ICeDT: Immune Cell Deconvolution in Tumor tissues

Wei Sun

3/7/2019

Introduction

This workfolow demonstrate the usage of ICeDT and also evaluate its performance using a simulated dataset.

```
library(ICeDT)
```

```
## Loading required package: alabama
```

```
## Loading required package: numDeriv
```

```
## Loading required package: gtools
```

```
dimX <- function(v){ if(is.null(dim(v))){r=length(v)} else{r=dim(v)}; r }
```

Simulate gene expressoin data

We simulate the expression of 250 genes from 135 bulk tumor samples as well as purified samples for 5 cell types, with 5 replicates per cell type.

```
set.seed(666)
data(mean_var_relation)
simData = simFunc(nS=135, nG=250, nrep=5, nCT=5, pctAb=20,
                  meanVar_Rel=mean_var_relation)
names(simData)
```

```
## [1] "MGene_Exp"    "PGene_Exp"    "cellType"     "aberrant"
## [5] "rho"          "tumor_purity"
```

```
lapply(simData, dimX)
```

```
## $MGene_Exp
## [1] 250 135
##
## $PGene_Exp
## [1] 250  25
##
## $cellType
## [1] 25
##
## $aberrant
## [1] 250
##
## $rho
## [1]   5 135
##
## $tumor_purity
## [1] 135
```

```
simData$tumor_purity[1:5]
```

```
## [1] 0.8513188 0.6029154 0.6582970 0.6808798 0.7493846
```

```
simData$PGene_Exp[1:2,1:7]
```

```
##              tumor        tumor        tumor        tumor        tumor
## [1,] 7.931848e-05 2.859189e-04 3.817089e-05 1.096758e-04 1.080461e-04
## [2,] 2.574509e-05 7.253033e-05 3.896733e-05 2.049412e-05 8.703739e-05
##              CT2        CT2
## [1,] 3034.01723 3893.93296
## [2,]   61.96612   33.01001
```

Estimate signature matrix

A sigmature matrix refers to a matrix of expected cell type specific expression for a set of genes across several cell types. These genes are usually selected due to their cell type-specific expression pattern.

We have simulated gene expression of tumor as well as four other cell types. However, in practice, we do not know the expression from tumor samples. Thus we only use the expression of non-tumor cell types when estimating signature matirx.

```
X = simData$PGene_Exp
dim(X)
```

```
## [1] 250  25
```

```
X[1:2,1:7]
```

```
##              tumor        tumor        tumor        tumor        tumor
## [1,] 7.931848e-05 2.859189e-04 3.817089e-05 1.096758e-04 1.080461e-04
## [2,] 2.574509e-05 7.253033e-05 3.896733e-05 2.049412e-05 8.703739e-05
##              CT2        CT2
## [1,] 3034.01723 3893.93296
## [2,]   61.96612   33.01001
```

```
X = X[,which(colnames(X) != "tumor")]

refE = refEstimate(X)
lapply(refE, dimX)
```

```
## $refMat
## [1] 250   4
##
## $refVar
## [1] 250   4
##
## $ctMu
## [1] 250   4
##
## $ctVar
## [1] 250   4
```

```
refE$refMat[1:2,]
```

```
##              CT2        CT3         CT4       CT5
## [1,] 3869.36411 992.272995 3102.820206 849.60168
## [2,]   78.45819   7.947046    7.470082  59.98775
```

```
refE$refVar[1:2,]
```

```
##              CT2       CT3       CT4       CT5
## [1,] 0.09148839 0.1196739 0.2716482 0.1365715
## [2,] 0.22901550 0.1410654 0.3117909 0.4490935
```

```
refE$ctMu[1:2,]
```

```
##           CT2     CT3      CT4      CT5
## [1,] 8.200511 6.83614 7.957212 6.678798
## [2,] 4.244652 1.96588 1.882645 3.948717
```

```
refE$ctVar[1:2,]
```

```
##             CT2       CT3       CT4       CT5
## [1,] 0.1206693 0.1277156 0.1657092 0.1319400
## [2,] 0.2358273 0.2138398 0.2565212 0.2908468
```

`refVar` is cell type-specific variance estiamte without borrowing information across cell types, while `ctVar` is the estimates after borrowing information across cell types. They have strong correlation.

```
cor(c(refE$refVar), c(refE$ctVar))
```

```
## [1] 0.8844
```

cell type composition estimation

```
Y = simData$MGene_Exp
Z = refE$refMat

ice1 = ICeDT(Y = Y, Z = Z, tumorPurity = NULL, refVar = NULL)
```

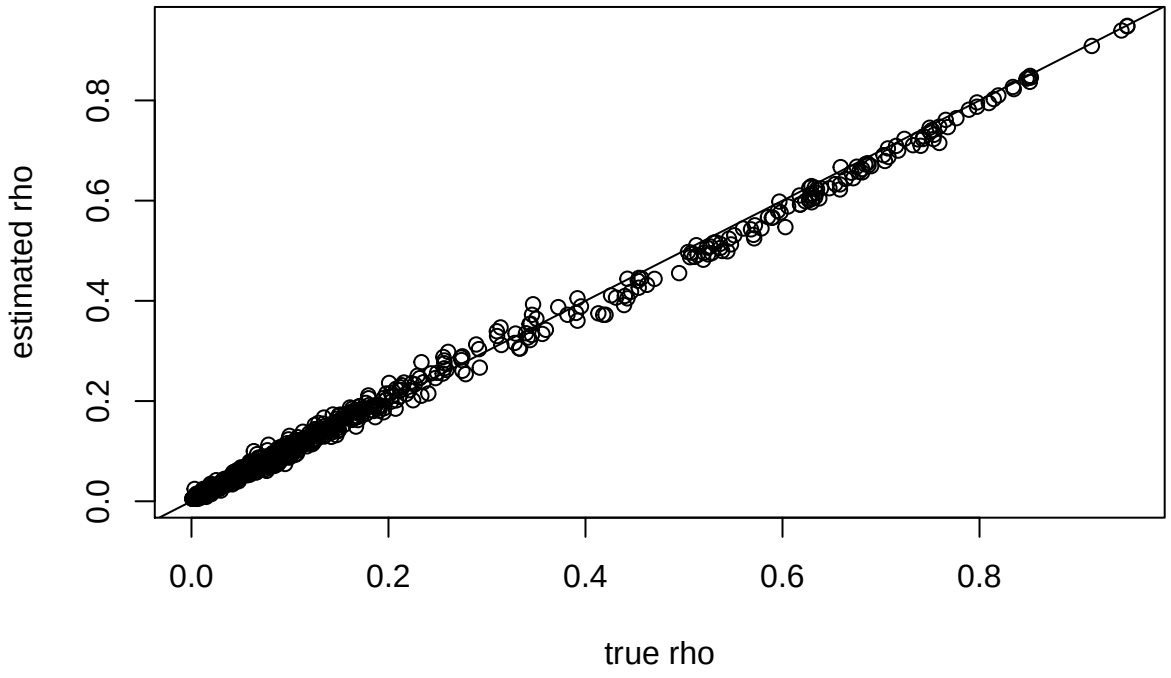
```
## Iter 1: max diff of rho: 0.526962138400084
```

```
## .........
## Iter 11: max diff of rho: 0.00552148595219343
## .........
## Iter 21: max diff of rho: 0.00228034794013091
## .....
```

```
lapply(ice1, dimX)
```

```
## $rho
## [1]   5 135
##
## $sigma2C
## [1] 135
##
## $sigma2A
## [1] 135
##
## $cProp
## [1] 135
##
## $cProb
## [1] 250 135
```

```
plot(simData$rho, ice1$rho, xlab="true rho", ylab="estimated rho")
abline(0,1)
```



Session information

```
sessionInfo()
```

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS  10.14.3
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ICeDT_0.99.1      gtools_3.8.1      alabama_2015.3-1  numDeriv_2016.8-1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.18    digest_0.6.15   rprojroot_1.3-2 backports_1.1.2
##  [5] magrittr_1.5    evaluate_0.11   stringi_1.2.4   rmarkdown_1.10
##  [9] tools_3.5.0     stringr_1.3.1   yaml_2.2.0      compiler_3.5.0
## [13] htmltools_0.3.6 knitr_1.20
```