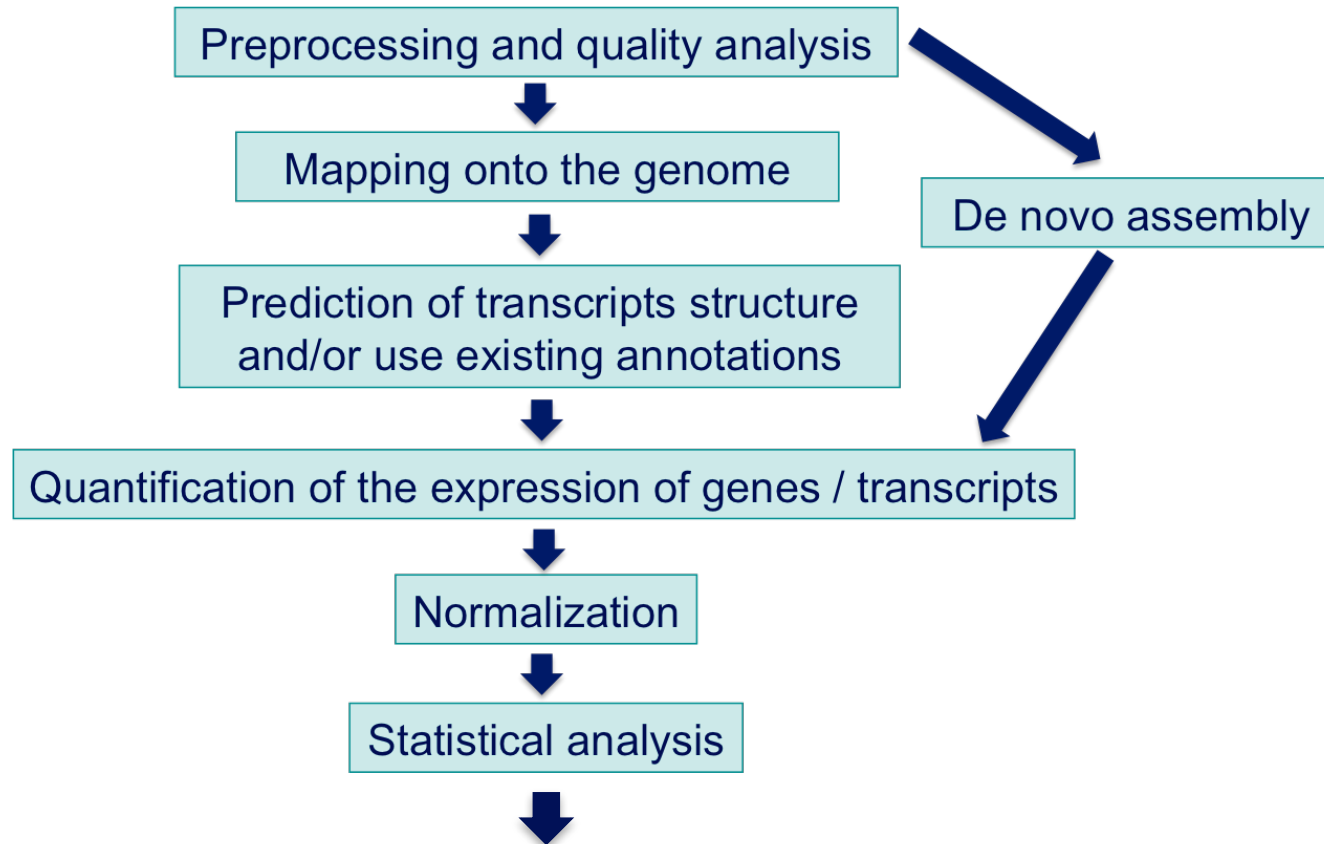


Functional analysis of RNA-seq data

Céline Keime
keime@igbmc.fr

Analysis of RNA-seq data



Functional enrichment analysis,
pathway analysis, integration with other
data, ...

Functional analysis

- A lot of functional analysis tools available
 - Initially developed for microarray data
 - e.g. GO tools listed in <https://omictools.com/search?q=gene+ontology>
 - Methods specific to RNA-seq data
 - goseq (Young et al., Genome Biology 2010;11:R14)
 - SeqGSEA (Wang et al. BMC Bioinformatics 2013, 14(Sup5):S16)
 - GSAASeqSP (Xiong et al Scientific Reports 2014; 4:6347)
- DAVID will be used for this practical session because
 - Graphical interface & free software
- DAVID
 - Database for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery
 - <https://david-d.ncifcrf.gov/>
 - A very interested article describing how to use DAVID : Huang et al. Nature Protocols 2009;4(1):44-57.

DAVID

Annotation Summary Results

Current Gene List: demolist1

Current Background: Homo sapiens

- ☒ Disease (1 selected)
- ☒ Functional_Categories (3 selected)
- ☒ Gene_Ontology (3 selected)
- ☒ General_Annotations (0 selected)
- ☒ Literature (0 selected)
- ☒ Main_Accessions (0 selected)
- ☒ Pathways (3 selected)
- ☒ Protein_Domains (3 selected)
- ☒ Protein_Interactions (0 selected)
- ☒ Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

- Functional Annotation Clustering
- Functional Annotation Chart
- Functional Annotation Table

Different sources of annotation

- Disease (OMIM)
- Gene Ontology
- Pathways (KEGG, Biocarta)
- Protein Domains (InterPro, SMART)
- Protein Interaction (BIND)
- ...

Different tools

- Functional Annotation Clustering
 - Cluster functionally similar terms associated with a gene list into groups
- Functional Annotation Chart
 - Identify enriched annotation terms associated with a gene list
- Functional Annotation Table
 - Query associated annotations for all genes from a list

Gene Ontology

- Defines concepts/classes used to describe gene function and relationships between these concepts
- Classifies functions along three aspects
 - Molecular function : molecular activities of gene products
 - Cellular component : where gene products are active
 - Biological process : pathways and larger processes made up of the activities of multiple gene products

Exercise : functional analysis

- Use DAVID to perform functional analysis of genes significantly over-expressed in siMitf vs siLuc samples
 1. Select over-expressed genes using the filter tool on GalaxEast
 - Proposed thresholds :
Adjusted p-value < 0.05 and $\log_2(\text{FoldChange}) > 1$
 2. Create a file with gene name for all these genes using the cut tool on GalaxEast
 3. Analyse this gene list using DAVID

1. Select over-expressed genes

- Among significantly differentially expressed genes, select genes with $\log_2(\text{FoldChange}) > 1$

Filter data on any column using simple expressions (Galaxy Version 1.1.0) Options

Filter

42: siMitfvssiLuc.up.annot.txt

Dataset missing? See TIP below.

With following condition

c14>1

Double equal signs, ==, must be used as shown above. To filter for an arbitrary string, use the Select tool.

Number of header lines to skip

1

Execute

⚠ Double equal signs, ==, must be used as "equal to" (e.g., c1 == 'chr22')

ℹ TIP: Attempting to apply a filtering condition may throw exceptions if the data type (e.a.. string, integer) in every

History

search datasets

RNAseq1709
23 shown, 19 deleted
290.36 MB

42:
siMitfvssiLuc.up.annot.txt
3,793 lines
format: tabular, database: hg38

12	13	14	15
ic siMitf	FoldChange	log2FoldChange	pva
19861	3.936	1.977	0
8763	4.932	2.302	0
5532	14.313	3.839	0
16741	4.324	2.112	0

43: Filter on data 42

612 lines

format: tabular, database: hg38

Filtering with c14>1,
kept 16.13% of 3793 valid lines
(3793 total lines).

1	2	3	4
Gene ID	siLuc2	siLuc3	siMitf3
ENSG00000018408	4640	5232	18689
ENSG00000081189	1686	1770	8339
ENSG00000124942	310	416	5136
ENSG00000143341	3663	3901	15667

2. Create a list of gene names

- Select associated gene names in the previous table

Cut columns from a table (Galaxy Version 1.0.1) Versions Options

Cut columns

Delimited by

From



44: Cut on data 43 View Edit Close

612 lines
format: **tabular**, database: **hg38**

1

Associated Gene Name

WWTR1

MEF2C

AHNAK

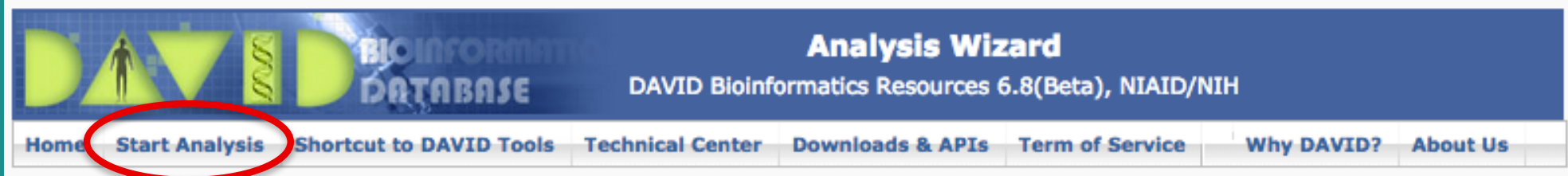
HMCN1



siMitfvssiLuc_upgenes_lfc1_padj005.txt file

3. Analyse your gene list using DAVID

- Go to <https://david-d.ncifcrf.gov>
- Click on Start Analysis



3. Start DAVID analysis

■ Enter your gene list

Upload List Background

Upload Gene List

[Demolist 1](#) [Demolist 2](#)
[Upload Help](#)

Step 1: Enter Gene List
A: Paste a list

Or
B: Choose From a File

Multi-List File

Step 2: Select Identifier

Step 3: List Type

Gene List
Background

Step 4: Submit List

■ Select species

Please note that multiple species have been detected in your gene list. You may select a specific specie(s) with the List Manager on the left side of the page by highlighting the specific specie(s) and pressing the "Select" button. As a default, all species in your list will be used for analysis. Also note that you may need to select an appropriate background under the "BACKGROUNDS" tab in the manager to the left. By default, the background corresponding to the first species in the list will be selected if an uploaded or Affymetrix background is not in use.

or more species [Help](#)

- Use All Species -
Homo sapiens(508)
Mus musculus(464)
Rattus norvegicus(433)


List Manager [Help](#)

siMitfvssiLuc_upgenes_lfc1_padj0

Select List to:

[View Unmapped Ids](#)

Exercise : functional analysis

1. What are the 10 most enriched functional annotation terms among annotations of the genes from your list ?
How many genes are annotated with each of these terms ?
Which genes are annotated with the most enriched term ?
2. As you see redundancy in previous results, it could be interesting to cluster functionally similar terms into groups.
Look at the results of this clustering. What is the first identified cluster ?
Click on  to visualize members of this cluster (genes and annotations).
3. *Claudin 15* gene is a member of this cluster.
What are all associated annotations for this gene ?
Among these annotations you will find the KEGG pathway “Cell adhesion molecules”.
Are other genes from your list member of this pathway ?