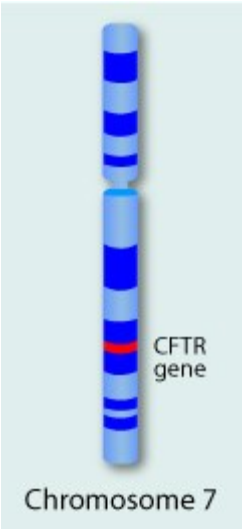
Systems Biology: A Personal View
XV. Network Medicine

Sitabhra Sinha
IMSc Chennai

http://learn.genetics.utah.edu/

Diseases, Genes and Networks



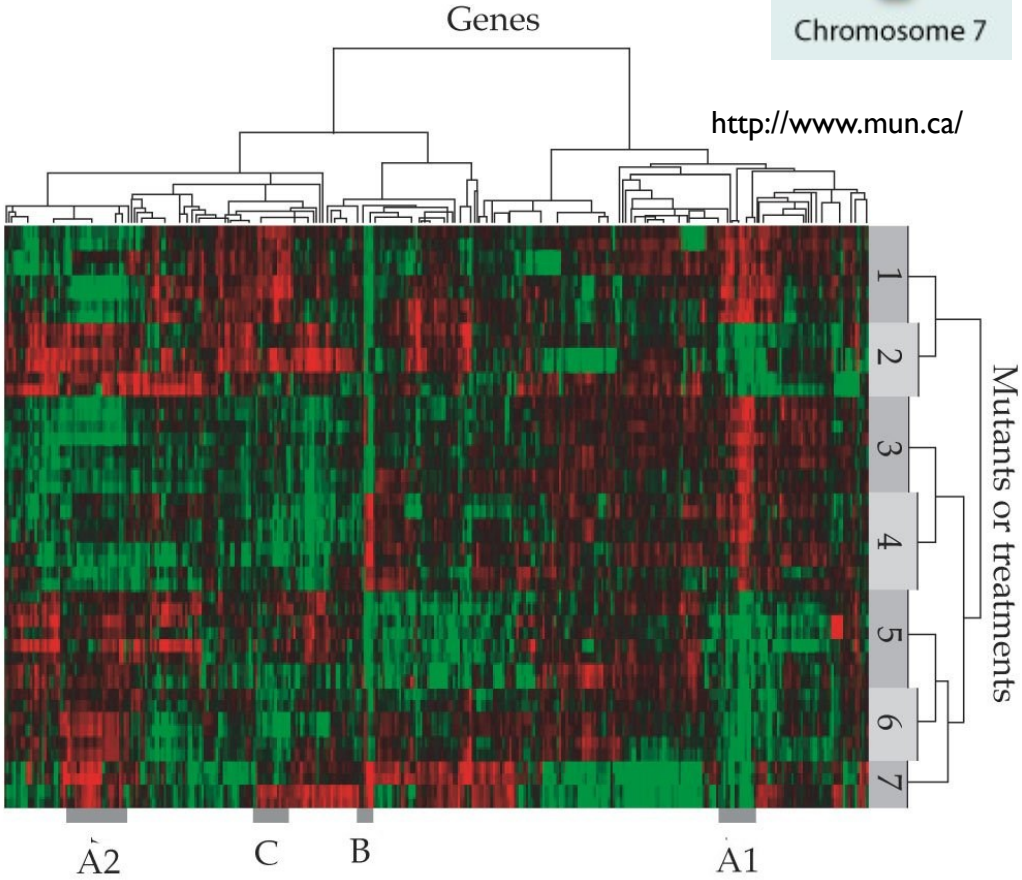
Now that we have the ability to sequence entire genomes, can we say determine the genetic basis of different diseases – and potentially come up with treatments ?

Unfortunately, there are very few diseases that can be linked to mutation of a single gene – such as cystic fibrosis caused by mutation in CTFR gene

In most cases, diseases are associated with many different genes

Possible to analyze the differential expression patterns in diseases through gene array expts

Can systems biology give a network perspective to understanding the genetic basis of diseases ?



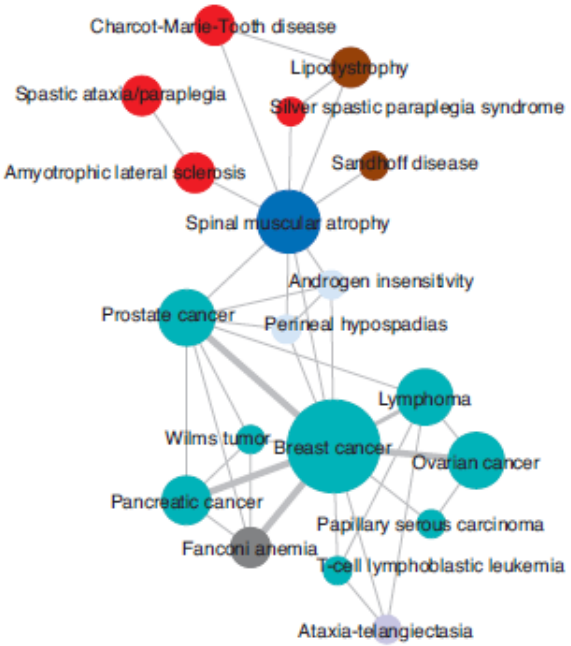
http://www.mun.ca/

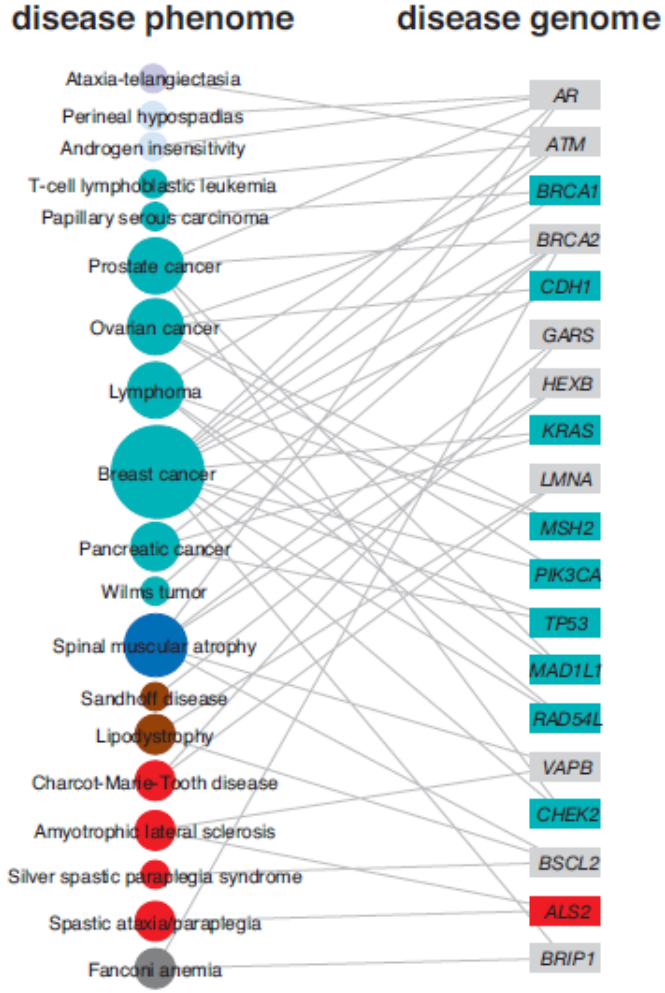
Complex differential gene expression patterns for different treatments seen from microarray studies

The human disease network

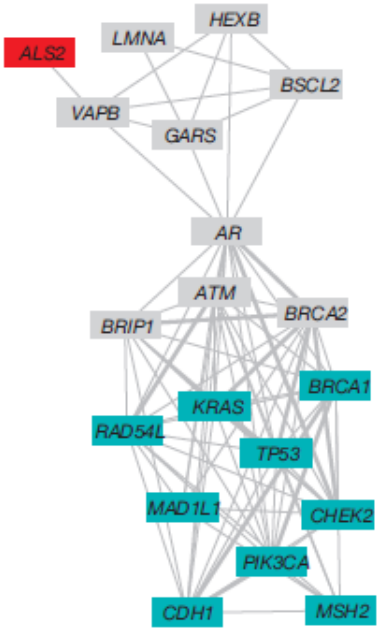
DISEASOME bipartite network

Human Disease Network (HDN)



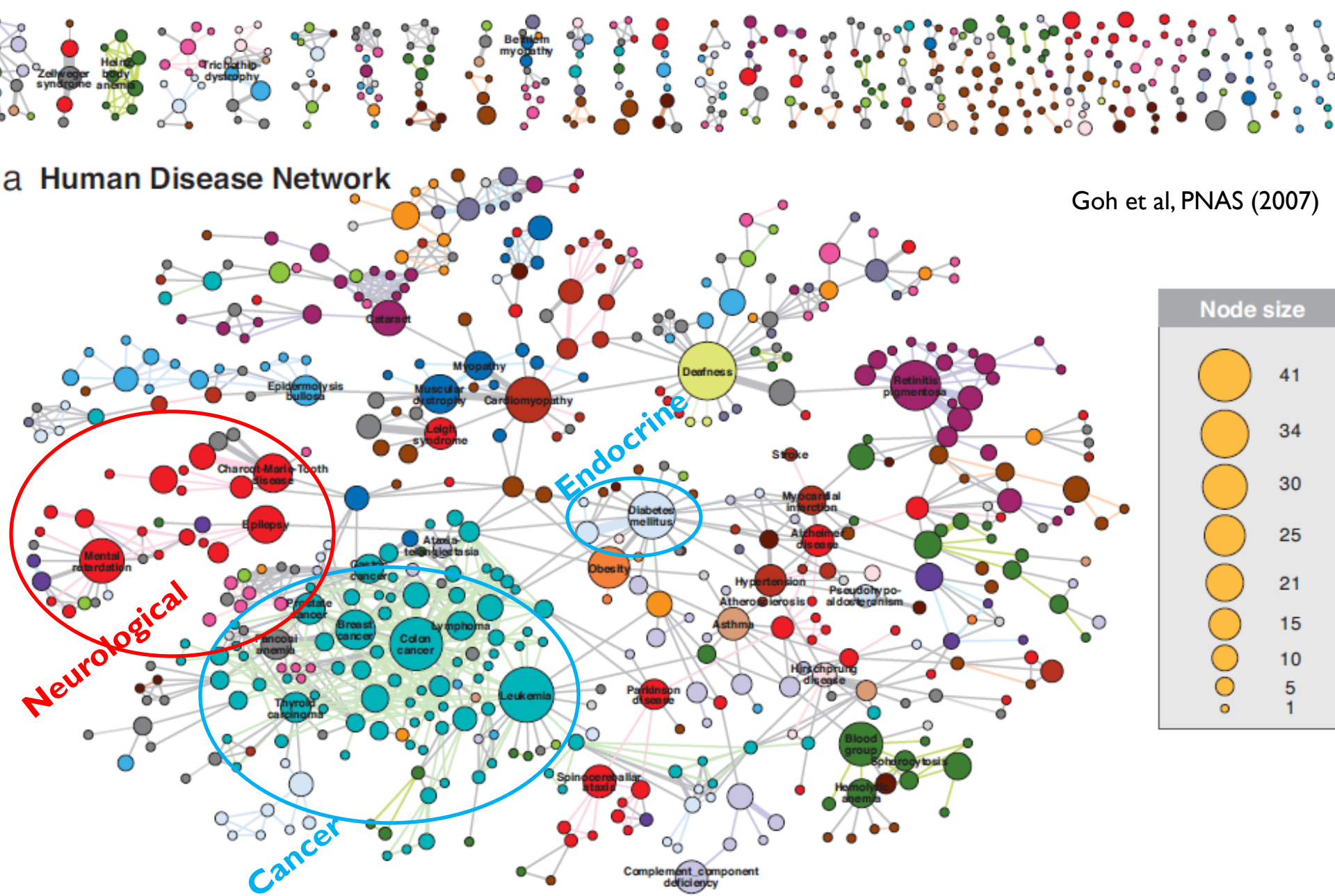


Disease Gene Network (DGN)



OMIM-based disorder–disease gene associations

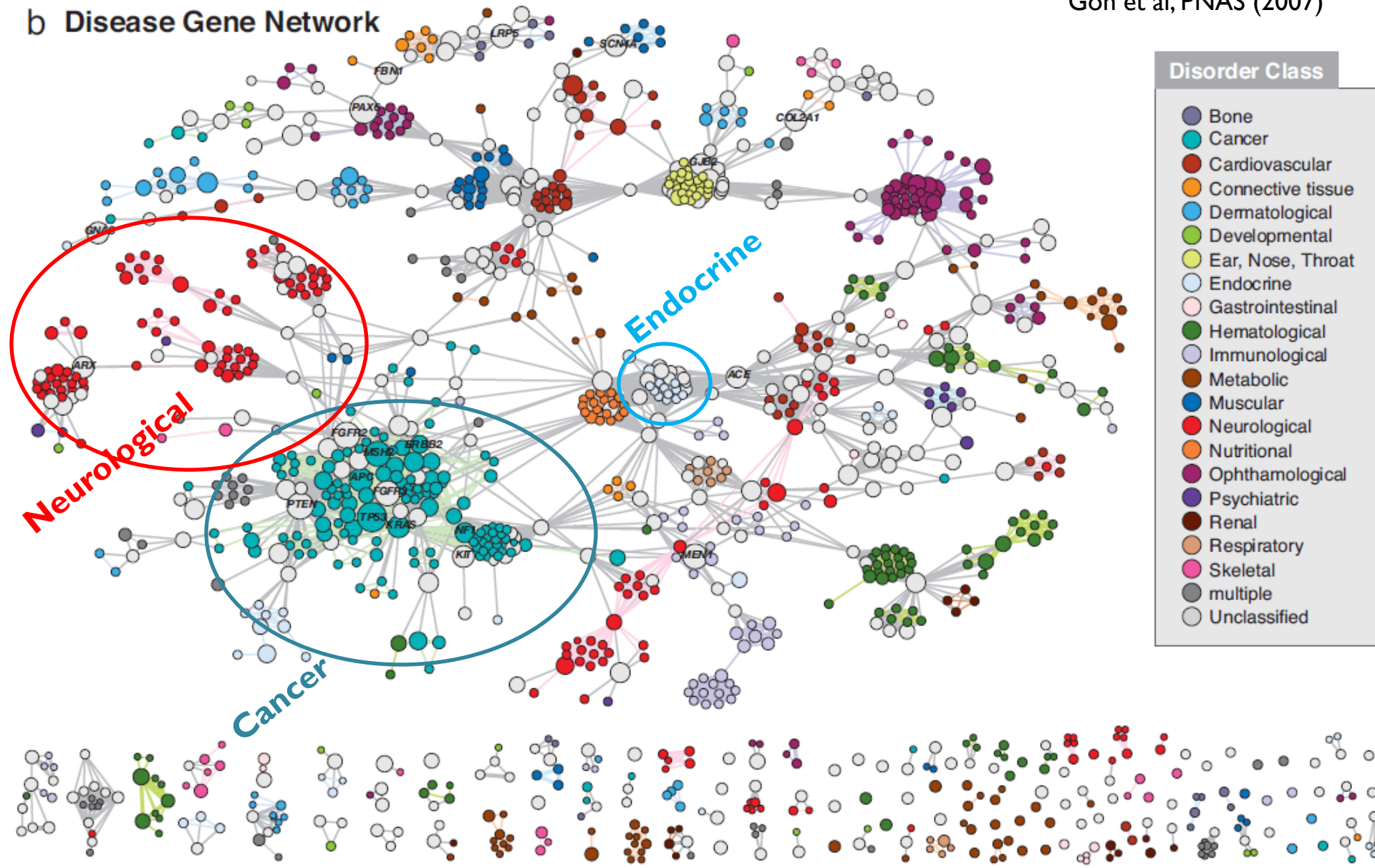
Goh et al, PNAS (2007)



Node: disorder (size proportional to number of genes associated with the disorder)

Link: Number of common genes associated with two disorders

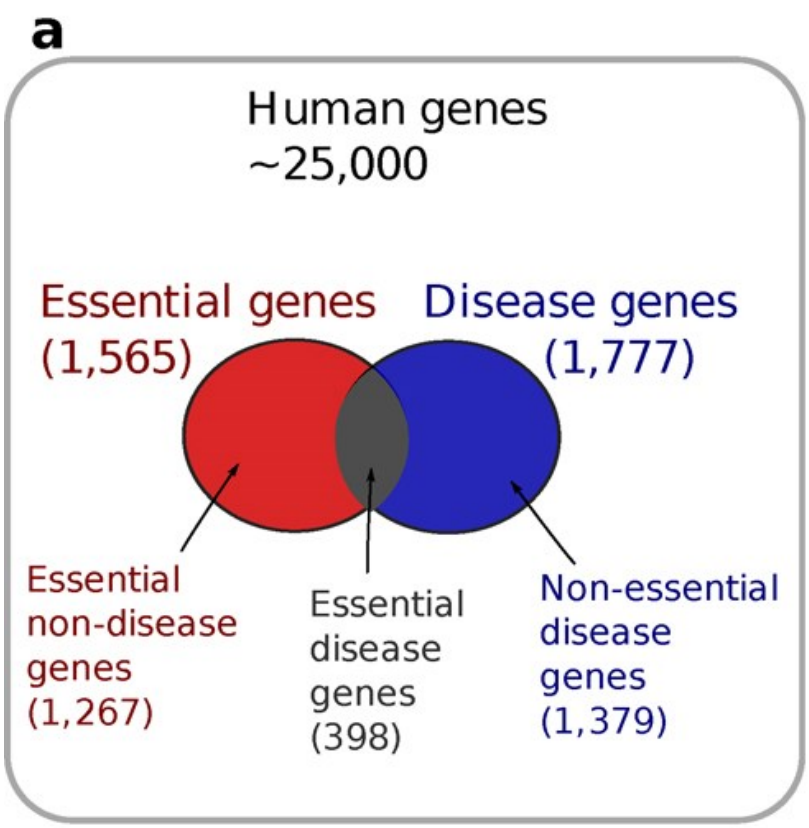
Goh et al, PNAS (2007)

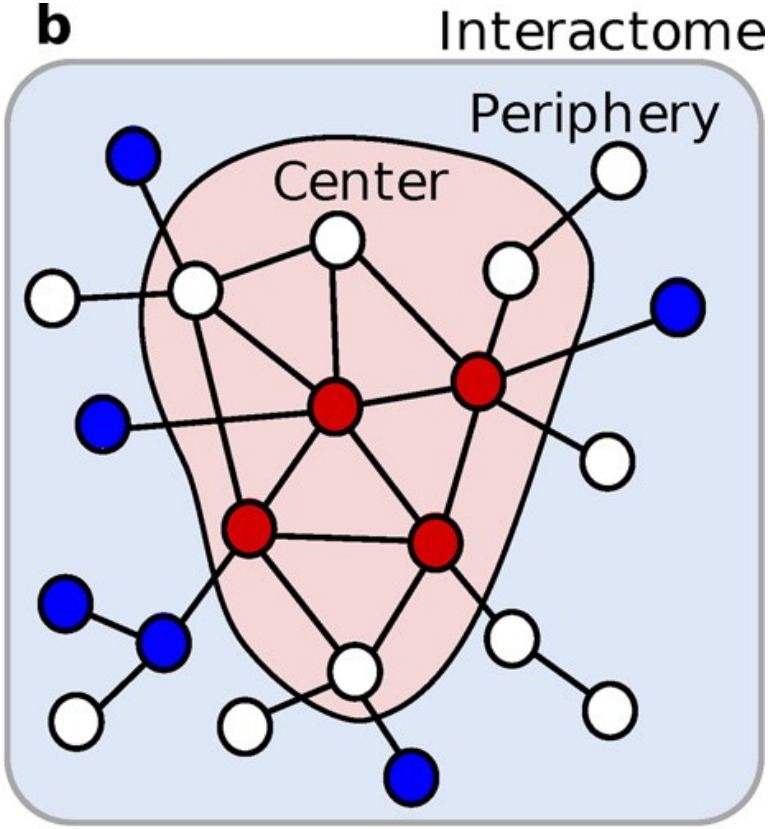


Node: gene (size proportional to number of disorders in which the gene is implicated)
Link: common disorder that two genes are associated with

Disease genes are not hubs

Barabasi et al, Nature Rev Genetics (2011)



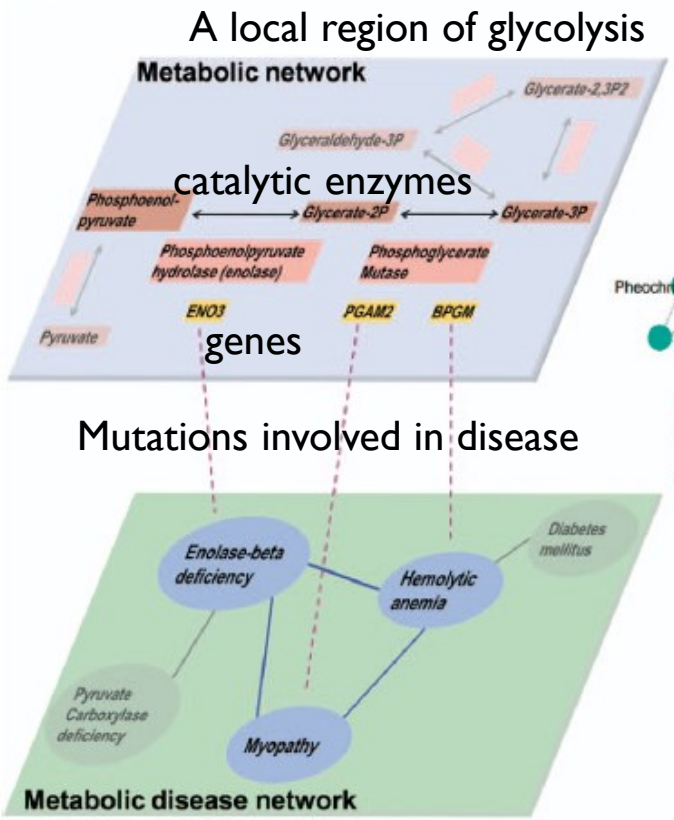


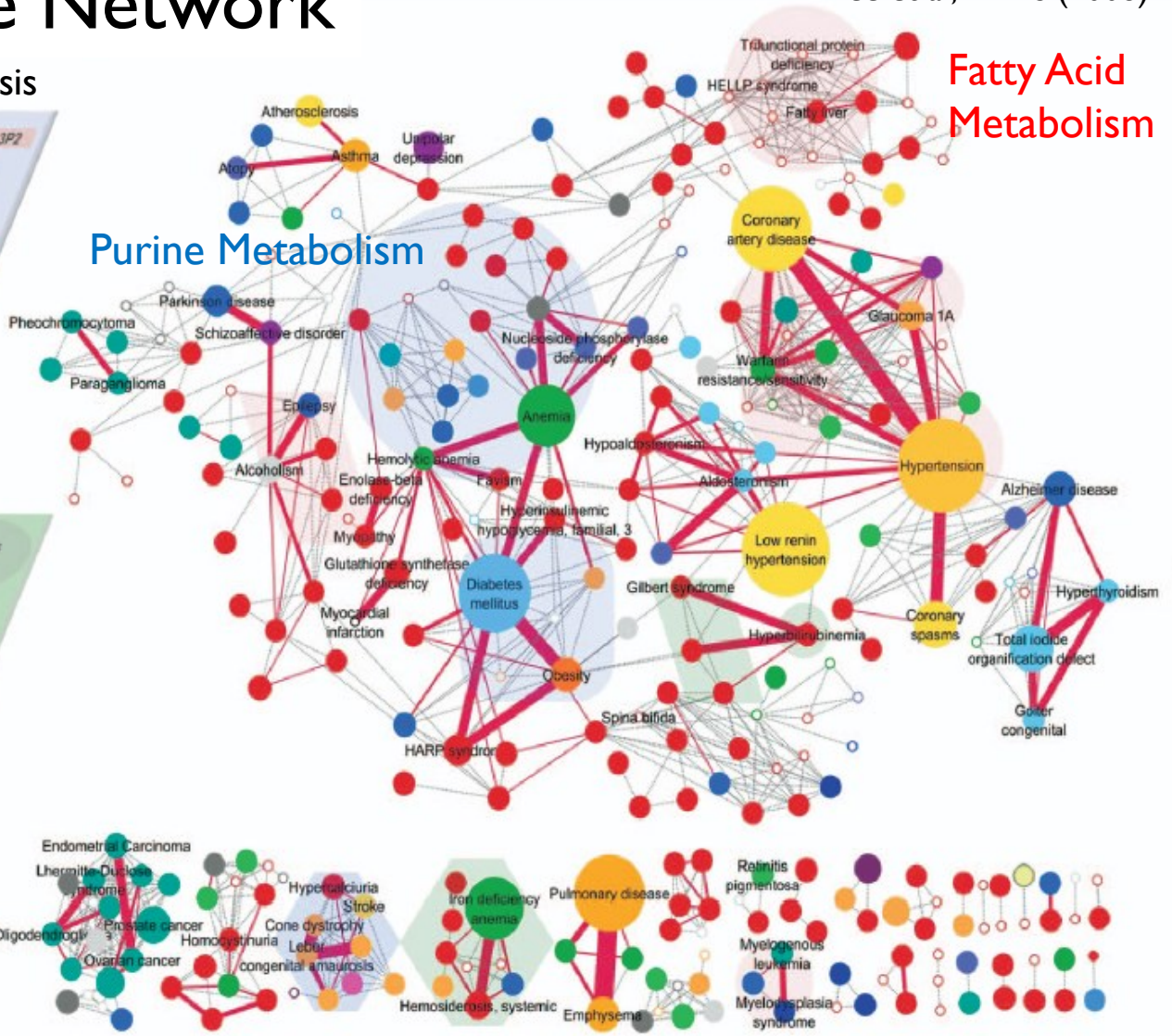
Of the approximately 25,000 human genes, (i) ~1,700 have been associated with specific diseases and (ii) 1,600 genes are known to be *in utero* essential, i.e., their absence is associated with embryonic lethality.

Non-essential disease genes are at network periphery while in utero essential genes tend to be at the center (encode hubs, expressed in many tissues) of the interactome.

Metabolic Disease Network

Lee et al, PNAS (2008)





MD Network: 308 nodes (non-isolated diseases) connected by 878 metabolic links [KEGG, OMIM]

Node color: disease class
Node size: disease prevalence in dataset
Link width: comorbidity of two diseases

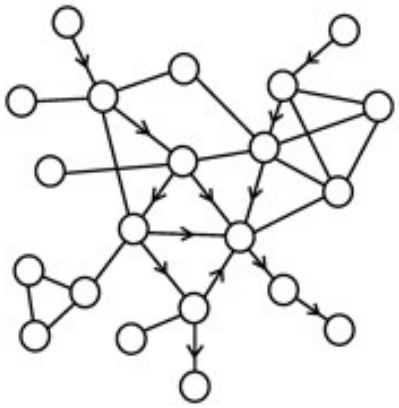
Comorbidity between metabolically linked diseases is higher than those that are not connected

Merges information on protein-protein interactions, co-complex memberships, regulatory interactions, and metabolic network

Identify functionally & topologically compact subgraph containing most disease components, representing potential disease module

Identify the specific molecular pathways whose disruption maybe responsible for disease phenotype (esp. if size of disease module is too large)

***i.* Interactome reconstruction**

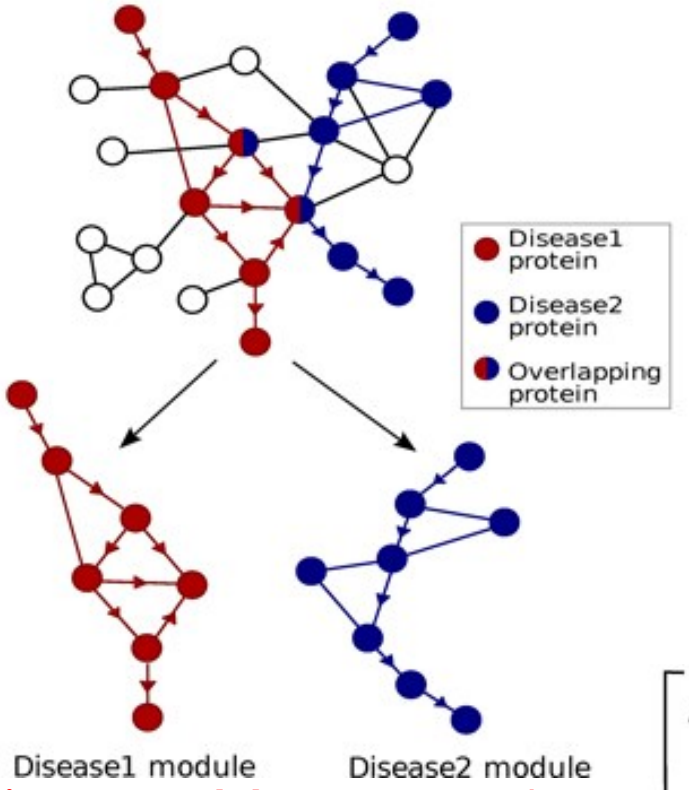


***ii.* Disease gene (seed) identification**

Potential sources:
i) OMIM
ii) GWAS
iii) Literature

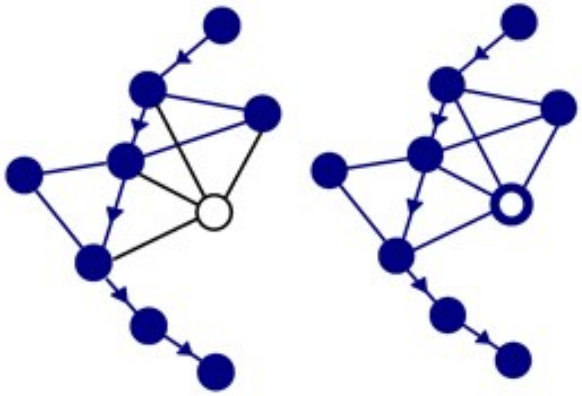
Known disease-associated genes obtained from linkage analysis, GWAS, etc.

***iii.* Disease module identification**



Disease module comprise nodes whose perturbation (mutations, deletions, copy number variations, or expression changes) can be linked to a particular disease phenotype

***iv.* Pathway identification**



Validation

Functional homogeneity
i) gene ontology
ii) tissue specificity
iii) phenotypic similarity

Dynamic homogeneity
i) coexpression
ii) genetic interactions
iii) drug response

Prediction
Disease genes
Disease pathways
Drug targets

Barabasi et al, Nature Rev Genetics (2011)

Cancer: a systems disease

Cancer (malignant tumor or neoplasm) is a group of diseases (comprising more than 200 distinct types) that involve abnormal cell growth with the potential for metastasis – i.e., spread from its original site of occurrence to other parts of the body [*adapted from wikipedia*]

Traditionally focus has been on the role of mutations of specific genes
Led to identification of

- **Proto-oncogenes** are genes that normally help cells grow. When they mutate or there are too many copies of it, it becomes a oncogene that causes uncontrolled cell growth leading to cancer.
- **Tumor suppressor genes** slow down cell division, repair DNA mistakes, or mediate apoptosis (programmed cell death) – when they don't work properly, uncontrolled cell growth occurs, leading to cancer.

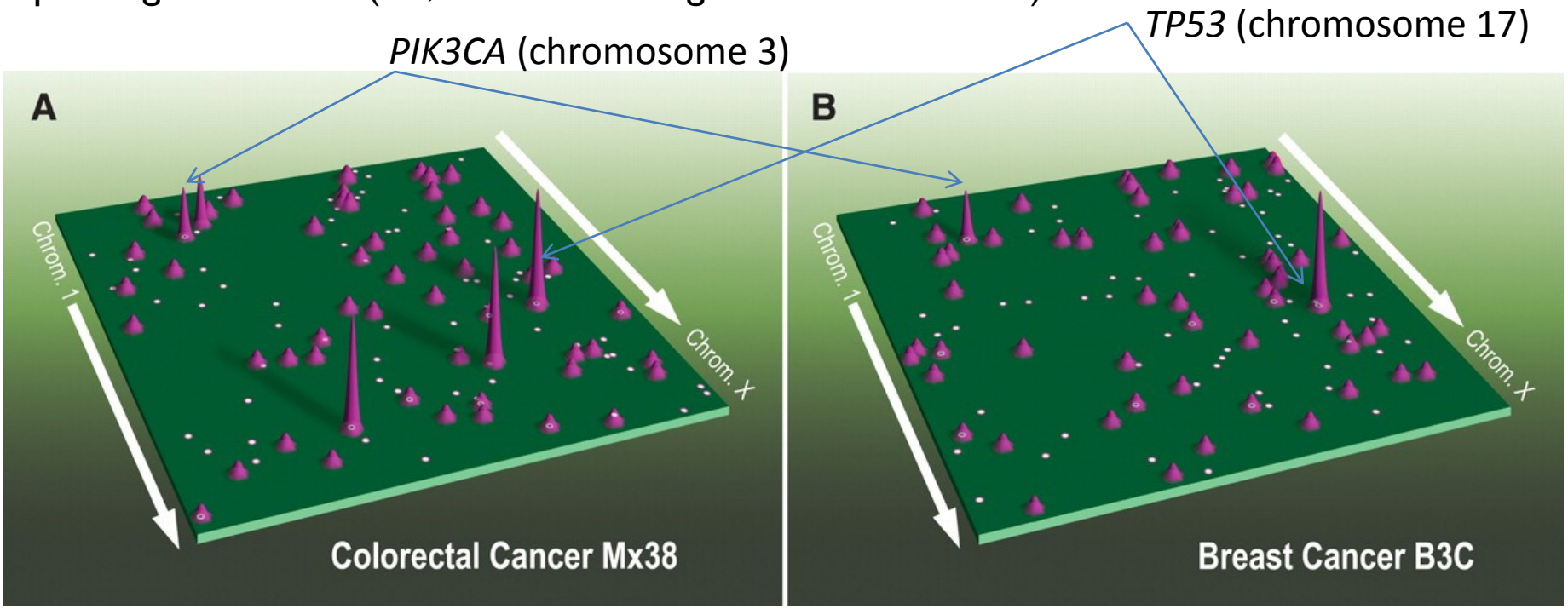
However, focusing on individual gene mutations does not give a global picture of the inter-relations of different genes and cancer phenotypes

"systems disease" involve large-scale disruption in the intricate arrangement of the signaling system responsible for intra-cellular command, control and communication

Cancer genome landscape suggests a pathway-centric view

2-dimensional map representing chromosomal position of genes – indicating the ones mutated in individual colorectal or breast tumors (dots)

Peak heights represent the cancer mutation prevalence (CaMP) score of genes for each tumor type – the likelihood that a gene is mutated at a frequency higher than the rate of passenger mutation (i.e., mutation having no effect on fitness)



Wood et al, Science (2007): "When all the mutations that occur in different tumors are summed, the number of potential driver genes is large. But this is likely to actually reflect changes in a much more limited number of pathways, numbering no more than 20."

The importance of p53 for cancer

Mechanism of inactivating p53	Typical tumours	Effect of inactivation
Amino-acid-changing mutation in the DNA-binding domain	Colon, breast, lung, bladder, brain, pancreas, stomach, oesophagus and many others	Prevents p53 from binding to specific DNA sequences and activating the adjacent genes
Deletion of the carboxy-terminal domain	Occasional tumours at many different sites	Prevents the formation of tetramers of p53
Multiplication of the MDM2 gene in the genome	Sarcomas, brain	Extra MDM2 stimulates the degradation of p53
Viral infection	Cervix, liver, lymphomas	Products of viral oncogenes bind to and inactivate p53 in the cell, in some cases stimulating p53 degradation
Deletion of the $p14^{ARF}$ gene	Breast, brain, lung and others, expecially when p53 itself is not mutated	Failure to inhibit MDM2 and keep p53 degradation under control
Mislocalization of p53 to the cytoplasm, outside the nucleus	Breast, neuroblastomas	Lack of p53 function (p53 functions only in the nucleus)

"…signalling pathways involving p53 — like cellular signalling pathways in general — cannot be understood by looking at isolated components. Instead, it is essential to consider the tangled networks into which these signalling components are integrated."

Vogelstein, Lane and Levine, Nature (2000)

p53 network

p53 is a highly connected 'hub'

Activation of network by stress…

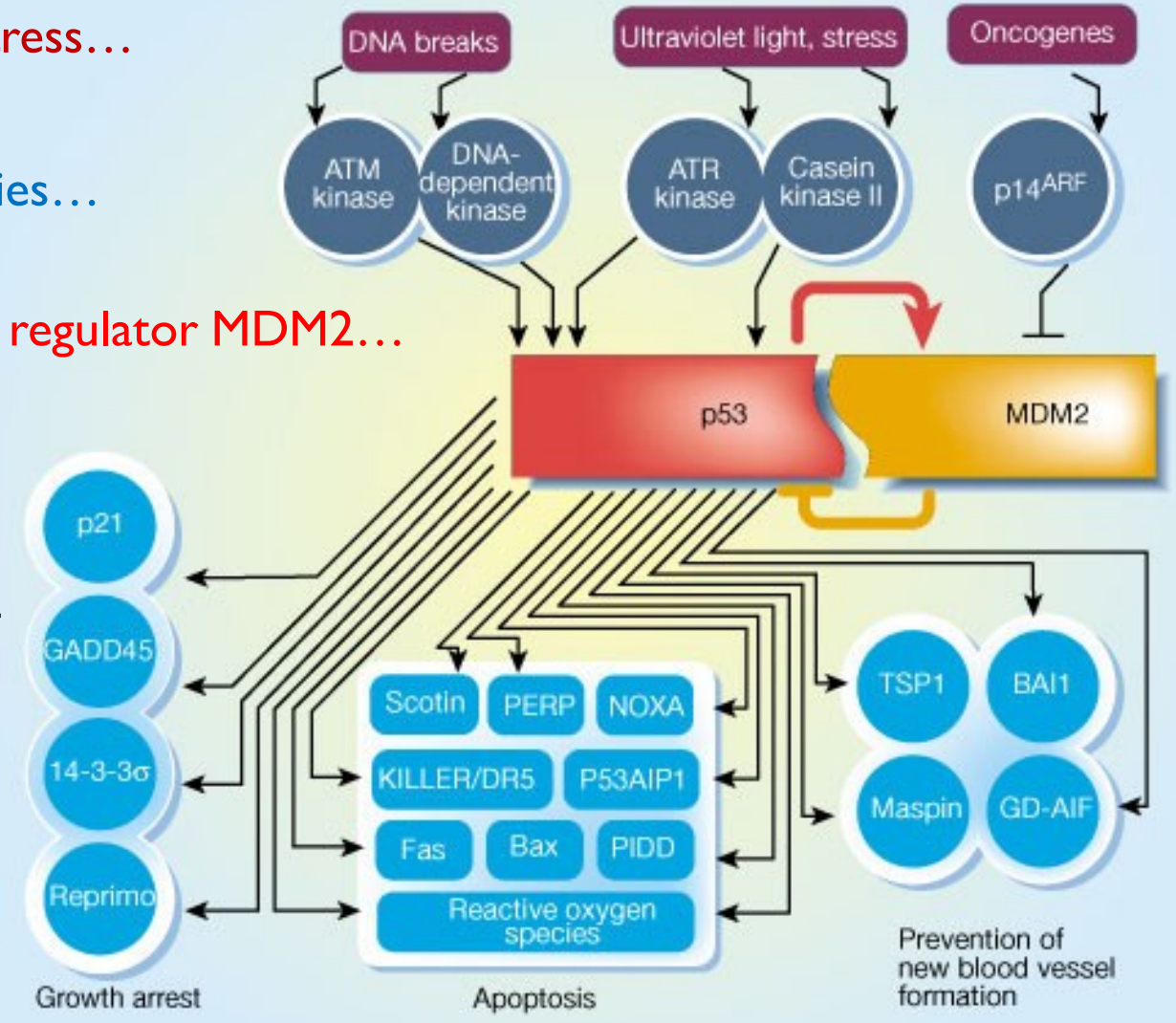
stimulates enzymatic activities…

modify p53 and its negative regulator MDM2…

increases levels of activated p53 protein.

Expression of several target genes is activated by binding of activated p53 to their regulatory regions.

Genes are part of processes that slow down development of tumours.



Vogelstein, Lane and Levine, Nature (2000)

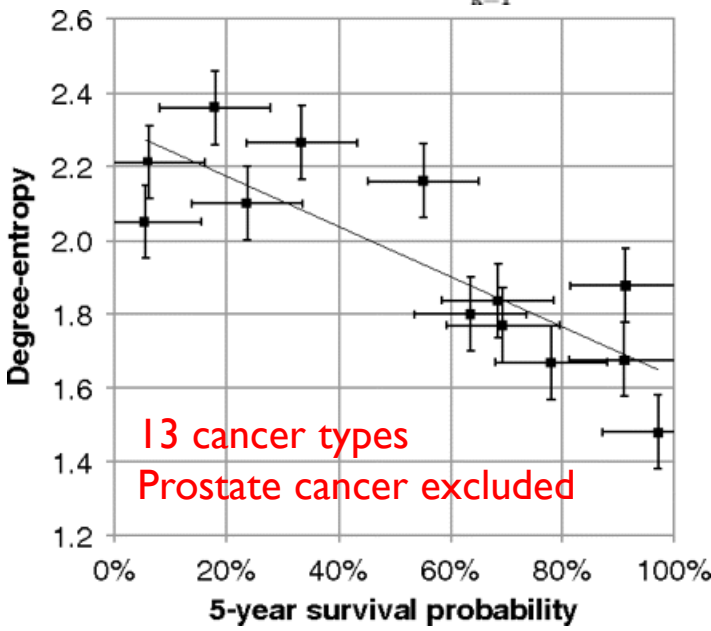
Relating cancer signaling network properties & disease survivability

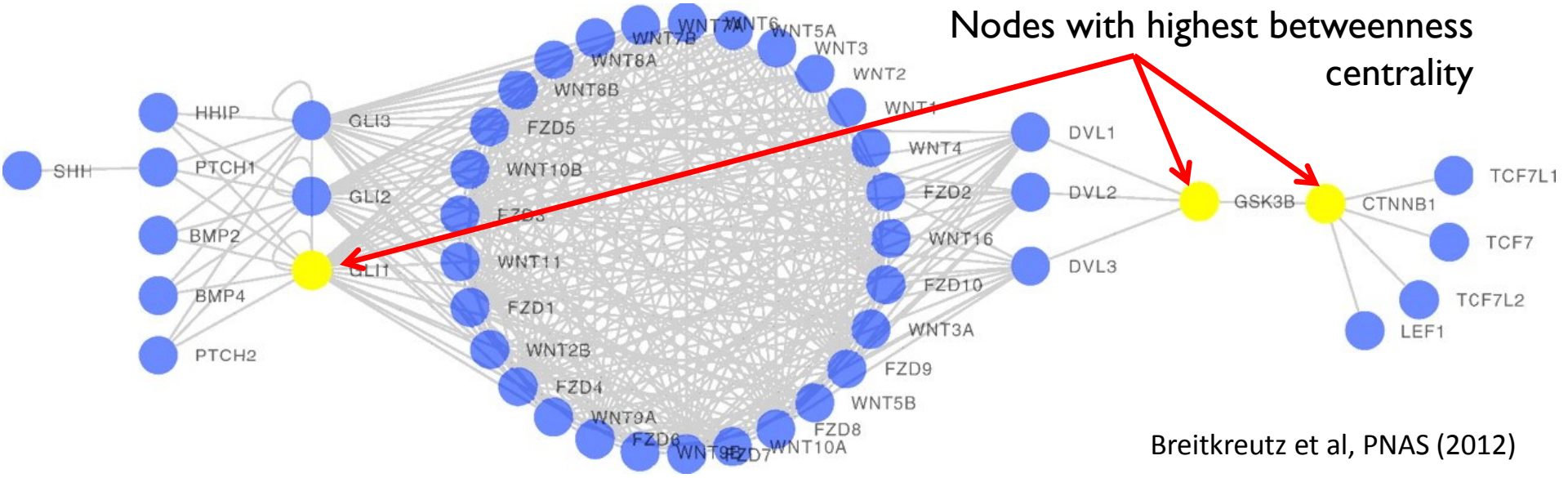
Cancer survival depends strongly on tumor type

E.g., 5-year survival probability for prostate cancer is >99% while for pancreatic cancer it is <6%

Breitkreutz et al (2012): correlation between patient survivability and degree entropy of the PPIN for a given cancer site

Degree entropy $H = -\sum_{k=1}^{N-1} p(k) \log p(k)$,





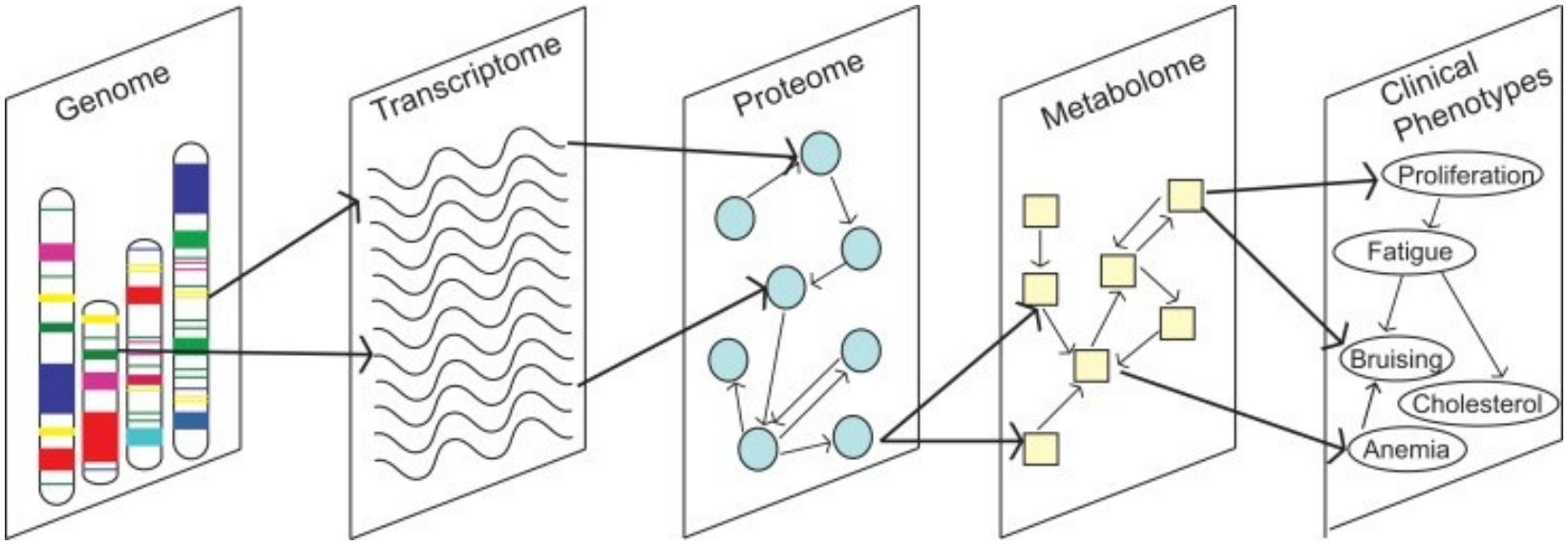
Breitkreutz et al, PNAS (2012)

Protein–protein interaction network for *basal cell carcinoma* (KEGG)

Challenge:

Data integration across different scales

"In Cancer Biology, data integration is of particular importance because of the complex interplay between genetics, cell signaling, and metabolic pathways … [however] integration [even] within omics data types [is] complicated [because of] differences in sample quality,…organism under investigation, tissue type, and experimental conditions (e.g., diet)." Blair et al, Frontiers in Physiology (2012)



Blair et al, Frontiers in Physiology (2012)

What are the inter-connections between the different biological domains and how do disruptions that can span levels give rise to a disease phenotype ?