

R script description:

Modeling tissue contamination to improve
molecular identification of the primary tumor site of
metastases

Martin Vincent, Katharina Perell, Finn Cilius Nielsen,
Gedske Daugaard and Niels Richard Hansen

October 21, 2013

Abstract

This document contains short descriptions of the R scripts applied
in the article *modeling tissue contamination to improve molecular identification of the primary tumor site of metastases*.

R packages need

For building predictors, running simulations and plotting:

- **msg1**
For multinomial group lasso predictors. Available on CRAN.
- **pamr**
For ANOVA+PAM predictors. Available on CRAN.
- **Biobase**
Available from <http://www.bioconductor.org/>.
- **subsamples**
Available from <http://www.math.ku.dk/~richard/msg1/>
- **ggplot2**, **reshape2** and **plyr**
For plotting. All available on CRAN.

Data

Data is available from the Gene Expression Omnibus with accession number GSE51429. For convenience relevant data is supplied as a bioconductor expression set, which is available from:

<http://www.math.ku.dk/~richard/msgl/> .

Script for validating predictors based solely on the core biopsies

- `loocv_of_cb_predictor.R`
Leave-one-out cross-validation of multinomial group lasso predictor.

Script for validating predictors based solely on the resections

- `validation_of_prim_msgl_predictors.R`
Using the multinomial group lasso method.
- `validation_of_prim_anova+pam.R`
Using the ANOVA+PAM method.
- `figure_3.R`
Script for producing a plot similar to Figure 3. The validation script must be run first.
- `figure_4.R`
Script for producing a plot similar to Figure 4. The validation script must be run first.

Scripts for validating predictors based on a combination of the resection and core biopsies

- `validation_of_prim+cb_msgl_predictors.R`
Using the multinomial group lasso method.
- `validation_of_prim+cb_anova+pam_predictors.R`
Using the ANOVA+PAM method.
- `figure_6.R`
Script for producing a plot similar to Figure 6. The validation script must be run first.

Principal component plot

- `figure_5.R`
Script for producing a plot similar to Figure 5.

General scripts

- `contamination_model_simulation.R`
Contamination model and simulation R-functions.
- `anova+pam_fit.R`
ANOVA+PAM method R-functions.
- `load_data.R`
Loading and normalizing data script.