Paper Notes

Friday 16 September 2016 16:50

Biology

Data

DATABASE of databases!!! http://www.pathguide.org

1. Microarray -> **Gene Expression Omnibus (GEO)**
2. CrispR screens
3. Chip-seq
4. RNAi
5. SGA
6. Yeast 2 hybrid
7. Mass spectrometry

Experimental techniques	Network type		Link	Signed	Direction	Computational methods	Database	Reference
Microarray	Relevance - co-expression gene interaction network		Correlation between gene expression levels				Gene Expression Omnibus	
Yeast 2 hybrid	Protein interaction network		Physical interaction of proteins				BioGRID	
Mass spectrometry	Protein interaction network		Physical interaction of proteins				BioGRID	

Assumptions	Comment

SUPPLEMENTARY INFORMATION

In format provided by Mitra *et al.* (OCTO

Supplementary information S1 (table). Sources of molecular interaction and 'omics' profiling data

	Interaction type(s)	Detection methodologies	Databases
Physical	Protein-protein	Yeast-2-hybrid (Y2H)[1-3], co-immuno-precipitation (Co-IP)[4], mass spectroscopy[5,6]}, affinity purification coupled with mass spectroscopy (AP/MS)[7,8]	BioGRID[9], IntAct[10], APID[11], STRIN MINT[13], DIP[14], HPRD[15], MIPS-MPI Netpath[17], DroiD[18]
	Protein-DNA (e.g., regulatory networks)	Yeast-1-hybrid (Y1H)[19], chromatin immuno-precipitation based methods (CHIP-CHIP)[20], DNA-footprinting[21]	TRANSFAC[22], UniProbe[23], DroiD[18] BioGRID[9], TcoF-DB[24], BIPA[25], hPD EDGEdb[27], NPIDB[28,29]
	Protein-RNA	RNA electro-mobility shift (RNA-EMSA)[30], RNA-pull down[31]	PRID[32], BIPA[25] NPInter[33], RBPDB[34] StarBase[36]
	Metabolic (e.g., enzyme-substrate, ligand-receptor)	Mass spectroscopy based selective reaction monitoring (SRM)[6,37], NMR[38], affinity purification[8], co-IP[3], fluorescence spectroscopy[39]	Reactome[40], KEGG [41], BioCyc and H HMDB[43,44], EcoCyc[45], HumanCyc[46], ConsensusPathDB[47]
	Protein/gene-compound (e.g., drug-target, chemical-protein)	Chemical structure[48,49], forward or reverse chemo-genomic/proteomic profiling[50-52], in silico predictions[53]	SuperTarget[54], Matador[54], DrugBank ChemProt[56], STITCH[57], AffinDB[58], MatrixDB[59], PSMDB[60], PDB-Ligan ChEMBL[62], ConsensusPathDB[47]
Functional	Genetic (gene-gene)	Synthetic genetic array (SGA)[63], Epistatic Miniarray Profiling (E-MAP)[64], co-expression profiling[65,66]	BioGRID[9], DRYGIN[67], CYGD[68], Dr ConsensusPathDB[47]
	Gene-Disease	Literature curation, clinical and sequence information	OMIM[69], HuDiNe[70], Diseasome[71]
Omics data type		**Detection methodologies**	**Databases**
Transcriptomics		Microarray[72], RNASeq[73]	GEO[74], SMD[75], TCGA[76], GXD[77], ONCOMINE[78], ArrayExpress[79],
RNAi (phenomics)		RNAi interference assay[80]	RNAiDB[81], GenomeRNAi[82], siReco

OBER 2013)

NG[12], PI[16],
, DI[26],
[4], PRIDB[35],
MetaCyc[42], ,
[55] d[61],
roID[18],
rds[83]

SGA	Gene interaction network based on phenotypes		Phenotype quantification, genetic interaction					
	Signaling network							
	Metabolic network							
Combined								

Assumptions of mRNA levels network inference

1. mRNA levels correlate to the protein levels
2. mRNA levels of transcription factors and their targets tend to be correlated
3. binary interactions
    1. This can be solved by using hypergraphs or the equivalent bipartite networks! See Ne introduction
4. Static representation of a nonlinear dynamic process
5. ????

Assumptions and biases of SGA phenotype networks (inference of genetic interactions)

i. Phenotype and function refers to growth measurements
ii. One condition, usually the optimal conditions -> Differential network biology
iii. Biased towards negative interactions

Biological Networks at the molecular level

Epigenomics	Methylation profiling[84]	DAnCER[85], DiseaseMeth[86], PubMet MethDB[88], MethCancerDB[89], Methy
Mutation / SNP	SNP Array[91], genome sequencing[92]	TCGA[76], dbSNP[93], dbQSNP[94], GWA [95], OMIM[69]
Proteomics	CHIP [11,2,96], Mass Spectrometry[5,6]	PDB[97], ExPASy[98], InterPro[99], World JASPAR[101]
Phosphorylation profile	Mass Spectrometry[5], literature curation	PhosphoGRID[102], PhospoELM[103], P

For more information and databases, visit www.pathguide.org

	Functional relations

wman networks an

th[87], yCancer[90]
AS Central
-2DPage[100],
HOSIDA[104]

1.

Biological evidence and framework

- GOAL: develop a complete map of biological modules underlying cellular architecture and fu
  Integrative approaches for finding modular structure in biological networks.
- Levels of biological - molecular networks -> **Genetic interaction networks/ better understa**
  - For a computer, the lower level of abstraction would contain details on the hard- ware level will represent the logic of the program. In agreement with this approach, a syste con- sider a biological system with all its complexity and identify, from the genomic se phenotype, different levels of abstractions. At the lower level of this conceptual struct several networks representing the physical structure and organization of the genome nodes could be genes/coding sequences, single-nucleotide polymorphisms (SNPs) or linked by edges representing their physical proximity and organization within chromo homology etc. (**Figure 2**, level I). The second level of abstraction would represent the genome into physical components: proteins and RNA. Edges between these elements they are co-expressed in different contexts or that their expression profiles throughou experimental conditions are highly correlated (**Figure 2**, level II; Ge et al., 2003; Vidal third level of abstraction would represent physical interactions between different elem protein (PPI), protein-DNA (PDI) or protein-RNA (PRI) interactions (**Figure 2**, level III; V The fourth level of abstraction will allow the visualization of the functional relationshi physical elements. This level would contain GI net- works, signaling and metabolic pat level IV). The fifth level would represent biological processes. This level would contain proteins implicated in the same biological process would be linked by an edge (**Figure** and last level of abstraction would represent phenotypes and show the relationships associated with similar phenotypes and diseases (**Figure 2**, level VI).
- Metabolic networks convert raw materials from the environment into value-added product dispose of intracellular materials. Regulatory networks alter the output of the genomic prog term regulatory network only encompasses transcriptional and translational regulation; hov regulatory features associated with all cellular processes. Signalling networks are the comm that connect cellular components and coordinate the phenotypic response to perturbations networks simultaneously monitor a wide variety of external and internal parameters, includ and DNA damage. The observations are continuously processed in an integrated fashion, an metabolic outputs are modulated to deliver the appropriate response. For simplicity's sake, on the molecular components of a cell and their interactions — intercellular interactions are environmental factors. -> **Towards genome-scale signalling- network reconstructions**
- The state of a cell is governed by the complex regulation of the expression of its genes. This many levels, ranging from chromatin remodelling to post translational modifications [60]. In understand gene regulation, a large and growing body of transcriptomic data can be used to between genes, and has prompted the development of a number of gene regulatory netwo reconstruction algorithms, including Graphical Gaussian models, Bayesian networks and rele

unction ->

nd to better predict
e while the higher
ems biologist will
equence to the
ture, we would find
. In these networks,
coding sequences
somes, their
expression of that
would indicate that
ut multiple
et al., 2011). The
ments – protein–
idal et al., 2011).
ps linking these
thways (**Figure 2**,
networks where
2, level V). The sixth
between elements

s and recycle or
gramme. Here, the
wever, there are
unication networks
s. Signalling
ing nutrient levels
nd the genomic and
this Review focuses
e treated as

regulation occurs at
n order better to
o infer interactions
rk (GRN)
evance networks ##

Information theory and signal transduction systems/ From molecular information process inference

- the aims of large-scale network inference, which are typically to highlight and explore depe sets—and thereby help to generate new hypotheses worthy of further investigation—rathe single grand unifying model of the system. -> **Systems biology (un)certainties**
- Models are simplified (but not simplistic) representations of real systems, and this is 8precis that makes them attractive to explore the consequences of our assumptions, and to identify understanding of the principles governing a biological system. Models are tools to uncover cannot be directly observed, akin to microscopes or nuclear magnetic resonance machines (interpreted appropriately, with due attention paid to inherent uncertainties, the mathemat computational modeling of biological systems allows the exploration of hypotheses. But the models depends on the ability to assess, communicate, and, ultimately, under- stand their u **Systems biology (un)certainties**
- For some model organisms, however, protein interaction data covers 20% of the proteins kn organism (ignoring multiple isoforms due to alternative splicing, etc.). This observation pose and important question of just how representative a random subnet is for the global netwo **scale-free networks are not scale-free: Sampling properties of networks**
- We emphasize that the degree distribution alone does not suffice to characterize a network networks, e.g., some with many cross-connections (loops) and others with ''tree-like'' form have the same degree distribution. -> **Subnets of scale-free networks are not scale-free: Sa of networks**
- Essential genes are those genes critical for cell viability under certain contexts -> **Network a essentiality in functional genomics experiments**
- However, negative feedback plays an important role in biological robustness and evolution5 played an equally prominent role in the origins of life. <- **Prebiotic network evolution/ six k**
- **A possible resolution to the instabilities that arise in net- work expansion and evolution is feedback, which enables growth, with negative feedback, which contributes to robustnes 2c).** <- **Prebiotic network evolution/ six key parameters**
- This reverse engineering is extremely difficult. Although an electrical engineer could design circuits that would amplify signals, he would find it difficult to deduce the circuit diagram of amplifier by correlating its outputs with its inputs. It is thus unlikely that we can deduce the level description of a module solely from genome- wide information about gene expression interactions between proteins. Solving this problem is likely to require additional types of in finding general principles that govern the structure and function of modules ## **From molec biology**
- But techniques for collecting information about the entire genome will be only as powerful available to analyse it, just as our ability to infer protein structure and function from protein increased with the sophistication of tools for sequence analysis. ## **From molecular to mod**
- Cell functioning requires the tight linking between multiple regulatory systems that include controlling initiation of gene transcription, RNA splicing, mRNA transport, translation initiati translational protein modifications, or the degradation of mRNA/protein -> **Michael Stumpf**

sing to network

ndencies in data
r than to uncover a

sely the property
y where we lack
mechanisms that
15). Used and
ical and
e relevance of these
uncertainties. ->

nown to exist in that
es the interesting
rk -> **Subnets of**

k: very different
(no loops at all), can
mpling properties

nalysis of gene

52 and may have
ey parameters
s to balance positive
s and stability (Fig.

many different
an unknown
circuity or a higher-
and physical
formation and
cular to modular cell

as the tools
n sequence data has
ular cell biology
mechanisms for
on, post-
f, David J. Balding,

Mark Girolami (editors)-Handbook of Statistical Systems Biology ## Chapter 12

- main obstacles is the difficulty of choosing the appropriate experimental data, together with computational approaches. -> **Handbook of Statistical Systems Biology ## Chapter 12**
- the goal for methodology to infer regulatory networks consists in modelling and recovering interactions such as 'protein i activates (or inhibits) the transcription of gene j'. -> **Handbook Systems Biology ## Chapter 12**
- From the biological side, it is not that clear that the identification of regulatory networks is concentrations of mRNA due to the role of additional sources of regulation such as micro RN **Networks 2007 ## chapter 3**
- What are the limits of the definition of a regulatory protein? We are interested here in prot subset of the genes of an organism. Other proteins also called regulatory play a more genera transcription (for example the eukaryotic transcription factor of type II). In principle, these g polymerase itself, are essential for the transcription of all genes that encode proteins. Howe action is non-specific, they are not covered in this Section ## RNA polymerase is necessary f so is abundant in the cell!!! **-> Biological Networks 2007 ## chapter 4**
- Transcriptional Networks represent only binary interactions between mRNA molecules(!!!!) bind to form protein complexes and then they interact with other proteins and/or catalyze *are not suitable when it comes to expressing interactions that involve more than two partne fashion. The formalism of hypergraphs allows to express non-binary interactions and could b context*. **-> Biological Networks 2007 ## chapter 4**
- because high-throughput datasets can be filled with technical and biological noise and have biases and coverage, improved statistics are needed to distinguish signal from noise, as well integration to annotate the biologically relevant relationships; because the logical interpreta network is not easily comprehensible to the human brain, computational modeling is neede output from the signal input or system perturbation; because no analysis and modeling met detailed targeted biological experiments are needed to validate mod- els before a hypothes approach health and medical problems -> **Understanding biological functions through mole** Problems of the high throuput data!!
- Small-scale gene-centric studies have delineated many valuable genetic and biochemical rel genes and pathways. This information forms the skeleton of the entire complete network an annotation standards. For now, such information covers only a tiny fraction of the full netw towards certain biological functions. For example, protein interactions in the literature cura (Human Protein Reference Database) [14] cover only 20 000 interactions out of a con- serva 000 interactions and are strongly biased for cancer-related processes (Supplementary in- fo Large-scale experiments and data mining are obviously more effective for mapping genetic **Understanding biological functions through molecular networks** ## Problems of the single approach!!
- However, the change to network biology does not simply entail hand- ing over biology to ph mathematicians. A good understanding of biology is needed to ask the right questions, to ch network analysis tools, and to con rm analysis results by solid experimentation. After all, ne biology. The fundamental goal of network biology is the same as molecular biology: to unde

h the appropriate

regulatory
k of Statistical

possible only using
NA. **-> Biological**

eins that regulate a
al role in initiating
generalists, like RNA
ever, since their
for all transcription

, but some proteins
reactions!!! *Graphs*
ers in an obligate
be used in the above

different technical
as better data
ation of the whole
ed to predict the
thod is perfect, more
sis can be used to
ecular networks ##

ationships between
nd can serve as
ork and is biased
ted by HPRD
ative estimate of 150
rmation, table S1).
networks ->
gene - process

hysicists and
hoose proper
twork biology is
erstand basic

biological processes and the mechanisms of human diseases. -> **Understanding biological f molecular networks**

- Furthermore, yeast two-hybrid assays produce many false-positive outcomes, and the curre pathway maps may be heavily biased towards connection to functionally important genes si have been popular targets for research -> **Kitano, Hiroaki Computational systems biology.p**
- despite the known poor overall correlation between mRNAs and their protein products, the assumption that differentially expressed mRNAs impact their respective experimental condi in proteins. However, to the best of our knowledge, this assumption has never been explicit **Relationship between di erentially expressed mRNA and mRNA- protein correlations in a x system**
- Although clustering analysis provides insight into the "correlation" among genes and biologi does not reveal the "causality" of regulatory relationships. Several methods have been prop automatically discover regulatory relationships solely on the basis of microarray data (*7–9*). methods use information derived from mRNA abundance, so there is limited scope to infer transcriptional regulation. Posttranscriptional and posttranslational mechanisms of regulati incorporated as large-scale data become available, but many properties have yet to be mea accuracy or in high throughput. -> **Systems Biology: A Brief Overview**
- There is a need for systematic approaches to infer causal relationships between interacting we refer to the "direction" (edge direction), "sign" (activation/inhibition) and "mode" (e.g. p ubiquitination) of signal flow in PPI networks. **-> Integrating protein-protein interaction ne phenotypes reveals signs of interactions**
- The signed network opens up new scope for network analysis, such as the application of str theory. Further integration of other information-flow properties such as edge-direction wou sophisticated flow-based network analysis. **-> Integrating protein-protein interaction netw phenotypes reveals signs of interactions**
- **Integrating protein-protein interaction networks with phenotypes reveals signs of interac**
    - Each RNAi screen identifies positive and negative regulators of a particular phen
        - How?
    - PPI is for the physical interactions, from gene expression one can reconstruct a differential expression. **Direct regulation relationships must include both physi activation - repression data.**
- Biological-network reconstructions provide us with a framework that allows us to identify th whole and, subsequently, understand the relationship between the molecular phenotype an organism phenotype. -> **Towards genome scale signalling network reconstructions**
- a minority of genes are essential, and these define hubs of activity that can in some cases ex given functional module to influence and even coordinate multiple cellular processes. It is n interactional complexity, that single genes rarely specify a phenotype in its entirety. The ou genetic network are now apparent, but a compel- ling case can be made for a deeper and m exploration of this model system as an exemplar for more complex eukaryotes. -> **Explorin interactions and networks with yeast**
- The physical-interaction map, generated by large-scale two-hybrid51,52 or affinity purificati mass spectrometry identification26,43,53,54, provides a view of the gene products that ass

unctions through

nt hand-crafted
imply because these
pdf
re is an implicit
tions via differences
ly tested. ->
xenograft model

ical phenomena, it
posed to
At present, such
causality based on
on must be
sured with sufficient

proteins, by which
phosphorylation,
tworks with

ucture balance
uld enable
works with

ctions
notype ?

network based on
ical interactions and

he context of the
nd the general

xtend beyond a
o wonder, given this
tlines of a yeast
nore complete
g genetic

ion followed by
emble into soluble

mass spectrometry identification26,43,53,54, provides a view of the gene products that ass
protein complexes and function together as biochemical machines. Rather than physical inf
genetic-interaction map provides functional information, largely identifying gene products t
functionally related pathways. Although genetic interactions overlap with protein–protein in
often than expected by chance, such overlap is relatively rare, occurring at a frequency of le
2). -> **Exploring genetic interactions and networks with yeast**

- Large-scale genetic analyses reveal that mutations in most eukaryotic genes have little disc
example, systematic gene deletion in *S. cerevisiae*, discussed in detail below, produced a re
only ~20% of yeast genes are essential for viability when deleted individually in haploids gro
laboratory conditions
- Comparisons between aggravating and alleviating effects revealed that, for most functional
tions were either largely aggravating or largely alleviat- ing, but not mixed, an asymmetrical
termed 'monochromatic'.
- the set of inter- actions that are observed for a particular query gene can be suggestive of it
position of a gene in a genetic-interaction network being highly predictive of its molecular r
genetic interactions and networks with yeast
- **Because most genetic interactions do not overlap with physical interactions, the two type
said to be largely orthogonal**
- Integration attempts in yeast, combining physical protein- protein interaction maps with co
revealed that interacting proteins are more likely to be encoded by genes with similar expre
noninteracting proteins (Ge et al., 2001; Grigoriev, 2001; Jansen et al., 2002; Kemmeren et a
observations were subsequently confirmed in many other organisms (Ge et al., 2003). Beyo
mental aspect of finding significant overlaps between interaction edges in interactome netw
coexpression edges in tran- scription profiling networks, these observations have been used
overall biological significance of interactome datasets -> Interactome Networks and Human
- Experimental artefacts, variability in cover- age across data sets, sampling bias towards well
limitations in screening power and inher- ent sensitivities in various assays can yield false po
negatives in interaction data -> **Integrative approaches for finding modular structure in bio**
- In addition, as reported by Zotenko et al. [17], essential proteins tend to form highly connec
than function independently -> **A new essential protein discovery method based on the int
protein interaction and gene expression data**
- However, stand- ard clustering is not ideal for PPI networks: proteins may have multiple fun
therefore the corresponding nodes may belong to more than one cluster; for example, 207
the CYC2008 hand-curated yeast complex data set3 participate in more than one complex. -
overlapping protein complexes in protein-protein interaction networks
- The existence of a GI between two genes does not necessarily imply that these two genes co
proteins or that the two genes are even expressed in the same cell. In fact, a GI only implies
share a functional relationship. These two genes may be involved in the same biological pro
they may also be involved in compensatory pathways with unrelated apparent function -> G
networks: better understand to better predict
- A network in the left represents a global PIN. A eukaryotic cell in the center can be divided i
compartments. Endoplasmic Cytoskeleton Golgi Cytosol Lysosome(or Vacuole) Mitochon

emble into soluble
ormation, the
hat operate in
nteractions more
ss than 1% (REF.

ernable effect. For
markable result:
owing in standard

groups, interac-
feature that they

s function, with the
ole. -> **Exploring**

s of interaction are

expression profiles,
ession profiles than
al., 2002). These
nd the funda-
works and
d to estimate the
Disease
-studied processes,
osi- tives and false
ological networks.
cted clusters rather
tegration of protein-

nctions, and
of 1,628 proteins in
-> detecting

ode for interacting
that the two genes
ocess or pathway; or
Genetic interaction

nto 11
ndrion, Endosome

compartments, Endoplasmic, Cytoskeleton, Golgi, Cytosol, Lysosome(or Vacuole), Mitochon
Plasma, Nucleus, Peroxisome and Extracellular, where Lysosome only exists in animal cells.
compartment, a PSLIN of this compartment is constituted by the proteins localized in this co
their interactions. With the subcellular localization information of proteins, the PSLINs can b
mapping the global PIN to each compartment separately. -> **Rechecking the Centrality-Leth
Scope of Protein Subcellular Localization Interaction Networks**

- In contrast to pathways, protein complexes are the functional units of proteome organizatio
dynamic assembly is fundamental to induce cellular responses to different internal and exte
Protein Complex–Based Analysis Framework for High-Throughput Data Sets
- The vast majority of proteins work as parts of assemblies composed of several elements, th
protein complexes as essential cellular functional units. Given the fundamental importance
interactions, proteome-wide "interactome" maps based on pairwise protein interactions us
hybrid (Y2H) system have been determined for several organisms (Giot et al., 2003; Ito et al
2004; Rual et al., 2005; Stanyon et al., 2004; Stelzl et al., 2005; Uetz et al., 2000). Alternative
isolation based on co-affinity purification combined with tandem mass spectrometry (coAP-
to generate protein complex maps at proteome-scale for *Saccharomyces cerevisiae* -> **A Pro
Network of Drosophila melanogaster**
- However, these abovementioned approaches for extracting dense subgraphs fail to take int
inherent organization. Recent analysis of experimentally detected protein complexes [23] h
complex consists of a core component and attachments. Core proteins are highly co-expres
functional similarity, and each attachment protein binds to a subset of core proteins to form
complex. Based -> **Identification of protein complexes from multi-relationship protein inte**
- we construct a multi-relationship protein interaction network (MPIN) by integrating PPI net
gene ontology (GO) annotation information.-> **Identification of protein complexes from mu
protein interaction networks**

Network construction methods and evaluation

- Information theoretic approaches to GRN reconstruction have two major strengths. The firs
information is able to capture nonlinear associations between variables, a feature seen in ex

uclear, Endosome, For each ompartment and be generated by **hality Rule in the**

on, and their ernal cues. ->

ereby defining of protein ing the yeast two- ., 2001; Li et al., ely, protein complex -MS) has been used **otein Complex**

o account the as revealed that a sed and share high m a biological **eraction networks** work topology with **ulti-relationship**

st is that mutual xpression data.

Second, it has been demonstrated that the use of the data processing inequality (DPI), whic noisy system X → Y → Z, knowledge of Z cannot give more information about X than Y can gi in distinguishing two genes regulated by a third from a trio of co-regulating genes. In recons GRNs, the combined approach using mutual information and DPI outperforms Bayesian and techniques in the precision and recall of direct regulatory links ## **Information theory and s systems/ From molecular information processing to network inference**

- Several alternative network inference approaches exist (*4*, *5*) that provide better, more robu networks, and can incorporate expert or domain knowledge. These methods may be used to regulatory relationships, and how interactions between genes change over time or differ be and healthy controls. -> **Systems biology (un)certainties**
- Genome-scale inference of transcriptional gene regulation has become possible with the ad throughput technologies such as microarrays and RNA sequencing, as they provide snapsho transcriptome under many tested experimental conditions. From these data, the challenge computationally predict direct regulatory interactions between a transcription factor and its aggregate of all predicted interactions comprises the gene regulatory network. -> **Wisdom robust gene network inference**
- Understanding the advantages and limitations of different network inference methods is cri effective application in a given biological context -> **Wisdom of crowds for robust gene netw**
- We conclude that there is no category of network inference methods that is inherently supe performance depends largely on the specific implementation of each individual method. -> **for robust gene network inference**
- We next analyzed how method-specific biases influenced the recovery of different connecti (network motifs), and we observed characteristic trends for different method categories (**Fi** feed-forward loops were recovered most reliably by mutual-information and correlation-ba whereas sparse-regression and Bayesian-network methods performed worse at this task. Th that the latter approaches preferentially select regulators that independently contribute to target genes. However, the assumption of independence is violated for genes regulated by transcription factors, as in the case of feed-forward loops. Indeed, linear cascades were mo predicted by regression and Bayesian-network methods. This shows that current methods e off between performance on cascades and performance on feed-forward loops. -> **Wisdom robust gene network inference**
- Network inference methods have complementary advantages and limitations under differe suggests that combining the results of multiple inference methods could be a good strategy predictions. -> **Wisdom of crowds for robust gene network inference**
- After identifying these modules24, we tested them for enrichment of Gene Ontology terms Note 7). Network modules are strongly enriched for very specific biological processes. This a unique functions to most of the identified modules in both networks
- When constructing a compendium of microarrays for global network inference, one should toward oversampling a narrow set of experimental conditions. -> **Wisdom of crowds for rob inference ### QUESTION!!!!**
- Bayesian- network methods exhibited below-average performance in this challenge, likely b heuristic searches, which are often too costly for systematic data resampling and may be be

n states that for a
ive about X [19,65],
structing simulated
l relevance networks
ignal transduction

ust candidate
o investigate gene-
etween disease cases

vent of high-
ts of the
is to
s target genes; the
of crowds for

tical for their
work inference
erior and that
Wisdom of crowds

vity patterns
g. 2c). For example,
sed methods,
ne reason for this is
the expression of
mutually dependent
re accurately
xperience a trade-
n of crowds for

nt contexts, which
for improving

(Supplementary
allowed us to assign

thus avoid any bias
bust gene network

because they use
etter suited for

heuristic searches, which are often too costly for systematic data resampling and may be be
smaller networks. Information theoretic methods performed better than correlation-based
two approaches had similar biases in predicting regulatory relationships. They also perform
regression and Bayesian- network methods on feed-forward loops, fan-ins and fan-outs (the
connected parts of the network), but they had an increased rate of false positives for cascad
crowds for robust gene network inference

- A fundamental assumption of network inference algorithms is that mRNA levels of transcrip
their targets tend to be correlated; we found that this is true for *E. coli*, but not for *S. cerevis*
lower coverage of *S. cerevisiae* gold standards may also play a role (*E. coli* has the best-knov
network of any free-living organism16), the poor correlation at the mRNA level in *S. cerevisi*
the increased regulatory complexity and prevalence of post-transcriptional regulation in eul
would suggest that accurate inference of eukaryotic regulatory networks requires additiona
promoter sequences and data sets for transcription-factor binding and chromatin modificat
crowds for robust gene network inference
- We will show in this chapter that there are many ways to express the reverse-modeling of re
as a machine learning problem. However, regardless of the adopted approach, the hardest p
identification of the network structure. This problem when expressed as a combinatorial on
NP-hard ->
- Transcription of DNA into RNA is the first — and often most regulated — step in gene expres
genes are regulated at the transcriptional level, and 5-10% of protein-encoding genes encod
proteins, the interactions between regulated genes and regulatory proteins constitute a cor
transcriptional network or genetic network. -> Biological Networks 2007 Chapter 4
- It is therefore likely that comprehensive mapping of the quantitative genetic interaction net
integration of a number datasets from different screening approaches, similar to the recent
the physical protein-protein interaction (PPI) networks in yeast and human -> **Quantitative**
interactions in yeast - Comparative evaluation and integrative analysis
- Compared to PPI networks, an additional challenge originates from the quantitative nature
interaction datasets; instead of comparing the overlap in binary terms, such as presence or
physical interaction, here we should take into account the full spectrum of genetic interactio
extreme cases of negative interactions (i.e., synthetic sick and lethality) to the positive class
pairs (e.g., masking and suppression subcategories) [2,3,17].

2. SGA relationships -> **Systematic Mapping of Genetic Interaction Networks**
   a. An aggravating (or negative) interaction occurs when a double mutant exhibits a pher
   severe than expected from the phenotypes of the individual mutants. This type of int
   that the gene products function in redundant parallel pathways, and highlights the ro
   molecular network in tolerating genetic variations [4].
   b. An alleviating (or positive) interaction occurs when a double mutant exhibits a phenot
   severe than expected from the phenotypes of individual mutants. This type of interac
   the gene products operate in concert or in series within the same pathway [4].

3. Genome-scale, quantitative analysis of genetic interactions. We consider a digenic interactio
mutant that shows a significant deviation in fitness compared with the expected multiplicati

methods, but the
ed better than
e more densely
des. -> **Wisdom of**

tion factors and
siae. Although the
wn regulatory
iae is likely due to
karyotes, which
l inputs, such as
ion7. -> **Wisdom of**

egulatory networks
point to solve is the
e is known to be

ssion. As most
de regulatory
mplex web called

tworks will require
efforts to complete
maps of genetic

of the genetic
absence of a
ons, ranging from
es of interacting

notype that is more
eraction indicates
bustness of the

type that is less
ction indicates that

on as a double
ive effect of

mutant that shows a significant deviation in fitness compared with the expected multiplicat
combining two single mutants (6). Negative interactions refer to a more severe fitness defec
with the extreme case being synthetic lethality; positive interactions refer to double mutant
fitness defect than expected. To quantitatively score genetic interactions in large-scale SGA
developed a mod- el to estimate fitness defects directly from double- mutant colony sizes ->
Landscape of a Cell

4. There have been some attempts to investigate the temporal properties for individual protei
interactions by integrating PPI data with time-course gene expression data [21-29]. <- Detec
protein complexes from dynamic protein-protein interaction networks

Centralities and node influence

- centrality is a quantitative measure that aims at revealing the importance of a node. -> **Axi**
- degree of a vertex alone, as a specific centrality measure, is not sufficient to distinguish leth
from viable ones (Wuchty (2002)), that in protein networks there is no relation between net
and robustness against amino- acid substitutions (Hahn et al. (2004)), and that for biologica
several centrality measures have to be considered (Wuchty and Stadler (2003); Koschützki a
(2004)) -> **Centrality analysis methods for biological networks and their application to gen
networks**
- Topological features of the protein networks have been demonstrated to reflect the functio
interacting genes. For example, essential genes in yeast tend to be well connected and glob
protein network (Jeong et al., 2001; Wuchty and Almaas, 2005). Furthermore, globally cente
more likely to be well conserved and serve as an evolutionary backbone for the network (W
2005). -> **Interactome-transcriptome analysis reveals the high centrality of genes differen
lung cancer tissues**
- Integration of protein network data may extend the reach of the established method of ana
the genes in a broader context. Based on this notion, we seek to reveal the biological signifi
differentially expressed genes in squamous cell lung cancer that is identified through our re
gene expression profiling study by using interactome-transcriptome analysis. We find high c
differentially induced genes, but not for the genes that are suppressed in cancer. -> **Interac
analysis reveals the high centrality of genes differentially expressed in lung cancer tissues**
- k-core analysis has been performed on the yeast essential genes and were shown to be glob
non-essential genes were not (Wuchty and Almaas, 2005). The study also indicates that thes
conserved throughout different species. -> **Interactome-transcriptome analysis reveals the
genes differentially expressed in lung cancer tissues**
- e problem of choosing initial nodes as source spreaders to achieve maximum scale of spread
*vence maximization problem*13. Our research focuses on the strategy of choosing set of s

ct than expected,
ts with a less severe
screens, we
> **The Genetic**

ins and protein
cting temporal

oms of centrality
al proteins clearly
twork connectivity
l network analysis
and Schreiber
e regulatory

nality of the
ally centered in the
ered interactions are
uchty and Almaas,
tially expressed in

lysis by considering
cance of
cent microarray
centrality in these
tome-transcriptome

bal hubs, whereas
se global hubs are
high centrality of

ding is de ned as *in*

*uence maximization problem*12. Our research focuses on the strategy of choosing a set of cr source spreaders in this report. -> **Identifying a set of influential spreaders in complex netw**

- Classical centrality measures in complex networks –like the degree or number of neighbors with4,5, betweenness centrality8 counting the number of shortest paths through a certain n centrality9 based on the idea that relations with more influential neighbors confer greater in shell decomposition10 that correlates with the outcome of supercritical spreading originatin nodes11–13 –rely only on topology, even if an underlying process can be indirectly associat contrast, the impact of individual elements in the global performance of the system inevitab specificities of the dynamics. -> **A measure of individual role in collective dynamics**
- Topological analysis of the yeast protein interaction network shows that the means of degre coefficient (0.16), and betweenness centrality (0.001) of essential genes in yeast are about t those in nonessential genes -> **Predicting essential genes based on network and sequence**
- many bottlenecks also tend to be hubs). Therefore, we further investigate which one of thes a better predictor of protein essentiality in both regulatory and interaction networks. **-> The Bottlenecks in Protein Networks/ Correlation with Gene Essentiality and Expression Dyna**
- we observed that bottlenecks (both nonhub–bottlenecks and hub–bottlenecks) have a stron products of essential genes, whereas hub– nonbottlenecks are surprisingly not essential. Th that it is the betweenness that is a stronger determinant of the essentiality of a protein in th network, not the degree **-> The Importance of Bottlenecks in Protein Networks/ Correlatio Essentiality and Expression Dynamics**
- An intriguing question in the analysis of biological networks is whether biological characteris such as essentiality, can be explained by its placement in the network, i.e., whether topolog implies biological importance. One of the first connections between the two in the context o interaction network, the so-called centrality-lethality rule, was observed by Jeong and collea **Hubs in the Yeast Protein Interaction Network Tend To Be Essential/ Reexamining the Con the Network Topology and Essentiality**
- Since then the correlation between degree and essentiality was confirmed by other studies recently there was no systematic attempt to examine the reasons for this correlation **->Why Yeast Protein Interaction Network Tend To Be Essential/ Reexamining the Connection bet Topology and Essentiality**
- We found that a significant portion of 2,144 mouse genes with yeast orthologs changed the between mouse and yeast (Fig. 1a). We arranged the orthologous pairs of yeast and mouse phenotypic groups based on their changing essentiality patterns. We found 91 genes are ess and mouse (E2E), 246 genes are nonessential in yeast but essential in mouse (N2E), 659 gen yeast but nonessential in mouse (E2N), and 1,149 genes are nonessential in both yeast and Network rewiring is an important mechanism of gene essentiality change
- currently three main types of experimental strategies for the genome-wide discovery of ess knockout (Giaever et al., 2002; Chen et al., 2015), gene knockdown (Harborth et al., 2001; J Roemer et al., 2003) and transposon mutagenesis (Gallagher et al., 2007; Langridge et al., 2 methods can generate accurate collections of essential genes, but they are expensive, time- laborious. Furthermore, these experimental methods are not suitable for some complex org for humans -> **Predicting Essential Genes and Proteins Based on Machine Learning and Ne**

itical nodes as
works
a node interacts
node, eigenvector
mportance, or the k-
ng in specific
ted in some cases. In
ply depends on the

ee (11.55), clustering
twice as large as
analysis #
se two quantities is
e Importance of
mics
ng tendency to be
us, we determined
he regulatory
on with Gene

stics of a protein,
gical prominence
of a protein
agues **->Why Do**
nnection between

[4–7], but until
y Do Hubs in the
tween the Network

ir essentialities
genes into four
sential in both yeast
nes are essential in
mouse (N2N) ->

ential genes: gene
i et al., 2001;
009). These
-consuming and
ganisms, especially
etwork Topological

for humans. -> **Predicting Essential Genes and Proteins Based on Machine Learning and Ne**
Features/ A Comprehensive Review

- Another observation in this search of the literature is that of the 28 articles reporting the ut topological features to predict essential genes and proteins, in 12 (43%) the computational those based on machine learning, a method in which computers make and improve predicti data through learning algorithms (see next sections for details). On the other hand, of the 1 the utilization of machine learning for the prediction of essential genes and proteins, 12 (86 utilization of network topological features as learning attributes (features that describe a ce details in the next section). Therefore, this brief analysis of these 34 papers suggests a stron machine learning and network topological features regarding the prediction of essential ger **Predicting Essential Genes and Proteins Based on Machine Learning and Network Topolog Comprehensive Review**
- despite the great progress that has been made in the machine learning-based prediction of using network topological features since the publication of the pioneering study in the field **Table1**), it is worth to mention that all these models are predictive of "constitutive" essentia genes that are essential regardless the growth condition. However, many genes considered certain growth condition might not play as critical role in another condition (Tong et al., 200 2011). These are the so-called conditionally essential genes, i.e., nonessential genes that be depending on the environment conditions. The development of machine learning approach of these conditionally essential genes using network topological features will be the main ch future considering that the network topological characteristics of these genes remain unexp essentiality VS conditional essentiality -> **Predicting Essential Genes and Proteins Based on and Network Topological Features/ A Comprehensive Review**
- Synthetic lethality : The situation in which two genes that are non-essential when individual lethality when they are combined as a double mutant. -> **Exploring genetic interactions and yeast**
- we estimate that a global network will contain ~200,000 synthetic-lethal interactions. To pu context, there are ~1,000 essential genes in yeast, for which a single mutation leads to a leth there are 200-fold more ways to generate a similar phenotype through a digenic synthetic-l This finding indicates that digenic interactions might underlie many inherited phenotypes, a why the analytical power of single-gene effects on many phenotypes has been so limited. -> **interactions and networks with yeast**
- Moreover, our approach of focusing on nonhub–bottle- necks is useful for finding proteins t different processes and are involved in cross-talk. -> The Importance of Bottlenecks in Prote Correlation with Gene Essentiality and Expression Dynamics
- This indicates the existence of a large number of nodes with high betweenness but low conn nodes). Importantly, such nodes are absent in computer- generated, random scale-free netw mention of bottlenecks: **High-Betweenness Proteins in the Yeast Protein Interaction Netwo**

twork Topological

ilization of network
methods used were
ons based on some
4 articles reporting
%) report the
rtain instance; see
g link between
nes and proteins. ->
gical Features/ A

essential genes
(Chen and Xu, 2005;
al genes, that is,
as essential under
04; Nichols et al.,
come essential
es for the prediction
hallenge in the near
lored. ## global
Machine Learning

lly mutated cause
l networks with

t this number in
hal phenotype, but
ethal interaction.
nd begins to explain
Exploring genetic

that mediate
ein Networks/

nectivity (HBLC
works ->First
ork 2005

Evaluation of results

- gene enrichment ## **Gene Ontology: tool for the unification of biology**
- New experiments to test new interactions that appeared from reverse engineering of netwo
    - Even a comparatively small subnetwork of this size is still a challenge to visualize insig authors assess its quality in two ways. **First**, *they determine whether the genes of the enriched for specific cellular process categories in the Gene Ontology database, which* and this is a wonderful strength of the paper—the authors experimentally validate so neighbors of *MYC* -> **Reverse engineering gene regulatory network**
    - Basso *et al.* demonstrate that as long as the available data explore a wide range in the of the system, biologically meaningful interactions can be recovered by computationa **Reverse engineering gene regulatory network** for the paper -> Reverse engineering o networks in human B cells
    - We note that these data support a direct regulatory effect of the tested transcription gene, but chromatin immunoprecipitation experiments would be required to deter- m binding. -> **Wisdom of crowds for robust gene network inference**
- Gene essentiality data of yeast were manually compiled from the Comprehensive Yeast Gen (http://mips.helmholtz-muenchen.de/genre/proj/yeast/) and large-scale experiments26. Th 1,178 essential and 4,904 nonessential yeast genes. -> Network rewiring is an important me essentiality change
- [currently, the available essential genes and protein databases are DEG (Zhang and Lin, 200 2013), OGEE (Chen et al., 2012), and EGGS (http://www.nmpdr.org/FIG/eggs.cgi)]. These da researchers to explore the features of essential genes and proteins and, through this explora features are associated with essentiality and, finally, develop computational methods propo essential genes and proteins. -> **Predicting Essential Genes and Proteins Based on Machine Network Topological Features/ A Comprehensive Review**
- There exist many different databases of a certain type of interaction (e.g., DIP, BioGRID and usually, these databases are regularly updated. Different databases or newer versions of a g have different sets of interactions that, in turn, will give rise to new networks with distinct s consequently, different values of network topological features. As an example, we can cite t Hwang et al. (2009) and Acencio and Lemke (2009). In both studies, PINs of *Saccharomyces* created; however, the interactions of the PIN constructed in the study by Hwang et al. (200 from the version ScereCR20070107 of DIP and the interactions of the PIN constructed in the and Lemke (2009) were gathered from the version 2.0.42 of the BioGRID database. Therefor performances of the models created by these authors cannot be reliably compared. ## The must be the same kind and the same data! That means not only the same database because regularly, but the exact same datasets.
- Thus, it seems that only network topological features are not enough to distinguish essentia genes and proteins. This raises the following question: **is the positive correlation between network topological features only an artifact of a possible bias (essential genes and protei**

orks!! e.g
htfully, so the
subnetwork are
n they are. **Second**—
me of the first

e 'expression space'
al algorithms ->
of regulatory

factor on the target
hine physical

nome Database
he dataset contained
echanism of gene

9), CEG (Ye et al.,
ata have enabled
ation, reveal which
osed to identify
e Learning and

IntAct for PPIs) and,
given database will
tructures and,
the studies by
cerevisiae were
9) were collected
e study by Acencio
re, the prediction
reference networks
e they are updated

al from non-essential
essentiality and
ins are the focus of

more studies and therefore tend to have higher values of network topological) present in derived from small scale experiments?

- Regardless the resolution of this debate, a large-scale study for evaluating how well essentia proteins can be predicted solely by network topological features is necessary to confirm this prediction performance.
- ROC curves in R http://blog.revolutionanalytics.com/2016/11/calculating-auc.html
- Dygraphs for r Interactive plots for HTML files

the networks mainly

al genes and
s moderate