marcus-seldin / QENIE Public

Code Issues Pull requests Actions Projects Wiki

master 1 branch 0 tags Go to file Code

70bd461 on 5 Sep 2018 85 commits

Inte...	Add files via ...	4 years ago
Mu...	Add files via ...	4 years ago
QQ...	Update QQ-...	3 years ago
R-s...	Update R-sc...	3 years ago
REA...	Update REA...	3 years ago
Sec...	Add files via ...	4 years ago
adi...	Add files via ...	4 years ago
liver...	Add files via ...	4 years ago

README.md

QENIE (Quantitative Endocrine Network Interaction Estimation)

About

Correlation-based approach for identification of endocrine interactions

Readme

2 stars

2 watching

6 forks

Releases

No releases published

Packages

No packages published

Contributors 2

marcus-seldin Marcus Seldin

calpan

Languages

R 100.0%

Correlation-based approach for identification of endocrine interactions

Required package

The Ssec ranking score and protein-specific pathway enrichment commands in R utilize functions from the package WGCNA

External data

In addition to tissue-specific expression arrays, this pipeline also imports information from Uniprot to filter for secreted proteins. This list has been uploaded for ease of pipeline execution, however is subjected to updates. Revised lists can be retrieved from the following sources:

UniProt <www.uniprot.org/> deposited annotations for Organism:mouse "Mus musculus (Mouse) [10090]" and subcellular localization:secreted :"Secreted [SL-0243]"

Citation: UniProt: the universal protein knowledgebase Nucleic Acids Res. 45: D158-D169 (2017)
These are provided as: Secreted_proteins_Uniprot

Tissue-specific expression

Filtering for tissue-specific expression was performed by manual inspection using BioGPS on the following mouse arrays http://biogps.org/dataset/BDS_00009/: GeneAtlas GNF1M, gcrma While this can be easily automated, we felt more confident with inspection due to the small number of samples per tissue and consequent high level of variation among expression

GEO code for arrays: GSE1133

Citation: McClurg P, Janes J, Wu C, Delano DL, Walker JR, Batalov S, Takahashi JS, Shimomura K, Kohsaka A, Bass J, Wiltshire T, Su AI (2007) Genomewide association analysis in diverse inbred mice: power and population structure. Genetics 176(1):675-83.

Data pretreatment:

Pretreatment of data is described below: Most mouse expression arrays were performed on a Affymetrix HT_MG-430A, unless otherwise indicated in the GEO accession

> GEO Accession for adipose and liver arrays: GSE64770
>
> Raw data deposited in GEO is RMA-normalized microarray data. From these, each numeric value represents a probe detected in each HMDP animal. Given that some of probes correspond to the same gene, we chose to start with a single expression value for each unique gene (averaged amongst probes). Also, these studies contain multiple mice per strain. Given that for each study, mice were fed the same diet and age-matched, we aggregated expression values to one average value per mouse strain. These averages are discussed in greater detail below:

> Our pipeline begins with gene expression arrays for liver and adipose tissue, where each gene is represented as an averaged value across probes and strains used in the study. These aggregate matrices are also provided in this repository. The arrays consisted of ~22,400 probes which were aggregated to averages for each gene (12,242). The expression values for each mouse were also averaged to reflect a single value per gene per strain (106). Therefore, each liver and adipose tissue expression matrix consists of 12,242 genes among 106 unique HMDP strains.
>
> To account for different arrays containing differing number of genes, a portion of the Ssec pipeline script divides the Ssec score to the total length of the target tissue genes detected. This step is meant to prevent inflation of these scores by a larger number of genes detected in each target tissue, thus providing an adequate comparison for values between cross-tissue circuits.

We subdivided the commands into the following steps (all listed in the R_script file):

1. Construct cross-tissue correlation and pvalue matrices
2. Compute rowsum -ln(pvalue) for ranking interactions and filter for factors proteins as secreted
3. Condition correlation matrix on a by-gene basis for pathway enrichment

Generation of QQ-Plots to infer secreted vs non-secreted factors

The script used to generate QQ-plots from the Ssec scores (step 2 above) was designed and provided by Simon Koplev and listed as: QQ-Plot_R-script.R

Acknowledgements

We would like to thank Arjun Sarathi and Manikandan (Mani) Narayanan at the Bioinformatics and Integrative Data Sciences (BIRDS) Lab at IIT Madras for their thorough troubleshooting and identifying a switch in the initially-described QQ-Plot Axes.

Any questions/comments, please contact mseldin3@gmail.com

Additional GEO Accessions for datasets used in the study are listed below:

chow aorta: GSE38120
chow heart: GSE77263
chow liver: GSE16780
chow adipose: GSE42890
high fat/high sucrose hypothalamus: GSE79551