Package 'ICEDT'

September 25, 2018

Type Package

Title Immune Cell Expression Deconvolution in Tumor Tissues (ICeD-T)

Version 0.99

Date 2018-06-08

Author Douglas Roy Wilson, Jr.

Maintainer Wei Sun <wsun@fredhutch.org>

Description
Immune cell Composition estimation in Tumor Samples accounting for Aberrant Gene Behavior.

Depends alabama

License GPL (>= 2)

R topics documented:

`ICEDT-package` *ICeD-T*

Description

ICeD-T

Details

The DESCRIPTION file: The ICeD-T package is designed to estimate cell type composition from tumor samples while allowing for aberrant gene expression behavior within tumor samples. Please see the manuscript referenced below for further details.

Author(s)

Douglas Roy Wilson, Jr.

Maintainer: Wei Sun <wsun@fredhutch.org>

References

Wilson, Douglas R; Ibrahim, Joseph G.; Sun, Wei "ICED-T Provides Accurate Estimates of Immune Cell Abundance in Tumor Samples by Allowing for Aberrant Gene Expression Patterns" «In Review», 2018.

`ICeDT_fit_noWgt_noRef` *ICeD-T Model Fit (No Weights, No Supplied Reference)*

Description

This function uses purified, immune reference sample data and mixed expression profiles to estimate the cell type composition of tumor samples. This function can utilize information regarding tumor purity if available.

Usage

```
ICeDT_fit_noWgt_noRef(Y, X, cellType, fixedCT = NULL,
fixedCT_rho = NULL, useRho = FALSE, borrow4SD = TRUE,
maxIter_prop = 100, maxIter_PP = 100, RhoConv_CO = 1e-04,
Subj_CO)
```

Arguments

`Y`	A matrix of normalized mixture expression profiles (e.g. FPKM, TPM) of dimensions nG by nS. This matrix contains one row per gene and one column per mixture expression profile.
`X`	A matrix of normalized reference sample expression profiles of dimensions nG by nP. This matrix contains one row per gene and one column per purified reference sample of immune cells. This dataset must contain at least two samples of each of the immune cell types one desires to estimate.
`cellType`	A vector indicating the immune cell identities of the expression profiles contained in X. It must be of length nP and its order must match the order of the columns of X.
`fixedCT`	This parameter dictates which label used by cellType indicates the cell type for which cell type proportion information is known.
`fixedCT_rho`	Vector of length nS which contains the cell type proportion for fixedCT for each sample. Order must match that given in the columns of Y.
`useRho`	useRho is a boolean variable (TRUE = use fixed cell type proportion information, FALSE = estimate fixed cell type proportion).
`borrow4SD`	Indicator (TRUE/FALSE) of whether or not to borrow information across cell types in order to estimate the standard deviations in pure reference samples.
`maxIter_prop`	Maximum number of iterations used for updating proportion parameters within a single EM iteration.
`maxIter_PP`	Maximum number of iterations used for updating
`RhoConv_CO`	The maximum difference in estimated cell type proportions for defining convergence.
`Subj_CO`	The number of subjects who must meet the RhoConv_CO cutoff in order to stop iterating the model.

Details

The following model is designed to compute ICeD-T model fits when purified reference sample data is available and when one does not wish to use previously defined references. In the vernacular established above, the fixed cell type is typically a tumor cell type and fixedCT_rho thus represents tumor purity.

In order to utilize this function, one must create at least two dummy "pure" tumor samples and append them to X with an appropriate addition to fixedCT.

Value

A list item containing:

- $IC_Abundance : One row per subject, one column per cell type containing ICeD-T estimates.

- $Fixed_CellType: The label defining the fixed cell type.

- $Sigma2M : The estimated variance parameter for consistent marker genes. One parameter for each subject in the order specified in IC_Abundance rows.

- $Sigma2A : The estimated variance parameter for Aberrant genes. One parameter for each subject in the order specified in IC_Abundance rows. If Sigma2A > Sigma2M for a subject, this indicates that these parameters must be switched, as must P_Consistent and PP_Consistent.

- $Z : The utilized reference matrix with one row per gene and one column per immune cell type in fixedCT.

- $CT_Var: The variance profiles of the established immune cell types from purified reference samples.

- $P_Consistent: The estimated proportion of consistent genes, one value for each subject in the order established by IC_Abundance. If Sigma2M > Sigma2A and one is interested in Aberrant genes, "P_Consistent" must be replaced by 1-P_Consistent for such subjects.

- $PP_Consistent: A matrix of conditional probabilities that a given gene is Aberrant given the model estimates (one row per gene, one column per subject). For subjects where Sigma2M > Sigma2A, the corresponding column of PP_Consistent must be replaced by 1-PP_Consistent.

Note

Please note the addition of "dummy" pure tumor expression profiles. Additionally, note the proper treatment of $P_Consistent and $PP_Consistent if Sigma2M > Sigma2A for any subject.

Author(s)

Douglas Roy Wilson, Jr.

References

Wilson, Douglas R; Ibrahim, Joseph G.; Sun, Wei "ICED-T Provides Accurate Estimates of Immune Cell Abundance in Tumor Samples by Allowing for Aberrant Gene Expression Patterns" «In Review», 2018.

ICeDT_fit_noWgt_suppRef

ICeD-T Model Fit (No Weights, Supplied Reference)

Description

This function uses a pre-computed reference expression profile for K immune cell types and mixed expression profiles to estimate the cell type composition of tumor samples. This function can utilize information regarding tumor purity if available.

Usage

```
ICeDT_fit_noWgt_suppRef(Y,fixedCT_rho=NULL,useRho=FALSE,RefMat,
                        maxIter_prop = 100,
                        maxIter_PP=100,RhoConv_CO = 1e-4, Subj_CO)
```

Arguments

Y	A matrix of normalized mixture expression profiles (e.g. FPKM, TPM) of dimensions nG by nS. This matrix contains one row per gene and one column per mixture expression profile.
fixedCT_rho	Vector of length nS which contains the cell type proportion for fixedCT for each sample. Order must match that given in the columns of Y.
useRho	useRho is a boolean variable (TRUE = use fixed cell type proportion information, FALSE = estimate fixed cell type proportion).
RefMat	A reference matrix containing one row per gene and one column for each assessed immune cell type (dimensions nG by K).
maxIter_prop	Maximum number of iterations used for updating proportion parameters within a single EM iteration.
maxIter_PP	Maximum number of EM iterations used for updating.
RhoConv_CO	The maximum difference in estimated cell type proportions for defining convergence.
Subj_CO	The number of subjects who must meet the RhoConv_CO cutoff in order to stop iterating the model.

Details

The following model is designed to compute ICeD-T model fits using no weights and a supplied reference. In the vernacular established above, the fixed cell type is typically a tumor cell type and fixedCT_rho thus represents tumor purity.

Value

A list item containing:

- $IC_Abundance : One row per subject, one column per cell type containing ICeD-T estimates.

- $Fixed_CellType: The label defining the fixed cell type.

- $Sigma2M : The estimated variance parameter for consistent marker genes. One parameter for each subject in the order specified in IC_Abundance rows.

- $Sigma2A : The estimated variance parameter for Aberrant genes. One parameter for each subject in the order specified in IC_Abundance rows. If Sigma2M > Sigma2A for a subject, this indicates that these parameters must be switched, as must P_Consistent and PP_Consistent.

- $P_Consistent: The estimated proportion of consistent genes, one value for each subject in the order established by IC_Abundance. If Sigma2M > Sigma2A and one is interested in Aberrant genes, "P_Consistent" must be replaced by 1-P_Consistent for such subjects.

- $PP_Consistent: A matrix of conditional probabilities that a given gene is Aberrant given the model estimates (one row per gene, one column per subject). For subjects where Sigma2M > Sigma2A, the corresponding column of PP_Consistent must be replaced by 1-PP_Consistent.

Note

Additionally, note the proper treatment of $P_Consistent and $PP_Consistent if Sigma2M > Sigma2A for any subject.

Author(s)

Douglas Roy Wilson, Jr.

References

Wilson, Douglas R; Ibrahim, Joseph G.; Sun, Wei "ICED-T Provides Accurate Estimates of Immune Cell Abundance in Tumor Samples by Allowing for Aberrant Gene Expression Patterns" «In Review», 2018.

ICeDT_fit_Wgt_suppRef *ICeD-T Model Fit (Weights, Supplied Reference)*

Description

This function uses a supplied immune reference and mixed expression profiles to estimate the cell type composition of tumor samples. This function can utilize information regarding tumor purity if available.

Usage

```
ICeDT_fit_Wgt_suppRef(Y,fixedCT_rho=NULL,useRho=FALSE,RefMat,RefVar,
varLog = FALSE, limitWgt = TRUE, limitQuant = c(0.10,0.90),
userWgt=NULL, useIQR=FALSE,maxIter_prop = 100,
            maxIter_PP=100,RhoConv_CO = 1e-4, Subj_CO)
```

Arguments

Y	A matrix of normalized mixture expression profiles (e.g. FPKM, TPM) of dimensions nG by nS. This matrix contains one row per gene and one column per mixture expression profile.
fixedCT_rho	Vector of length nS which contains the cell type proportion for fixedCT for each sample. Order must match that given in the columns of Y.
useRho	useRho is a boolean variable (TRUE = use fixed cell type proportion information, FALSE = estimate fixed cell type proportion).

RefMat	A reference matrix containing one row per gene and one column for each assessed immune cell type (dimensions nG by K).
RefVar	A reference variance matrix containing one row per gene and one column for each assessed immune cell type (dimensions nG by K). Different from RefMat which characterizes the "average expression", this reference contains the "variance in expression." Order of columns and rows must match that in RefMat.
varLog	A boolean variable (TRUE=RefVar is computed for log-scale expression, FALSE=RefVar is computed from normal-scale expression).
limitWgt	A boolean variable to specify whether or not to control variance weights for robustness considerations (TRUE = constrain model weights, FALSE = do not constrain model weights)
limitQuant	A vector of length two specifying the upper and lower quantiles used to limit model weights. The first element specifies the lower quantile to use for weight replacment (Bottom limitQuant[1]*100% replaced). The second element specifies the upper quantile to use for weight replacement (Weights above limitQuant[2]*100% replaced).
userWgt	If user specifies useIQR = 6, then user specified model weights are used. This variable contains these weights.
useIQR	A categorical variable (0 or 6) which gives the form of the weights to use. useIQR = 0 utilizes the weight structure suggested in the ICeD-T manuscript. useIQR = 6 utilizes user supplied weights.
maxIter_prop	Maximum number of iterations used for updating proportion parameters within a single EM iteration.
maxIter_PP	Maximum number of EM iterations allowed.
RhoConv_CO	The maximum difference in estimated cell type proportions for defining convergence.
Subj_CO	The number of subjects who must meet the RhoConv_CO cutoff in order to stop iterating the model.

Details

The following model is designed to compute ICeD-T model fits using no weights and a supplied reference. In the vernacular established above, the fixed cell type is typically a tumor cell type and fixedCT_rho thus represents tumor purity.

Value

A list item containing:

- $IC_Abundance : One row per subject, one column per cell type containing ICeD-T estimates.

- $Fixed_CellType: The label defining the fixed cell type.

- $Sigma2M : The estimated variance parameter for consistent marker genes. One parameter for each subject in the order specified in IC_Abundance rows.

- $Sigma2A : The estimated variance parameter for Aberrant genes. One parameter for each subject in the order specified in IC_Abundance rows. If Sigma2M > Sigma2A for a subject, this indicates that these parameters must be switched, as must P_Consistent and PP_Consistent.

- $P_Consistent: The estimated proportion of consistent genes, one value for each subject in the order established by IC_Abundance. If Sigma2M > Sigma2A and one is interested in Aberrant genes, "P_Consistent" must be replaced by 1-P_Consistent for such subjects.

- $PP_Consistent: A matrix of conditional probabilities that a given gene is Aberrant given the model estimates (one row per gene, one column per subject). For subjects where Sigma2M > Sigma2A, the corresponding column of PP_Consistent must be replaced by 1-PP_Consistent.

Note

Additionally, note the proper treatment of $P_Consistent and $PP_Consistent if Sigma2M > Sigma2A for any subject.

Author(s)

Douglas Roy Wilson, Jr.

References

Wilson, Douglas R; Ibrahim, Joseph G.; Sun, Wei "ICED-T Provides Accurate Estimates of Immune Cell Abundance in Tumor Samples by Allowing for Aberrant Gene Expression Patterns" «In Review», 2018.

Index