20181002b Gene signature assessment

cheungngo

2 October 2018

### 

## Gene signature assessment

### The 70-gene diagnostic signature for breast cancer progression risk

### 

library(BiocInstaller)   
biocLite("genefu")  
# "R package providing various functions relevant for gene expression analysis with emphasis on breast cancer"

### 

library(genefu)   
data(sig.gene70)   
dim(sig.gene70)

## [1] 70 9

head(sig.gene70)[,1:6]

## probe correlation average.good.prognosis.profile  
## NM\_003748 NM\_003748 -0.420671 0.12350000  
## NM\_003862 NM\_003862 -0.410964 0.05159091  
## Contig32125\_RC Contig32125\_RC -0.409054 0.05409091  
## U82987 U82987 -0.407002 0.06150000  
## AB037863 AB037863 -0.402335 0.06334091  
## NM\_020974 NM\_020974 -0.399987 -0.06231818  
## EntrezGene.ID NCBI.gene.symbol HUGO.gene.symbol  
## NM\_003748 8659 ALDH4A1 ALDH4A1  
## NM\_003862 8817 FGF18 FGF18  
## Contig32125\_RC NA <NA> <NA>  
## U82987 27113 BBC3 BBC3  
## AB037863 NA <NA> <NA>  
## NM\_020974 57758 SCUBE2 SCUBE2

### 

### How many components of the signature have a missing value for the associated NCBI gene symbol?

### 

sum(is.na(sig.gene70$NCBI.gene.symbol))

## [1] 14

### Kinases in the 70 gene signature

#### You can use grep on the Description field of the sig.gene70 data.frame to search for substrings of long gene names

#### How many of the members of the 70-gene signature are genes coding for kinases?

### 

length(grep('kinase',sig.gene70$Description))

## [1] 4

### 

### Some exploration in the genefu

### 

data(package = 'genefu')

names(sig.gene70)

## [1] "probe" "correlation"   
## [3] "average.good.prognosis.profile" "EntrezGene.ID"   
## [5] "NCBI.gene.symbol" "HUGO.gene.symbol"   
## [7] "Cytoband" "Alternative.symbols"   
## [9] "Description"