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	Refer ce
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	

en	Assumptions	Comment

SUPPLEMENTARY INFORMATION

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

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], S MINT[13], DIP[14], HPRD[15], MIPS 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], Dro BioGRID[9], TcoF-DB[24], BIPA[25], EDGEdb[27], NPIDB[28,29]
	Protein-RNA	RNA electro-mobility shift (RNA-EMSA)[30], RNA-pull down[31]	PRID[32], BIPA[25] NPInter[33], RBP 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 HMDB[43,44], EcoCyc[45], HumanC 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], Drug ChemProt[56], STITCH[57], AffinD MatrixDB[59], PSMDB[60], PDB-L ChEMBL[62], ConsensusPathDB
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 ConsensusPathDB[47]
	Gene-Disease	Literature curation, clinical and sequence information	OMIM[69], HuDiNe[70], Diseasom
Omics data type		**Detection methodologies**	**Databases**
Transcriptomics		Microarray[72], RNASeq[73]	GEO[74], SMD[75], TCGA[76], GXD ONCOMINE[78], ArrayExpress[79]
RNAi (phenomics)		RNAi interference assay[80]	RNAiDB[81], GenomeRNAi[82], si

OCTOBER 2013)

TRING[12], -MPPI[16],
biD[18], , hPDI[26],
PDB[34], PRIDB[35],
and MetaCyc[42], Cyc[46],
Bank[55] B[58], igand[61], [47]
[68], DroID[18],
e[71]
[77], ,
Records[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 I 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

- GOAL: develop a complete map of biological modules underlying cellular architecture and
  Integrative approaches for finding modular structure in biological networks.

		Functional relations

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

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

Newman networks an

)

d function ->

bMeth[87], MethyCancer[90]
GWAS Central
World-2DPage[100],
[103], PHOSIDA[104]

- Levels of biological - molecular networks -> **Genetic interaction networks/ better unders**
    - For a computer, the lower level of abstraction would contain details on the hard- w level will represent the logic of the program. In agreement with this approach, a sy con- sider a biological system with all its complexity and identify, from the genomic phenotype, different levels of abstractions. At the lower level of this conceptual str several networks representing the physical structure and organization of the genor nodes could be genes/coding sequences, single-nucleotide polymorphisms (SNPs) linked by edges representing their physical proximity and organization within chror homology etc. (**Figure 2**, level I). The second level of abstraction would represent th genome into physical components: proteins and RNA. Edges between these elemer they are co-expressed in different contexts or that their expression profiles throug experimental conditions are highly correlated (**Figure 2**, level II; Ge et al., 2003; Vid third level of abstraction would represent physical interactions between different e protein (PPI), protein-DNA (PDI) or protein-RNA (PRI) interactions (**Figure 2**, level II The fourth level of abstraction will allow the visualization of the functional relation physical elements. This level would contain GI net- works, signaling and metabolic level IV). The fifth level would represent biological processes. This level would cont proteins implicated in the same biological process would be linked by an edge (**Figu** and last level of abstraction would represent phenotypes and show the relationshi associated with similar phenotypes and diseases (**Figure 2**, level VI).
- Metabolic networks convert raw materials from the environment into value-added produ dispose of intracellular materials. Regulatory networks alter the output of the genomic p term regulatory network only encompasses transcriptional and translational regulation; h regulatory features associated with all cellular processes. Signalling networks are the com that connect cellular components and coordinate the phenotypic response to perturbatio networks simultaneously monitor a wide variety of external and internal parameters, incl and DNA damage. The observations are continuously processed in an integrated fashion, metabolic outputs are modulated to deliver the appropriate response. For simplicity's sa on the molecular components of a cell and their interactions — intercellular interactions 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. Th many levels, ranging from chromatin remodelling to post translational modifications [60] understand gene regulation, a large and growing body of transcriptomic data can be used between genes, and has prompted the development of a number of gene regulatory netw reconstruction algorithms, including Graphical Gaussian models, Bayesian networks and **Information theory and signal transduction systems/ From molecular information proc inference**
- the aims of large-scale network inference, which are typically to highlight and explore de sets—and thereby help to generate new hypotheses worthy of further investigation—rat single grand unifying model of the system. -> **Systems biology (un)certainties**
- Models are simplified (but not simplistic) representations of real systems, and this is 8pre

stand to better predict
ware while the higher
stems biologist will
c sequence to the
ructure, we would find
me. In these networks,
or coding sequences
mosomes, their
he expression of that
nts would indicate that
hout multiple
dal et al., 2011). The
elements – protein–
I; Vidal et al., 2011).
ships linking these
pathways (**Figure 2**,
ain networks where
ure 2, level V). The sixth
ps between elements

ucts and recycle or
rogramme. Here, the
however, there are
mmunication networks
ons. Signalling
luding nutrient levels
and the genomic and
ke, this Review focuses
are treated as

his regulation occurs at
. In order better to
d to infer interactions
work (GRN)
relevance networks ##
essing to network

pendencies in data
ther than to uncover a

ecisely the property

that makes them attractive to explore the consequences of our assumptions, and to iden
understanding of the principles governing a biological system. Models are tools to uncov
cannot be directly observed, akin to microscopes or nuclear magnetic resonance machine
interpreted appropriately, with due attention paid to inherent uncertainties, the mathem
computational modeling of biological systems allows the exploration of hypotheses. But t
models depends on the ability to assess, communicate, and, ultimately, under- stand thei
Systems biology (un)certainties

- For some model organisms, however, protein interaction data covers 20% of the proteins
organism (ignoring multiple isoforms due to alternative splicing, etc.). This observation po
and important question of just how representative a random subnet is for the global netw
scale-free networks are not scale-free: Sampling properties of networks
- We emphasize that the degree distribution alone does not suffice to characterize a netwo
networks, e.g., some with many cross-connections (loops) and others with ''tree-like'' for
have the same degree distribution. -> **Subnets of scale-free networks are not scale-free:**
of networks
- Essential genes are those genes critical for cell viability under certain contexts -> **Networ**
essentiality in functional genomics experiments
- However, negative feedback plays an important role in biological robustness and evolutio
played an equally prominent role in the origins of life. <- **Prebiotic network evolution/ si**
- **A possible resolution to the instabilities that arise in net- work expansion and evolution**
feedback, which enables growth, with negative feedback, which contributes to robustn
2c). <- **Prebiotic network evolution/ six key parameters**
- This reverse engineering is extremely difficult. Although an electrical engineer could desig
circuits that would amplify signals, he would find it difficult to deduce the circuit diagram
amplifier by correlating its outputs with its inputs. It is thus unlikely that we can deduce t
level description of a module solely from genome- wide information about gene expressio
interactions between proteins. Solving this problem is likely to require additional types of
finding general principles that govern the structure and function of modules ## **From mo**
biology
- But techniques for collecting information about the entire genome will be only as powerf
available to analyse it, just as our ability to infer protein structure and function from prot
increased with the sophistication of tools for sequence analysis. ## **From molecular to m**
- Cell functioning requires the tight linking between multiple regulatory systems that inclu
controlling initiation of gene transcription, RNA splicing, mRNA transport, translation initi
translational protein modifications, or the degradation of mRNA/protein -> **Michael Stum**
Mark Girolami (editors)-Handbook of Statistical Systems Biology ## Chapter 12
- main obstacles is the difficulty of choosing the appropriate experimental data, together w
computational approaches. -> **Handbook of Statistical Systems Biology ## Chapter 12**
- the goal for methodology to infer regulatory networks consists in modelling and recoveri
interactions such as 'protein i activates (or inhibits) the transcription of gene j'. -> **Handb**
Systems Biology ## Chapter 12
- From the biological side, it is not that clear that the identification of regulatory networks

tity where we lack
er mechanisms that
es (*15*). Used and
natical and
the relevance of these
ir uncertainties. ->

s known to exist in that
oses the interesting
work -> **Subnets of**

ork: very different
rm (no loops at all), can
: Sampling properties

k analysis of gene

on52 and may have
x key parameters
n is to balance positive
ness and stability (Fig.

gn many different
of an unknown
he circuity or a higher-
on and physical
f information and
lecular to modular cell

ful as the tools
ein sequence data has
odular cell biology
de mechanisms for
iation, post-
mpf, David J. Balding,

with the appropriate

ng regulatory
ook of Statistical

is possible only using

- From the biological side, it is not that clear that the identification of regulatory networks concentrations of mRNA due to the role of additional sources of regulation such as micro **Networks 2007 ## chapter 3**
- What are the limits of the definition of a regulatory protein? We are interested here in pr subset of the genes of an organism. Other proteins also called regulatory play a more ger transcription (for example the eukaryotic transcription factor of type II). In principle, thes polymerase itself, are essential for the transcription of all genes that encode proteins. Ho action is non-specific, they are not covered in this Section ## RNA polymerase is necessar 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 catalyz *are not suitable when it comes to expressing interactions that involve more than two part fashion. The formalism of hypergraphs allows to express non-binary interactions and cou context*. **-> Biological Networks 2007 ## chapter 4**
- because high-throughput datasets can be filled with technical and biological noise and ha biases and coverage, improved statistics are needed to distinguish signal from noise, as w integration to annotate the biologically relevant relationships; because the logical interpr network is not easily comprehensible to the human brain, computational modeling is nee output from the signal input or system perturbation; because no analysis and modeling m detailed targeted biological experiments are needed to validate mod- els before a hypoth approach health and medical problems -> **Understanding biological functions through m** Problems of the high throuput data!!
- Small-scale gene-centric studies have delineated many valuable genetic and biochemical genes and pathways. This information forms the skeleton of the entire complete network annotation standards. For now, such information covers only a tiny fraction of the full ne towards certain biological functions. For example, protein interactions in the literature cu (Human Protein Reference Database) [14] cover only 20 000 interactions out of a con- se 000 interactions and are strongly biased for cancer-related processes (Supplementary in- Large-scale experiments and data mining are obviously more effective for mapping genet **Understanding biological functions through molecular networks** ## Problems of the sing approach!!
- However, the change to network biology does not simply entail hand- ing over biology to mathematicians. A good understanding of biology is needed to ask the right questions, to network analysis tools, and to con rm analysis results by solid experimentation. After all, biology. The fundamental goal of network biology is the same as molecular biology: to un biological processes and the mechanisms of human diseases. -> **Understanding biologic molecular networks**
- Furthermore, yeast two-hybrid assays produce many false-positive outcomes, and the cu pathway maps may be heavily biased towards connection to functionally important gene have been popular targets for research -> **Kitano, Hiroaki Computational systems biolog**
- despite the known poor overall correlation between mRNAs and their protein products, t assumption that differentially expressed mRNAs impact their respective experimental co

is possible only using
RNA. **-> Biological**

roteins that regulate a
neral role in initiating
se generalists, like RNA
owever, since their
y for all transcription

!!), but some proteins
ze reactions!!! *Graphs*
tners in an obligate
ld be used in the above

ave different technical
well as better data
retation of the whole
eded to predict the
method is perfect, more
nesis can be used to
olecular networks ##

relationships between
k and can serve as
twork and is biased
urated by HPRD
rvative estimate of 150
formation, table S1).
tic networks ->
gle gene - process

physicists and
o choose proper
network biology is
nderstand basic
al functions through

rrent hand-crafted
s simply because these
y.pdf
there is an implicit
nditions via differences

assumption that differentially expressed mRNAs impact their respective experimental co
in proteins. However, to the best of our knowledge, this assumption has never been expli
Relationship between di erentially expressed mRNA and mRNA- protein correlations in system

- Although clustering analysis provides insight into the "correlation" among genes and biol does not reveal the "causality" of regulatory relationships. Several methods have been pr automatically discover regulatory relationships solely on the basis of microarray data (*7*– methods use information derived from mRNA abundance, so there is limited scope to inf transcriptional regulation. Posttranscriptional and posttranslational mechanisms of regul incorporated as large-scale data become available, but many properties have yet to be m accuracy or in high throughput. -> **Systems Biology: A Brief Overview**
- There is a need for systematic approaches to infer causal relationships between interacti we refer to the "direction" (edge direction), "sign" (activation/inhibition) and "mode" (e.g ubiquitination) of signal flow in PPI networks. **-> Integrating protein-protein interaction phenotypes reveals signs of interactions**
- The signed network opens up new scope for network analysis, such as the application of theory. Further integration of other information-flow properties such as edge-direction w sophisticated flow-based network analysis. **-> Integrating protein-protein interaction ne phenotypes reveals signs of interactions**
- **Integrating protein-protein interaction networks with phenotypes reveals signs of inte**
    - Each RNAi screen identifies positive and negative regulators of a particular ph
        - How?
    - PPI is for the physical interactions, from gene expression one can reconstruct differential expression. **Direct regulation relationships must include both ph activation - repression data.**
- Biological-network reconstructions provide us with a framework that allows us to identify whole and, subsequently, understand the relationship between the molecular phenotype 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 given functional module to influence and even coordinate multiple cellular processes. It is interactional complexity, that single genes rarely specify a phenotype in its entirety. The genetic network are now apparent, but a compel- ling case can be made for a deeper and exploration of this model system as an exemplar for more complex eukaryotes. -> **Explor interactions and networks with yeast**
- The physical-interaction map, generated by large-scale two-hybrid51,52 or affinity purific mass spectrometry identification26,43,53,54, provides a view of the gene products that a protein complexes and function together as biochemical machines. Rather than physical genetic-interaction map provides functional information, largely identifying gene product functionally related pathways. Although genetic interactions overlap with protein–protei often than expected by chance, such overlap is relatively rare, occurring at a frequency of 2). -> **Exploring genetic interactions and networks with yeast**
- Large-scale genetic analyses reveal that mutations in most eukaryotic genes have little d

nditions via differences
icitly tested. ->
a xenograft model

ogical phenomena, it
roposed to
9). At present, such
er causality based on
ation must be
easured with sufficient

ng proteins, by which
g. phosphorylation,
networks with

structure balance
would enable
etworks with

ractions
henotype ?

t a network based on
ysical interactions and

y the context of the
e and the general

s extend beyond a
s no wonder, given this
outlines of a yeast
d more complete
ring genetic

cation followed by
assemble into soluble
information, the
ts that operate in
n interactions more
f less than 1% (REF.

iscernable effect. For

example, systematic gene deletion in *S. cerevisiae*, discussed in detail below, produced a only ~20% of yeast genes are essential for viability when deleted individually in haploids laboratory conditions

- Comparisons between aggravating and alleviating effects revealed that, for most functio tions were either largely aggravating or largely alleviat- ing, but not mixed, an asymmetri termed 'monochromatic'.
- the set of inter- actions that are observed for a particular query gene can be suggestive o position of a gene in a genetic-interaction network being highly predictive of its molecula **genetic interactions and networks with yeast**
- **Because most genetic interactions do not overlap with physical interactions, the two ty said to be largely orthogonal**
- Integration attempts in yeast, combining physical protein- protein interaction maps with revealed that interacting proteins are more likely to be encoded by genes with similar ex noninteracting proteins (Ge et al., 2001; Grigoriev, 2001; Jansen et al., 2002; Kemmeren observations were subsequently confirmed in many other organisms (Ge et al., 2003). Be mental aspect of finding significant overlaps between interaction edges in interactome n coexpression edges in tran- scription profiling networks, these observations have been u overall biological significance of interactome datasets -> Interactome Networks and Hum
- Experimental artefacts, variability in cover- age across data sets, sampling bias towards w limitations in screening power and inher- ent sensitivities in various assays can yield false negatives in interaction data -> **Integrative approaches for finding modular structure in**
- In addition, as reported by Zotenko et al. [17], essential proteins tend to form highly con than function independently -> **A new essential protein discovery method based on the protein interaction and gene expression data**
- However, stand- ard clustering is not ideal for PPI networks: proteins may have multiple therefore the corresponding nodes may belong to more than one cluster; for example, 2 the CYC2008 hand-curated yeast complex data set3 participate in more than one comple overlapping protein complexes in protein-protein interaction networks
- The existence of a GI between two genes does not necessarily imply that these two gene proteins or that the two genes are even expressed in the same cell. In fact, a GI only impl share a functional relationship. These two genes may be involved in the same biological they may also be involved in compensatory pathways with unrelated apparent function networks: better understand to better predict
- A network in the left represents a global PIN. A eukaryotic cell in the center can be divide compartments, Endoplasmic, Cytoskeleton, Golgi, Cytosol, Lysosome(or Vacuole), Mitoch Plasma, Nucleus, Peroxisome and Extracellular, where Lysosome only exists in animal cel compartment, a PSLIN of this compartment is constituted by the proteins localized in this their interactions. With the subcellular localization information of proteins, the PSLINs ca mapping the global PIN to each compartment separately. -> **Rechecking the Centrality-L Scope of Protein Subcellular Localization Interaction Networks**

remarkable result:
growing in standard

nal groups, interac-
cal feature that they

f its function, with the
r role. -> **Exploring**

pes of interaction are

coexpression profiles,
pression profiles than
et al., 2002). These
eyond the funda-
etworks and
sed to estimate the
an Disease
ell-studied processes,
posi- tives and false
biological networks.
nected clusters rather
integration of protein-

functions, and
07 of 1,628 proteins in
x. -> detecting

s code for interacting
ies that the two genes
process or pathway; or
> Genetic interaction

ed into 11
nondrion, Endosome,
ls. For each
s compartment and
n be generated by
Lethality Rule in the

Network construction methods and evaluatio

- Information theoretic approaches to GRN reconstruction have two major strengths. The f information is able to capture nonlinear associations between variables, a feature seen in Second, it has been demonstrated that the use of the data processing inequality (DPI), w noisy system X → Y → Z, knowledge of Z cannot give more information about X than Y car in distinguishing two genes regulated by a third from a trio of co-regulating genes. In reco GRNs, the combined approach using mutual information and DPI outperforms Bayesian a techniques in the precision and recall of direct regulatory links ## **Information theory and systems/ From molecular information processing to network inference**
- Several alternative network inference approaches exist (*4*, *5*) that provide better, more ro networks, and can incorporate expert or domain knowledge. These methods may be used regulatory relationships, and how interactions between genes change over time or differ and healthy controls. -> **Systems biology (un)certainties**
- Genome-scale inference of transcriptional gene regulation has become possible with the throughput technologies such as microarrays and RNA sequencing, as they provide snaps transcriptome under many tested experimental conditions. From these data, the challeng computationally predict direct regulatory interactions between a transcription factor and aggregate of all predicted interactions comprises the gene regulatory network. -> **Wisdo robust gene network inference**
- Understanding the advantages and limitations of different network inference methods is effective application in a given biological context -> **Wisdom of crowds for robust gene n**
- We conclude that there is no category of network inference methods that is inherently su 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 conne (network motifs), and we observed characteristic trends for different method categories feed-forward loops were recovered most reliably by mutual-information and correlation- whereas sparse-regression and Bayesian-network methods performed worse at this task. that the latter approaches preferentially select regulators that independently contribute target genes. However, the assumption of independence is violated for genes regulated b

n

first is that mutual
n expression data.
hich states that for a
n give about X [19,65],
onstructing simulated
nd relevance networks
d signal transduction

obust candidate
d to investigate gene-
between disease cases

advent of high-
shots of the
ge is to
d its target genes; the
m of crowds for

critical for their
etwork inference
uperior and that
-> **Wisdom of crowds**

ectivity patterns
(**Fig. 2c**). For example,
-based methods,
. The reason for this is
to the expression of
by mutually dependent

transcription factors, as in the case of feed-forward loops. Indeed, linear cascades were r predicted by regression and Bayesian-network methods. This shows that current method off between performance on cascades and performance on feed-forward loops. -> **Wisd robust gene network inference**

- Network inference methods have complementary advantages and limitations under diffe suggests that combining the results of multiple inference methods could be a good strate predictions. -> **Wisdom of crowds for robust gene network inference**
- After identifying these modules24, we tested them for enrichment of Gene Ontology terr Note 7). Network modules are strongly enriched for very specific biological processes. Th unique functions to most of the identified modules in both networks
- When constructing a compendium of microarrays for global network inference, one shou toward oversampling a narrow set of experimental conditions. -> **Wisdom of crowds for inference ### QUESTION!!!!**
- Bayesian- network methods exhibited below-average performance in this challenge, likel heuristic searches, which are often too costly for systematic data resampling and may be smaller networks. Information theoretic methods performed better than correlation-bas two approaches had similar biases in predicting regulatory relationships. They also perfo regression and Bayesian- network methods on feed-forward loops, fan-ins and fan-outs (connected parts of the network), but they had an increased rate of false positives for cas **crowds for robust gene network inference**
- A fundamental assumption of network inference algorithms is that mRNA levels of transc their targets tend to be correlated; we found that this is true for *E. coli*, but not for *S. cere* lower coverage of *S. cerevisiae* gold standards may also play a role (*E. coli* has the best-kn network of any free-living organism16), the poor correlation at the mRNA level in *S. cerev* the increased regulatory complexity and prevalence of post-transcriptional regulation in would suggest that accurate inference of eukaryotic regulatory networks requires additio promoter sequences and data sets for transcription-factor binding and chromatin modifi **crowds for robust gene network inference**
- We will show in this chapter that there are many ways to express the reverse-modeling o as a machine learning problem. However, regardless of the adopted approach, the harde identification of the network structure. This problem when expressed as a combinatorial NP-hard ->
- Transcription of DNA into RNA is the first — and often most regulated — step in gene exp genes are regulated at the transcriptional level, and 5-10% of protein-encoding genes en proteins, the interactions between regulated genes and regulatory proteins constitute a transcriptional network or genetic network. -> Biological Networks 2007 Chapter 4
- It is therefore likely that comprehensive mapping of the quantitative genetic interaction integration of a number datasets from different screening approaches, similar to the rece the physical protein-protein interaction (PPI) networks in yeast and human -> **Quantitati interactions in yeast - Comparative evaluation and integrative analysis**
- Compared to PPI networks, an additional challenge originates from the quantitative natu

more accurately
ls experience a trade-
om of crowds for

erent contexts, which
egy for improving

ms (Supplementary
is allowed us to assign

uld thus avoid any bias
robust gene network

y because they use
better suited for
ed methods, but the
rmed better than
the more densely
cades. -> **Wisdom of**

cription factors and
evisiae. Although the
nown regulatory
visiae is likely due to
eukaryotes, which
onal inputs, such as
cation7. -> **Wisdom of**

f regulatory networks
st point to solve is the
one is known to be

pression. As most
code regulatory
complex web called

networks will require
ent efforts to complete
ve maps of genetic

re of the genetic

interaction datasets; instead of comparing the overlap in binary terms, such as presence
physical interaction, here we should take into account the full spectrum of genetic intera
extreme cases of negative interactions (i.e., synthetic sick and lethality) to the positive cl
pairs (e.g., masking and suppression subcategories) [2,3,17].

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

2. Genome-scale, quantitative analysis of genetic interactions. We consider a digenic intera
mutant that shows a significant deviation in fitness compared with the expected multiplic
combining two single mutants (6). Negative interactions refer to a more severe fitness de
with the extreme case being synthetic lethality; positive interactions refer to double mut
fitness defect than expected. To quantitatively score genetic interactions in large-scale S
developed a mod- el to estimate fitness defects directly from double- mutant colony size
Landscape of a Cell
3. There have been some attempts to investigate the temporal properties for individual pro
interactions by integrating PPI data with time-course gene expression data [21-29]. <- De
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. -> A
- degree of a vertex alone, as a specific centrality measure, is not sufficient to distinguish le
from viable ones (Wuchty (2002)), that in protein networks there is no relation between
and robustness against amino- acid substitutions (Hahn et al. (2004)), and that for biolog
several centrality measures have to be considered (Wuchty and Stadler (2003); Koschütz
(2004)) -> **Centrality analysis methods for biological networks and their application to g
networks**
- Topological features of the protein networks have been demonstrated to reflect the func
interacting genes. For example, essential genes in yeast tend to be well connected and gl
protein network (Jeong et al., 2001; Wuchty and Almaas, 2005). Furthermore, globally co

or absence of a
ctions, ranging from
asses of interacting

henotype that is more
interaction indicates
robustness of the

notype that is less
raction indicates that

ction as a double
cative effect of
efect than expected,
ants with a less severe
GA screens, we
s -> **The Genetic**

oteins and protein
tecting temporal

Axioms of centrality
ethal proteins clearly
network connectivity
ical network analysis
ki and Schreiber
gene regulatory

ctionality of the
obally centered in the
ntered interactions are

protein network (Jeong et al., 2001; Wuchty and Almaas, 2005). Furthermore, globally ce
more likely to be well conserved and serve as an evolutionary backbone for the network
2005). -> **Interactome-transcriptome analysis reveals the high centrality of genes differ
lung cancer tissues**

- Integration of protein network data may extend the reach of the established method of a
the genes in a broader context. Based on this notion, we seek to reveal the biological sign
differentially expressed genes in squamous cell lung cancer that is identified through our
gene expression profiling study by using interactome-transcriptome analysis. We find hig
differentially induced genes, but not for the genes that are suppressed in cancer. -> **Inter
analysis reveals the high centrality of genes differentially expressed in lung cancer tissu**
- k-core analysis has been performed on the yeast essential genes and were shown to be g
non-essential genes were not (Wuchty and Almaas, 2005). The study also indicates that t
conserved throughout different species. -> **Interactome-transcriptome analysis reveals t
genes differentially expressed in lung cancer tissues**
- e problem of choosing initial nodes as source spreaders to achieve maximum scale of spr
*uence maximization problem*12. Our research focuses on the strategy of choosing a set o
source spreaders in this report. -> **Identifying a set of influential spreaders in complex n**
- Classical centrality measures in complex networks –like the degree or number of neighbo
with4,5, betweenness centrality8 counting the number of shortest paths through a certai
centrality9 based on the idea that relations with more influential neighbors confer greate
shell decomposition10 that correlates with the outcome of supercritical spreading origina
nodes11–13 – rely only on topology, even if an underlying process can be indirectly assoc
contrast, the impact of individual elements in the global performance of the system inevi
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 de
coefficient (0.16), and betweenness centrality (0.001) of essential genes in yeast are abou
those in nonessential genes -> **Predicting essential genes based on network and sequen**
- many bottlenecks also tend to be hubs). Therefore, we further investigate which one of t
a better predictor of protein essentiality in both regulatory and interaction networks. **-> T
Bottlenecks in Protein Networks/ Correlation with Gene Essentiality and Expression Dy**
- we observed that bottlenecks (both nonhub–bottlenecks and hub–bottlenecks) have a st
products of essential genes, whereas hub– nonbottlenecks are surprisingly not essential.
that it is the betweenness that is a stronger determinant of the essentiality of a protein in
network, not the degree **-> The Importance of Bottlenecks in Protein Networks/ Correla
Essentiality and Expression Dynamics**
- An intriguing question in the analysis of biological networks is whether biological charact
such as essentiality, can be explained by its placement in the network, i.e., whether topo
implies biological importance. One of the first connections between the two in the contex
interaction network, the so-called centrality-lethality rule, was observed by Jeong and co
**Hubs in the Yeast Protein Interaction Network Tend To Be Essential/ Reexamining the C
the Network Topology and Essentiality**
- Since then the correlation between degree and essentiality was confirmed by other studi

ntered interactions are
(Wuchty and Almaas,
entially expressed in

analysis by considering
nificance of
recent microarray
h centrality in these
ractome-transcriptome
ues
lobal hubs, whereas
hese global hubs are
the high centrality of

eading is de ned as *in*
f critical nodes as
etworks
ors a node interacts
in node, eigenvector
er importance, or the k-
ating in specific
ciated in some cases. In
tably depends on the

gree (11.55), clustering
ut twice as large as
nce analysis #
hese two quantities is
The Importance of
namics
rong tendency to be
Thus, we determined
n the regulatory
ation with Gene

eristics of a protein,
logical prominence
xt of a protein
lleagues **->Why Do**
Connection between

ies [4–7], but until

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 between Topology and Essentiality**

- We found that a significant portion of 2,144 mouse genes with yeast orthologs changed their essentiality between mouse and yeast (Fig. 1a). We arranged the orthologous pairs of yeast and mouse genes into four phenotypic groups based on their changing essentiality patterns. We found 91 genes are essential in both yeast and mouse (E2E), 246 genes are nonessential in yeast but essential in mouse (N2E), 659 genes are essential in yeast but nonessential in mouse (E2N), and 1,149 genes are nonessential in both yeast and mouse. Network rewiring is an important mechanism of gene essentiality change
- currently three main types of experimental strategies for the genome-wide discovery of essential genes: knockout (Giaever et al., 2002; Chen et al., 2015), gene knockdown (Harborth et al., 2001; Boutros and Roemer et al., 2003) and transposon mutagenesis (Gallagher et al., 2007; Langridge et al., 2009). These methods can generate accurate collections of essential genes, but they are expensive, time-consuming and laborious. Furthermore, these experimental methods are not suitable for some complex organisms such as for humans. -> **Predicting Essential Genes and Proteins Based on Machine Learning and Network Topological Features/ A Comprehensive Review**
- Another observation in this search of the literature is that of the 28 articles reporting the use of topological features to predict essential genes and proteins, in 12 (43%) the computational methods were those based on machine learning, a method in which computers make and improve predictions from data through learning algorithms (see next sections for details). On the other hand, of the 34 articles reporting the utilization of machine learning for the prediction of essential genes and proteins, 12 (35%) reported the utilization of network topological features as learning attributes (features that describe a node, as will be detailed in the next section). Therefore, this brief analysis of these 34 papers suggests a strong relationship between machine learning and network topological features regarding the prediction of essential genes and proteins. **Predicting Essential Genes and Proteins Based on Machine Learning and Network Topological Features/ A Comprehensive Review**
- despite the great progress that has been made in the machine learning-based prediction of essential genes using network topological features since the publication of the pioneering study in the field (see **Table1**), it is worth to mention that all these models are predictive of "constitutive" essentiality, i.e., genes that are essential regardless the growth condition. However, many genes considered essential in a certain growth condition might not play as critical role in another condition (Tong et al., 2004; Liu et al., 2011). These are the so-called conditionally essential genes, i.e., nonessential genes that become essential depending on the environment conditions. The development of machine learning approaches for the prediction of these conditionally essential genes using network topological features will be the main challenge in the future considering that the network topological characteristics of these genes remain unexplored. essentiality VS conditional essentiality -> **Predicting Essential Genes and Proteins Based on Machine Learning and Network Topological Features/ A Comprehensive Review**
- Synthetic lethality : The situation in which two genes that are non-essential when individually mutated cause lethality when they are combined as a double mutant. -> **Exploring genetic interactions and essentiality in yeast**
- we estimate that a global network will contain ~200,000 synthetic-lethal interactions. To

Why Do Hubs in the
between the Network

their essentialities
use genes into four
essential in both yeast
genes are essential in
nd mouse (N2N) ->

essential genes: gene
1; Ji et al., 2001;
, 2009). These
me-consuming and
organisms, especially
Network Topological

utilization of network
nal methods used were
ictions based on some
e 14 articles reporting
(86%) report the
certain instance; see
rong link between
genes and proteins. ->
logical Features/ A

of essential genes
eld (Chen and Xu, 2005;
ntial genes, that is,
ed as essential under
2004; Nichols et al.,
become essential
aches for the prediction
n challenge in the near
explored. ## global
on Machine Learning

ually mutated cause
and networks with

put this number in

context, there are ~1,000 essential genes in yeast, for which a single mutation leads to a
there are 200-fold more ways to generate a similar phenotype through a digenic synthet
This finding indicates that digenic interactions might underlie many inherited phenotype
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 protei
different processes and are involved in cross-talk. -> The Importance of Bottlenecks in Pr
Correlation with Gene Essentiality and Expression Dynamics
- This indicates the existence of a large number of nodes with high betweenness but low c
nodes). Importantly, such nodes are absent in computer- generated, random scale-free n
mention of bottlenecks: **High-Betweenness Proteins in the Yeast Protein Interaction Ne**

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 net
    - Even a comparatively small subnetwork of this size is still a challenge to visualize in
    authors assess its quality in two ways. **First**, *they determine whether the genes of th*
    enriched for specific cellular process categories in the Gene Ontology database, wh
    and this is a wonderful strength of the paper—the authors experimentally validate
    neighbors of *MYC* -> **Reverse engineering gene regulatory network**
    - Basso *et al.* demonstrate that as long as the available data explore a wide range in
    of the system, biologically meaningful interactions can be recovered by computatio
    Reverse engineering gene regulatory network for the paper -> Reverse engineerin
    networks in human B cells
    - We note that these data support a direct regulatory effect of the tested transcripti
    gene, but chromatin immunoprecipitation experiments would be required to deter
    binding. -> **Wisdom of crowds for robust gene network inference**
- Gene essentiality data of yeast were manually compiled from the Comprehensive Yeast G
(http://mips.helmholtz-muenchen.de/genre/proj/yeast/) and large-scale experiments26.
1,178 essential and 4,904 nonessential yeast genes. -> Network rewiring is an important
essentiality change
- [currently, the available essential genes and protein databases are DEG (Zhang and Lin, 2
2013), OGEE (Chen et al., 2012), and EGGS (http://www.nmpdr.org/FIG/eggs.cgi)]. These
researchers to explore the features of essential genes and proteins and, through this exp
features are associated with essentiality and, finally, develop computational methods pro
essential genes and proteins. -> **Predicting Essential Genes and Proteins Based on Mach**
Network Topological Features: A Comprehensive Review

.
lethal phenotype, but
ic-lethal interaction.
s, and begins to explain
. -> **Exploring genetic**

ns that mediate
otein Networks/

onnectivity (HBLC
etworks ->First
twork 2005

tworks!! e.g
sightfully, so the
he subnetwork are
hich they are. **Second**—
some of the first

the 'expression space'
onal algorithms ->
g of regulatory

on factor on the target
- mine physical

Genome Database
. The dataset contained
mechanism of gene

009), CEG (Ye et al.,
data have enabled
loration, reveal which
pposed to identify
hine Learning and

Network topological features) A comprehensive review

- There exist many different databases of a certain type of interaction (e.g., DIP, BioGRID a usually, these databases are regularly updated. Different databases or newer versions of have different sets of interactions that, in turn, will give rise to new networks with distinc consequently, different values of network topological features. As an example, we can cit Hwang et al. (2009) and Acencio and Lemke (2009). In both studies, PINs of *Saccharomyc* created; however, the interactions of the PIN constructed in the study by Hwang et al. (2 from the version ScereCR20070107 of DIP and the interactions of the PIN constructed in and Lemke (2009) were gathered from the version 2.0.42 of the BioGRID database. There performances of the models created by these authors cannot be reliably compared. ## Th must be the same kind and the same data! That means not only the same database becau regularly, but the exact same datasets.
- Thus, it seems that only network topological features are not enough to distinguish esser genes and proteins. This raises the following question: **is the positive correlation betwee network topological features only an artifact of a possible bias (essential genes and pro more studies and therefore tend to have higher values of network topological)** present derived from small scale experiments?
- Regardless the resolution of this debate, a large-scale study for evaluating how well esse proteins can be predicted solely by network topological features is necessary to confirm t prediction performance.
- ROC curves in R http://blog.revolutionanalytics.com/2016/11/calculating-auc.html
- Dygraphs for r Interactive plots for HTML files

nd IntAct for PPIs) and,
a given database will
ct structures and,
ce the studies by
ces cerevisiae were
009) were collected
the study by Acencio
efore, the prediction
he reference networks
use they are updated

ntial from non-essential
en essentiality and
oteins are the focus of
in the networks mainly

ntial genes and
this moderate