(), axis.text = element_blank()) + scale_fill_manual(values=ct.colors)
  labs(fill="class", x=NULL, y=NULL) + facet_grid(lr ~ t) + theme(text = element_text(size = 7))
pss
```



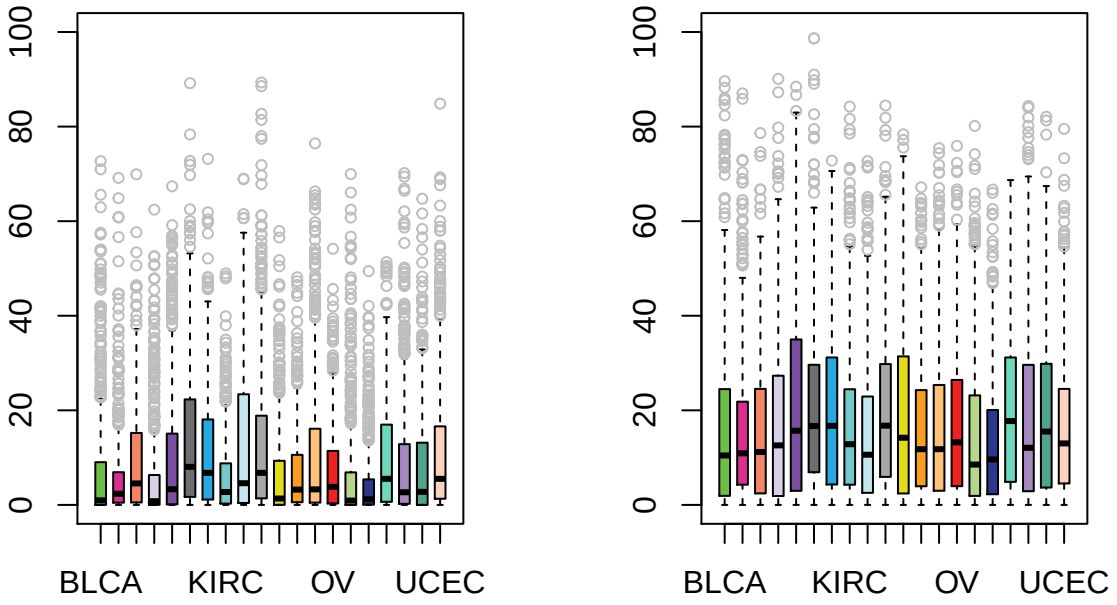
Expression Immune checkpoint ligands and receptors

```
par(mfrow=c(2,3))
for(g in c('PDCD1','CD274', 'PDCD1LG2', 'CD86', 'CTLA4')) {
  ce = unlist(expr[g,paste0(names(tumor.colors),'_C')])
  se = unlist(expr[g,paste0(names(tumor.colors),'_S')])
  lim = c(0,max(ce,se))
  plot(ce,se,ylim=lim,xlim=lim,main=g,col=tumor.colors,cex=1.6)
  abline(0,1,lty='dashed')
}
plot(1,1)
legend('center',names(tumor.colors),col=tumor.colors,pch=21,cex=0.8)
```



Tumor type specific LR interactions

```
# require min 1 RPKM expression of both ligand and receptor
df.filt = df %>% filter(c_l > 1 |  s_l > 1 | n_l > 1) %>% filter(c_r > 1 |  s_r > 1 | n_r > 1)

par(mfrow=c(1,2))
pan.cc = spread(df.filt[,c('t','cc','lr')],'t','cc')
rownames(pan.cc) = pan.cc[,1]
pan.cc = pan.cc[,setdiff(colnames(pan.cc),c('lr','MEDIAN'))]
boxplot(pan.cc,horizontal = F,boxwex=0.6,cex=0.7,col=tumor.colors[colnames(pan.cc)],ylim=c(0,100),outcol

pan.sc = spread(df.filt[,c('t','sc','lr')],'t','sc')
rownames(pan.sc) = pan.sc[,1]
pan.sc = pan.sc[,setdiff(colnames(pan.sc),c('lr','MEDIAN'))]
boxplot(pan.sc,horizontal = F,boxwex=0.6,cex=0.7,col=tumor.colors[colnames(pan.sc)],ylim=c(0,100),outcol
```




```
# list top-2 for each cancer type
apply(pan.cc,2,function(x) row.names(pan.cc)[order(-x)[1:2]])
```

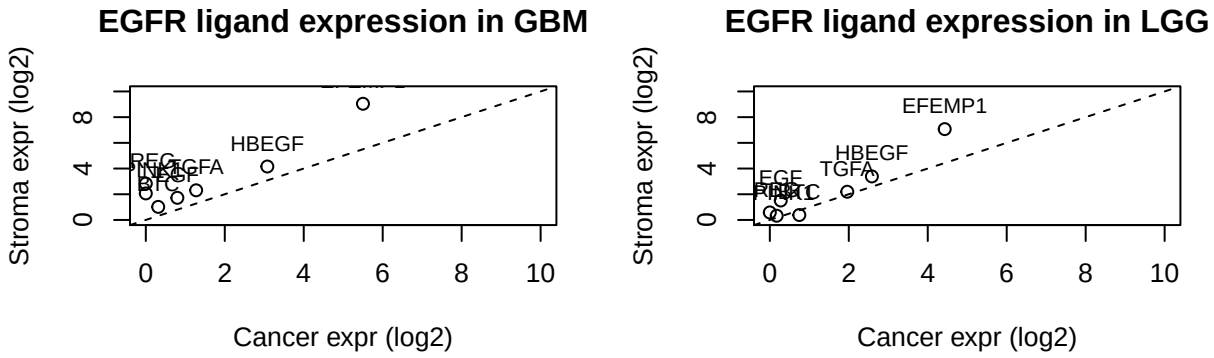
```
##      BLCA           BRCA              CESC            CRC
## [1,] "FGF12_FGFR3" "ARTN_GFRA1"      "BMP8B_BMPR1B" "AREG_ERBB3"
## [2,] "EFNB3_EPHB6" "HSP90AA1_FGFR3" "FGF12_FGFR4"  "EFNA3_EPHA1"
##      ESCA           GBM            HNSC           KIRC          KIRP
## [1,] "BMP7_ACVR2B" "DLL1_NOTCH1" "FGF12_FGFR4" "NRG1_ERBB3" "BMP8B_ACVR2B"
## [2,] "EFNB3_EPHB3" "BMP7_ACVR2B" "FGF12_FGFR2" "NRG1_ERBB2" "WNT7B_FZD1"
##      LGG            LIHC           LUAD           LUSC           OV
## [1,] "BMP2_ACVR2B" "EFNA1_EPHA1" "GSTP1_TRAF2" "FGF12_FGFR2" "BMP7_ACVR2B"
## [2,] "DLL1_NOTCH1" "EFNA3_EPHA1" "FGF12_FGFR3" "CNTN1_NRCAM" "BMP7_BMPR1B"
##      PAAD           PRAD            SKCM             STAD
## [1,] "EFNA3_EPHA4" "BMP6_BMPR1B"  "BMP7_BMPR1B"   "FGF12_FGFR3"
## [2,] "EFNA1_EPHA4" "BMP8B_BMPR1B" "SEMA6A_PLXNA2" "BMP7_ACVR2B"
##      THCA           UCEC
## [1,] "FGF12_FGFR3" "EFNA3_EPHB1"
## [2,] "EDN2_EDNRB"  "BMP7_ACVR2A"
```

```
apply(pan.sc,2,function(x) row.names(pan.sc)[order(-x)[1:2]])
```

```
##      BLCA           BRCA           CESC          CRC           ESCA
## [1,] "FGF1_FGFR3"  "BMP2_BMPR1B" "APOE_SORL1" "FGF1_FGFR4" "TF_TFRC"
## [2,] "NRG1_ERBB3" "BMP6_BMPR1B" "TF_TFRC"    "FGF2_FGFR4" "INHBA_BAMBI"
##      GBM            HNSC           KIRC           KIRP         LGG
## [1,] "SPINK1_EGFR" "WNT2_FZD7"    "PTHLH_PTH1R" "HGF_MET"    "CLCF1_CNTFR"
## [2,] "AREG_EGFR"   "COL1A1_DDR1" "COL1A1_DDR1" "FN1_ITGB8" "EFEMP1_EGFR"
##      LIHC           LUAD           LUSC           OV              PAAD
## [1,] "AREG_ERBB3" "LTB_LTBR"     "APOE_LRP8"    "INHBA_ACVR2B" "APOE_LRP5"
## [2,] "FGF1_FGFR4" "L1CAM_ERBB3" "COL1A1_DDR1" "APOE_SORL1"   "WNT5A_LRP5"
##      PRAD           SKCM            STAD           THCA            UCEC
## [1,] "IGFBP4_FZD8" "INHBA_ACVR2B" "FGF2_FGFR3"  "IL24_IL22RA1" "SHH_BOC"
## [2,] "RSPO3_FZD8"  "CLCF1_CNTFR"  "INHBA_BAMBI" "APOE_VLDLR"   "SHH_CDON"
```

stromal EGFR ligands in GBM

```
par(mfrow=c(2,2))
egfr.ligands = c('AREG','BTC','EGF','HBEGF','SPINK1','TGFA','EFEMP1')
egfr.gbm = expr[egfr.ligands,c('GBM_C','GBM_S')]
plot(egfr.gbm[,1],egfr.gbm[,2],xlab='Cancer expr (log2)',ylab='Stroma expr (log2)',xlim=c(0,10),ylim=c(0
text(egfr.gbm[,1],egfr.gbm[,2],egfr.ligands,pos=3,cex=0.8)
abline(0,1,lty='dashed')
egfr.ligands = c('AREG','BTC','EGF','HBEGF','SPINK1','TGFA','EFEMP1')
egfr.gbm = expr[egfr.ligands,c('LGG_C','LGG_S')]
plot(egfr.gbm[,1],egfr.gbm[,2],xlab='Cancer expr (log2)',ylab='Stroma expr (log2)',xlim=c(0,10),ylim=c(0
text(egfr.gbm[,1],egfr.gbm[,2],egfr.ligands,pos=3,cex=0.8)
abline(0,1,lty='dashed')
```

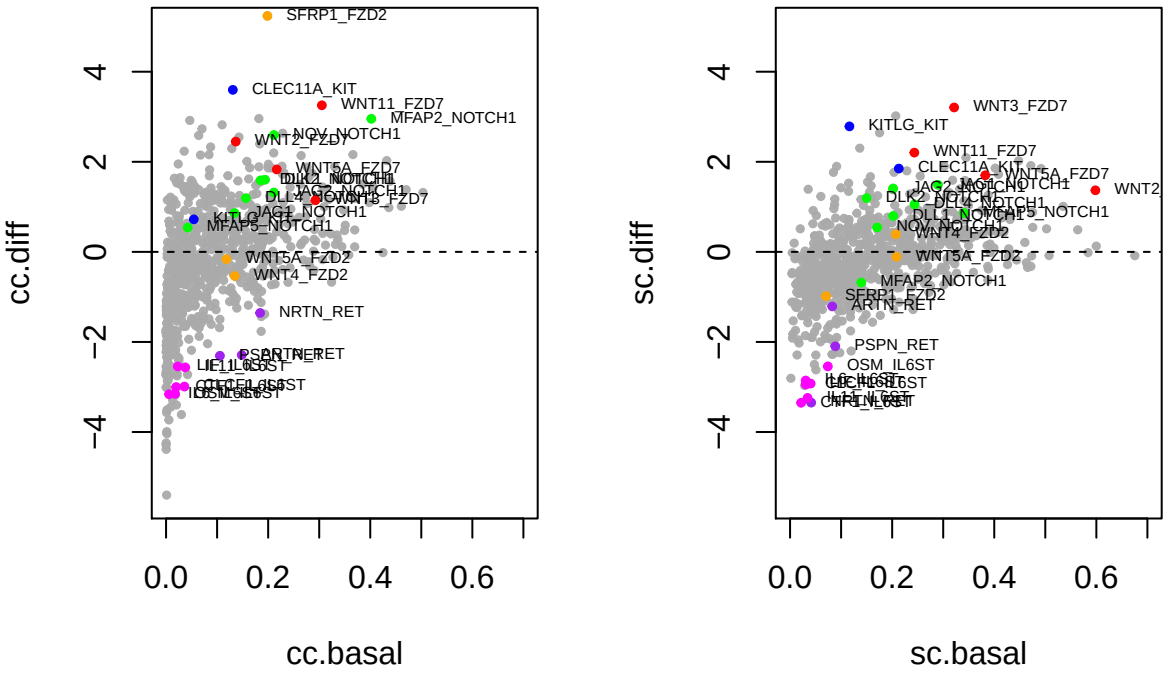


BRCA subtypes crosstalk analysis

Enrichment of crosstalk in basal subtype

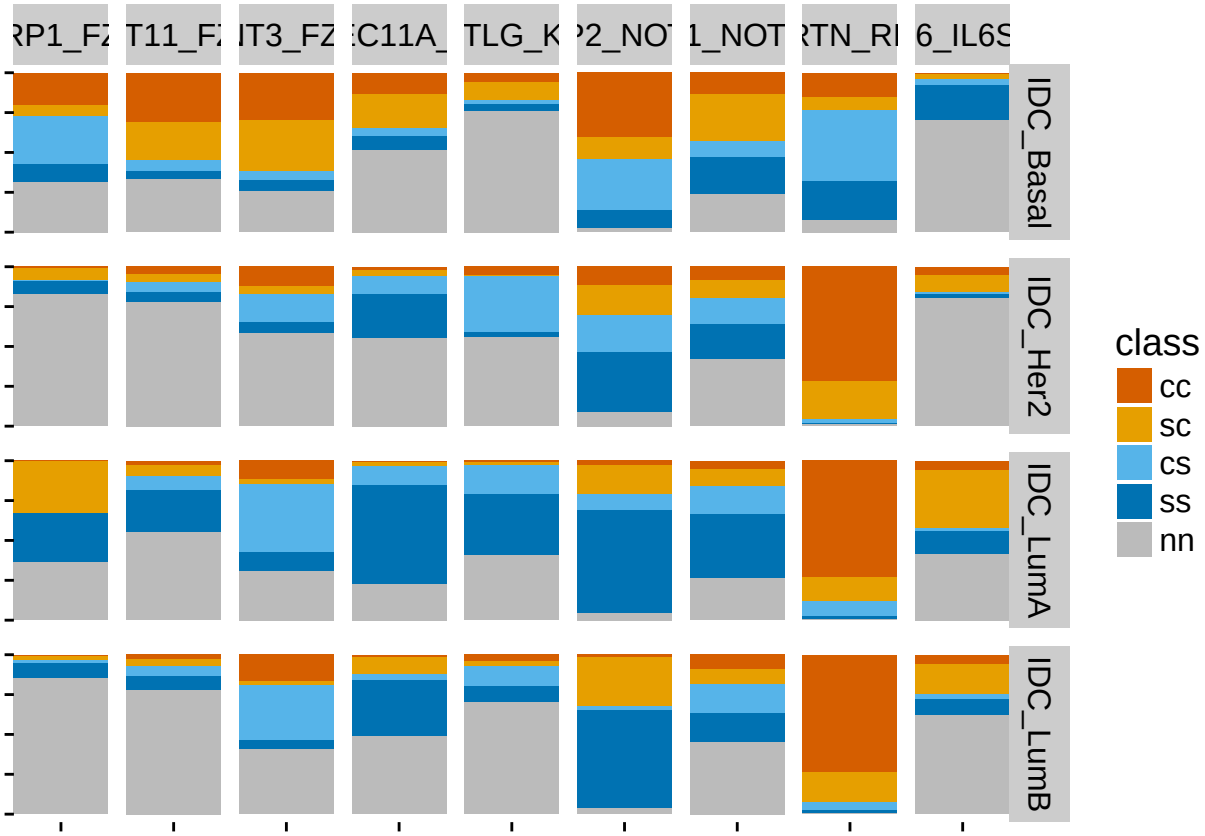
```
df.brca = read.csv('data/BRCA_product_score.csv')
df.brca$r = sub('.+_',"\\1",df.brca$lr)
df.brca$l = sub('_.+',"\\1",df.brca$lr)
df.brca = df.brca %>% filter(t == 'IDC_Her2' | t == 'IDC_LumA' | t == 'IDC_LumB' | t == 'IDC_Basal') %>%
# remove duplicated rows
df.brca = df.brca[!duplicated(df.brca),]
# 749 expressed LR pairs
#table(df.brca$t)

# Add pseudocounts to RC scores
df = df.brca
df$sum = (df$c_l+1)*(df$c_r+1)+(df$s_l+1)*(df$c_r+1)+(df$c_l+1)*(df$s_r+1)+(df$s_l+1)*(df$s_r+1)+(df$n_
df$cc = (df$c_l+1)*(df$c_r+1)/df$sum
df$sc = (df$s_l+1)*(df$c_r+1)/df$sum
df$cs = (df$c_l+1)*(df$s_r+1)/df$sum
df$ss = (df$s_l+1)*(df$s_r+1)/df$sum
df$nn = (df$n_l+1)*(df$n_r+1)/df$sum

cc.basal = (df %>% filter(t == 'IDC_Basal'))$cc
names(cc.basal) = (df %>% filter(t == 'IDC_Basal'))$lr
cc.others = apply(cbind(her2=(df %>% filter(t == 'IDC_Her2'))$cc,luma=(df %>% filter(t == 'IDC_LumA'))$
```

```
cc.diff = log2((cc.basal)/(cc.others))
sc.basal = (df %>% filter(t == 'IDC_Basal'))$sc
names(sc.basal) = (df %>% filter(t == 'IDC_Basal'))$lr
sc.others = apply(cbind(her2=(df %>% filter(t == 'IDC_Her2'))$sc,her2=(df %>% filter(t == 'IDC_LumA'))$s
sc.diff = log2((sc.basal)/(sc.others))

par(mfrow=c(1,2))
plot(cc.basal,cc.diff,pch=20,xlim=c(0.0,0.7),ylim=c(-5.5,5),cex=0.75,col=adjustcolor('#AAAAAA', alpha.f
abline(h=0,lty='dashed')
label.lr = function(key,col='red') {
  ids = grep(key,names(cc.basal))
  points(cc.basal[ids],cc.diff[ids],pch=20,xlim=c(0.0,0.7),ylim=c(-5.5,5),cex=0.8,col=adjustcolor(col, a
  text(cc.basal[ids],cc.diff[ids],names(cc.basal)[ids],cex=0.5,pos=4,col='black')
}
label.lr('_NOTCH1','green')
label.lr('_KIT','blue')
label.lr('_FZD7','red')
label.lr('_FZD2','orange')
label.lr('_RET','purple')
label.lr('_IL6ST','magenta')

plot(sc.basal,sc.diff,pch=20,xlim=c(0.0,0.7),ylim=c(-5.5,5),cex=0.75,col=adjustcolor('#AAAAAA', alpha.f
abline(h=0,lty='dashed')
label.lr = function(key,col='red') {
  ids = grep(key,names(sc.basal))
  points(sc.basal[ids],sc.diff[ids],pch=20,xlim=c(0.0,0.7),ylim=c(-5.5,5),cex=0.8,col=adjustcolor(col, a
  text(sc.basal[ids],sc.diff[ids],names(sc.basal)[ids],cex=0.5,pos=4,col='black')
}
label.lr('_NOTCH1','green')
label.lr('_KIT','blue')
label.lr('_FZD7','red')
label.lr('_FZD2','orange')
label.lr('_RET','purple')
label.lr('_IL6ST','magenta')
```



RC scores for selected LR pairs in BRCA subtypes

```
lrs = c('SFRP1_FZD2','WNT11_FZD7','WNT3_FZD7','CLEC11A_KIT','KITLG_KIT','MFAP2_NOTCH1','JAG1_NOTCH1','AL
df.basal.lrs = gather(df %>% filter(lr %in% lrs),'type','value', cc,sc,cs,ss,nn)
df.basal.lrs$type <- factor(df.basal.lrs$type, levels = c('cc','sc','cs','ss','nn'))
df.basal.lrs$lr <- factor(df.basal.lrs$lr, levels = lrs)
p3g = ggplot(df.basal.lrs, aes(x = "", y=value, fill = factor(type))) + geom_bar(width = 1, stat = "iden
  theme(axis.line = element_blank(), axis.text = element_blank()) + scale_fill_manual(values=ct.colors)
  labs(fill="class", x=NULL, y=NULL) + facet_grid(t ~ lr)
p3g
```



Expression of selected ligands and receptors in BRCA subtypes

```
subtype_colors = c(wes_palette(n=4, name="GrandBudapest1"),"#BBBBBB")

show = c('SFRP1','FZD7','NOTCH1','KIT','IL6ST','RET')
tmp = distinct(df.basal.lrs[df.basal.lrs$r %in% show,c('t','c_r','s_r','r')])
colnames(tmp) = c('t','c_l','s_l','l')
tmp = rbind(tmp,distinct(df.basal.lrs[df.basal.lrs$l %in% show,c('t','c_l','s_l','l')]))
colnames(tmp) = c('t','c_r','s_r','r')
tmp$r = factor(tmp$r, levels = show)

# add normal expression for each gene
for (rec in show) {
  ne = (df.basal.lrs %>% filter(r == rec) %>% select(n_r))[1,]
  if (is.na(ne)) {ne = (df.basal.lrs %>% filter(l == rec) %>% select(n_l))[1,]}
  tmp = rbind(tmp,data.frame(t='normal',c_r=as.numeric(ne),s_r=ne,r=rec))
}

p3ha = ggplot() +
  geom_bar(data = tmp, aes(x=t, y=c_r, fill=t),stat = "identity") +
  scale_fill_manual(values=subtype_colors) + facet_grid(. ~ r) + scale_y_log10(lim =c(1, 250), breaks =

p3hb = ggplot() +
  geom_bar(data = tmp, aes(x=t, y=s_r, fill=t),stat = "identity") +
  scale_fill_manual(values=subtype_colors) + facet_grid(. ~ r) + scale_y_log10(lim =c(1, 250), breaks =
```

```
pg = plot_grid(p3ha, p3hb, labels = c("a", "b"),ncol=1,align='v')
```

```
## Warning: Removed 7 rows containing missing values (position_stack).
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```

```
pg
```