Inference of predictive gene interaction networks

Benjamin Haibe-Kains

DFCI/HSPH

December 15, 2011

The Computational Biology
and Functional Genomics
Laboratory
at the Dana-Farber Cancer Institute and Harvard School of Public Health

Background

- Phenotypes result from biological **networks**, not individual genes

- New biotechnologies allow us to analyze multiple genes in parallel:
  - next generation sequencing
  - **gene expression profiling**
  - Chip-seq
  - ...

- Understand the complex interactions between genes and the behavior of a network is fundamental
  - to bring new biological insights
  - to make "useful" predictions

Relevance

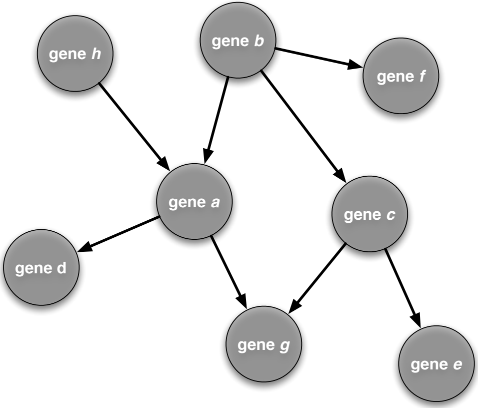
Aim: **Infer reliable predictive gene interaction networks from gene expression data**

Beyond biological understanding, such networks would be efficient tools for:

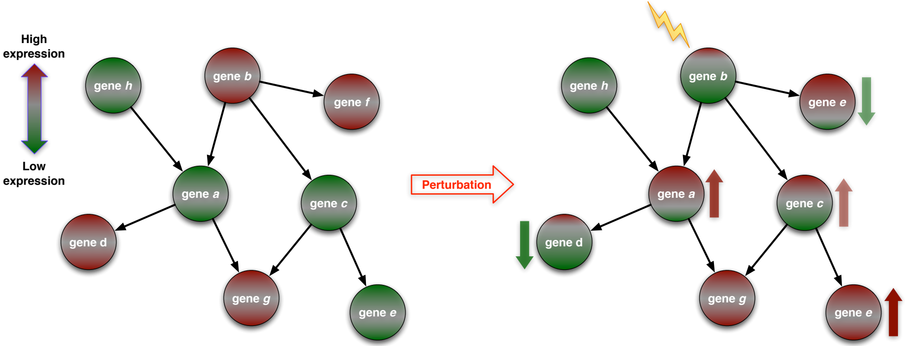
- predicting the response of an organism (cancer patient) to perturbations (targeted therapies)
- identifying the key genes to target for significantly decreasing a pathway activity
- optimizing combination of drugs and therapy regiments

Gene Interaction Network

- Genes are represented as "nodes"
- Interactions are represented by "edges"
- Edges can be directed to show "causal" interactions
- Edges are not *necessarily* direct interactions



Gene Interaction Network and Perturbation



Challenges in network inference

. . . and how we address them

1. *Problem complexity*: seed the search for the "best" network by using prior biological knowledge about gene interactions
   - ➟ **Predictive Networks** web application

2. *Curse of dimensionality*: development of a local regression-based network inference to enable analysis of hundreds of genes in parallel

3. *Lack of validation*: development of performance criteria to assess the quality of network models

4. *Lack of software*: implementation of network inference methods and related tools in R
   - ➟ **predictionet** package

Challenges in network inference

. . . and how we address them

1. *Problem complexity*: seed the search for the "best" network by using prior biological knowledge about gene interactions
   - ➟ **Predictive Networks** web application

2. *Curse of dimensionality*: development of a local regression-based network inference to enable analysis of hundreds of genes in parallel

3. *Lack of validation*: development of performance criteria to assess the quality of network models

4. *Lack of software*: implementation of network inference methods and related tools in R
   - ➟ **predictionet** package

Challenges in network inference

. . . and how we address them

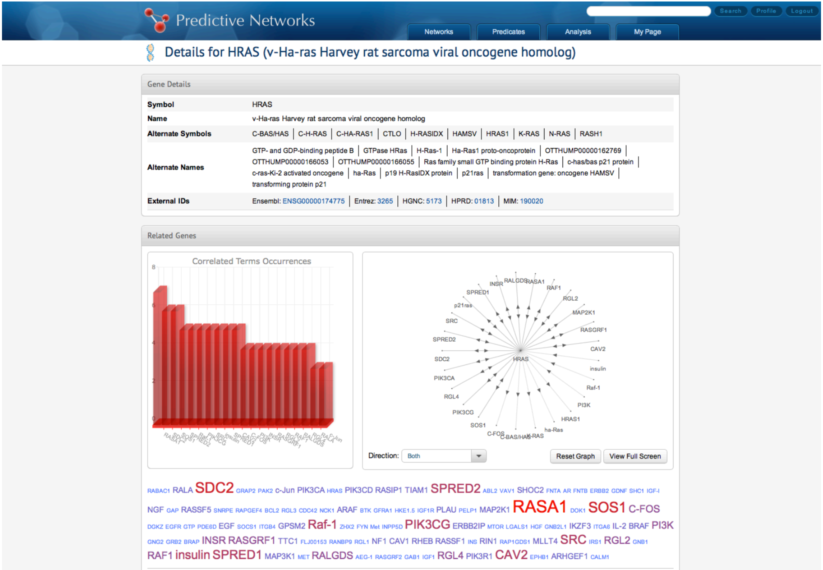
1. *Problem complexity*: seed the search for the "best" network by using prior biological knowledge about gene interactions
   - ➟ **Predictive Networks** web application
2. *Curse of dimensionality*: development of a local regression-based network inference to enable analysis of hundreds of genes in parallel
3. *Lack of validation*: development of performance criteria to assess the quality of network models
4. *Lack of software*: implementation of network inference methods and related tools in R
   - ➟ **predictionet** package

Challenges in network inference

. . . and how we address them

1. *Problem complexity*: seed the search for the "best" network by using prior biological knowledge about gene interactions
   - ➟ **Predictive Networks** web application
2. *Curse of dimensionality*: development of a local regression-based network inference to enable analysis of hundreds of genes in parallel
3. *Lack of validation*: development of performance criteria to assess the quality of network models
4. *Lack of software*: implementation of network inference methods and related tools in R
   - ➟ **predictionet** package

Predictive Networks web application



Details for HRAS (v-Ha-ras Harvey rat sarcoma viral oncogene homolog)

Gene Details

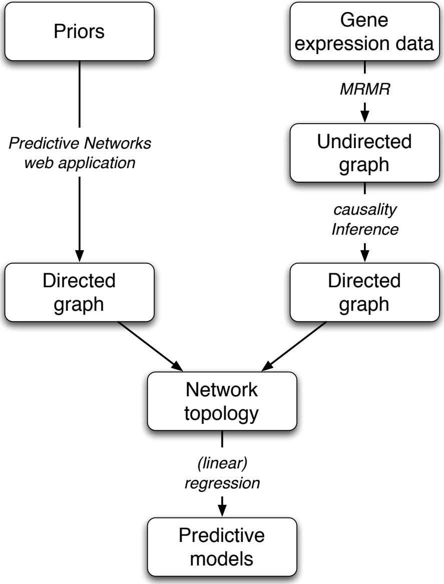
Symbol	HRAS
Name	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
Alternate Symbols	C-BAS/HAS | C-H-RAS | C-HA-RAS1 | CTLO | H-RASIDX | HAMSV | HRAS1 | K-RAS | N-RAS | RASH1
Alternate Names	GTP- and GDP-binding peptide B | GTPase HRas | H-Ras-1 | Ha-Ras1 proto-oncoprotein | OTTHUMP00000162769 | OTTHUMP00000166053 | OTTHUMP00000166055 | Ras family small GTP binding protein H-Ras | c-has/bas p21 protein | c-ras-Ki-2 activated oncogene | ha-Ras | p19 H-RasIDX protein | p21ras | transformation gene: oncogene HAMSV | transforming protein p21
External IDs	Ensembl: ENSG00000174775 | Entrez: 3265 | HGNC: 5173 | HPRD: 01813 | MIM: 190020

Related Genes

RABAC1 RALA SDC2 GRAP2 PAK2 c-Jun PIK3CA HRAS PIK3CD RASIP1 TIAM1 SPRED2 ABL2 VAV1 SHOC2 FNTA AR FNTB ERBB2 GDNF SHC1 IGF-I NGF GAP RASSF5 SNRPE RAPGEF4 BCL2 RGL3 CDC42 NCK1 ARAF BTK GFRA1 HKE1.5 IGF1R PLAU PELP1 MAP2K1 RASA1 DOK1 SOS1 C-FOS DGKZ EGFR GTP PDE6D EGF SOCS1 ITGB4 GPSM2 Raf-1 ZHX2 FYN Met INPP5D PIK3CG ERBB2IP MTOR LGALS1 HGF GNB2L1 IKZF3 ITGA6 IL-2 BRAF PI3K GNG2 GRB2 BRAP INSR RASGRF1 TTC1 FLJ00153 RANBP9 RGL1 NF1 CAV1 RHEB RASSF1 INS RIN1 RAP1GDS1 MLLT4 SRC IRS1 RGL2 GNB1 RAF1 insulin SPRED1 MAP3K1 MET RALGDS AEG-1 RASGRF2 GAB1 IGF1 RGL4 PIK3R1 CAV2 EPHB1 ARHGEF1 CALM1

`https://compbio.dfci.harvard.edu/predictivenetworks/`

Regression-based network inference

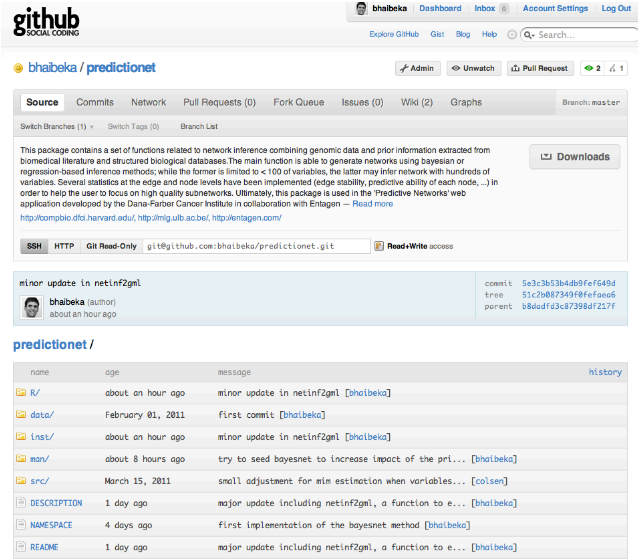


Performance criteria for network inference

We implement a cross-validation framework to assess:

- edge-specific stability: what are the interactions inferred in most of the cross-validation folds?
- gene-specific prediction score: what are the genes whose expression can be well predicted by their parent/source genes?

Predictionet R package



This package contains a set of functions related to network inference combining genomic data and prior information extracted from biomedical literature and structured biological databases.The main function is able to generate networks using bayesian or regression-based inference methods; while the former is limited to < 100 of variables, the latter may infer network with hundreds of variables. Several statistics at the edge and node levels have been implemented (edge stability, predictive ability of each node, ...) in order to help the user to focus on high quality subnetworks. Ultimately, this package is used in the 'Predictive Networks' web application developed by the Dana-Farber Cancer Institute in collaboration with Entagen — Read more

http://compbio.dfci.harvard.edu/, http://mlg.ulb.ac.be/, http://entagen.com/

`git@github.com:bhaibeka/predictionet.git`

```
minor update in netinf2gml
```

bhaibeka (author)
about an hour ago

```
commit  5e3c3b53b4db9fef649d
tree    51c2b087349f0fefaea6
parent  b8dadfd3c87398df217f
```

predictionet /

name	age	message	history
R/	about an hour ago	minor update in netinf2gml [bhaibeka]	
data/	February 01, 2011	first commit [bhaibeka]	
inst/	about an hour ago	minor update in netinf2gml [bhaibeka]	
man/	about 8 hours ago	try to seed bayesnet to increase impact of the pri... [bhaibeka]	
src/	March 15, 2011	small adjustment for mim estimation when variables... [colsen]	
DESCRIPTION	1 day ago	major update including netinf2gml, a function to e... [bhaibeka]	
NAMESPACE	4 days ago	first implementation of the bayesnet method [bhaibeka]	
README	1 day ago	major update including netinf2gml, a function to e... [bhaibeka]	

https://github.com/bhaibeka/predictionet

Acknowledgements

Computational Biology and Functional Genomics Laboratory

Dana-Farber Cancer Institute, USA

Renee Rubio
Kathleen Fleming
Niall Prendergast
Raktim Sinha
Daniel Schlauch
Alejandro Quiroz
Megha Padi
Stefan Bentink
John Quackenbush

Machine Learning Group

Université Libre de Bruxelles, Belgium

Catharina Olsen
Gianluca Bontempi

Ontario Cancer Institute

Universty of Toronto, Canada

Amira Djebbari

Entagen

Entagen, USA

Christopher Bouton
Erik Bakke
James Hardwick