

ESTIMATE vignette

Kosuke Yoshihara

August 15, 2013

1 Introduction

This vignette shows an example of using the `estimate` package in R. Normal cells in tumor tissue not only influence the tumor signal in molecular studies but also play an important role in cancer biology. The `estimate` package predicts the presence of stromal and immune cells in tumor tissue using gene expression data.

2 Example

We first use our data obtained from 10 ovarian cancer samples using Affymetrix U133Plus2.0 platform. It has gene-level expression data consisting 17,256 genes (row) and 10 samples (column). Second, we unify different number of genes per microarray platforms to 10,412 common genes. This algorithm uses HUGO GeneSymbols or Entrez gene IDs. Therefore, the row name in an input data must be gene symbols or Entrez gene IDs.

```
> library(estimate)
> OvarianCancerExpr <- system.file("extdata", "sample_input.txt", package="estimate")
> filterCommonGenes(input.f=OvarianCancerExpr, output.f="OV_10412genes.gct", id="GeneSymbol")

[1] "This dataset includes 10412genes."
[1] "0genes were mismatched."
```

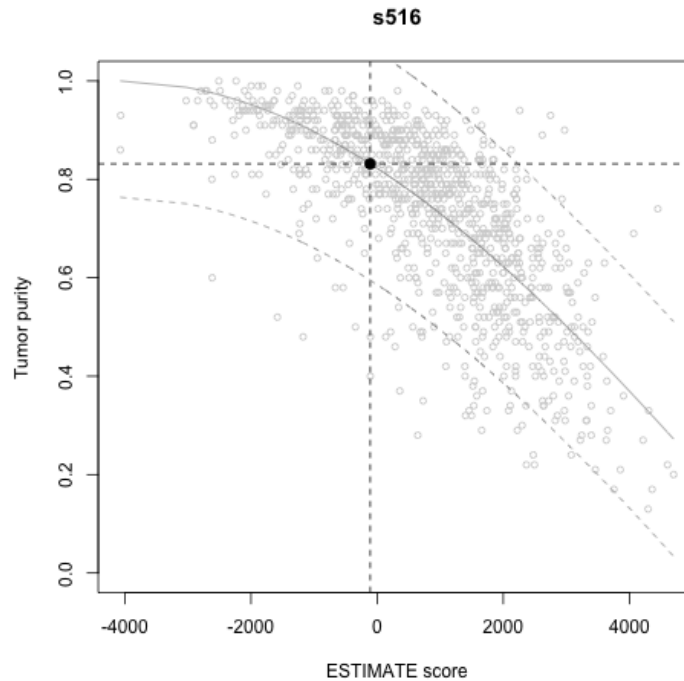
Third, we calculate stromal and immune scores that represent the presence of stromal and immune cells in tumor tissue, respectively. In addition, this function measures raw ESTIMATE score that are positively correlated with DNA copy number based tumor purity. This method is based on single sample gene set enrichment analysis (ssGSEA) algorithm.

```
> estimateScore("OV_10412genes.gct", "OV_estimate_score.gct", platform="affymetrix")

[1] "1 gene set: StromalSignature overlap= 141"
[1] "2 gene set: ImmuneSignature overlap= 141"
```

Once we calculate tumor purity, we can view the association between raw ESTIMATE score and tumor purity.

```
> plotPurity(scores="OV_estimate_score.gct", samples="s516", platform="affymetrix")
```



3 Running time

Finally, we show running time of our functions using ten ovarian cancer data.

Function `unifyCommonGenes`

```
> ptm <- proc.time()
> filterCommonGenes(input.f=OvarianCancerExpr, output.f="OV_10412genes.gct", id="GeneSymbol")
```

```
[1] "This dataset includes 10412genes."
```

```
[1] "0genes were mismatched."
```

```
> proc.time()-ptm
```

```
   user  system elapsed
1.313   0.028   1.342
```

Function `estimateScore`

```
> ptm <- proc.time()
> estimateScore("OV_10412genes.gct", "OV_estimate_score.gct", platform="affymetrix")

[1] "1 gene set: StromalSignature overlap= 141"
[1] "2 gene set: ImmuneSignature overlap= 141"

> proc.time()-ptm

      user  system elapsed
0.426  0.013   0.442
```