Package 'pvac'

April 7, 2025

Type Package

Title PCA-based gene filtering for Affymetrix arrays

Version 1.55.0

Date 2010-12-30

Author Jun Lu and Pierre R. Bushel

Maintainer Jun Lu <`jlu276@gmail.com`>, Pierre R. Bushel <`bushel@niehs.nih.gov`>

Description The package contains the function for filtering genes by the proportion of variation accounted for by the first principal component (PVAC).

License LGPL (>= 2.0)

Imports affy (>= 1.20.0), stats, Biobase

Depends R (>= 2.8.0)

Suggests pbapply, affydata, ALLMLL, genefilter

LazyLoad yes

biocViews Microarray, OneChannel, QualityControl

git_url https://git.bioconductor.org/packages/pvac

git_branch devel

git_last_commit 372aad1

git_last_commit_date 2024-10-29

Repository Bioconductor 3.21

Date/Publication 2025-04-06

Contents

`pvacFilter` *Filter genes by the proportion of variation accounted for by the first principal component (PVAC)*

Description

Compute the PVAC scores, derive a filtering threshold value, and return the names of probesets that have passed the filter

Usage

```
pvacFilter(abatch, pct=0.99)
```

Arguments

`abatch`	an instance of `AffyBatch` from the function call `ReadAffy`
`pct`	the percentile value of the emperical distribution of PVAC scores of a set of "non-expressed" genes. Used to select the filtering threshold. The default value is 0.99.

Details

This function implements a new filtering method for Affymetrix GeneChips, based on principal component analysis (PCA) on the probe-level expression data. Given that all the probes in a probeset are designed to target one or a common cluster of transcripts, the measurements of probes in a probeset should be correlated. The degree of concordance of gene expression among probes can be approximated by the proportion of variation accounted by the first principal component (PVAC). Using a wholly defined spike-in dataset, we have shown that filtering by PVAC provides increased sensitivity in detecting truly differentially expressed genes while controlling the false discoveries. The filtering threshold value is chosen from the PVAC score distribution in a set of "non-expressed" gene (those with absent calls in all samples).

Value

A list with the following components,

`aset`	Names of the probesets that have passed the filter
`nullset`	Names of the presumably "non-expressed" probesets (those with absent calls across all the study samples)
`pvac`	A named vector containing the PVAC scores of all probesets
`cutoff`	The PVAC cutoff value. The maximum is set to 0.5 (which corresponds to 50% of the total variation in a probeset)

Author(s)

Jun Lu

Examples

```
    if ( require(affydata) ) {
    data(Dilution)
    res = pvacFilter(Dilution)
    res$aset[1:5]  # 5 probesets that have passed the filter
 }
```

Index