

Package ‘NAPPA’

March 3, 2015

Type Package

Title Performs the Processing and Normalisation of Nanostring miRNA and mRNA Data

Version 2.0.1

Date 2014-06-02

Author Chris Harbron, Mark Wappett

Maintainer Mark Wappett <m.a.wappett@gmail.com>

Depends R (>= 3.0.1), plyr (>= 1.8), NanoStringNorm (>= 1.1)

Description Enables the processing and normalisation of the standard mRNA data output from the Nanostring nCounter software.

License GPL-2

Repository CRAN

NeedsCompilation no

Date/Publication 2015-03-03 20:38:56

R topics documented:

NAPPA-package	2
calibrate.sigmoid.curve	2
NAPPA	3
NS.Dilution	6
NS.Lung	7
nsolver.Dilution	7
Index	8

NAPPA-package

NAPPA: NAnostring Pre-Processing Algorithm.

Description

Performs positive control, negative control and housekeeping normalisation of Nanostring mRNA data. Enables multiple cartridges from the same codeset to be normalised as a single dataset.

Details

Package: NAPPA
Type: Package
Version: 2.0
Date: 2014-07-15
License: GPL-2

Author(s)

Author: Chris Harbron, Mark Wappett

Maintainer: Mark Wappett <mark.wappett@astrazeneca.com>

`calibrate.sigmoid.curve`*calibrate.sigmoid.curve is a curve fitting utility function.*

Description

Calibrate.sigmoid.curve is a utility function used to fit the parameters for the sigmoidparameters option within NAPPA controlling the degree of shrinkage applied to the housekeeping normalisation. This may be useful if applying NAPPA to a new sample type. It is recommended that this is only performed once a sizeable data corpus has been collected over multiple separate experiments.

Usage

```
calibrate.sigmoid.curve(  
  data,  
  plotit=T,  
  plotspline=F,  
  plotbyds=F)
```

Arguments

data	This should be the output RCC files from the Nanostring nCounter software, saved as a tab delimited text file. If there are multiple cartridges from the same experiment - simply merge together into a single file to run. In the first column, housekeeping genes should be indicated by changing the standard 'Endogenous,' call generated by nCounter to 'Housekeeping.'
plotit	Produce a scatterplot of gene slope against average gene expression, showing the fitted sigmoid curve.
plotspline	Add a spline fit to check the quality of fit of the sigmoid curve.
plotbyds	Add individual curves for each individual dataset in the list to check for heterogeneity between datasets.

Details

For each endogenous gene within each dataset a linear regression is performed of the expression level after positive control normalisation but before the housekeeping normalisation step in NAPPA against the housekeeping normalisation factor. The estimated slopes with their standard errors are recorded along with the average expression level of the gene. A sigmoid curve is fitted over all observations, modelling slope between zero and one as a function of average gene expression, with observations weighted by the inverse of the standard error of regression.

Value

A vector of length two giving the fitted scale and slope parameters for the fitted sigmoid curve which can be passed to the sigmoidparameters argument in NAPPA.

Author(s)

Chris Harbron

NAPPA

NAPPA: A novel statistical method for the processing and normalisation of mRNA data output from the Nanostring nCounter software

Description

Enables the processing and normalisation of the mRNA data output from the Nanostring nCounter software. Performs an adjustment based on the observed field of view for each lane. Performs a background correction using the truncated Poisson distribution adjustment. Performs a positive control normalisation using the E2 value. Performs a housekeeper normalisation by estimating the slope multiplier from sigmoidal curve fit.

Usage

```

NAPPA(
  data,
  tissueType = c("tumour", "cells"),
  NReferenceSamples = sampleNumber,
  sampleNumber = ncol(data) - 3,
  scaleFOV = T,
  background.method = c("poisson", "subtract", "poisson.global",
    "subtract.global", "subtract.max", "subtract.globalmax",
    "subtract.mean2sd", "subtract.globalmean2sd", "none"),
  nposcontrols = 4,
  poscontrol.method = c("average", "weighted.average",
    "geometric.mean", "average.prebc"),
  hk.method = c("shrunk.correct", "shrunk.subtract",
    "subtract", "correct"),
  betas = NULL,
  hknormfactor.mean = NULL,
  sigmoidparameters = NULL,
  addconstant = 10,
  imputezeroes.method = c("min", "min.retro", "none"),
  raise.low.counts=2,
  output = NULL)

```

Arguments

<code>data</code>	This should be the output RCC files from the Nanostring nCounter software, saved as a tab delimited text file. If there are multiple cartridges from the same experiment - simply merge together into a single file to run. In the first column, housekeeping genes should be indicated by changing the standard 'Endogenous,' call generated by nCounter to 'Housekeeping.'
<code>tissueType</code>	Should be indicated as either 'tumour' or 'cells' depending on the sample type used.
<code>NReferenceSamples</code>	The number of samples which will be used to determine the mean expression level in the calculation of the normalisation shrinkage parameters and housekeeping correction factor. The first NReferenceSamples sample columns in the input file are used. The default is to use all samples.
<code>sampleNumber</code>	A synonym for NReferenceSamples
<code>scaleFOV</code>	Logical flag indicating whether to normalise for the number of successfully imaged fields of view. By default set to TRUE.
<code>background.method</code>	The method used for background correction. The default is poisson which performs a Truncated Poisson Correction using the average background counts for each lane (sample). Other options are poisson.global which performs a Truncated Poisson Correction using a global average background over all lanes, subtract, subtract.max and subtract.mean2sd which subtract the mean, maximum and mean plus 2 standard deviations of the background count within

	each lane, and their global versions <code>subtract.global</code> , <code>subtract.globalmax</code> and <code>subtract.globalmean2sd</code> which subtract the same statistics calculated over all lanes.
<code>nposcontrols</code>	The number of the positive control probes to use in calculating a positive control normalisation factor. By default set to 4.
<code>poscontrol.method</code>	The method used to combine the positive controls to generate a single positive control value for each lane. The default option is <code>average</code> , other options are <code>weighted.average</code> , <code>geometric mean</code> and <code>average.prebc</code> (an average based on the values before background correction, effectively reversing the order of the background and positive control steps)
<code>hk.method</code>	The method used for the housekeeping normalisation step. The default is <code>shrunken.correct</code> , performing a shrunken correction. Other options are <code>shrunken.subtract</code> , <code>correct</code> , and <code>subtract</code> which perform either housekeeper subtraction or correction using shrunken or standard subtraction.
<code>betas</code>	The shrinkage parameters for shrunken housekeeping normalisation. These may be taken from a previous run of NAPPA using the <code>output="Betas"</code> option. By default these are calculated for each gene based on the first <code>NReferenceSamples</code> samples.
<code>hknormfactor.mean</code>	The correction term for the housekeeping normalisation factors. This may be taken from a previous run of NAPPA using the <code>output="HousekeepingFactor"</code> option. By default this is calculated as the average housekeeping factor within the data.
<code>sigmoidparameters</code>	A vector of length two containing the location and slope parameters for the sigmoid curve used to calculate the shrinkage parameters for the shrunken housekeeping normalisation. By default these are determined from the <code>tissueType</code> argument.
<code>addconstant</code>	Constant added to the final expression levels to present them on a more user-friendly scale. By default set to 10.
<code>imputezeroes.method</code>	The method used to impute values for zero raw counts. The default option <code>min</code> imputes the lowest normalised value observed for that gene in the normalised data set. The option <code>none</code> does no imputation, and the option <code>min.retro</code> is a legacy option allowing back-compatibility with earlier versions of NAPPA.
<code>raise.low.counts</code>	Minimum value to raise all raw counts to in the initial pre-processing step. By default set to 2. To omit this step set to a value of zero.
<code>output</code>	Values to be returned from the function. By default the function returns just a gene expression matrix. If <code>output</code> is set to a non-null value then a list is returned containing the requested components as detailed in the value section.

Details

The RCC file that is output by the Nanostring nCounter software contains two empty rows when output. These rows must be removed prior to analysis. Multiple cartridges that use the same codeset

can be merged together into a single file (there is no upper limit to sample size). It is recommended that as many samples as possible are used to calculate the gene means for the housekeeping normalisation (12 samples is the recommended minimum).

Value

The value of the NAPPA function is determined by the output option. By default a matrix is returned containing the gene expression values with genes as rows and lanes (samples) as columns. If output is set to a non-null value then a list is returned containing the gene expression matrix as the item `geneexpression` and those of the following items listed in output:

Housekeeping	A matrix of the housekeeping genes
HousekeepingFactor	A vector of the housekeeping normalisation factor for each lane, and also the housekeeping correction factor, <code>HousekeepingFactor.Mean</code>
Betas	A vector of the shrinkage factors used in the housekeeping normalisation for each gene
Backgrounds	A vector of the average background for each lane
PosFactor	A vector of the positive control factors used for each lane
Description	The parameters used by the NAPPA function

If output is set to "All", a list containing all components is returned.

Author(s)

Chris Harbron Mark Wappett

Examples

```
# Read in RCC file and perform NAPPA normalisation
data(NS.Lung)
output <- NAPPA(data=NS.Lung, tissueType='tumour', NReferenceSamples=50)
```

NS.Dilution	<i>Nanostring nCounter data matrix</i>
-------------	--

Description

An example data matrix. The housekeeping genes have been indicated in the left-most column.

Usage

```
data(NS.Dilution)
```

Format

A data frame with 513 observations on the following 27 variables. The first three columns contain information about the gene and codeset. The remaining 24 columns contain the raw code counts for each of the 24 input samples.

NS.Lung	<i>Nanostring nCounter data matrix</i>
---------	--

Description

An example data matrix. The housekeeping genes have been indicated in the left-most column.

Usage

```
data(NS.Lung)
```

Format

A data frame with 513 observations on the following 53 variables. The first three columns contain information about the gene and codeset. The remaining 50 columns contain the raw code counts for each of the 50 input samples.

nsolver.Dilution	<i>Nanostring nCounter data matrix</i>
------------------	--

Description

A data matrix processed using the nSolver method.

Usage

```
data(nsolver.Dilution)
```

Format

A data frame with 492 observations on the following 24 variables.

Index

- *Topic **Nanostring**
 - calibrate.sigmoid.curve, 2
 - NAPPA, 3
 - *Topic **Normalisation**
 - NAPPA, 3
 - *Topic **datasets**
 - NS.Dilution, 6
 - NS.Lung, 7
 - nsolver.Dilution, 7
 - *Topic **package**
 - NAPPA-package, 2
- calibrate.sigmoid.curve, 2
- NAPPA, 3
- NAPPA-package, 2
- NS.Dilution, 6
- NS.Lung, 7
- nsolver.Dilution, 7