



Integrative Bioinformatics and Systems Biology



Dr. Mohamed Hamed

LECTURE PLANNING

Lecture: 12 lectures, 3 hrs each.

Total workload: 42 hrs : 36 hrs of lectures and tutorials and 6 hrs of self studies.

Entrance requirements: basic knowledge of biology and computer science.

Literature: Lecture slides, tutorial handouts and problem sets will be provided.

INTEGRATIVE BIOINFORMATICS AND SYSTEMS BIOLOGY

COURSE ABBREVIATION: INT-BIO

LANGUAGE: ENGLISH

USED MEDIA: POWERPOINT PRESENTATION

Module: Lecture and tutorial.

By: Dr. Mohamed Hamed

Head of the Integrative OMICs Analysis Group in Rostock University medical center, Rostock University, Germany.

MOTIVATION AND COURSE OBJECTIVES

The main challenge of modern systems biology is unraveling the holistic picture of the complex molecular interactions that occur on different molecular levels (genomic, transcriptomic, epigenomic, proteomic, etc). Therefore, the needs to integrate/jointly analyze biological data from different high-throughput technologies emerged in order to identify biomarkers for early diagnosis and prognosis of complex diseases and facilitating the development of novel treatment approaches.

This course aims at teaching students how to perform data-specific computational analyses as well as integrative analysis approaches, combining knowledge from different OMICs-based datasets. Both, theoretical and practical aspects will be covered. Students will have the opportunity to work both independently during tutorials and in teams during the research project.

COMPETENCES TO BE DEVELOPED

Students will get practical and extensive hands-on experience on:

- R scripting language and bioconductor packages.
- Data-specific computational analysis and pipelines for the vast amounts of biological data produced using high-throughput technologies.
- Developing and applying integrative bioinformatics methods that could be utilized in all biology-related areas of interest.
- Basics of machine learning methods as tools for integrating biological features from heterogeneous Omics data.
- Students will be developing their own research projects, interpreting the obtained results, writing a manuscript, scientifically discussing the results.

ASSESSMENT

- Students need to finalize a research project applying all/ most of the learned methods and skills during the course. Novelty and extending the learned methods is highly encouraged and will be well graded.
- The outcomes of each research project should be compiled in a high scientific quality research article that is ready for submission in a peer-review journal.
- All projects will be presented, discussed and scientifically reviewed in the last lecture.

R language mini-course 1

-Introduction to the course

-Basics of R language and statistical methods.

-R studio IDE

R language mini-course 2

-Advanced R statistics

-Bio-conductor packages

R language mini-course 3

-Case study:

Microarray analysis using R

Introduction to integrative bioinformatics

-Importance of data integration

-Different methods for biological data integration

-OMICs data types, and TCGA repository

-Databases of diseases-related genes and miRNAs

Transcriptomic analysis

-From microarrays to RNA-seq

-RNA-seq analysis

-Linking to Ontologies and pathways

-Drug signature databases: CMAP and LINKS

Non-coding RNAs

-Small and long non-coding RNAs

-miRNA sequencing analysis

-miRNA databases

-lncRNAs analysis

Network- based integrative methods

-TFmiR analysis

-Network motif analysis

-Central hubs identifications

- Network visualization (Cytoscape)

Epigenetics

-Introduction to the epigenetic landscapes of normal and tumor cells

-DNA methylation, Co-methylation analysis

-DMRs identifications

Chip-seq experiments and/or GWAS

-Computational analysis of Chip-Seq data and/or

-Downstream analysis of genetic variants

Integrative analysis based on machine learning.

-Introduction to machine learning in bioinformatics.

-Unsupervised methods: Clustering biological data

-PCA analysis

Supervised machine learning methods

-Classification and regression analysis

-Model selection and evaluation of learning methods

-Outlook at deep neural networks applications in bioinformatics

PROJECTS DISCUSSION AND CLOSURE

-Projects presentation.

-Reviews of the potential manuscripts

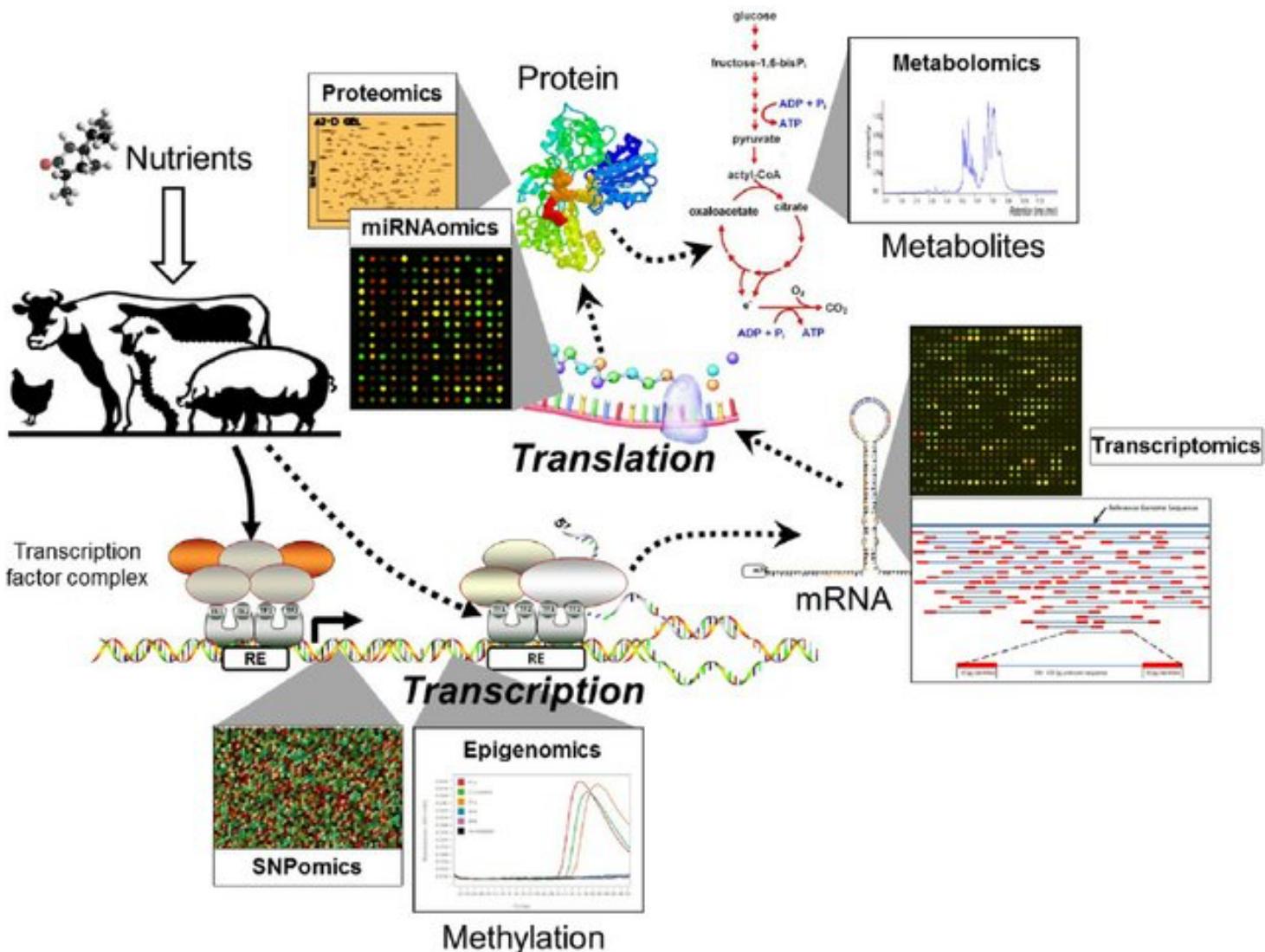
Goal of Data integration

The main goal of any **data integration** methodology is

- to extract additional **biological** knowledge from multiple datasets that cannot be gained from any single dataset alone. ...

Thus, these methods often suffer from various limitations when applied to multiple **data** types.

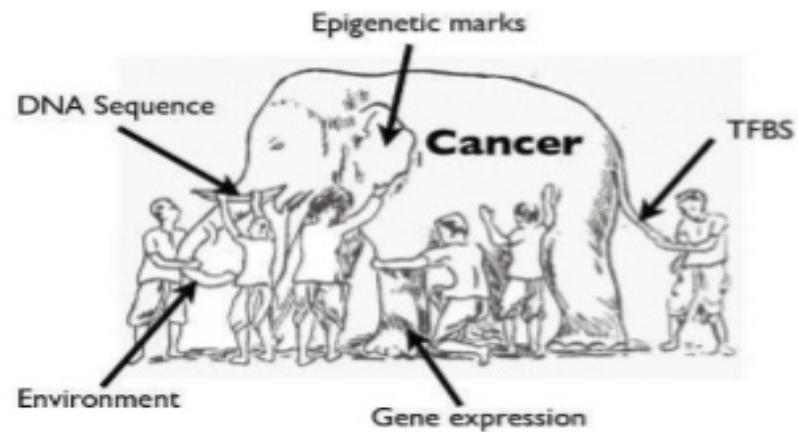
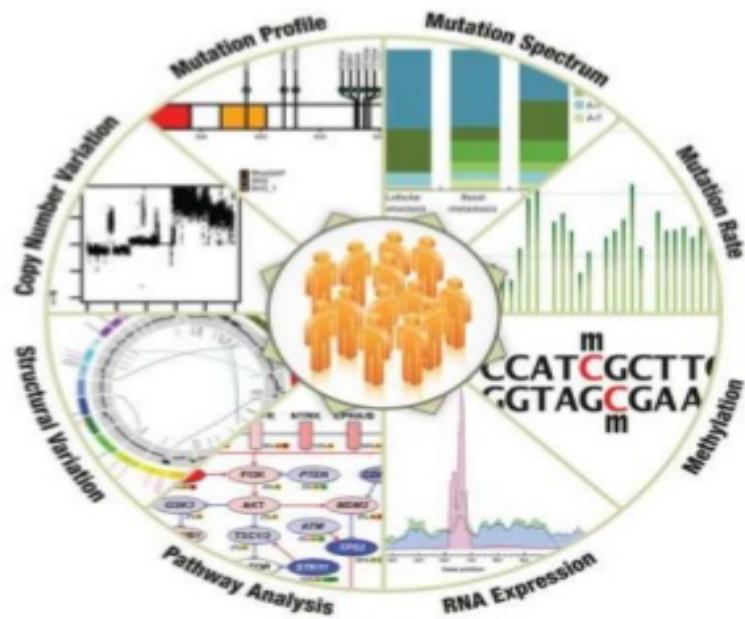
OMICs Data



Biological Data Integration

5

Why data integration is required in for cancer studies ?

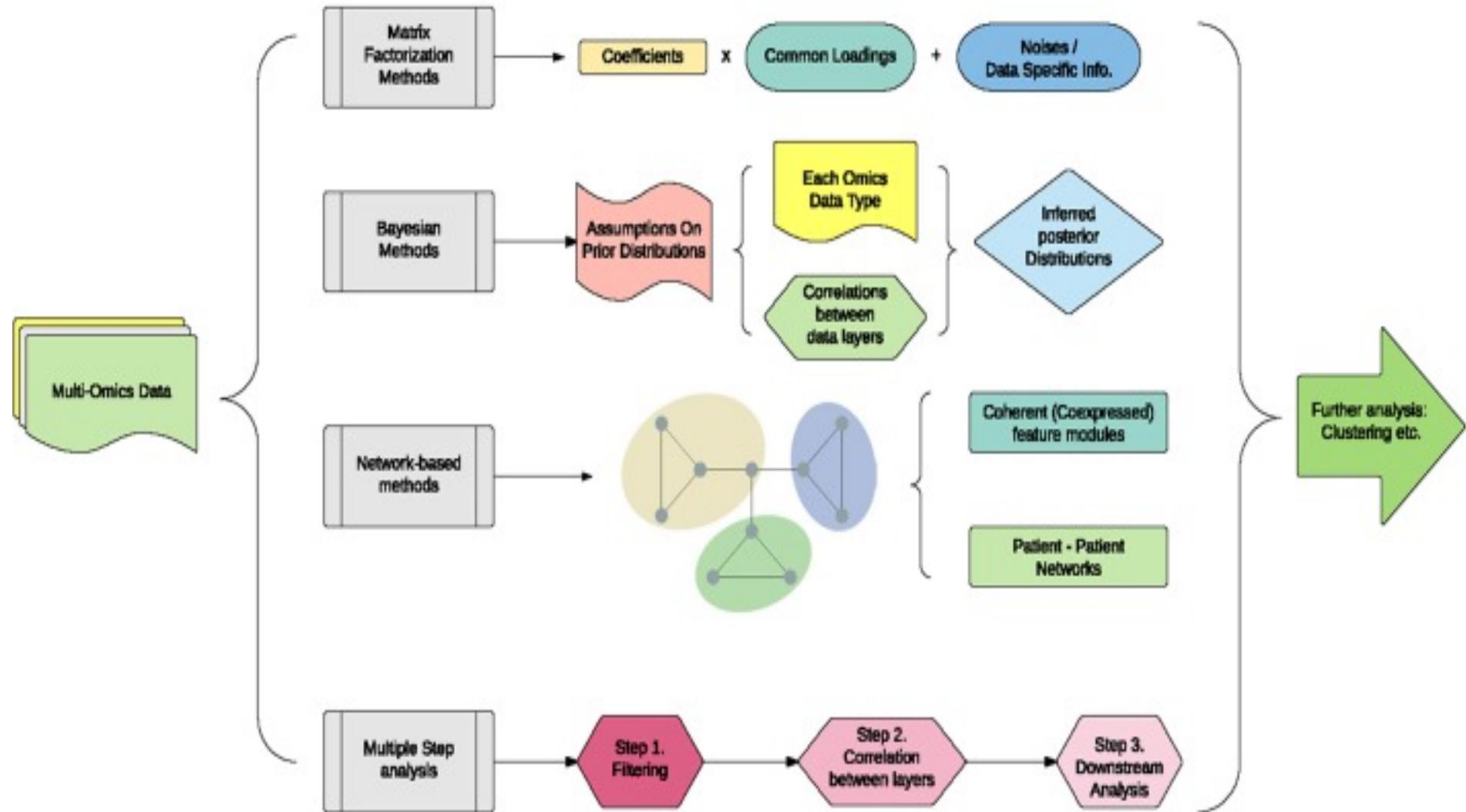


Studying cancer dataset in isolation
will produce an incomplete story

Ding et al. Hum. Mol. Genet. 2010

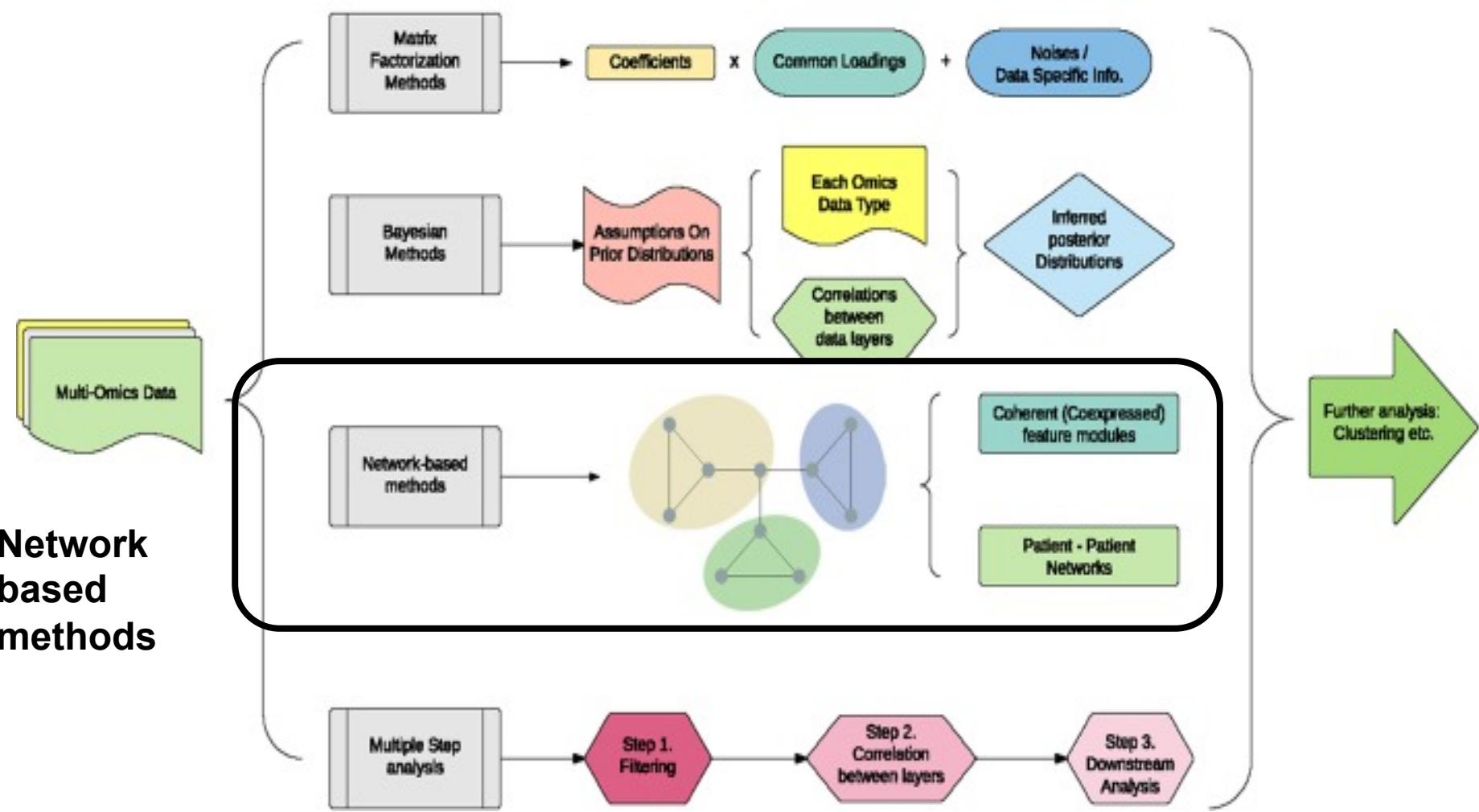
Biological Data Integration methods

6



Unsupervised methods for Biological Data Integration

7



Unsupervised methods for Biological Data Integration

8



Machine
learning
based
methods

Association Databases

Gene-Process association database

Gene Ontology Consortium

Home Documentation Downloads Tools About Contact us

Enrichment analysis

Your gene IDs here...

biological process

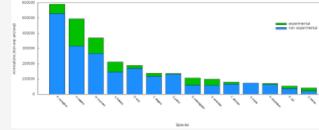
Homo sapiens

Submit

Help

Powered by PANTHER

Statistics



Other GOC tools

Explore other GOC tools in the AmiGO software suite.

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

Filter classes

Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It 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.

[more](#)

Annotations

Download annotations (standard files)

Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

Gene-Tissue association database

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs |

Download a sequence or region

Export data

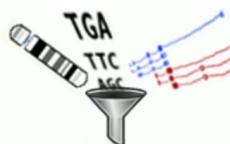
CAGAATGAT
AAATGTTCT
AAA**G**AAGCA
CTGTCATGC
ATAAAAGAA
AGTGATACT

Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

Customise your download



Custom datasets can be retrieved using the BioMart data-mining tool.

You may find exploring this web-based query tool easier than extracting information direct from our databases.

GTEX Portal

2018-04-04
New Multi-Gene Query
Read More >>

Current Release
Latest Version: V7
[Release Info](#) | [Download](#) | [Current Release Statistics](#) | [How to cite GTEX?](#)

The Genotype-Tissue Expression (GTEx) project is an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation. Samples were collected from 53 non-diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq. Remaining samples are available from the GTEx Biobank. The GTEx Portal provides open access to data including gene expression, QTLs, and histology images.

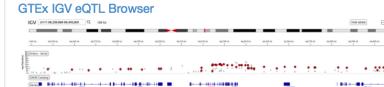
The current release is V7 including 11,688 samples, 53 tissues and 714 donors.

Genetic Association

Single-Tissue eQTLs

Search eQTL by gene or SNP ID

GTEX IGV eQTL Browser



GTEX Gene-eQTL Visualizer



View eQTL data of a gene...

GTEX eQTL Calculator

Gene-Disease association database



The DisGeNET logo features the word "DisGeNET" in a large, stylized font where the letters "D", "i", "s", "G", "e", and "N" are in pink and "E" and "T" are in blue. To the right of the text is a graphic representation of a network or graph, consisting of several blue and pink circular nodes connected by thin grey lines.

Home About Search Browser Downloads Cytoscape RDF Help

DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated to *et al.*, 2016; Piñero *et al.*, 2015). DisGeNET integrates data from expert curated repositories, GWAS catalogues, animal models and DisGeNET data are homogeneously annotated with controlled vocabularies and community-driven ontologies. Additionally, several tools are provided to assist the prioritization of genotype–phenotype relationships.

The current version of DisGeNET (v5.0) contains 561,119 gene-disease associations (GDAs), between 17,074 genes and 20,362 traits, and clinical or abnormal human phenotypes, and 135,588 variant-disease associations (VDAs), between 83,002 SNPs and 20,362 phenotypes.

The information in DisGeNET can be accessed in several ways:

- The web interface, through the [Search](#) and [Browse](#) functionalities
- The Resource Description Framework ([DisGeNET-RDF](#)) representation via the [SPARQL endpoint](#), and the [Faceted Browser](#)
- The [DisGeNET Cytoscape App](#)
- Scripts in the most commonly used programming languages
- The [disgenet2r](#) package.
- The [SQLite database](#)
- Tab separated files. See [downloads](#) section

miRNA-Disease association database

HMDD v3.0: the Human microRNA Disease Database version3.0

Home

Browse

Search

miR-Target Network

Download

Submit

Help

Welcome to the HMDD v3.0

HMDD (the Human microRNA Disease Database) is a database that curated experiment-supported evidence for human microRNA (miRNA) and disease associations. miRNAs are one class of important regulatory RNAs, which mainly repress gene express at the post-transcriptional level. Increasing reports have shown that miRNAs play important roles in various critical biological processes. For their importance, the dysfunctions of miRNAs are associated with a broad spectrum of diseases. The first version of HMDD was built on December 2007 and represents the first miRNA disease database in the world according to literature record. Each entry in HMDD v1.0 has four items for annotation; they are miRNA name, disease name, the reference PubMed ID, and the evidence supporting the miRNA-association from the original paper. During the past five years, we updated HMDD for more than 30 times. HMDD v2.0 presents more detailed and comprehensive annotations to the human miRNA-disease association data, including miRNA-disease association data from the evidence of genetics, epigenetics, circulating miRNAs, and miRNA-target interactions. In addition, a "submission" function was implemented in the version 3. HMDD v3.0

Statistics:

Currently, in HMDD v3.0, we manually collected 32281 miRNA-disease association entries which include 1102 miRNA genes, 850 diseases from 17412 papers.

History:

June 28, 2018, HMDD v3.0 was released.

June 20, 2013, HMDD v2.0 was released.

January, 2012, HMDD has been updated for 27 times during the past four years.

January, 2011, HMDD has been updated for 19 times during the past three years.

October, 2008, an analysis and modeling paper based on the miRNA-disease association data in the HMDD database was published on PLoS ONE.

December, 2007, the original HMDD database was released.

Contact us

HMDD is funded by the German Research Foundation (DFG).

For mouse mapping: (MGI)

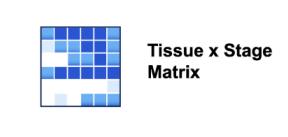
 [About](#) [Help](#) [FAQ](#)

[Search](#) ▾ [Download](#) ▾ [More Resources](#) ▾ [Submit Data](#) [Home](#) [Genes](#) [Phenotypes](#) [Human Disease](#) [Expression](#) [Recombinases](#) [Function](#) [Strains / SNPs](#) [Homology](#)

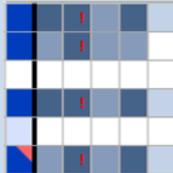
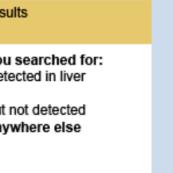
[Find Mice \(IMSR\)](#)  [Contact Us](#) [Browsers](#)


GXD collects and integrates the gene expression information in MGI. Its primary emphasis is on endogenous gene expression during mouse development.

GXD is funded by grant HD062499
 Eunice Kennedy Shriver National Institute of Child Health and Human Development

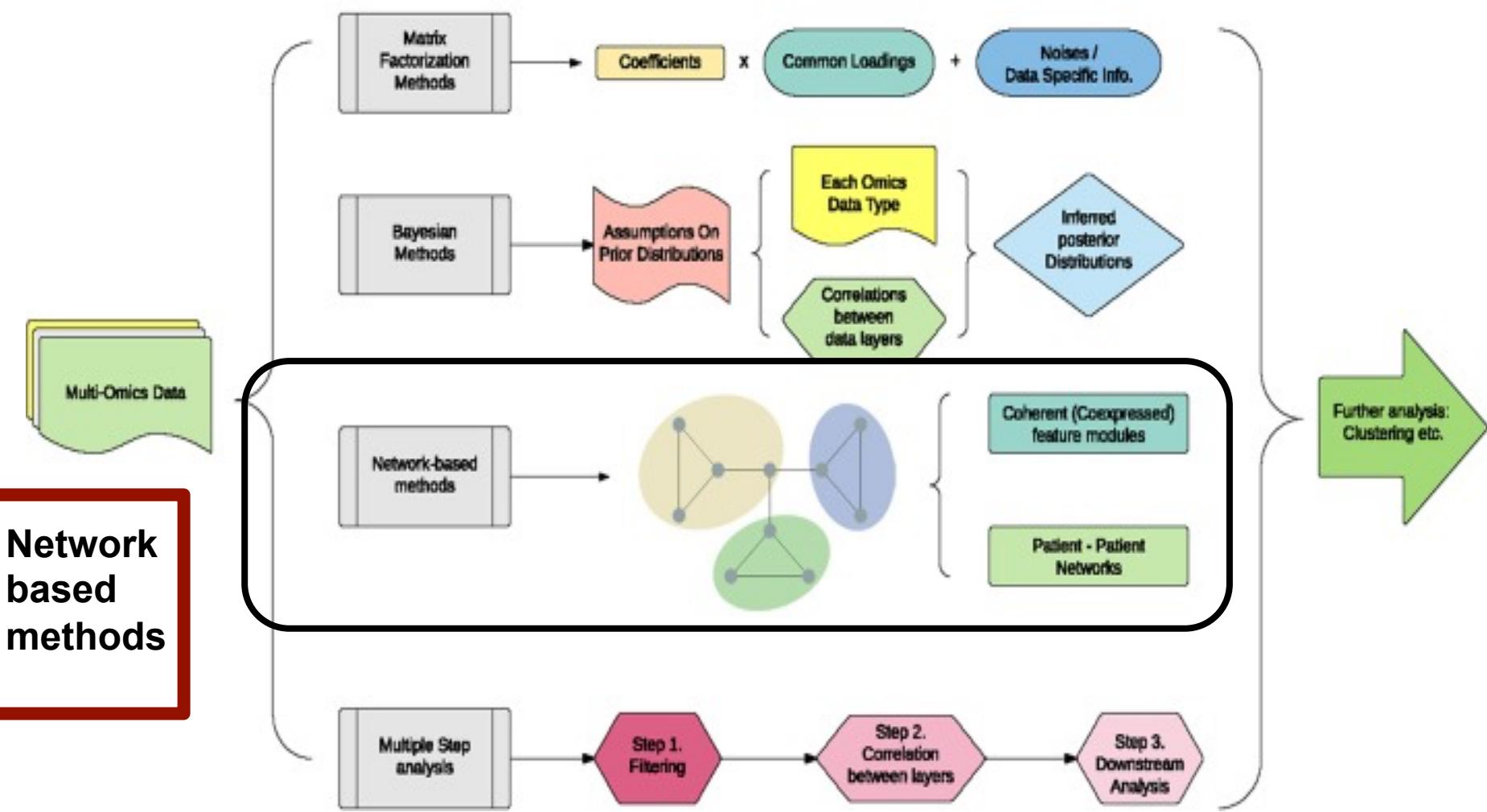
 Expression Data and Image Search  Differential Expression Search 
 Developmental Anatomy Browser  Tissue x Stage Matrix 
 GXD Statistics  MouseMine API Data Access 

HIGHLIGHTS

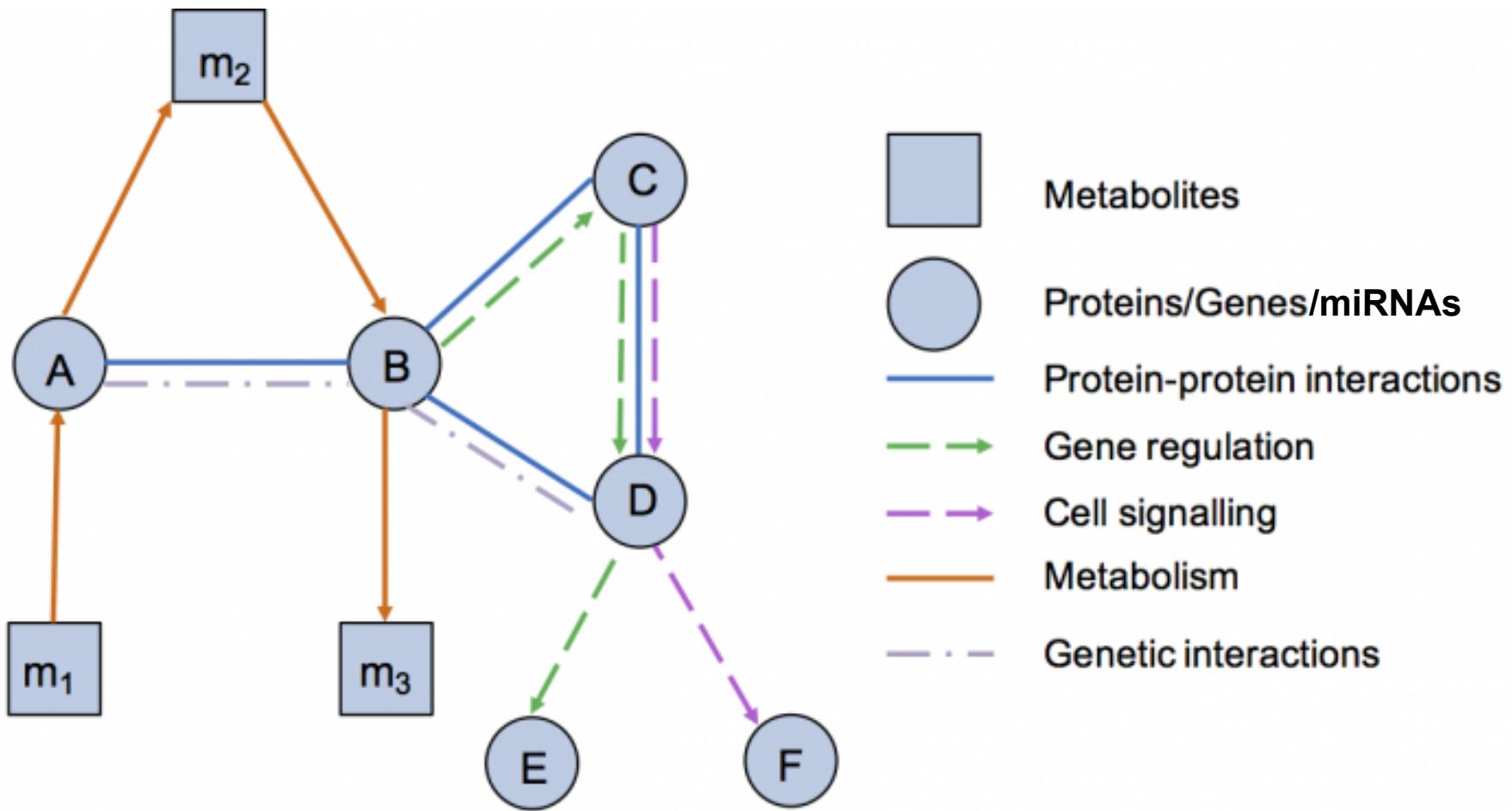
 Gene Expression & Phenotype Matrix  Expression Images *Nkx6-1*  Recently added: *Chal J et al.*  Results
You searched for:
Detected in liver
but not detected
anywhere else  Updated Differential Expression Search  Expression + Recombinase Activity

 About GXD  Fast Track Your Data  Contact Us

Unsupervised methods for Biological Data Integration



Types of biological networks/interactions

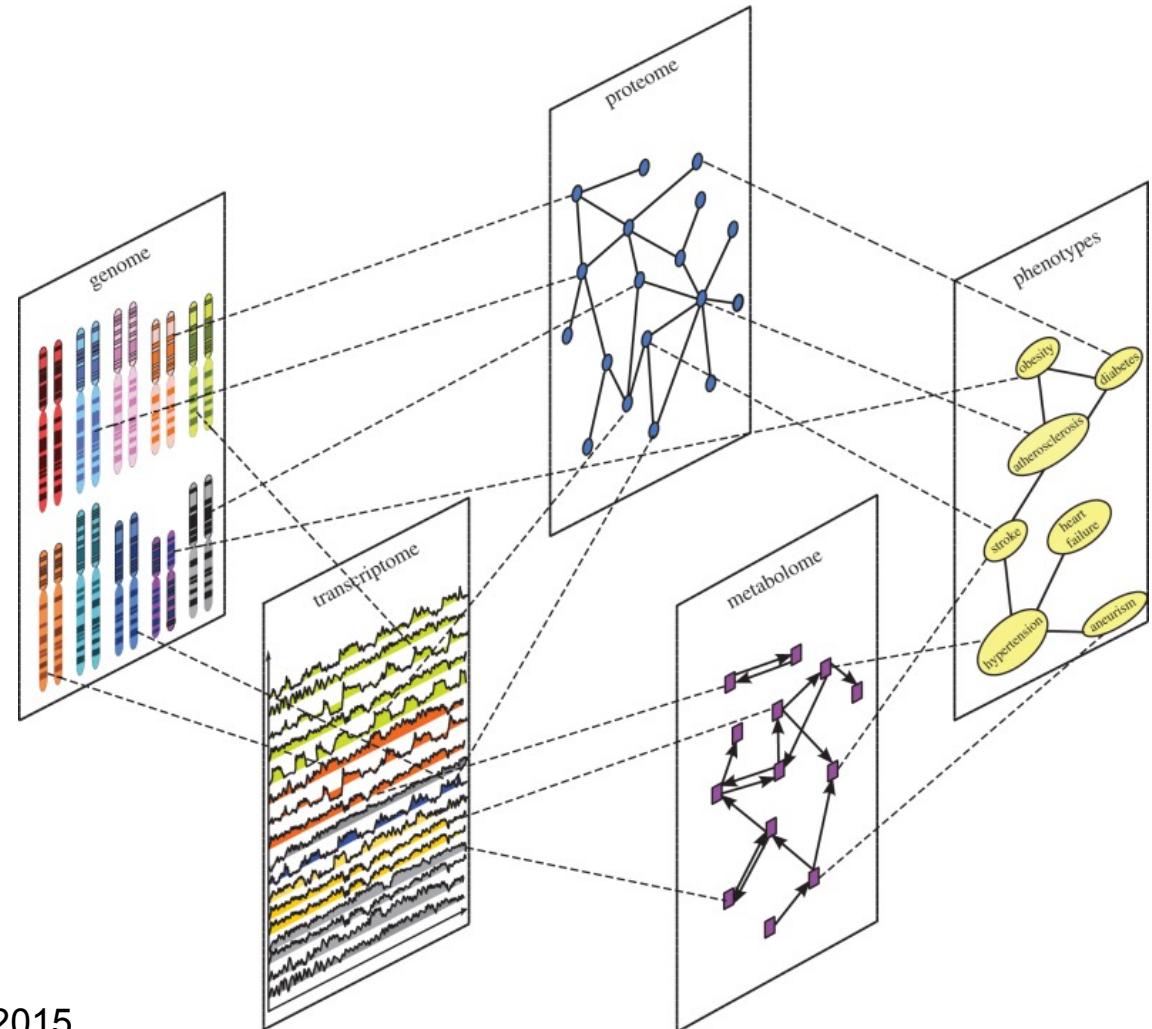


Integration of multi-Omics data

1

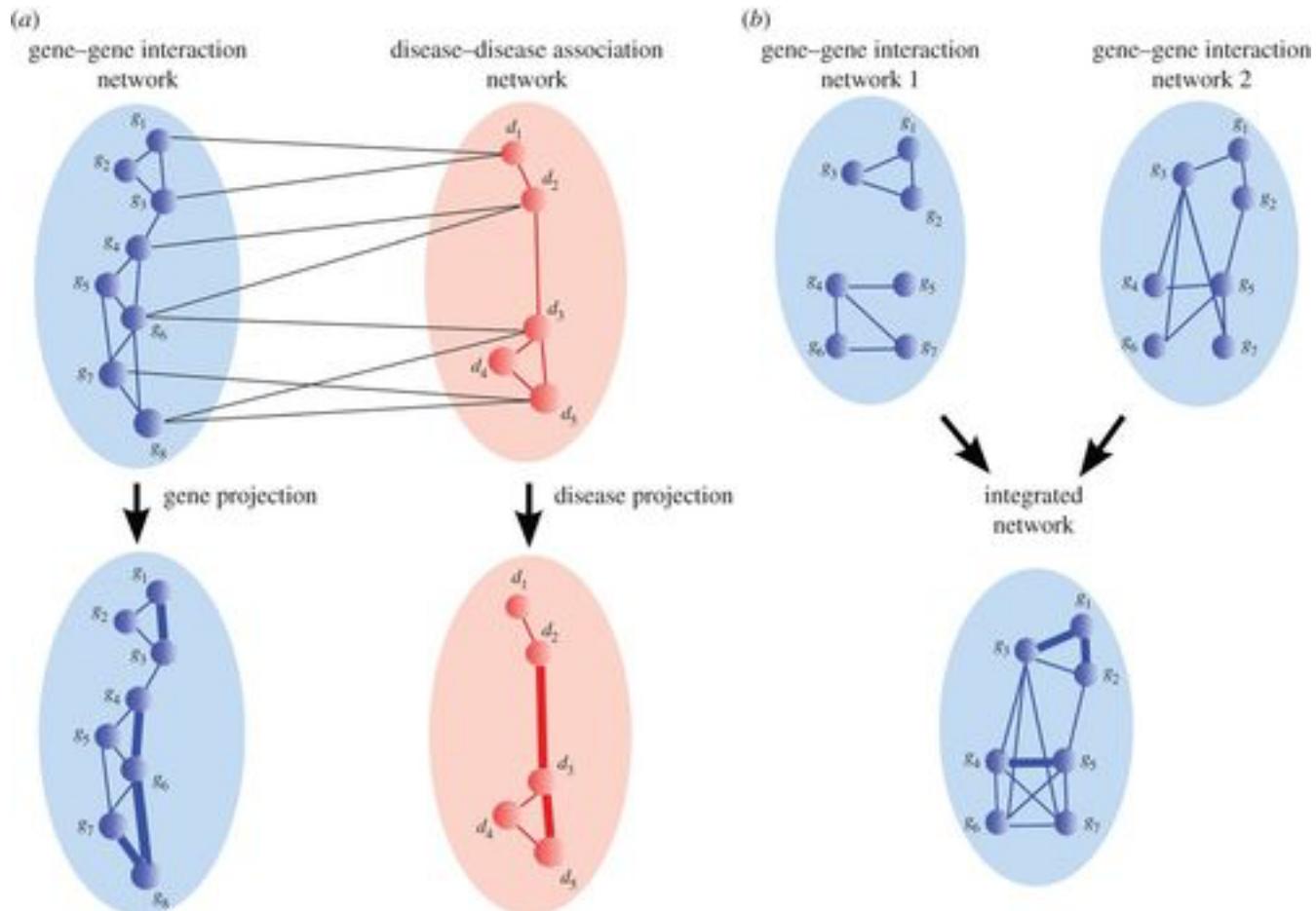
They are not multiple networks...

but one high dimensional network



Gligorijević et al 2015

Predictions based on network integration



Gligorijević et al 2015

1- The TFmiR Web Server

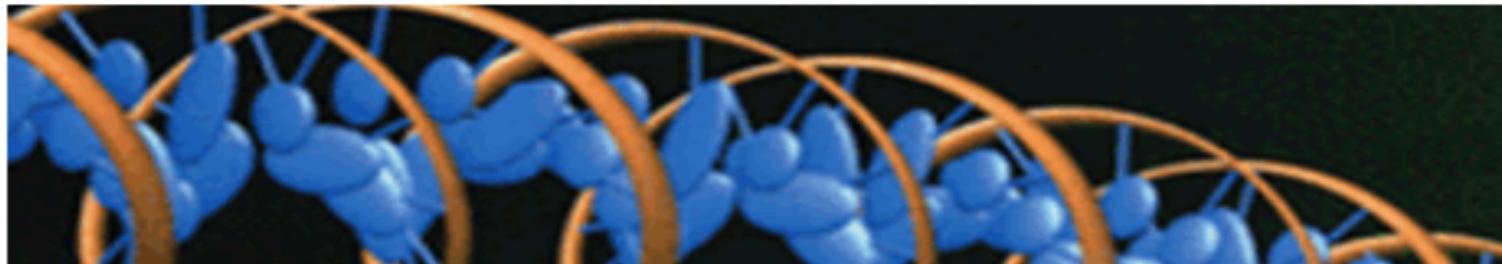
- TFs and miRNAs are essential key players for regulating gene expression.
- Integrative analysis of 4 combinatorial regulatory interactions between TFs, miRNAs, and target genes that are involved in disease pathogenesis.
- Enriched with downstream analysis toolkit for discovering various functional and topological features in the disease network.
- Allows two input user scenarios.

Nucleic Acids Research

[ABOUT THIS JOURNAL](#)[CONTACT THIS JOURNAL](#)[SUBSCRIPTIONS](#)[CUR](#)

Institution: Saarlaendische Universitaets-uLandesbibliothek/ Med. Abteilung [Sign In](#)

[Oxford Journals](#) > [Science & Mathematics](#) > [Nucleic Acids Research](#) > [Volume 43, Issue W1](#) >



TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks



Mohamed Hamed, Christian Spaniol, Maryam Nazarieh and Volkhard Helms*

+ Author Affiliations

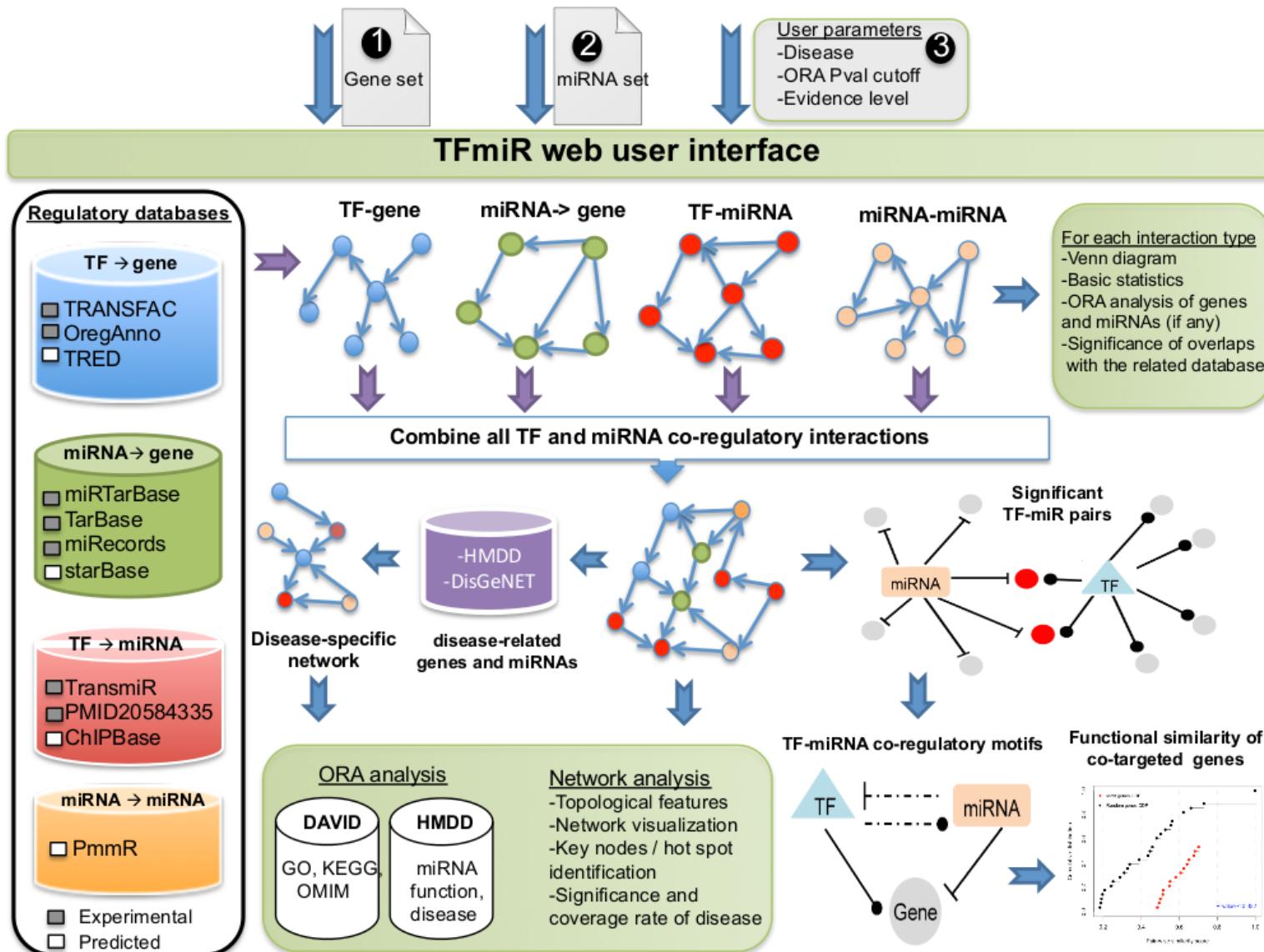
 *To whom correspondence should be addressed. Tel: +49 681 302 70701; Fax: +49 681 302 70702; Email: volkhard.helms@bioinformatik.uni-saarland.de

Received February 27, 2015.

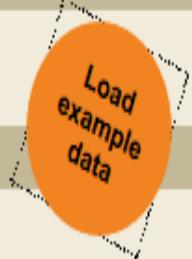
Revision received April 16, 2015.

Accepted April 18, 2015.

TFmiR System Architecture



Step 1: Input selection

miRNA miRNA.bc.txt No file chosenmRNA mRNA.bc.txt No file chosen Load
example
data

Step 2: Configuration

p-Value threshold

0.05

Related disease

Breast Neoplasms



Evidence

Both



Step 3: Go!

 Start processing

Log

```
[14:43:57] mRNA received, ready to start processing
[14:43:57] The file mRNA.bc.txt has been uploaded
[14:43:57] The file miRNA.bc.txt has been uploaded
[14:43:57] Example files for breast cancer have been loaded into your session. Check your p-Value and experimental evidence and click the processing button to start analysis.
```

Step 1: Input selection

miRNA	miRNA.bc.txt	<input type="button" value="Choose File"/> No file chosen
mRNA	mRNA.bc.txt	<input type="button" value="Choose File"/> No file chosen

Load example data

Step 2: Configuration

p-Value threshold

Related disease

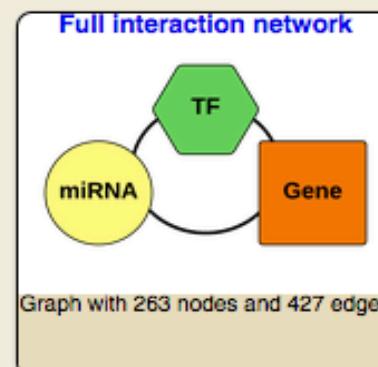
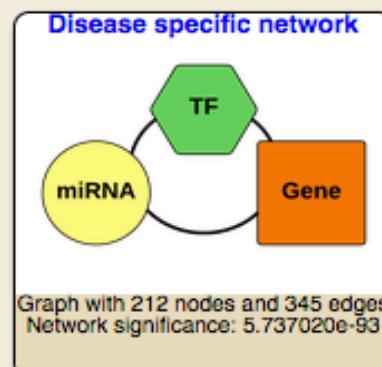
Evidence

Step 3: Go!

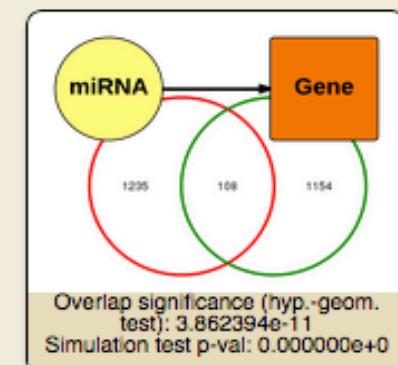
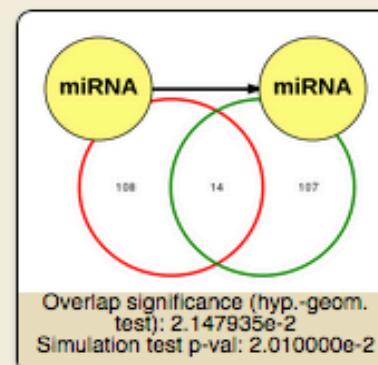
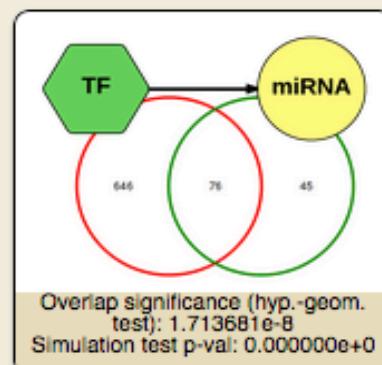
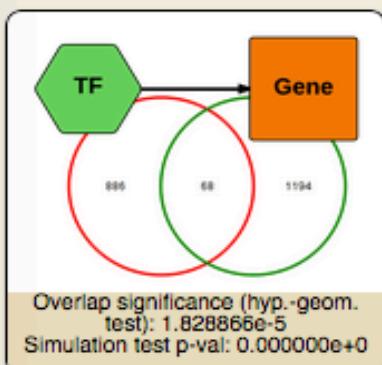
Processing Complete
Click to restart

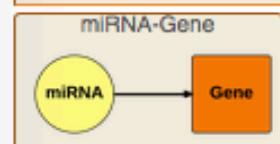
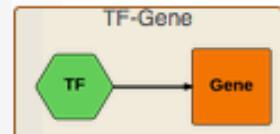
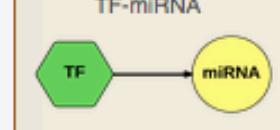
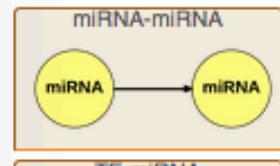
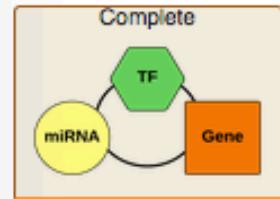
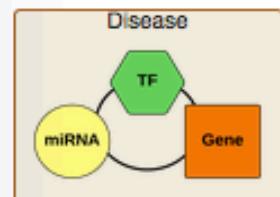
Step 4: Review result sets

(a) Created networks



(b) Interaction types

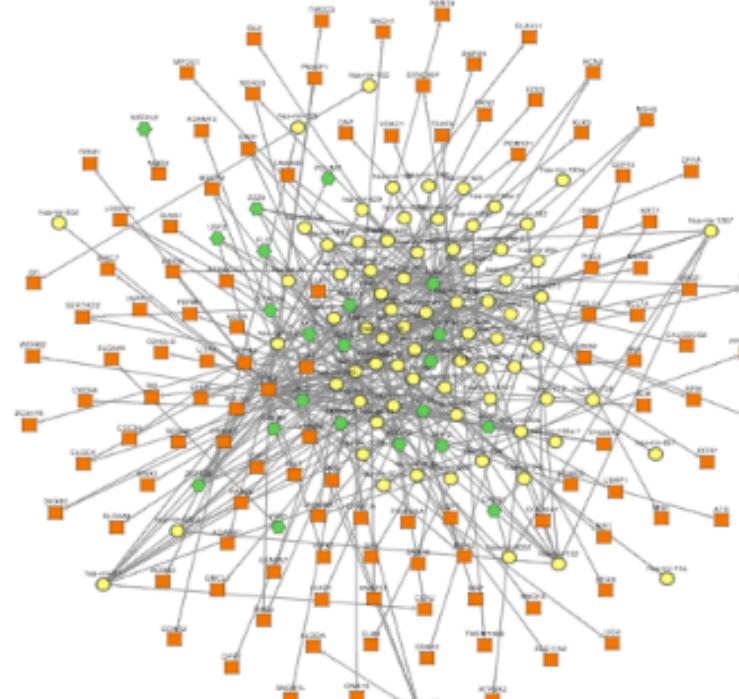



[Network](#) [Interactions](#) [ORA](#) [Motifs](#)
[Motifs](#) [Search for motifs](#)

Attention: this may take a while! low co-regulated

[Layout](#) [Reset](#)

cose

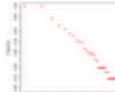
[Extend neighbours](#)
[Deselect all](#)
[Export image](#)
[Hotspot nodes:](#) [Dominating set](#)
[Degree](#)
[Closeness](#)
[Betweenness](#)
[Eigenvector](#)
[Common](#)
[Union](#)

Selected Nodes:

Hypergeometric Test p-Value for disease 5.73701966458094e-93

Edge coverage for disease 100

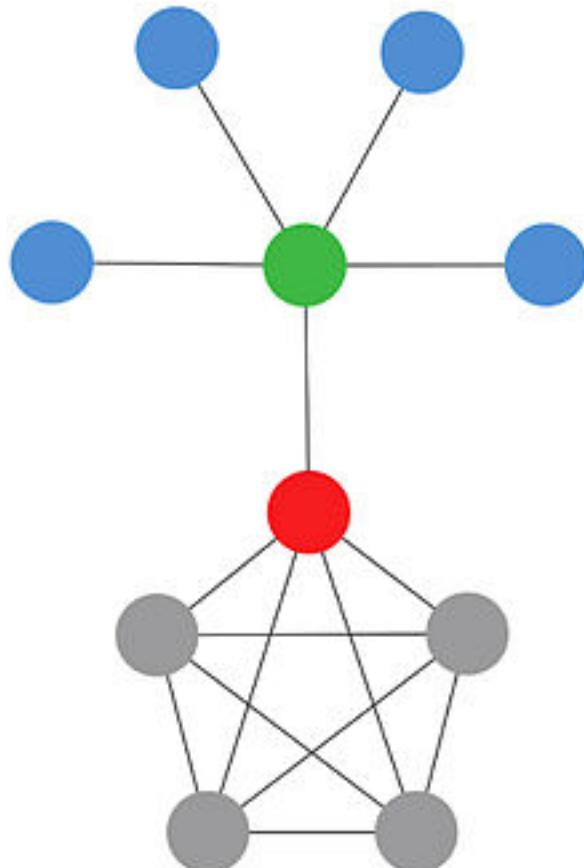
Node coverage for disease 100

Degree distribution



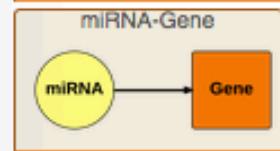
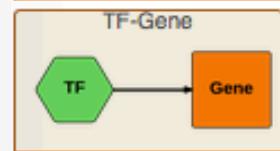
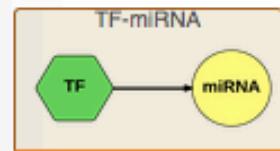
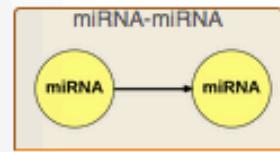
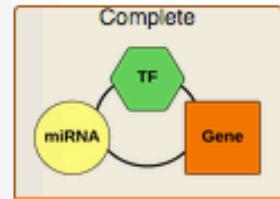
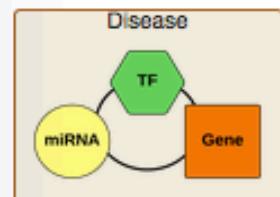
Identifying Hotspot genes/ potential driver genes

Degree centrality



CENTRALITY INTERPRETATIONS

Centrality Measures	Interpretation 1	Interpretation 2
Degree	<i>how many people can this person reach directly ?</i>	<i>in network of investors and start up: how many start up has this company invested to ?</i>
Betweenness	<i>how likely is this person to be the most direct route between two people in the network ?</i>	<i>in network of knowledge: who is the employee though whom most of the confidential information is likely to flow ?</i>
Closeness	<i>how fast can this person reach everyone in the network ?</i>	<i>in network of information dissemination: how fast this word-of-mouth issue spread from this person to the rest of the network ?</i>
Eigenvector	<i>how well is this person connected to other well-connected people ?</i>	<i>in network of paper citations: who is the author that is most cited by other well-cited authors ?</i>



Available networks

Network Interactions ORA Motifs

Motifs Search for motifs

Attention: this may take a while! low co-regulated

Layout Reset

cose

Extend neighbours

Deselect all

Export image

Hotspot nodes: Dominating set

Degree

Closeness

Betweenness

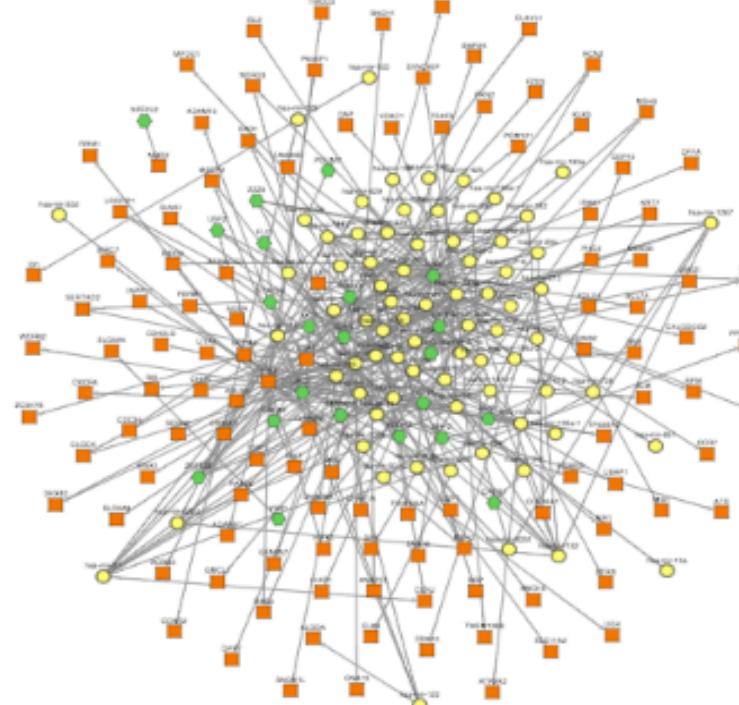
Eigenvector

Common

Union

Motif detection

Hotspot nodes



Interactive network
visualization

Selected Nodes:

Hypergeometric Test p-Value for disease 5.73701966458094e-93

Edge coverage for disease 100

Node coverage for disease 100

Degree distribution

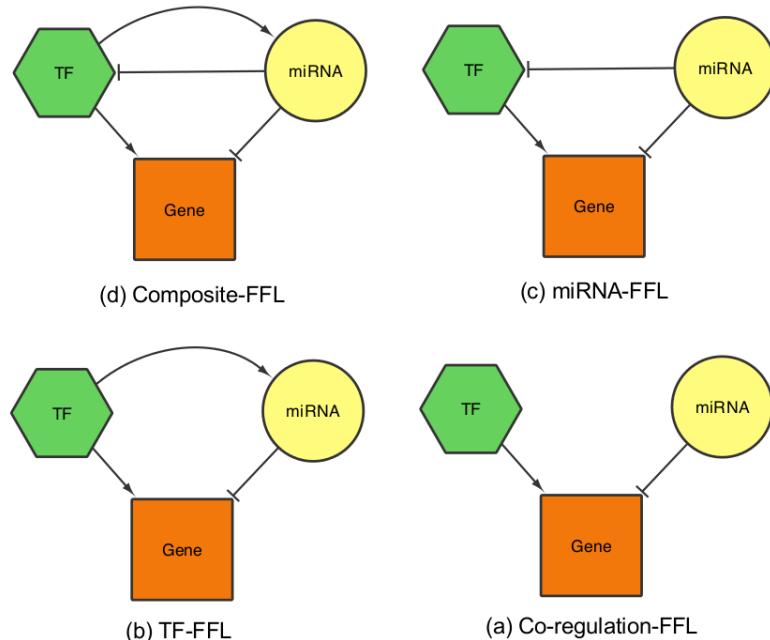


Basic statistics

TF-miRNA motif detection

Identifying significant TF-miRNA co-occurring pairs

Where $P\text{-value} = 1 - \sum_{i=0}^x \frac{\binom{k}{i} \binom{M-k}{N-i}}{\binom{M}{N}}$



- k is the number of target genes of a certain miRNA.
- N is the number of genes regulated by a certain TF.
- x is the number of common target genes between these TF and miRNAs.
- M is the number of genes in the union of all human genes targeted by human miRNAs and all human genes regulated by all human TFs in our databases.

Significance of the FFL motifs

We compare how often motifs appear in the real network to the number of times they appear in randomized ensembles preserving the same node degrees. The random networks were constructed 100 times and compared to the real network.

$$P\text{-value} = \frac{N_h}{N_r}$$

N_h : number of random times that a certain motif type is acquired more than or equal to its number in the real network, and $N_r = 100$.

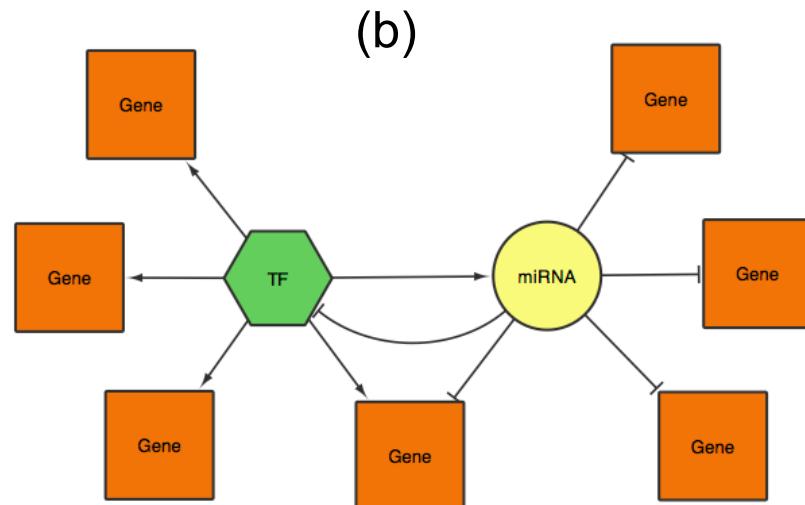
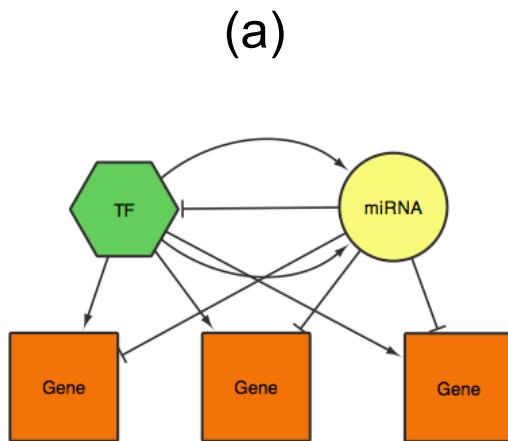
Z-score for each motif type to examine by how many standard deviations the observed real motif was above or below the mean of the random ones.

$$Z\text{score} = \frac{N_o - N_m}{\sigma}$$

Assessing the discovered motifs

- Functional homogeneity

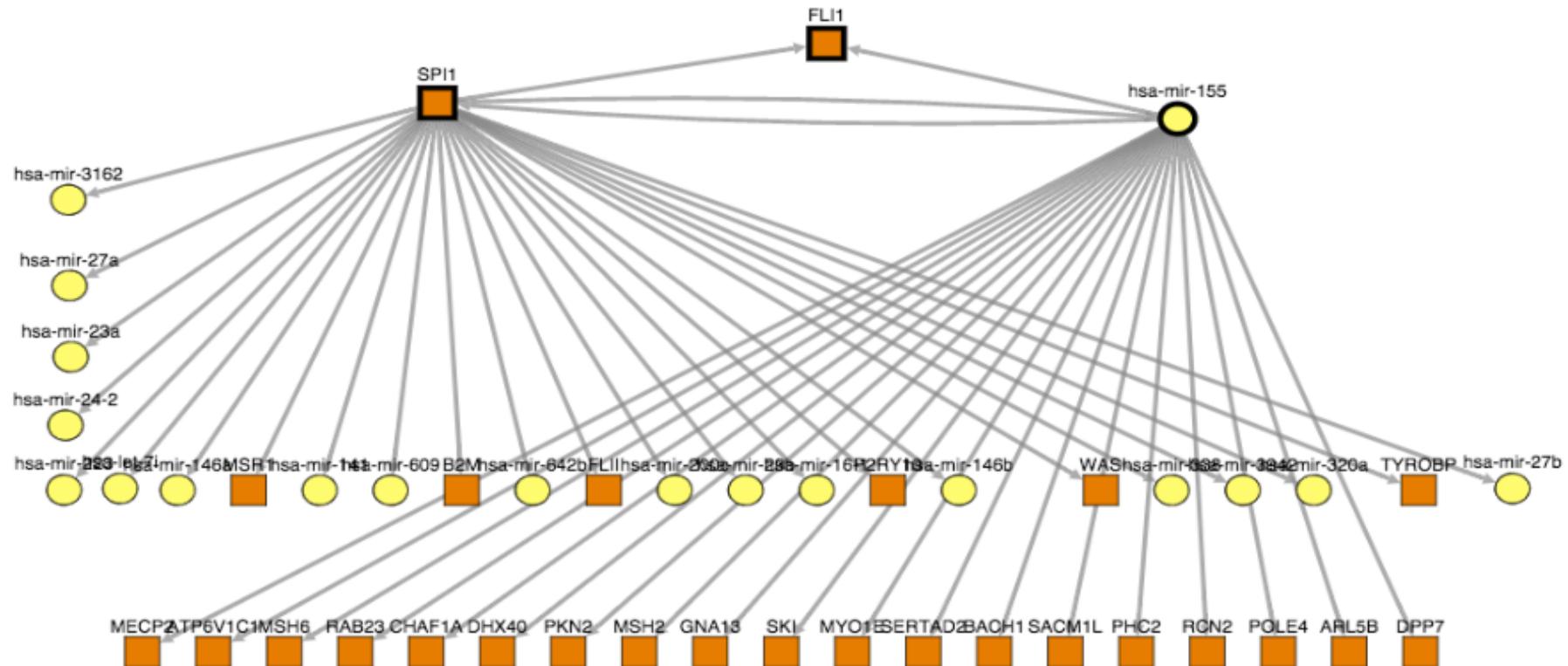
-TFmiR allows the user to analyze the GO semantic similarity for all pairs of genes targeted by the same TF and miRNA pair or for all pairs of genes regulated by the TF and the miRNAs of that TF-miRNA pair.



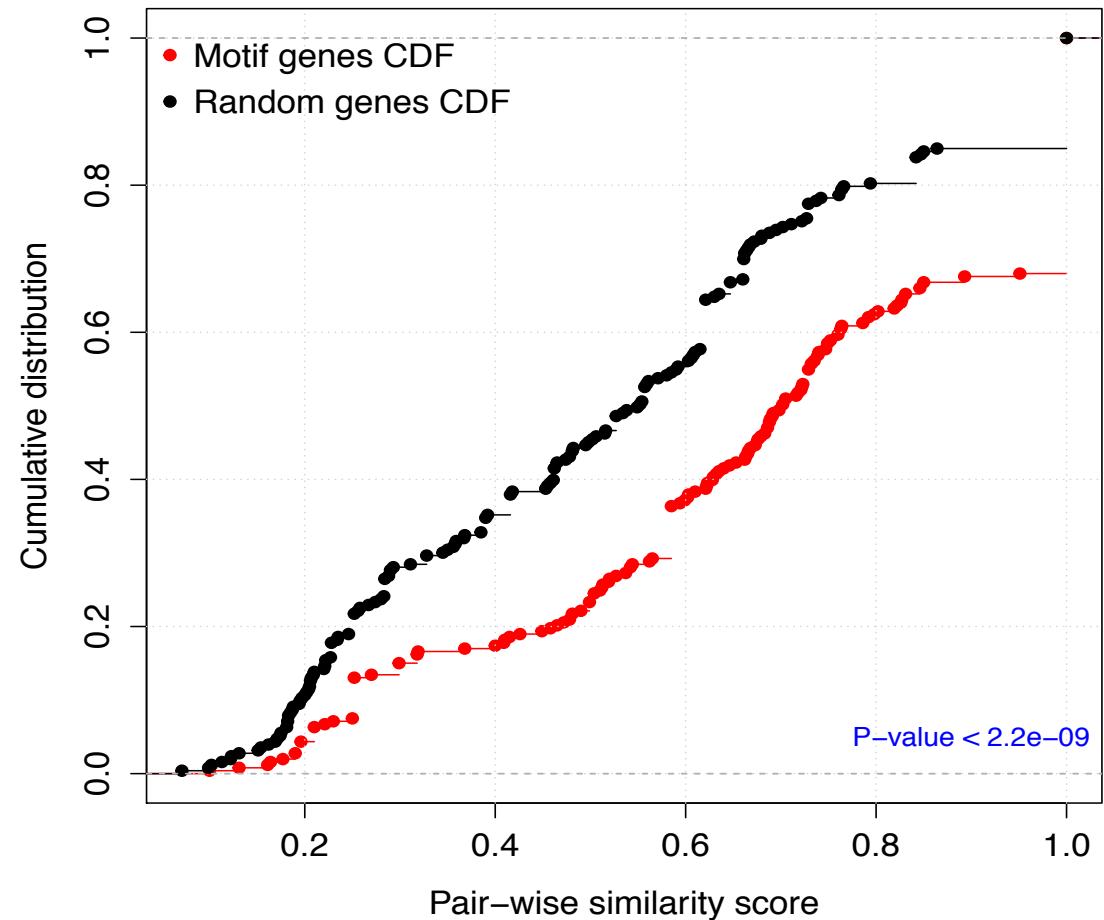
A promising FFL motif on BRCA

A detected new composite FFL motif involves the TF *SPI1*, the miRNA *hsa-mir-155*, and the target gene *FLI1*. Cooperative functional module in breast cancerogenesis.

Co-regulated subnetwork for TF: SPI1, miRNA: hsa-mir-155, Gene: FLI1



Functional Homogeneity



Cumulative distributions of GO functional semantic scores of gene pairs of co-regulated genes in the examined motif (red) versus randomly selected genes (black).

Same principle of associations

- Disease specific network
- Why not ?
- Tissue specific network
- Process specific network
- Tissue and process
- Tissue and disease
- Process and disease

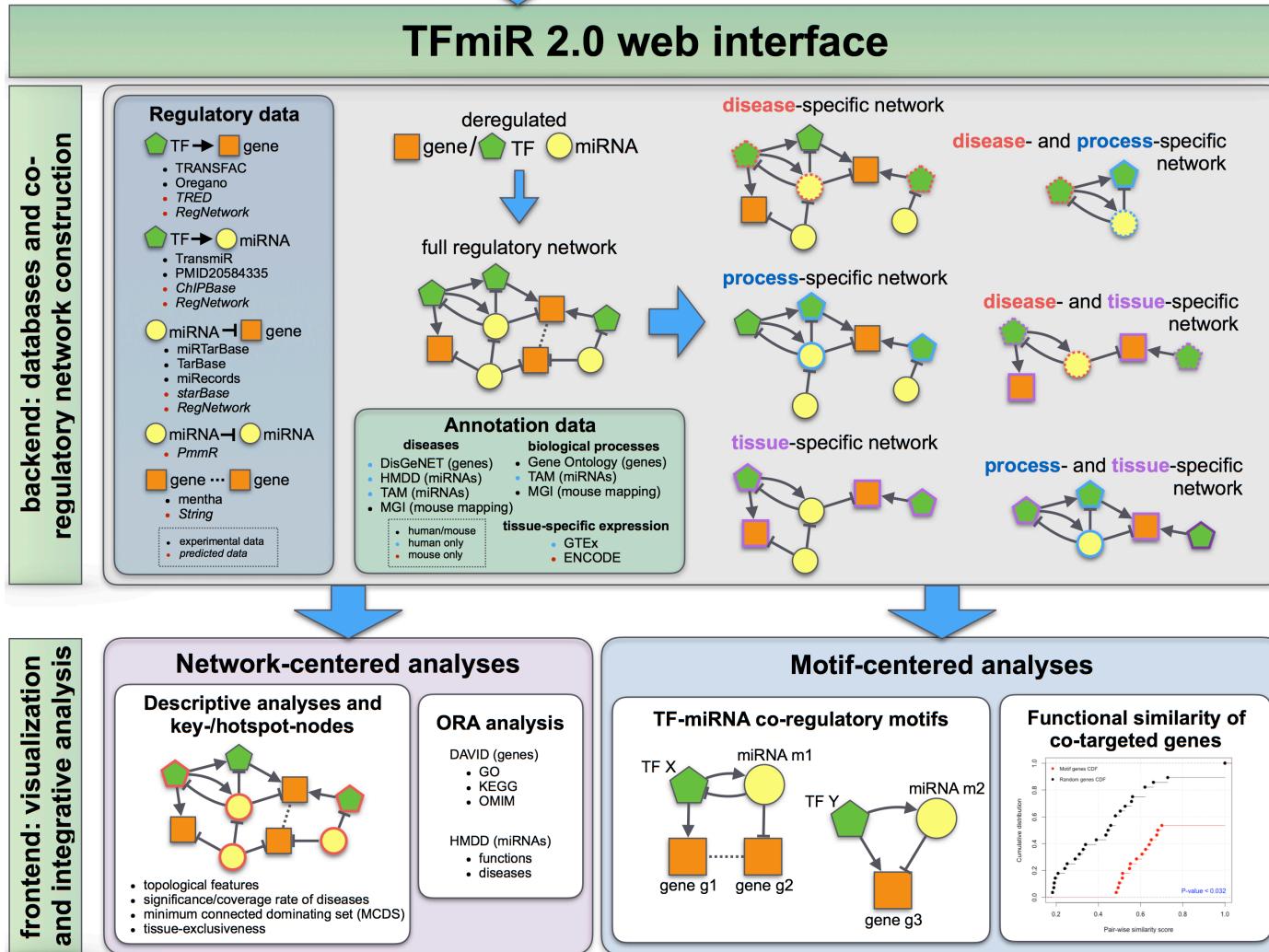
TFmiR v2

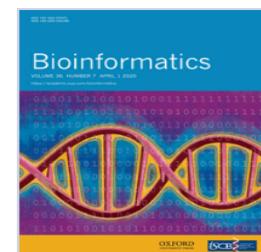
User

data on human or mouse: deregulated



interactive workflow





Volume 36, Issue 7
1 April 2020

Article Contents

- Abstract
 - 1 Introduction
 - 2 Functionality of TFmiR2
 - 3 Case study
 - 4 Summary
 - Funding
 - References
- < Previous Next >

TFmiR2: constructing and analyzing disease-, tissue- and process-specific transcription factor and microRNA co-regulatory networks

Maryam Nazarieh, Mohamed Hamed, Christian Spaniol, Thorsten Will, Volkhard Helms [✉](#)

Bioinformatics, Volume 36, Issue 7, 1 April 2020, Pages 2300–2302,

<https://doi.org/10.1093/bioinformatics/btz871>

Published: 20 November 2019 Article history ▾

Views ▾

Cite

Permissions

Share ▾

Abstract

Summary

TFmiR2 is a freely available web server for constructing and analyzing integrated transcription factor (TF) and microRNA (miRNA) co-regulatory networks for human and mouse. TFmiR2 generates tissue- and biological process-specific networks for the set of deregulated genes and miRNAs provided by the user. Furthermore, the service can now identify key driver genes and miRNAs in the constructed networks by utilizing the graph theoretical concept of a minimum connected dominating set. These putative key players as well as the newly implemented four-node TF-miRNA motifs yield novel insights that may assist in developing new therapeutic approaches.

Availability and implementation

The TFmiR2 web server is available at <http://service.bioinformatik.uni-saarland.de/tfmir2>.

TFmiR v2

TFmiR2

Constructing and analyzing disease-, tissue- and process-specific transcription factor and miRNA co-regulatory networks v2.0

[Help Page](#)



Step 1: Input selection

miRNA	mirNA.bc.txt	Choose File	no file selected
mRNA	mRNA.bc.txt	Choose File	no file selected

Load example data

Step 2: Configuration

miRNA-target enrichment p-value threshold	0.05	ORA p-value threshold	0.05
Species	Human	Related process	Akt pathway Angiogenesis Apoptosis Bone regeneration Cardiogenesis Cell cycle related Cell differentiation
Related disease	Breast Neoplasms	Evidence	Experimental
Related tissue	Breast - Mammary Tissue	Randomization Method	conserved
Protein-Protein/Gene-Gene interaction cutoff	0.8		

Step 3: Go!

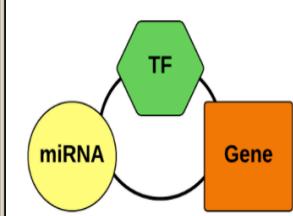
Please stand by...

TFmiR v2

Step 4: Review result sets

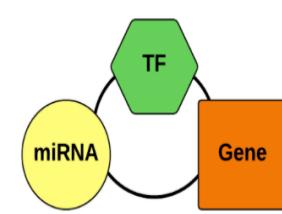
(a) Created networks

Disease-specific network



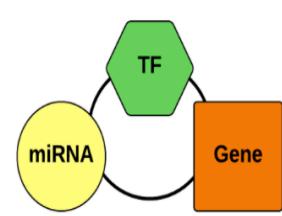
Graph with 152 nodes and 183 edges
Network significance: 0.000000e+0

Disease & Process-specific network



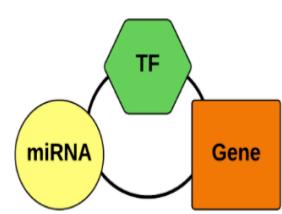
Graph with 31 nodes and 33 edges
Network significance: 0.000000e+0

Process-specific network



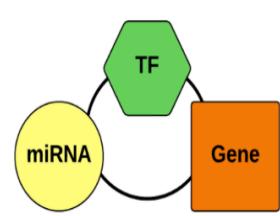
Graph with 113 nodes and 133 edges
Network significance: 0.000000e+0

Tissue & Process-specific network



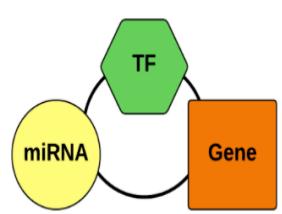
Graph with 25 nodes and 25 edges
Network significance: 0.000000e+0

Tissue-specific network



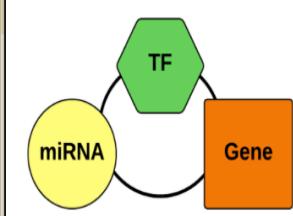
Graph with 208 nodes and 242 edges
Network significance: 0.000000e+0

Disease & Tissue-specific network



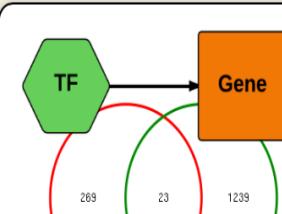
Graph with 27 nodes and 28 edges
Network significance: 0.000000e+0

Full interaction network

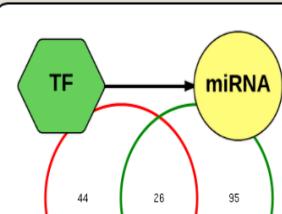


Graph with 217 nodes and 253 edges

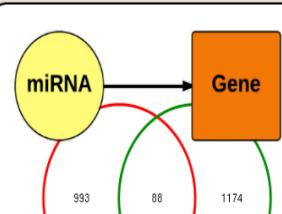
(b) Interaction types



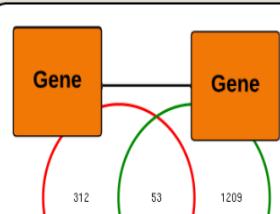
Overlap significance (hyp.-geom. test): 1.000000e+0
Simulation test p-val: 1.000000e+0



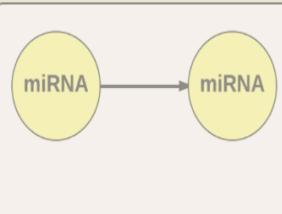
Overlap significance (hyp.-geom. test): 8.185871e-15
Simulation test p-val: 0.000000e+0



Overlap significance (hyp.-geom. test): 1.000000e+0
Simulation test p-val: 1.000000e+0



Overlap significance (hyp.-geom. test): 1.000000e+0
Simulation test p-val: 1.000000e+0



Network based integration methods

2- Similarity Network Fusion

nature > nature methods > articles > article

a naturerresearch journal

MENU ▾

nature|methods

Article | Published: 26 January 2014

Similarity network fusion for aggregating data types on a genomic scale

Bo Wang, Aziz M Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains & Anna Goldenberg 

Nature Methods 11, 333–337 (2014) | Download Citation 

Abstract

Recent technologies have made it cost-effective to collect diverse types of genome-wide data. Computational methods are needed to combine these data to create a comprehensive view of a given disease or a biological process. Similarity network fusion (SNF) solves this problem by constructing networks of samples (e.g., patients) for each



Search



E-alert



Submit



My Account

211
Citations

25
Altmetric

[Article metrics >](#)

Sections

Figures

References

Abstract

References

Acknowledgements

Author information

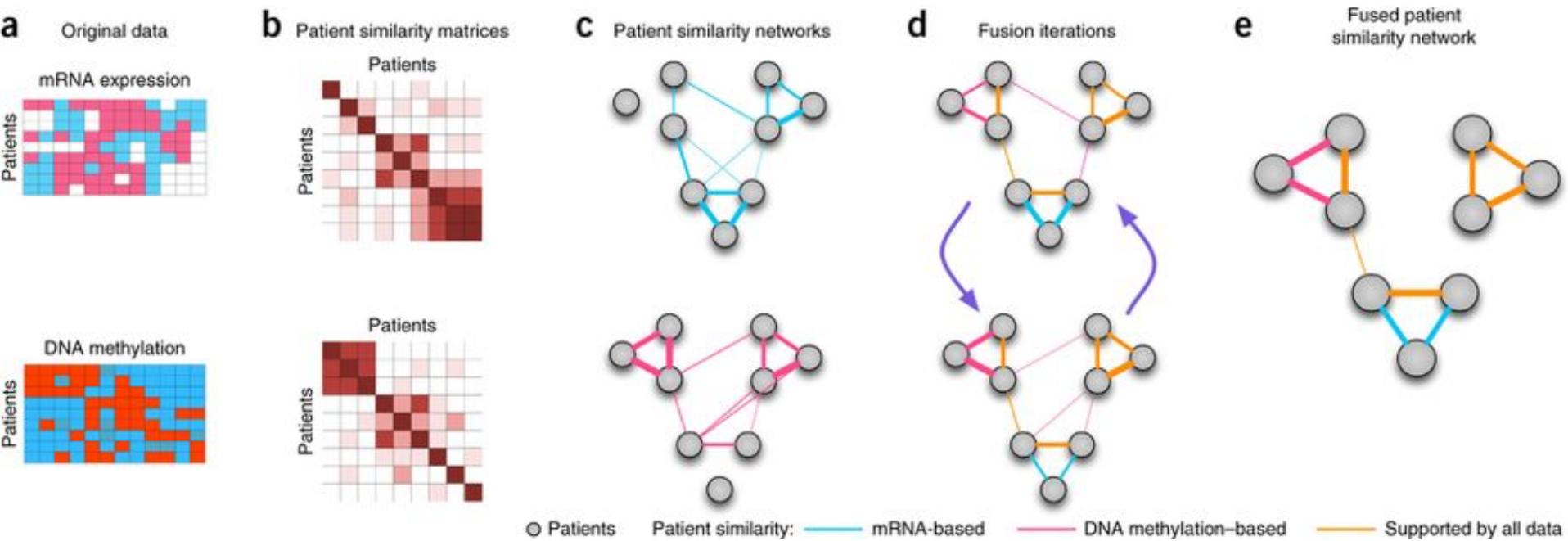
Supplementary information

Rights and permissions

About this article

Further reading

2-Similarity Network Fusion



(a) Example representation of mRNA expression and DNA methylation data sets for the same cohort of patients. (b) Patient-by-patient similarity matrices for each data type. (c) Patient-by-patient similarity networks, equivalent to the patient-by-patient data. Patients are represented by nodes and patients' pairwise similarities are represented by edges. (d) Network fusion by SNF iteratively updates each of the networks with information from the other networks, making them more similar with each step. (e) The iterative network fusion results in convergence to the final fused network. Edge color indicates which data type has contributed to the given similarity.

To Do

- 1- Prepare your list of differentially expressed genes / miRNAs according to the required formats

How exactly has a file to look like?

An input file (for mRNA, miRNA analogous) looks like:

```
AATK      -1
ABCB8     1
ABCG4     1
ABHD10    1
ABLIM1    -1
ABT1      1
etc...
```

To Do

- 1- Prepare your list of differentially expressed genes / miRNAs according to the required formats

How exactly has a file to look like?

An input file (for mRNA, miRNA analogous) looks like:

```
AATK      -1
ABCB8     1
ABCG4     1
ABHD10    1
ABLIM1    -1
ABT1      1
etc...
```

To Do

- 2-upload your files to TFmiR v1 or v2 and perform all possible analysis

TFmiR2
Constructing and analyzing disease-, tissue- and process-specific transcription factor and miRNA co-regulatory networks v2.0

Help Page CBI CENTER FOR BIOINFORMATICS

Step 1: Input selection

miRNA Choose File no file selected

mRNA Choose File no file selected

Load example data

Step 2: Configuration

miRNA-target enrichment p-value threshold

Species

Related disease

ORA p-value threshold

Related process

Related tissue

Evidence

Protein-Protein/Gene-Gene interaction cutoff

Randomization Method

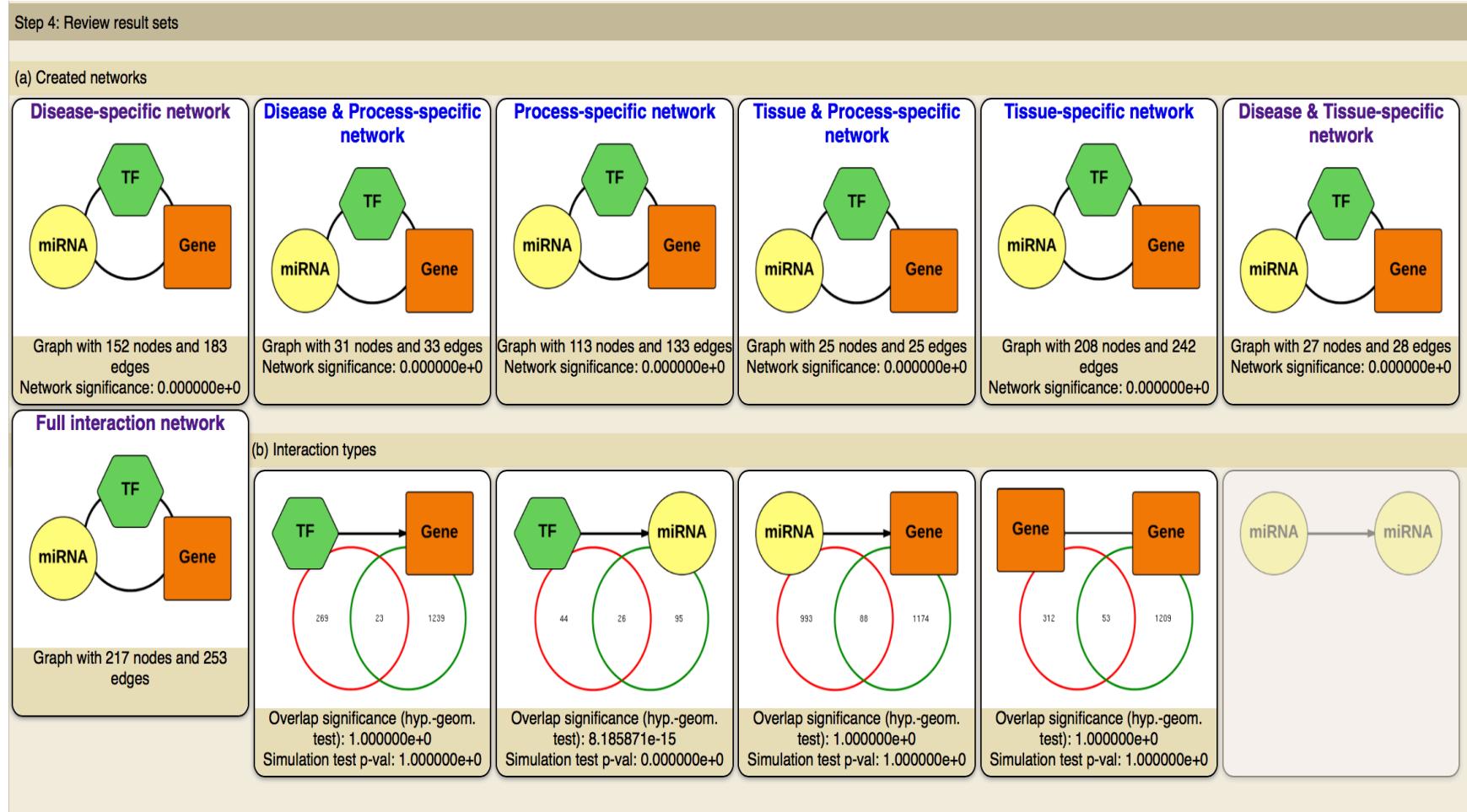
Step 3: Go!

Please stand by...

The screenshot shows the TFmiR2 web application interface. It has three main sections: Step 1 (Input selection) with fields for miRNA and mRNA files; Step 2 (Configuration) with settings for enrichment p-value threshold (0.05), species (Human), related disease (Breast Neoplasms), and related processes (Akt pathway, Angiogenesis, Apoptosis, Bone regeneration, Cardiogenesis, Cell cycle related, Cell differentiation); and Step 3 (Go!) which is currently processing with a yellow button saying "Please stand by...". A "Load example data" button is highlighted with an orange circle. The CBI logo is in the top right.

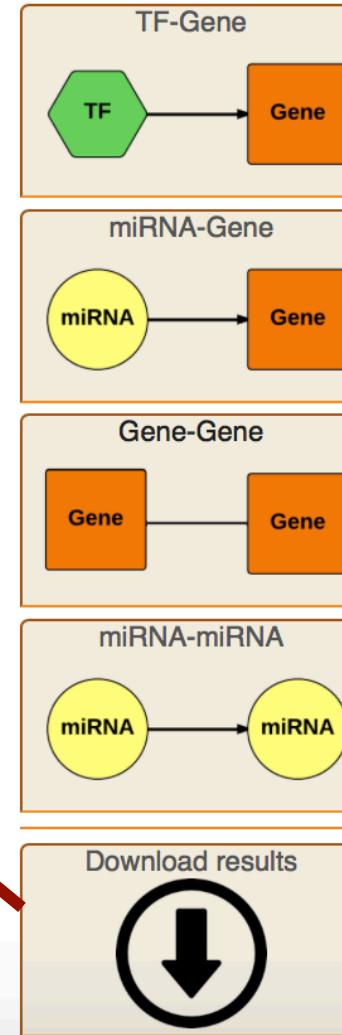
To Do

- 4- Visualize all networks, identify motifs and their similarity tests



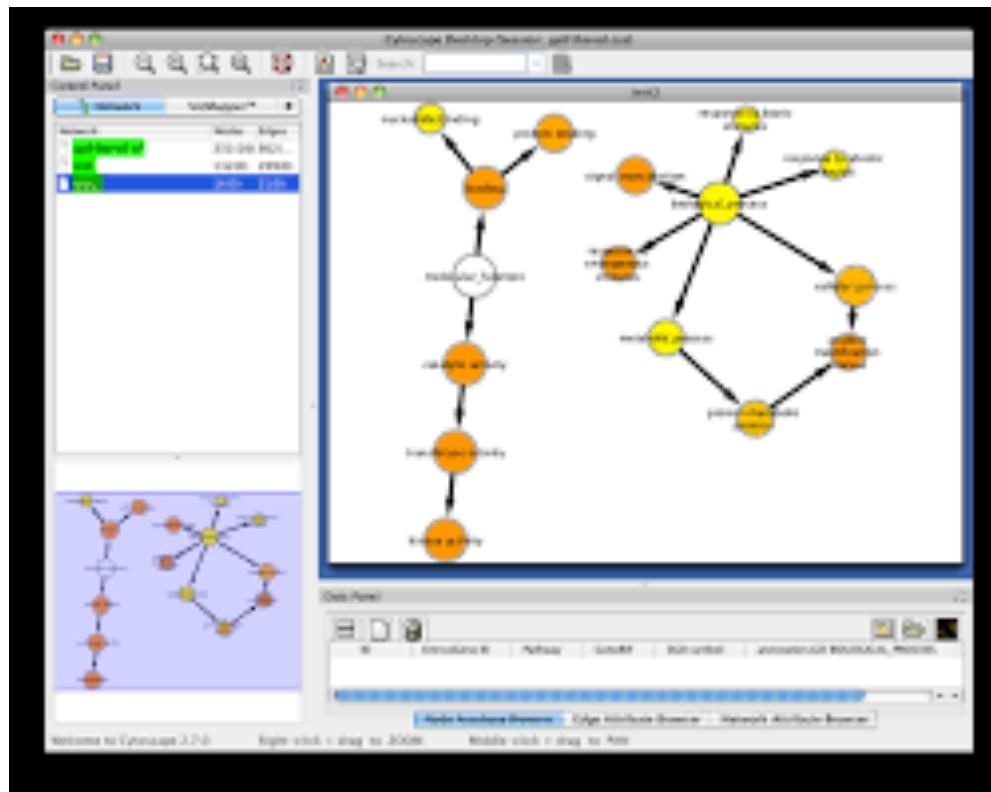
To Do

- 3-download your results and start to look at them



To Do

- 4-downlaod cytoscape and start to visualize your network and highlight the identified hot spots



**YOUR TURN
START WITH THE TUTORIALS**