“Systems Biology” and biomedical applications

1) Diseases– introduction
2) Networks in biomedicine – introduction
3) Application: The human diseasome
4) Application: Comorbidity
5) Innate immunity: introduction and applications
6) Inflammation: introduction and applications
7) Tumors: introduction and applications
8) P4 medicine

NETWORKS AND CANCER

NEOPLASM

MASS of CELLS
ORIGINATES from ONE SINGLE CELL
GENETIC ALTERATION WHICH IS TRANSMITTED TO OTHER CELLS

HISTOLOGY/CYTOLOG: CELLS CAN BE VERY SIMILAR OR VERY DIFFERENT FORM THE ORIGIN (NORMAL) CELL

FUNCTION: SAME/DYSFUNCTION/LOSS OF FUNCTION

GROWTH:
- **VARIABILE**
- **PROGRESSIVE**
- **AUTONOMOUS**
- **ATYPICAL**

TUMORS: BENIGN and MALIGN

• ***Benign tumor:*** **when its microscopic and gross characteristics are considered to be relatively innocent, implying that it will remain localized and is amenable to local surgical removal.**

- Affected patients generally survive.

• ***Malignant tumor:*** **as applied to a neoplasm, implies that the lesion can invade and destroy adjacent structures and spread to distant sites (metastasize) to cause death.**

- **C**ollectively referred to as *cancers (*derived from the Latin word for “crab”—that is, they adhere to any part that they seize in an obstinate manner, similar to a crab’s behavior.

-Affected patients may not survive

Most frequent causes of cancer

- 1. Diet (obesity)
- 2. Smoking
- 3. Alcohol consumption
- 4. Reproductive history (estrogen stimulation)
- 5. Infectious agents

The main types of cancer leading to overall cancer mortality each year are:

- **lung (1.3 million deaths/year)**
- **stomach (803 000 deaths)**
- **colorectal (639 000 deaths)**
- **liver (610 000 deaths)**
- **breast (519 000 deaths)**

The most frequent types of cancer worldwide are:

- Among men - lung, stomach, liver, colorectal, oesophagus and prostate

- Among women - breast, lung, stomach, colorectal and cervical

WHO:

Key risk factors for cancer that can be avoided are:

-**tobacco use** - responsible for 1.8 million cancer deaths per year (60% of these deaths occur in low- and middle-income countries);
-**being overweight**, obese or physically inactive - together responsible for 274 000 cancer deaths per year;
-**harmful alcohol use** - responsible for 351 000 cancer deaths per year;
-**sexually transmitted human papilloma virus** (HPV) infection - responsible for 235 000 cancer deaths per year; and
-**occupational carcinogens** - responsible for at least 152 000 cancer deaths per year.

-Cancer prevention is an essential component of all cancer control plans **because about 30% of all cancer deaths can be prevented.**

More than 30% of cancer could be prevented by modifying or avoiding key risk factors, according to a 2005 study by international cancer collaborators[1].

Risk factors include:

1. **tobacco use**
2. **being overweight or obese**
3. **low fruit and vegetable intake**
4. **physical inactivity**
5. **alcohol use**
6. **sexually transmitted HPV-infection**
7. **urban air pollution**
8. **indoor smoke from household use of solid fuels**

Environmental factors potentially leading to DNA changes:

- Lifestyle factors (nutrition, tobacco use, physical activity, etc)
- Naturally occurring exposures (ultraviolet light, radon gas, infectious agents, etc.)
- Medical treatments (radiation and medicines including chemotherapy, hormone drugs, drugs that suppress the immune system, etc.)
- Workplace exposures
- Household exposures
- Pollution

Cigarette smoke increases the risk on many types of tumors:

Cancer type	Odds ratio
All cancer types	ND
Small cell lung cancer	111.3
Lung squamous	103.5
Lung adenocarcinoma	21.9
Larynx	13.2
Pharynx	6.6
Oral cavity	4.2
Esophagus squamous	3.9
Esophagus adenocarcinoma	3.9
Bladder	3.8
Liver	2.9
Stomach	2.1
Acute myeloid leukemia	2.0
Ovary	1.9
Cervix	1.8
Kidney	1.7
Pancreas	1.6
Colorectal	1.3

Lung tumors: >30cig./day

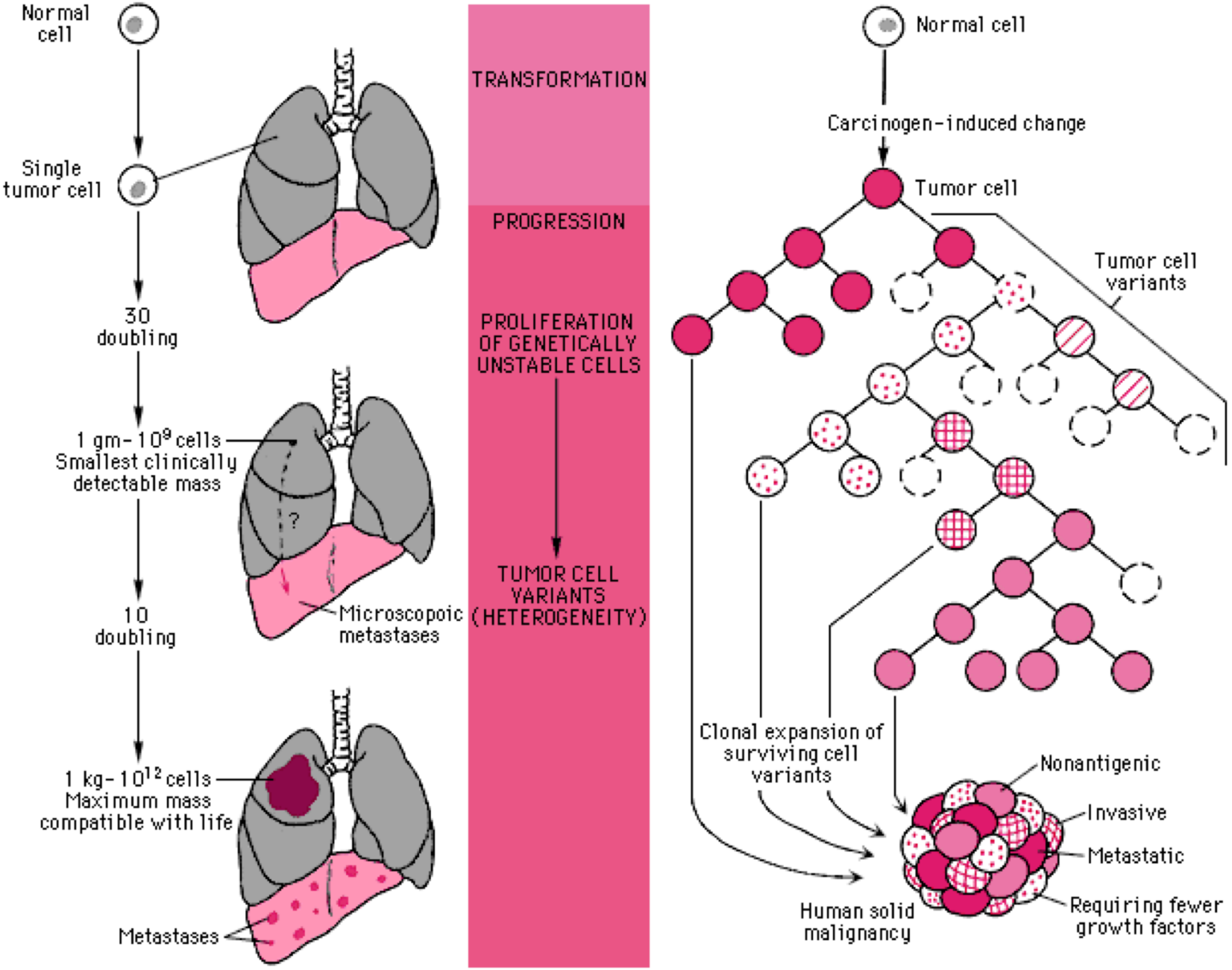
Definizione di cancerogeno (J & E Miller)

Dicesi cancerogeno un agente che, somministrato a un animale previamente non trattato, induce, per azione genotossica diretta, un incremento statisticamente significativo dell'incidenza di una data neoplasia rispetto agli animali di controllo (non esposti all'azione dell'agente in questione); ciò indipendentemente dal fatto se, nella popolazione animale di riferimento, l'incidenza spontanea della neoplasia in oggetto sia alta o bassa.

DIFFERENCE BETWEEN BENIGN AND MALIGNANT TUMORS

CARATTERISTICHE	BENIGNE	MALIGNE
Growth velocity	low mitosis percentage normal mitosis mitosi normali normal nucleoli	high mitosis percentage abnormal mitosis nucleoli ingranditi
Differentiation	similar to normal maintainement of normal functions	often low or altered (lost) functions
Diffusion	encapsulated no invasion no metastasis	non encapsulated local invasion metastasis

THE BIOLOGY OF TUMOR GROWTH



Tumors:

- **Local invasion**: progressive infiltration, invasion, and destruction of surrounding tissues

- **Metastasis**: spread of a tumor to sites that are physically discontinuous with the primary tumor
- marks a tumor as malignant

METASTASIS

DETACHMENT

MIGRATION: DISSEMINATION

- **SEEDING WITHIN BODY CAVITIES**
- **HEMATOGENOUS SPREAD**
- **LYMPHATIC SPREAD**
-

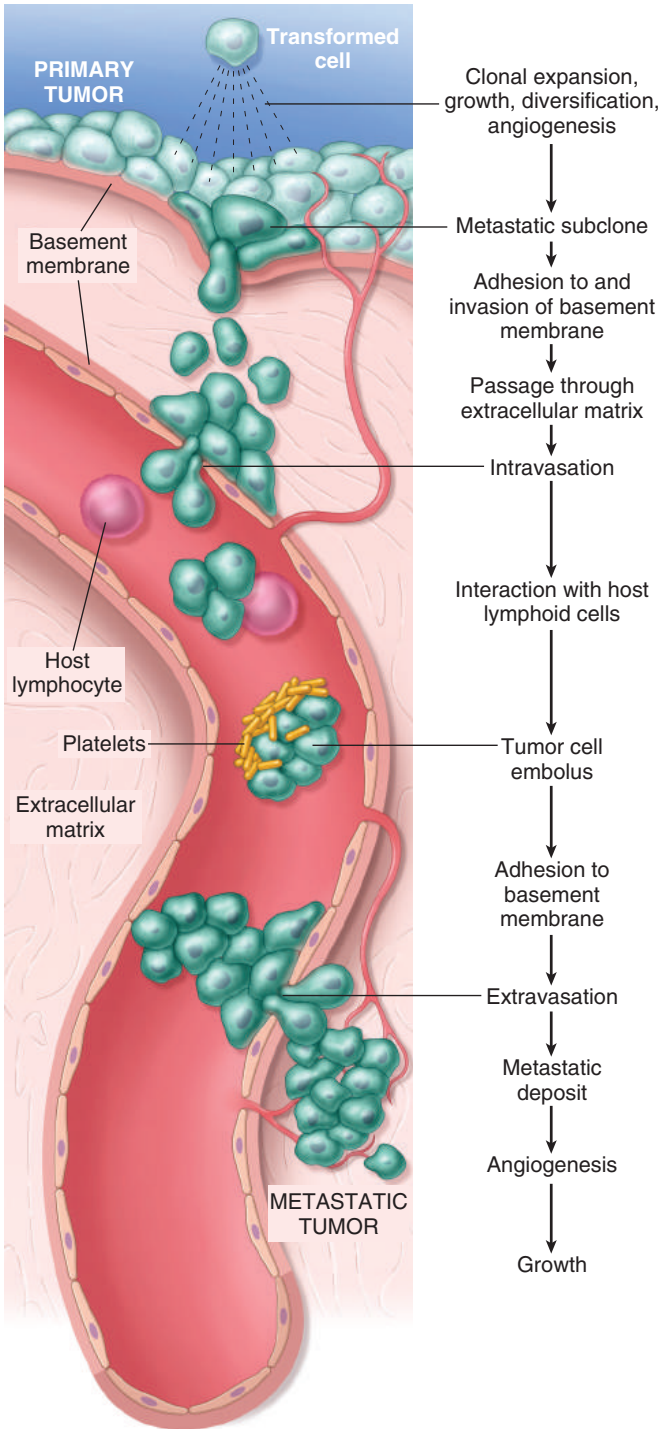
ARREST

IMPLANTATION

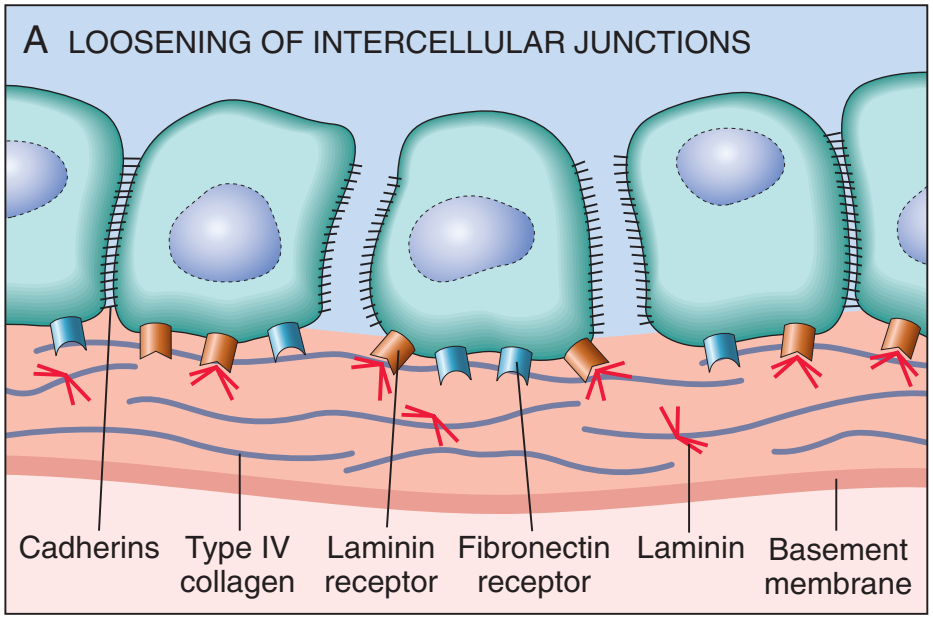
GROWTH

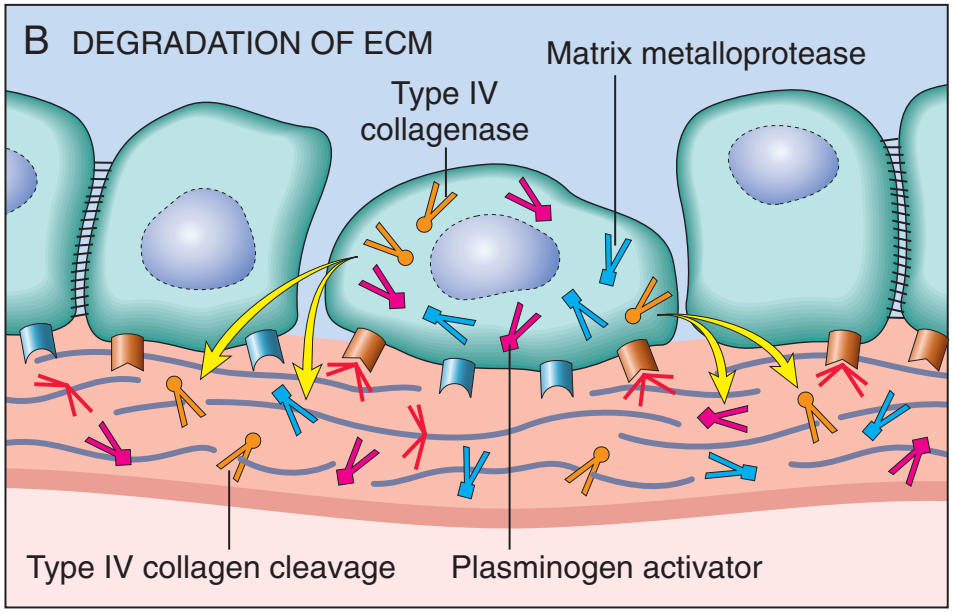
FILTER ORGANS (LUNGS, LIVER)
PREFFERENTIAL ORGANS (Es: Prostate carcinima gives bone metastasis)

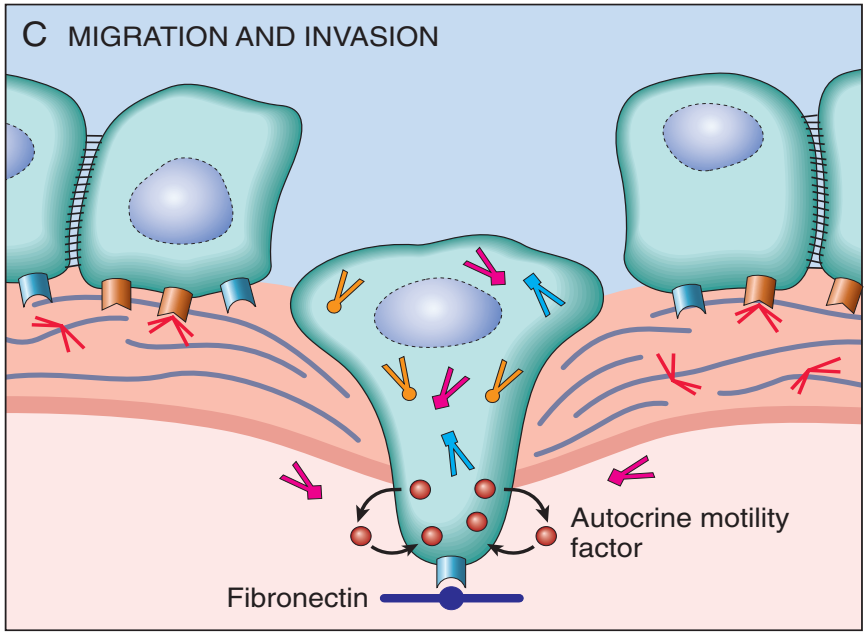
Metastatic cascade



Sequence of basal membrane invasion by tumor cells

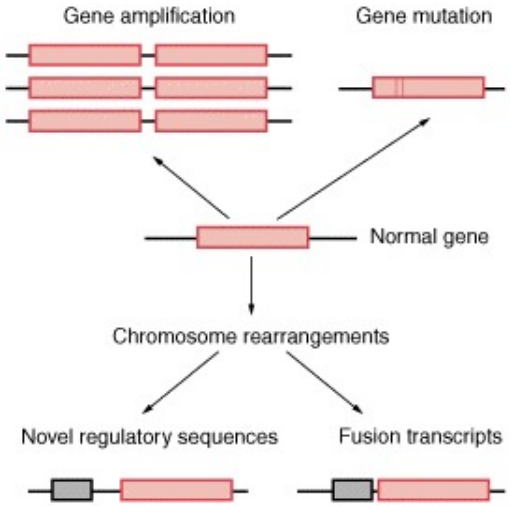






GENETIC ALTERATIONS IN CANCER

- **Amplification:**
  - **c-myc (adenocr, sarcomas)**
  - **c-erbB2 (adenocr.)**
  - **N-myc (adenocr., neuro-end. tumors)**
  - **c-erbB-1 (cr. with squamous cells)**
- **Mutations:**
  - **N-ras (AML)**
  - **K-ras (Adenocr.)**
  - **TP53**
- **Gene rearrangemens**
  - **fusion genes**
  - **overexpression**
- **Deletions**
- **Chromosomal translocation**
  - **c-myc (B-cell lymphoma)**
  - **c-abl (CML)**



CANCER GENES

ALTERATION OF GENES THAT IS "FIXED" IN THE DNA AND TRANSMITTED TO DAUGHTER CELLS

1) GENES THAT POSITIVELY REGULATE PROLIFERATION:

PROTOONCOGENES ⟶ ONCOGENES

MECHANISM: MUTATIONS, GENIC FUSION, AMPLIFICATION

2) GENES THAT NEGATIVELY REGULATE PROLIFERATION:

ONCOSUPPRESSOR GENES

MECHANISM: GENE INACTIVATION of both alleles (mutation, deletion, etc)

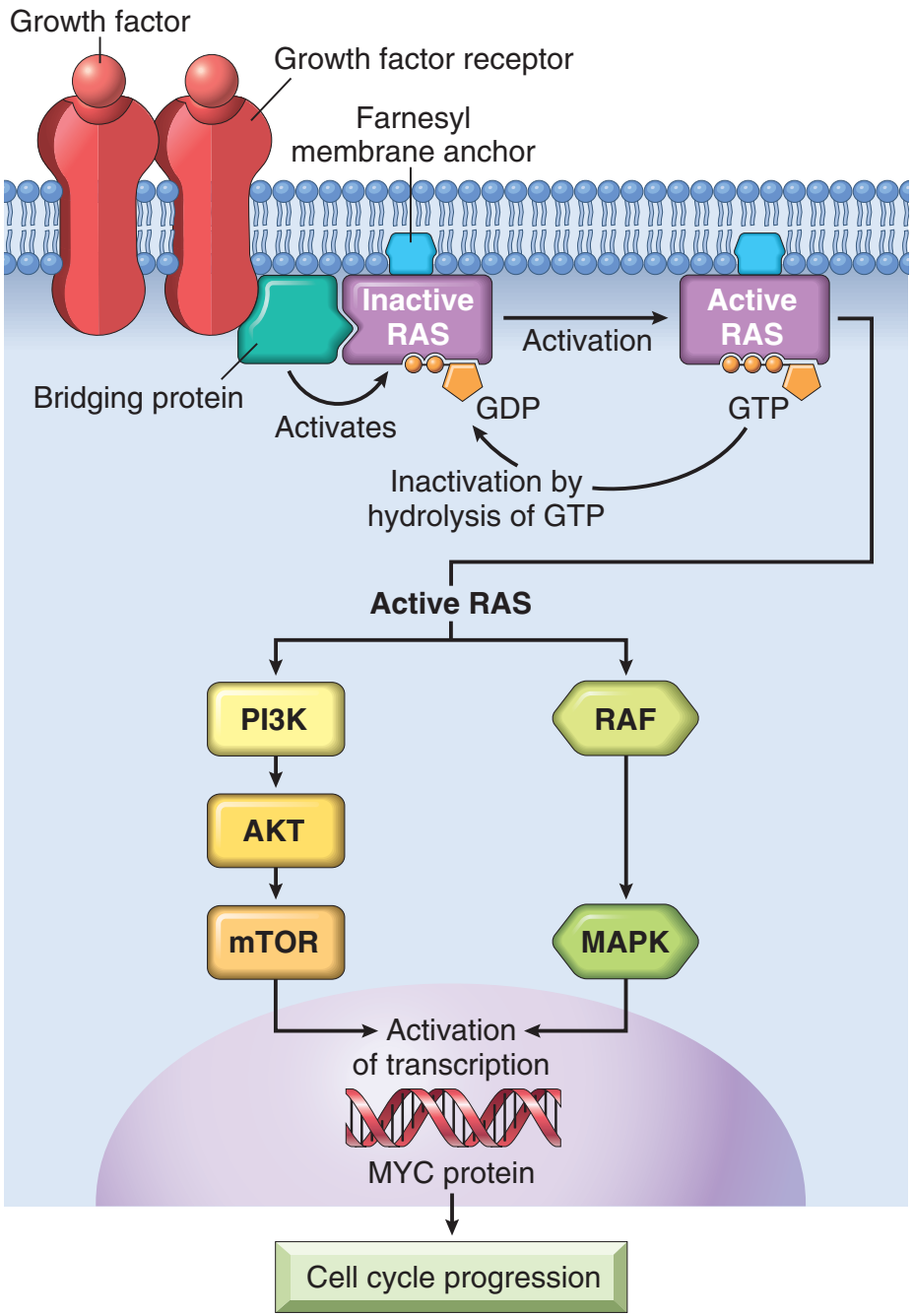
- *Proto-oncogenes:* normal cellular genes whose products promote cell proliferation
- *Oncogenes:* mutant or overexpressed versions of proto- oncogenes that function autonomously without a requirement for normal growth-promoting signals

ONCOGENES

- Classe I: Growth factors (sis)
- Classe II: Receptors for growth factors
  (erbB, fms, trk)
- Classe III: Signal treansduction factors
  (src, ras, abl, raf, gsp)
- Classe IV: Nuclear transcription factors
  (jun, fos, myc)

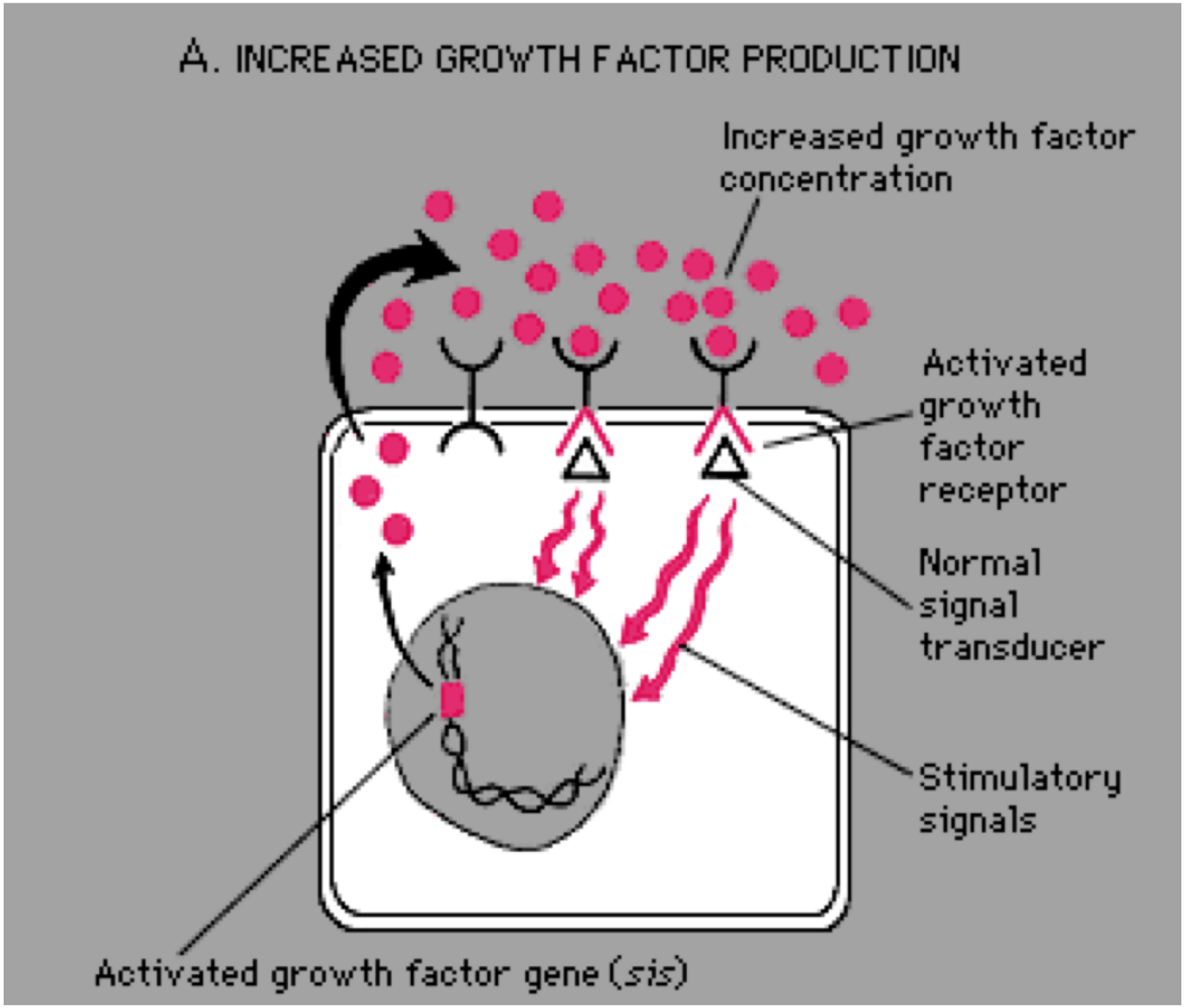
Table 7-6. SELECTED ONCOGENES, THEIR MODE OF ACTIVATION, AND ASSOCIATED HUMAN TUMORS

CATEGORY	PROTO-ONCOGENE	MECHANISM	ASSOCIATED HUMAN TUMOR
Growth Factors			
PDGF-β chain	*sis*	Overexpression	Astrocytoma
			Osteosarcoma
Fibroblast growth factors	*hst*-1	Overexpression	Stomach cancer
	int-2		Bladder cancer
			Breast cancer
			Melanoma
Growth Factor Receptors			
EGF-receptor family	*erb*-B1	Overexpression	Squamous cell carcinomas of lung
	erb-B2	Amplification	Breast, ovarian, lung, and stomach cancers
	erb-B3	Overexpression	Breast cancers
CSF-1 receptor	*fms*	Point mutation	Leukemia
Proteins Involved in Signal Transduction			
GTP-binding	*ras*	Point mutations	A variety of human cancers, including lung, colon, pancreas; many leukemias
Non-receptor tyrosine kinase	*abl*	Translocation	Chronic myeloid leukemia
			Acute lymphoblastic leukemia
Nuclear Regulatory Proteins			
Transcriptional activators	*myc*	Translocation	Burkitt´s lymphoma
	N-*myc*	Amplification	Neuroblastoma
			Small cell carcinoma of lung
	L-*myc*	Amplification	Small cell carcinoma of lung


Growth factor
Growth factor receptor
Farnesyl membrane anchor
Inactive RAS
Activation
Active RAS
Bridging protein
Activates
GDP
GTP
Inactivation by hydrolysis of GTP
Active RAS
PI3K
RAF
AKT
mTOR
MAPK
Activation of transcription
MYC protein
Cell cycle progression

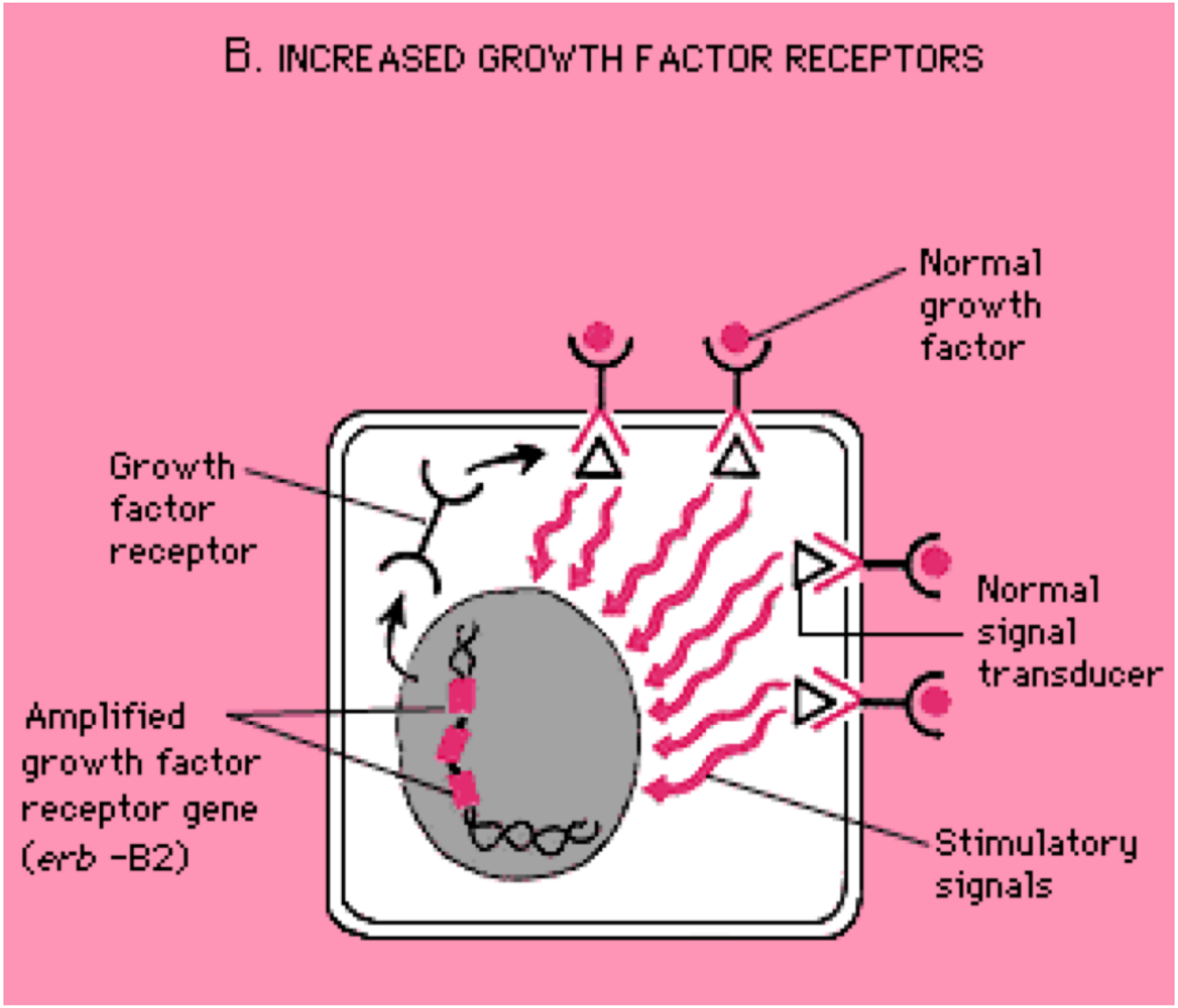
Mechanisms controlling tumor induction by oncogenes

Increased production of growth factor

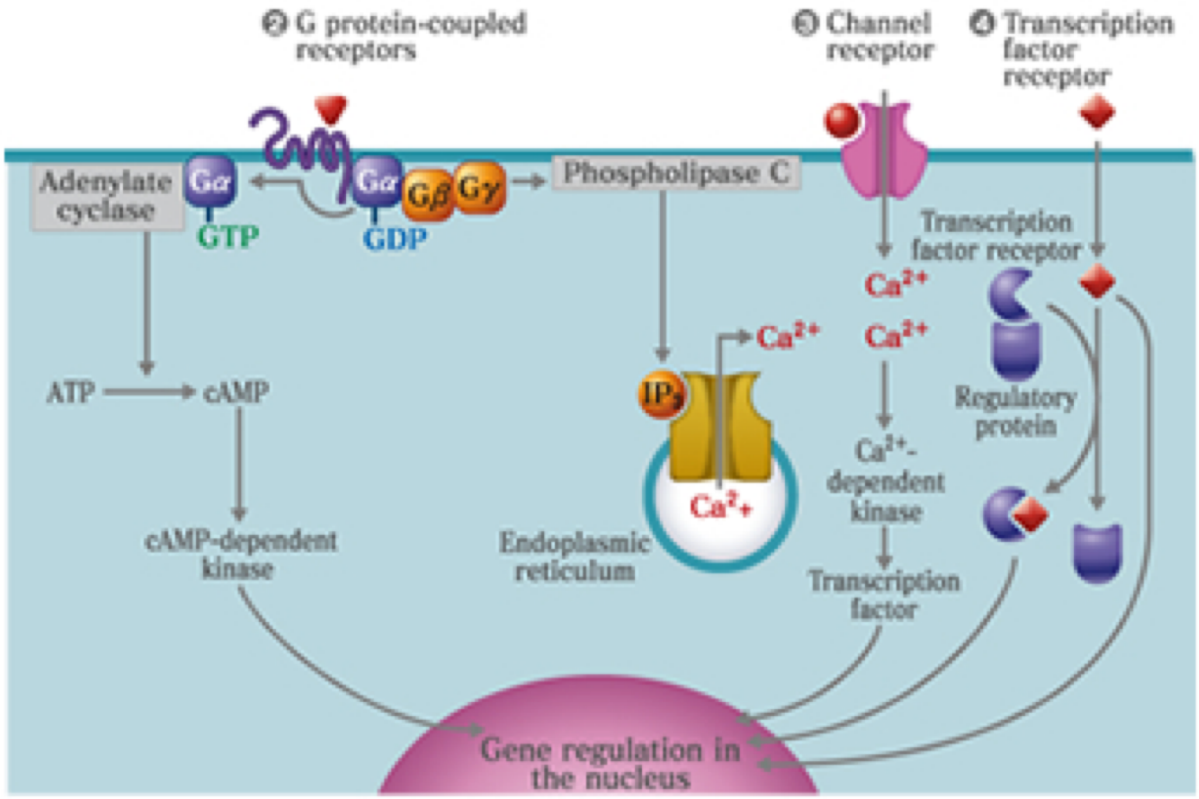


Mechanisms controlling tumor induction by oncogenes

Increase of growth factor receptor



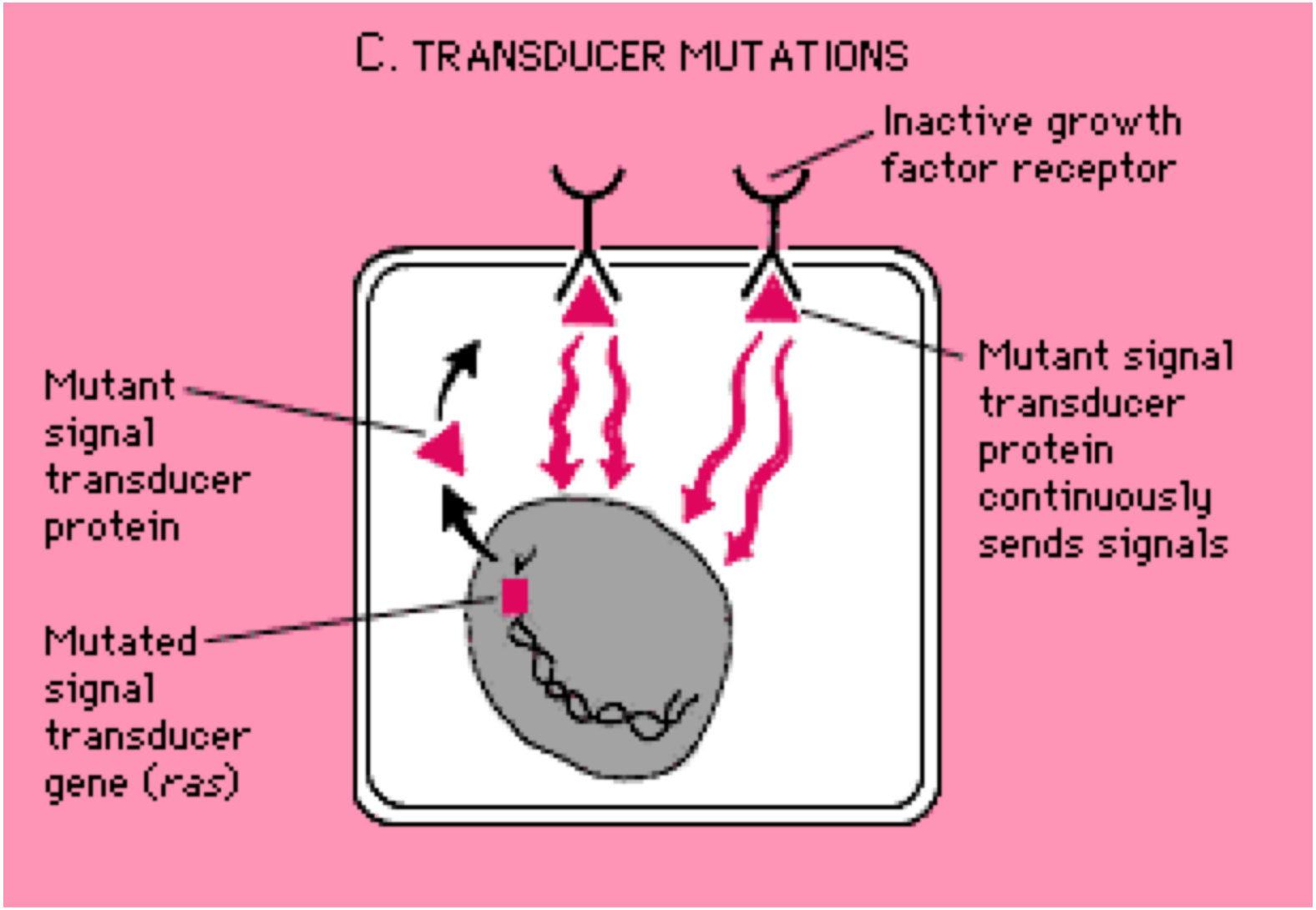
EXAMPLE OF TRANSDUCTION PATHWAY



ZOOM

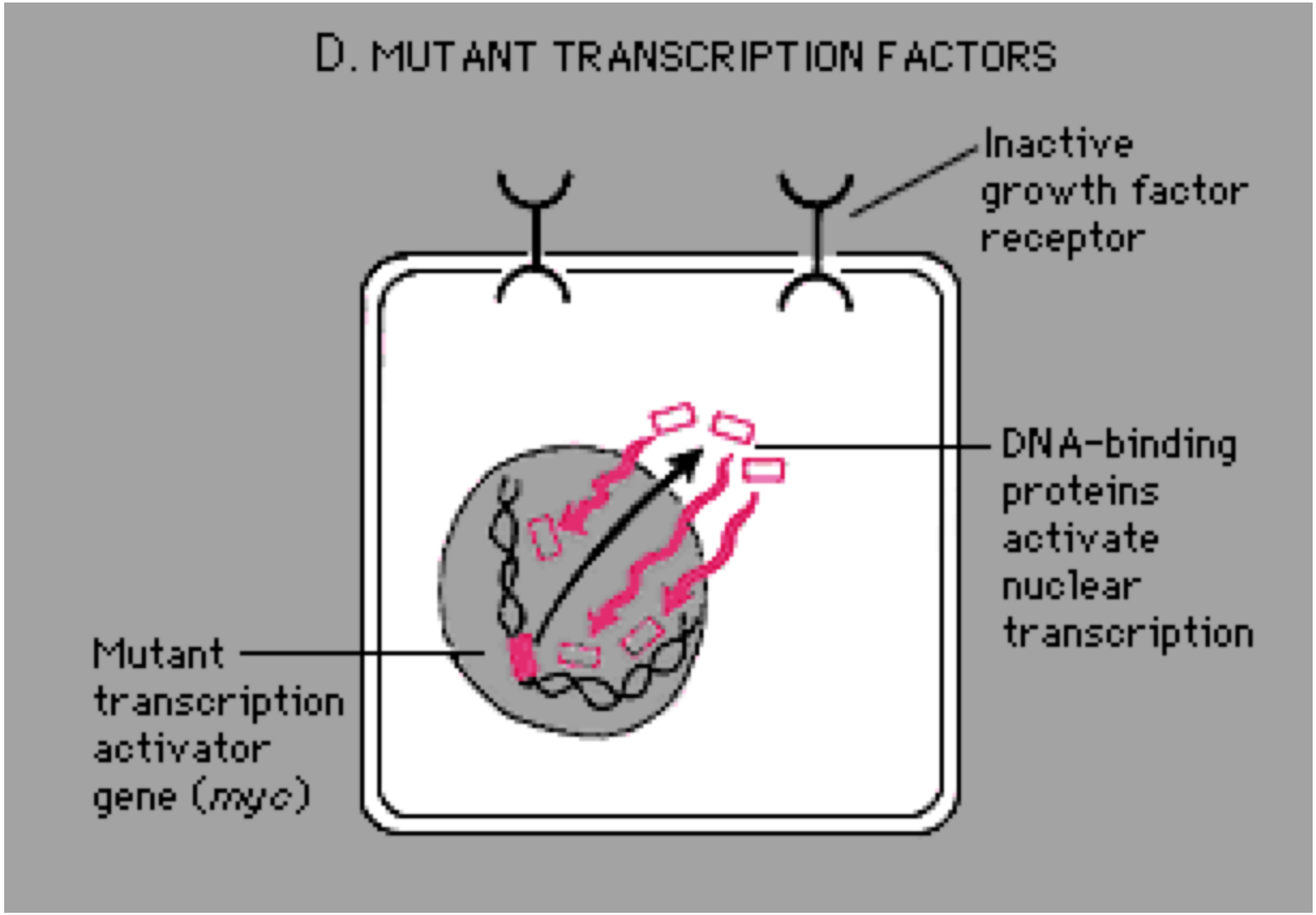
Mechanisms controlling tumor induction by oncogenes

Mutation of signal transduction genes



Mechanisms controlling tumor induction by oncogenes

Mutations of genes encoding for transcription factors



CANCER GENES

ALTERATION OF GENES THAT IS "FIXED" IN THE DNA AND TRANSMITTED TO DAUGHTER CELLS

1) GENES THAT POSITIVELY REGULATE PROLIFERATION:

PROTOONCOGENES ⟶ ONCOGENES

MECHANISM: MUTATIONS, GENIC FUSION, AMPLIFICATION

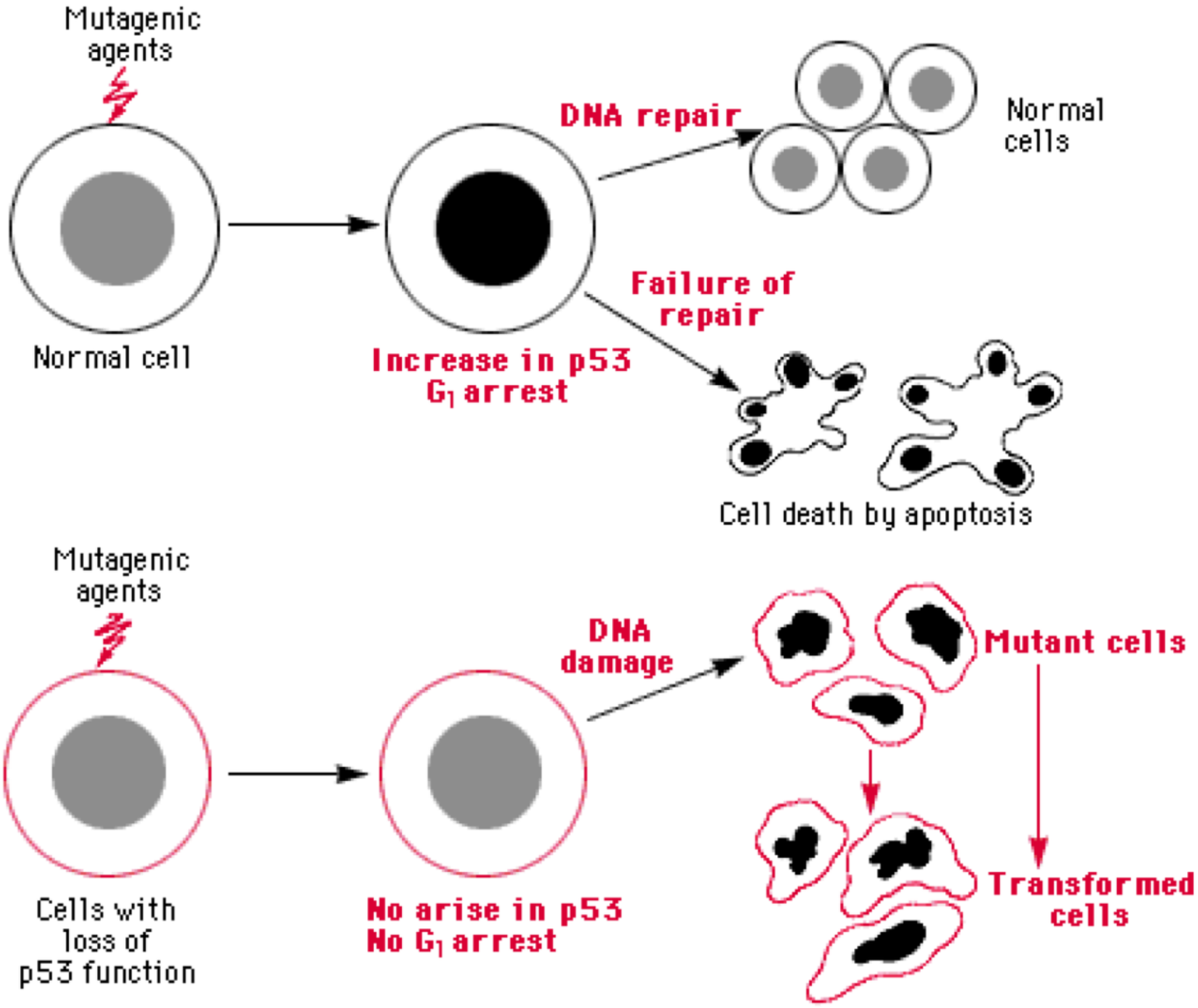
2) GENES THAT NEGATIVELY REGULATE PROLIFERATION:

ONCOSUPPRESSOR GENES

MECHANISM: GENE INACTIVATION of both alleles (mutation, deletion, etc)

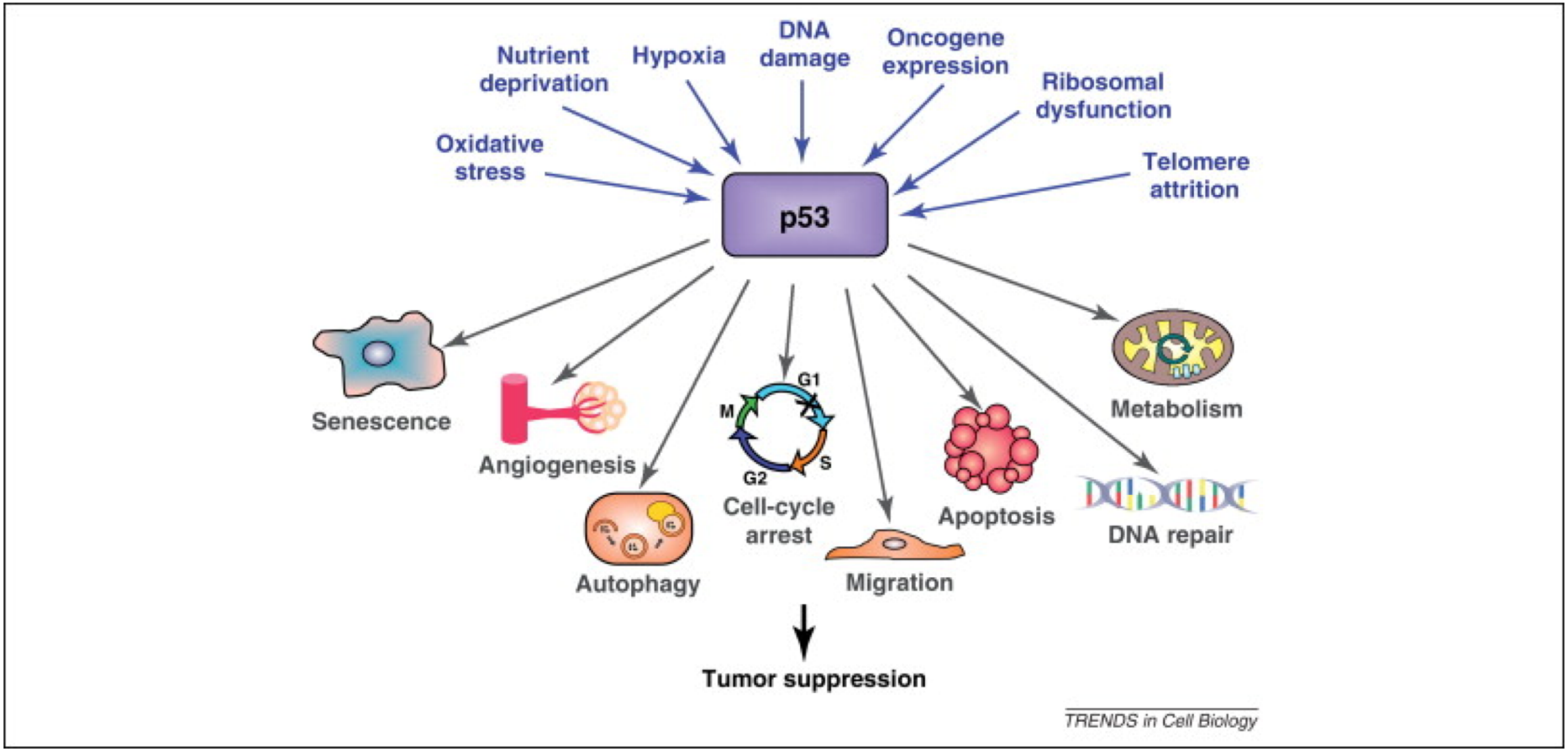
- *Proto-oncogenes:* normal cellular genes whose products promote cell proliferation
- *Oncogenes:* mutant or overexpressed versions of proto- oncogenes that function autonomously without a requirement for normal growth-promoting signals

Onco-suppressor genes
Normal and mutated p53 gene



Oncosuppressor genes

Roles of p53 protein



NETWORKS AND CANCER:

Predizione dell'esito del cancro al seno in base alla modularità dinamica delle reti di interazione proteina-proteina

NATURE BIOTECHNOLOGY VOLUME 27 NUMBER 2 FEBRUARY 2009

LETTERS

nature biotechnology

Dynamic modularity in protein interaction networks predicts breast cancer outcome

Ian W Taylor[1,2], Rune Linding[1,3], David Warde-Farley[4,5], Yongmei Liu[1], Catia Pesquita[6], Daniel Faria[6], Shelley Bull[1,7], Tony Pawson[1,2], Quaid Morris[4,5] & Jeffrey L Wrana[1,2]

Impact Factors for journals published by Nature Publishing Group

2013 Impact Factors – released July 2014

At NPG we are committed to serving the needs of scientists and their science. We do this best by selecting and communicating the most important and valuable scientific information to the broadest possible audience. The 2013 Impact Factors reflect NPG's success at doing this, and the exceptional authors and referees that we are privileged to work with. For a summary, please read our **press release**.

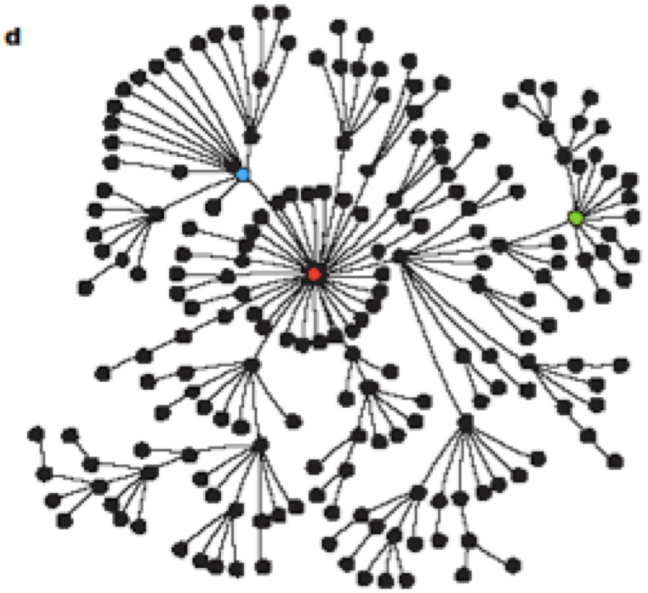
The table lists the 2013 Impact Factor and category ranks for journals published by NPG. Data is taken from the ***2013 Journal Citation Report***, Science Edition (Thomson Reuters, 2014).

A number of journals are listed in more than one category in the *Journal Citation Report*. In these cases, the category in which the journal has highest rank is listed.

Journal	Impact Factor	RANK (by Impact Factor)	CATEGORY
Nature	42.351	1/55	Multidisciplinary Sciences
Nature Communications	10.742	3/55	Multidisciplinary Sciences
Scientific Reports	5.078	5/55	Multidisciplinary Sciences
Scientific American	1.328	15/55	Multidisciplinary Sciences
Nature research journals			
Nature Biotechnology	39.08	1/165	Biotechnology & Applied Microbiology
Nature Cell Biology	20.058	7/185	Cell Biology
Nature Chemistry	23.297	4/148	Chemistry, Multidisciplinary
Nature Chemical Biology	13.217	10/291	Biochemistry and Molecular Biology
Nature Climate Change	15.295	2/215	Environmnetal Sciences
Nature Genetics	29.648	2/164	Genetics & Heredity
Nature Geoscience	11.668	1/173	Geosciences, Multidisciplinary
Nature Immunology	24.973	3/144	Immunology
Nature Materials	36.425	1/136	Physics, Applied
Nature Medicine	28.054	1/122	Medicine, Research & Experimental
Nature Methods	25.953	1/78	Biochemical Research Methods
Nature Nanotechnology	33.265	1/73	Nanoscience &, Nanotechnology
Nature Neuroscience	14.976	6/251	Neurosciences
Nature Photonics	29.958	1/82	Optics
Nature Physics	20.603	3/77	Physics, Multidisciplinary
Nature Protocols	7.782	3/78	Biochemical Research Methods
Nature Structural and Molecular Biology	11.633	2/74	Biophysics

TYPES OF HUBS

Scale-free network
(Barabasi)



Party hubs
(intramodular)

Date hubs
(intermodular)

- 'party' hubs function inside modules and coordinate specific cellular processes
- 'date' hubs link together rather different processes and organize the interactome

BACKGROUND: PREVIOUS STUDIES HAVE SHOWN THAT:

- **Gene expression is altered in tumor cells comparing to normal cells**
- **Proteins encoded by genes which are upregulated in cancer have a higher degree (lung carcinoma with squamous cells)**
- **Proteins involved in cancer have higher number of interactions**

QUESTION: does the altered gene expression in cancer affect interactome organization and thus influences disease prognosis

APPLICATION PHASES:

1. **General identification and characterization of hubs in PPI networks**

2. **Evaluation of the general importance of hubs in PPI networks**

3. **Characterization of hubs in cancer**

4. **Prediction of cancer evolution through dynamic properties of PPI networks**

5. **Conclusions**

1. General identification and characterization of hubs in PPI networks

a. **Identification of hubs** in 3 databases of protein-protein interactions:

1. OPHID
2. MINT
3. STRING

Hub = nodes with at least 5 links/interactors

b. **Study of dynamic modularity of the network: Quantification of hub co-expression with its direct** neighbours/partners using genome-wide co-expression in 79 human tissues

OPHID – I2D database

OPHID database 2005

BIOINFORMATICS ORIGINAL PAPER

Vol. 21 no. 9 2005, pages 2076–2082
doi:10.1093/bioinformatics/bti273

Databases and ontologies

Online Predicted Human Interaction Database

Kevin R. Brown[1,2] and Igor Jurisica[1,2,3,*]

[1]Division of Cancer Informatics, Ontario Cancer Institute, [2]Department of Medical Biophysics, and [3]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

Received on September 23, 2004; revised on January 10, 2005; accepted on January 11, 2005
Advance Access publication January 18, 2005

ABSTRACT

Motivation: High-throughput experiments are being performed at an ever-increasing rate to systematically elucidate protein–protein interaction (PPI) networks for model organisms, while the complexities of higher eukaryotes have prevented these experiments for humans.

Results: The Online Predicted Human Interaction Database (OPHID) is a web-based database of predicted interactions between human proteins. It combines the literature-derived human PPI from BIND, HPRD and MINT, with predictions made from *Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster* and *Mus musculus*. The 23 889 predicted interactions currently listed in OPHID are evaluated using protein domains, gene co-expression and Gene Ontology terms. OPHID can be queried using single or multiple IDs and results can be visualized using our custom graph visualization program.

Availability: Freely available to academic users at http://ophid.utoronto.ca, both in tab-delimited and PSI-MI formats. Commercial users, please contact I.J.

Contact: juris@ai.utoronto.ca

Supplementary information: http://ophid.utoronto.ca/supplInfo.pdf

providing ready access to the known human interactions, they do little to expand the knowledge of the interactome. Several databases have also been published that make predictions about the functional relationships between proteins based on a variety of *in silico* methods (Predictome, STRING, Prolinks, POINT) (Bowers *et al*., 2004; Huang *et al*., 2004; Mellor *et al*., 2002; von Mering *et al*., 2003).

The Online Predicted Human Interaction Database (OPHID) was designed to extend the human interactome using model organism data and to provide a repository for already known, experimentally derived human PPIs. While these predictions should be thought of as hypotheses until experimentally validated, there is increasing evidence that PPIs are conserved through evolution (Pagel *et al*., 2004; Wuchty *et al*., 2003). OPHID catalogs 16 034 known human PPIs obtained from BIND, MINT and HPRD, and makes predictions for 23 889 additional interactions.

Multiple types of evidence have been used in the literature both to support experimentally derived PPIs and to predict interactions *in silico*. Examples include domain–domain co-occurrence (Deng *et al*., 2002; Sprinzak and Margalit, 2001), gene co-expression (Bader

I2D 2011

http://ophid.utoronto.ca/ophidv2.204/

Interologous Interaction Database (I2D)

Welcome to I2D!

To faciliate experimentation and integrated computational analysis with model organism PPI networks, we have integrated known, experimental and predicted PPIs for five model organisms and human in the I2D database.

I2D is developed and maintained by **Jurisica Lab** at Ontario Cancer Institute, PMH. I2D will continue to expand as new protein-protein interaction data becomes available.

Statistics

Source Interactions:	308,402
Predicted Interactions:	386,847
Total Interactions:	681,404

INTEROLOGOUS INTERACTION DATABASE

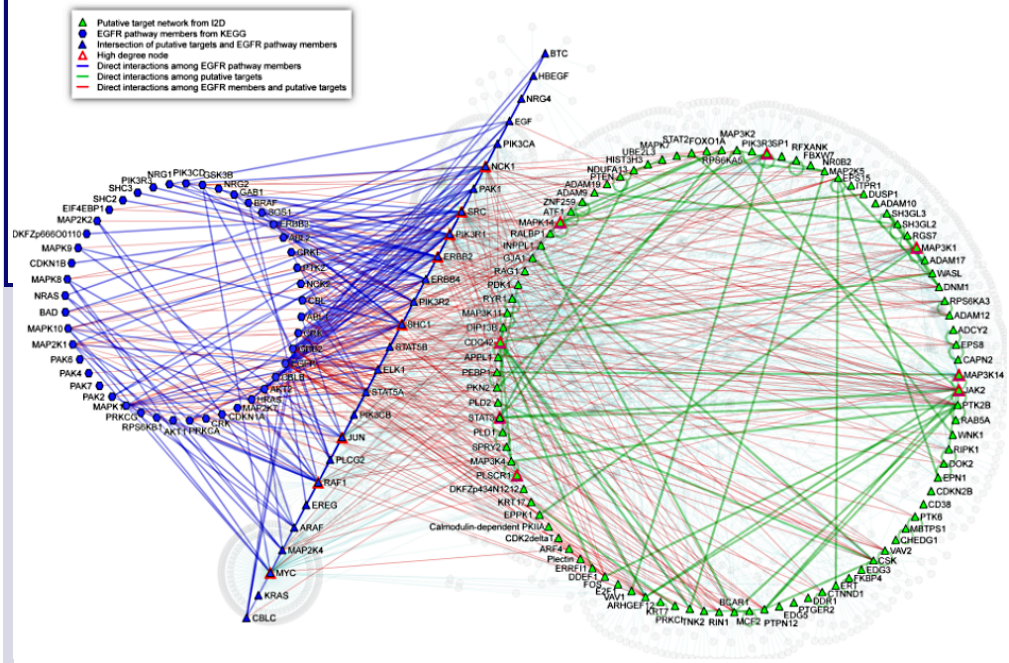
Powered by IBM technology

2015

Welcome to I2D!

To faciliate experimentation and integrated computational analysis with model organism PPI networks, we have integrated known, experimental and predicted PPIs for five model organisms and human in the I2D database.

I2D is developed and maintained by **Jurisica Lab** at Ontario Cancer Institute, PMH. I2D will continue to expand as new protein-protein interaction data becomes available.



Savas, S., Geraci, J., **Jurisica, I.**, Liu, G. A comprehensive catalogue of functional genetic variations in the EGFR pathway: Protein-protein interaction analysis reveals novel genes and polymorphisms important for cancer research. *Int J Cancer*,125(6): 1257-65, 2009.

Statistics

Source Interactions:	463,346
Predicted Interactions:	460,948
Total Interactions:	900,529

Database Access

The latest I2D version 2.3 is available for download in its entirety.

Download

I2D can also be queried online via a web interface.

Search

Visualization

NAViGaTOR is a powerful graphing application for the 2D and 3D visualization of biological networks

When I2D is queried, it can output data in several formats one of which is a NAViGaTOR compatible file. This file can be opened up in NAViGaTOR for visualization and further analysis. I2D can also be queried from within NAViGaTOR.

Visualize with NAViGaTOR!

Molecular INTeraction database

https://mint.bio.uniroma2.it/

go to: **HomoMINT**: an inferred human network | **Domino**: a domain peptide interactions database | **VirusMINT**: a virus protein interactions database

MINT | Home | Search | Curation | Statistics | Download | Contacts/Links/Linking

Statistics:
- 199787 interactions
- 33494 proteins
- 4564 pmids

FEBS Letters special issue: **the Digital, Democratic Age of Scientific Abstracts**

The spreadsheet for data submission to the FEBS Letters experiment: is available here

Welcome to MINT, the Molecular INTeraction database. MINT focuses on **experimentally verified protein-protein interactions** mined from the scientific literature by expert curators. The full MINT dataset can be freely downloaded.

The curated data can be analyzed in the context of the high throughput data and viewed graphically with the **'MINT Viewer'**.

MINT has signed the **IMEx agreement** (http://www.imexconsortium.org/) to share curation efforts and supports the Protein Standard Initiative (PSI) recommendation.

FEBS Letters and the FEBS Journal in collaboration with MINT enhance the content of their articles with the addition of Structured Digital Abstracts

Please, in any articles making use of the data extracted from MINT, refer to *MINT, the molecular interaction database: 2009 update. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G. Nucleic Acids Res. 2010 Jan;38(Database issue):D532-9. Epub 2009 Nov 6.*[Abstract]

Posted by Admin on 2011/03/15:
- **Added 2011.03 UniProt API version**
- **Added Psicquic query results to MINT search output**

A! non confondere MINT con MIPS!

The Institute of Bioinformatics and Systems Biology (IBIS) is part of the **Helmholtz Zentrum München** - German Research Center for Environmental Health and hosts the **Munich Information Center for Protein Sequences (MIPS)**

http://mips.helmholtz-muenchen.de/proj/ppi/

The MIPS Mammalian Protein-Protein Interaction Database

The MIPS Mammalian Protein-Protein Interaction Database is a collection of manually curated high-quality PPI data collected from the scientific literature by expert curators. We took great care to include only data from individually performed experiments since they usually provide the most reliable evidence for physical interactions.

Search the database

To suit different users needs we provide a variety of interfaces to search the database:

- Expert interface – Simple but powerful boolean query language.
- PPI search form – Easy to use PPI search
- Protein search – Just find proteins of interest in the database

String database di interazioni proteina-proteina

http://string-db.org/

Home · Download · Help/Info

STRING 9.0

STRING - Known and Predicted Protein-Protein Interactions

search by name | search by protein sequence | multiple names | multiple sequences

protein name: (examples: #1 #2 #3)

(STRING understands a variety of protein names and accessions; you can also try a random entry)

organism:

auto-detect

interactors wanted:

COGs | Proteins

Reset | GO !

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions.
The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context | High-throughput Experiments | (Conserved) Coexpression | Previous Knowledge

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

More Info Funding / Support Acknowledgements Use Scenarios

STRING *(Search Tool for the Retrieval of Interacting Genes/Proteins)* is being developed at CPR, EMBL, SIB, KU, TUD and UZH.
STRING references: Szklarczyk et al. 2011 / 2009 / 2007 / 2005 / 2003 / Snel et al. 2000.
Miscellaneous: Access Statistics, Robot Access Guide, STRING/STITCH Blog, Supported Browsers.

What's New? This is version 9.0 of STRING - now covering more than 1100 organisms (and counting) !
Sister Projects: check out STITCH and eggNOG - two sister projects built on STRING data!
Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our old releases.

ABOUT

STRING Database — Content

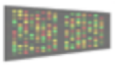
STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:

Genomic Context Predictions

High-throughput Lab Experiments



(Conserved) Co-Expression

Automated Textmining

Previous Knowledge in Databases

Coverage

The STRING database currently covers 9'643'763 proteins from 2'031 organisms.

SIB - Swiss Institute of Bioinformatics

CPR - NNF Center for Protein Research

EMBL - European Molecular Biology Laboratory

A gene atlas of the mouse and human protein-encoding transcriptomes

Andrew I. Su*†, Tim Wiltshire*†, Serge Batalov*†, Hilmar Lapp*, Keith A. Ching*, David Block*, Jie Zhang*, Richard Soden*, Mimi Hayakawa*, Gabriel Kreiman*‡, Michael P. Cooke*, John R. Walker*, and John B. Hogenesch*§¶

*The Genomics Institute of the Novartis Research Foundation, 10675 John J. Hopkins Drive, San Diego, CA 92121; and §Department of Neuropharmacology, The Scripps Research Institute, 10550 North Torrey Pines Road, San Diego, CA 92037

Edited by Peter K. Vogt, The Scripps Research Institute, La Jolla, CA, and approved March 2, 2004 (received for review February 3, 2004)

The tissue-specific pattern of mRNA expression can indicate important clues about gene function. High-density oligonucleotide arrays offer the opportunity to examine patterns of gene expression on a genome scale. Toward this end, we have designed custom arrays that interrogate the expression of the vast majority of protein-encoding human and mouse genes and have used them to profile a panel of 79 human and 61 mouse tissues. The resulting data set provides the expression patterns for thousands of predicted genes, as well as known and poorly characterized genes, from mice and humans. We have explored this data set for global trends in gene expression, evaluated commonly used lines of evidence in gene prediction methodologies, and investigated patterns indicative of chromosomal organization of transcription. We describe hundreds of regions of correlated transcription and show that some are subject to both tissue and parental allele-specific expression, suggesting a link between spatial expression and imprinting.

sion patterns of previously uncharacterized protein-encoding genes and *de novo* gene predictions from the mouse and human genome projects. Using custom-designed whole-genome gene expression arrays that target 44,775 human and 36,182 mouse transcripts, we have built a more extensive gene atlas using a panel of RNAs derived from 79 human and 61 mouse tissues. This data set constitutes one of the largest quantitative evaluations of gene expression of the protein-encoding transcriptome to date.

Building on our previous analyses, these expression patterns were examined for global trends in gene expression. We also provide experimental validation of thousands of gene predictions and use these data to determine which of the commonly used types of evidence for gene prediction most accurately correlates with expressed genes. In addition, we used this data set to search for chromosomal regions of correlated transcription (RCTs), which may indicate higher-order mechanisms of tran-

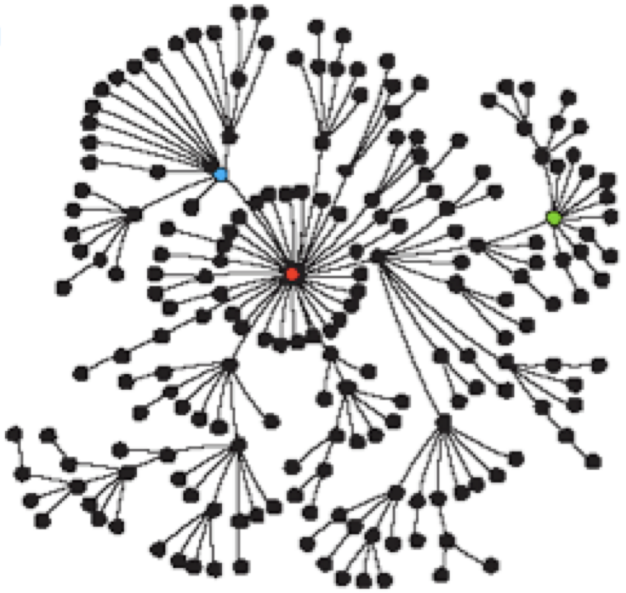
Pearson correlation coefficient (PCC) = descriptor of the degree of linear association between two variables.

The correlation coefficient ranges from −1 to 1. A value of 1 implies that a linear equation describes the relationship between X and Y perfectly, with all data points lying on a line for which Y increases as X increases. A value of −1 implies that all data points lie on a line for which Y decreases as X increases. A value of 0 implies that there is no linear correlation between the variables.

Correlation	Negative	Positive
None	–0.09 to 0.0	0.0 to 0.09
Small	–0.3 to –0.1	0.1 to 0.3
Medium	–0.5 to –0.3	0.3 to 0.5
Large	–1.0 to –0.5	0.5 to 1.0

How are hubs co-expressed with their direct interactors?

Scale-free network
(Barabasi)



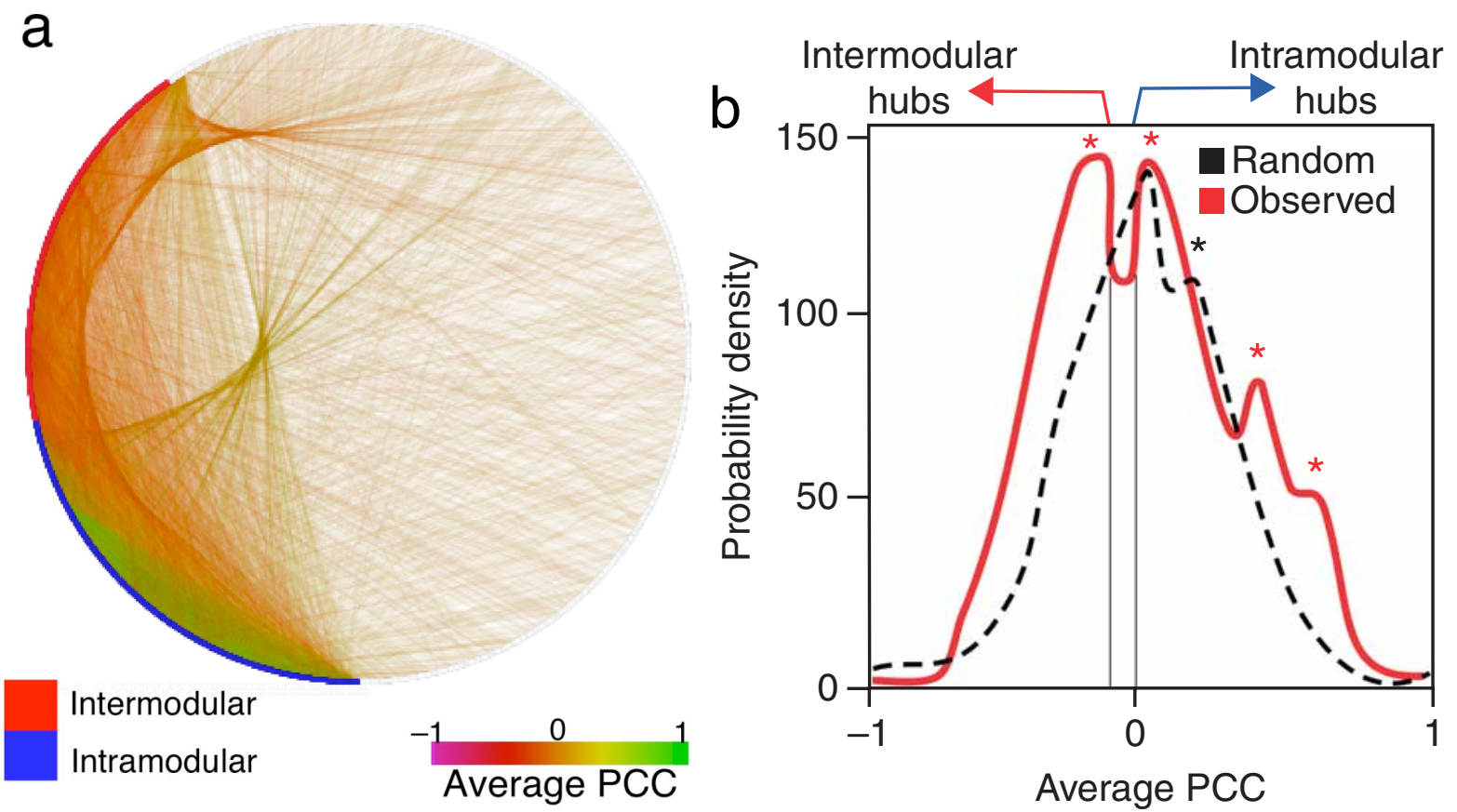
Party hubs
(intramodulari)

Date hubs
(intermodulari)

1. General identification and characterization of hubs in PPI networks

Dynamic modularity of human interactome: multimodal (bimodal) distribution of hubs

(Modularità dinamica dell'interactoma umano: distribuzione multimodale degli hubs)



E' stata calcolata la co-espressione tra ogni hub e i suoi interattori. Sono stati utilizzati dati del database OPHID e dell'atlante di transcrittomica di Su A.I. et al., *Proc. Natl. Acad. Sci., 2004*

Gli edges nella rete organizzata come circonferenza (a) sono colorati in base al valore del PCC

Explanation of the previous slide

Evidence of dynamic modularity in the human interactome (**a**) Network graph of the dynamic modular nature of the human interactome. Intramodular hubs (blue) and intermodular hubs (red) are arranged around the circumference, with interactions shown as edges that are coloured according to the PCC of co-expression of the partner proteins as shown. (**b**) The probability density of the average PCC of co- expression for human hub proteins with their interactors across 79 human tissues (red line) is shown.

A bimodal distribution is apparent for the observed data whereas a randomization of the same data result in a unimodal distribution (dashed black line).

In matematica, una **funzione di densità di probabilità** (o pdf dall'inglese *probability density function*) è la funzione di probabilità di una variabile casuale nel caso in cui la variabile casuale sia continua, cioè l'insieme dei possibili valori ha la potenza del continuo. Essa descrive la "densità" di probabilità in ogni punto nello spazio campionario.

Una distribuzione di probabilità è, in sostanza, una funzione matematica che, per ogni valore della variabile, fornisce la probabilità che venga osservato quel valore.

La distribuzione di probabilità *continua*: il risultato cade in un certo intervallo finito di valori, compreso, ad esempio, fra *a* e *b*. Una tale probabilità, *P(a, b)* si esprime come un integrale:

$$P(a,b) = \int_a^b \phi(x)\,dx$$

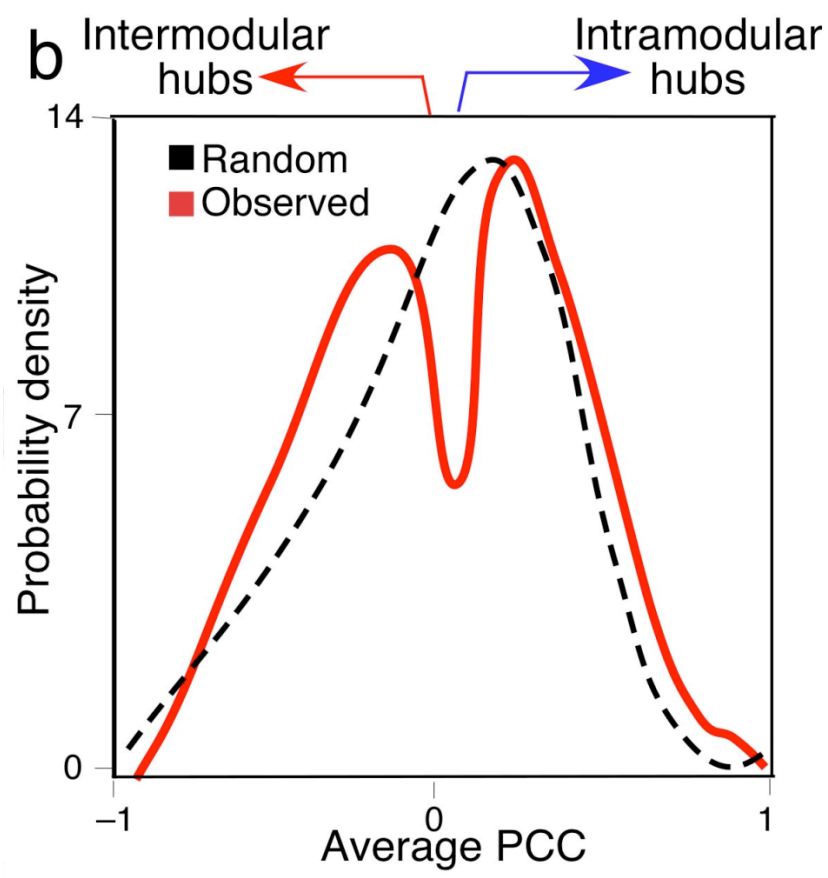
$\phi(x)$ = densità di probabilità = probabilità che il risultato cada in un intervallo *infinitesimamente piccolo* attorno al valore *x* divisa per l'ampiezza di questo intervallo.

http://www.thch.unipg.it/~franc/i/node4.html

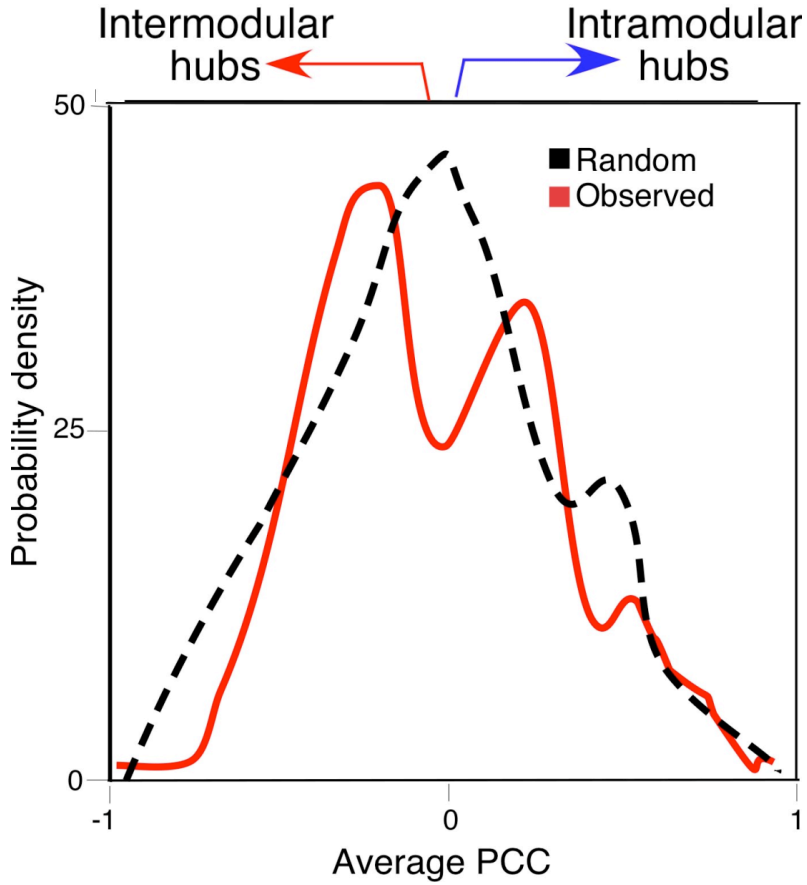
1. General identification and characterization of hubs in PPI networks

Multimodal distribution of hubs identified using databases MINT and STRING

MINT



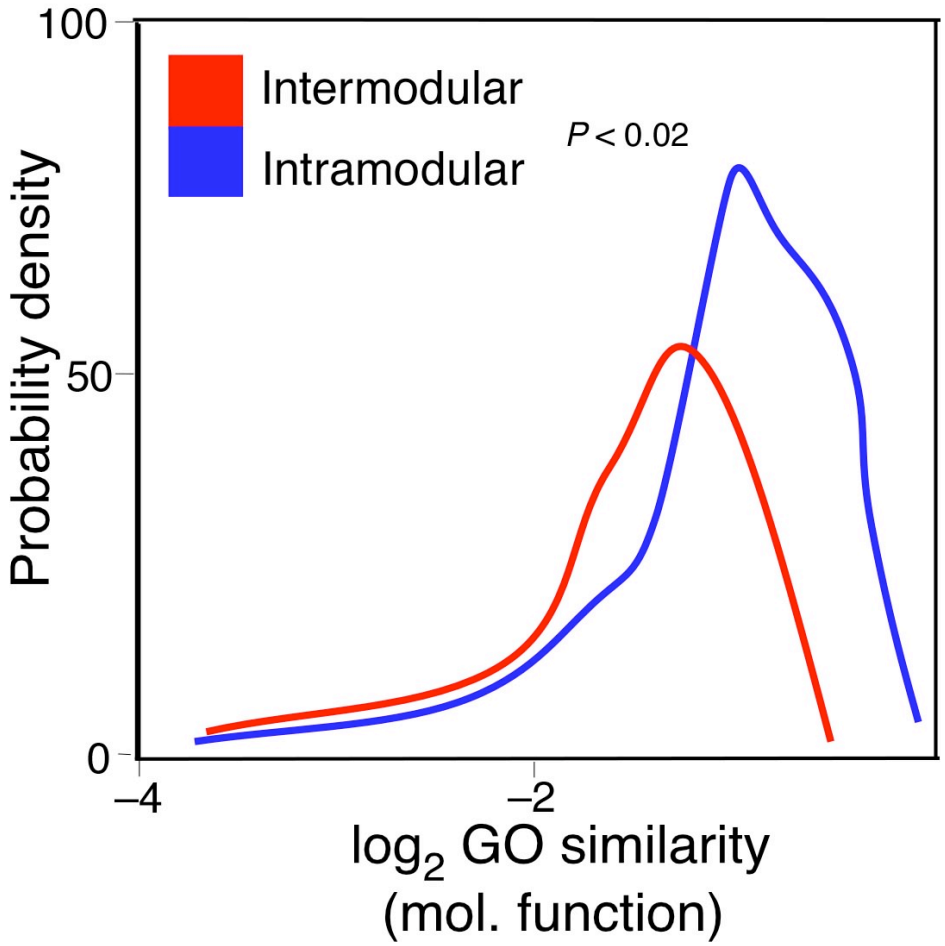
STRING



1. General identification and characterization of hubs in PPI networks

Intramodular hubs have greater GO molecular function similarity with their partners than do intermodular hubs

Gene Ontology (GO) molecular function of either intermodular hubs (red line) or intramodular hubs (blue line) and their partners



Probability density of the semantic similarity (LinGO13) Gene Ontology (GO) molecular function of either intermodular hubs (red line) or intramodular hubs (blue line) is shown.

1. General identification and characterization of hubs in PPI networks

Human interactome has two type of hubs:

Intramodular Hubs (party hubs) = high co-expression in the same tissues with their direct interactors; the interactions are constitutive

Intermodular Hubs (date hubs) = low co-expression in the same tissues with their direct interactors; the interactions depend on the biological context

1. General identification and characterization of hubs in PPI networks

Continuation from previous slide

1. **Multi-modal distribution of** hubs based on co-expression PCC

2. Modular Architetture of the interactome

3. Modular architecture leads to **higher-order functions (funzioni di livello superiore)**: intermodular hubs confer a temporal and spatial links between intramodular hubs, which represent specific functions

In mathematics and computer science, **higher-order functions** are functions which do at least one of the following:
- take one or more functions as an input
- output a higher order function

APPLICATION PHASES:

1. **General identification and characterization of hubs in PPI networks**

2. **Evaluation of the general importance of hubs in PPI networks**

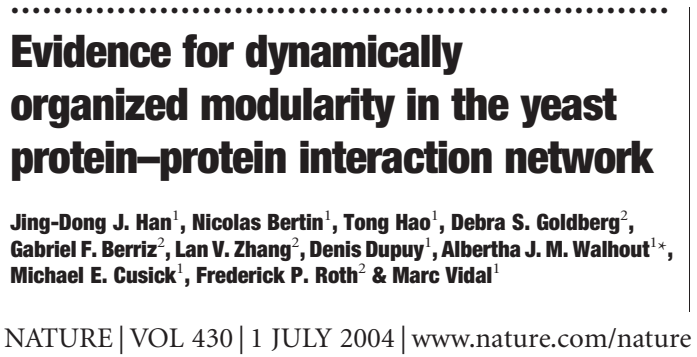
3. **Characterization of hubs in cancer**

4. **Prediction of cancer evolution through dynamic properties of PPI networks**

5. **Conclusions**

2. Evaluation of the importance of hubs in PPI networks

Background: Previous studies suggested a critical role for **Intermodular hubs** in the connectivity of PPI network in yeast



Evidence for dynamically organized modularity in the yeast protein–protein interaction network

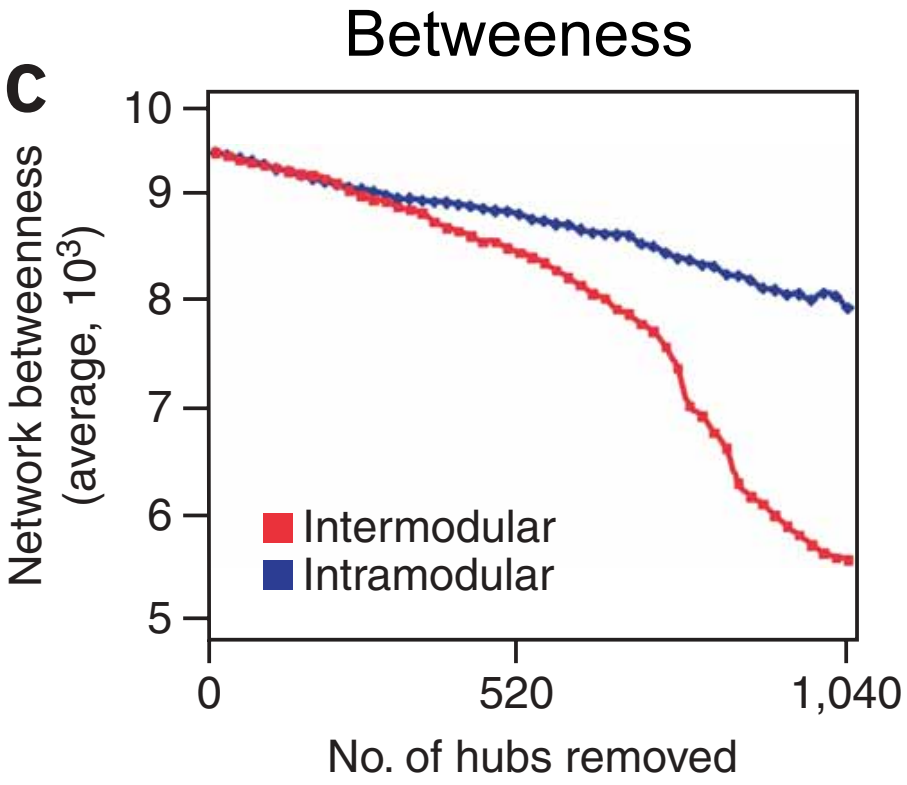
Jing-Dong J. Han[1], Nicolas Bertin[1], Tong Hao[1], Debra S. Goldberg[2], Gabriel F. Berriz[2], Lan V. Zhang[2], Denis Dupuy[1], Albertha J. M. Walhout[1]*, Michael E. Cusick[1], Frederick P. Roth[2] & Marc Vidal[1]

NATURE | VOL 430 | 1 JULY 2004 | www.nature.com/nature

Question: which is the role of intermodular hubs in the global PPI network with relevance for humans?

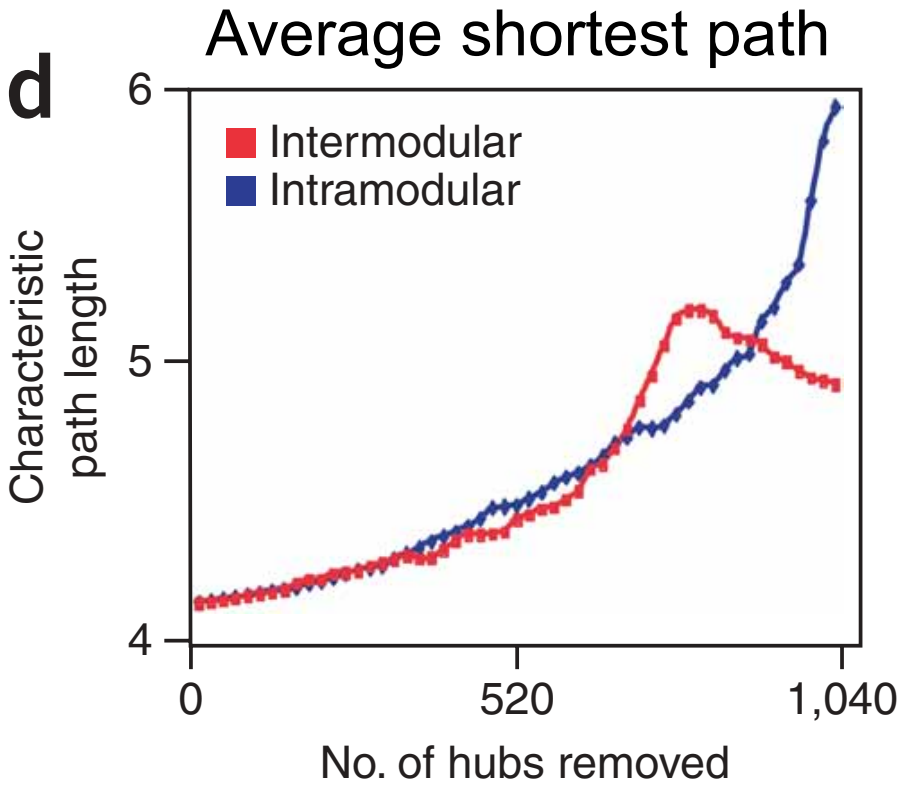
2. Evaluation of the importance of hubs in PPI networks

L'effetto della rimozione *in silico* degli hubs intermodulari dimostra la loro importanza nella connettività globale della rete (valutata misurando due indici di connettività: betweeness e shortest path)



2. Evaluation of the importance of hubs in PPI networks

L'effetto della rimozione *in silico* degli hubs intermodulari dimostra la loro importanza nella connettività globale della rete (valutata misurando due indici di connettività: betweeness e shortest path)



Topological network analysis. Betweeness and shortest path of networks were calculated using algorithms implemented by **the tYNA web interface**. When assessing network robustness to hub removal, an equivalent number of intermodular and intramodular hubs were removed from the network in order of descending clustering coefficient.

Yip, K.Y., Yu, H., Kim, P.M., Schultz, M. & Gerstein, M. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. Bioinformatics 22, 2968–2970 (2006).

tYNA = topnet-like Yale Network

2. Evaluation of the importance of hubs in PPI networks

Higher dependency of betweeness and shortest path on the removal of intermodular hubs suggests that:

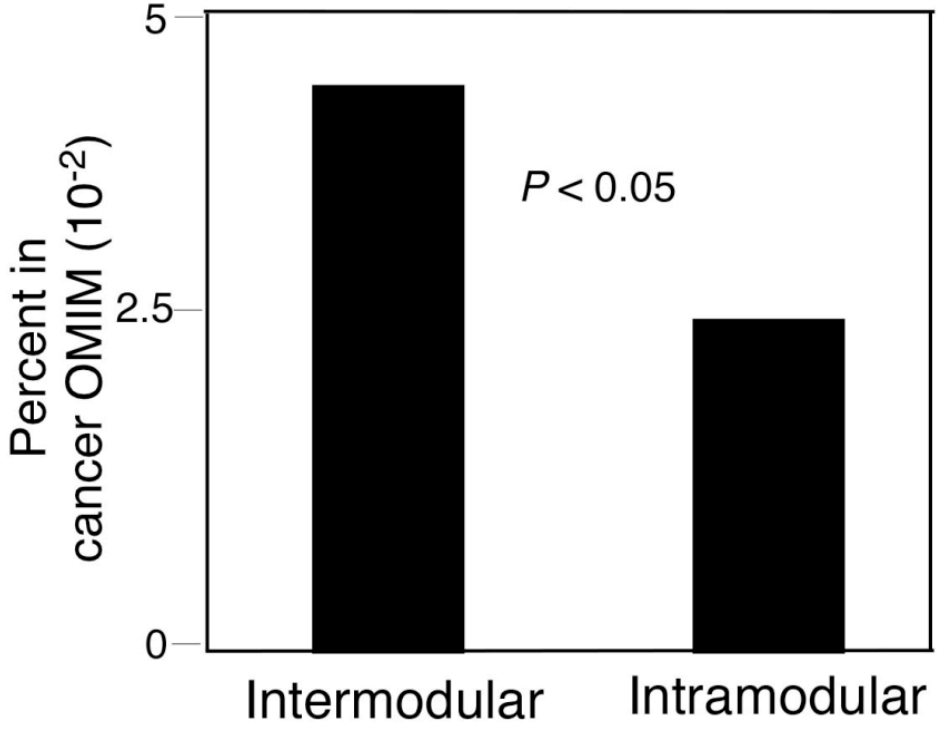
1. Human interactome is modular

2. Intermodular hubs connect functional modules controlled by intramodular hubs

3. Intermodular hubs maintain network architecture

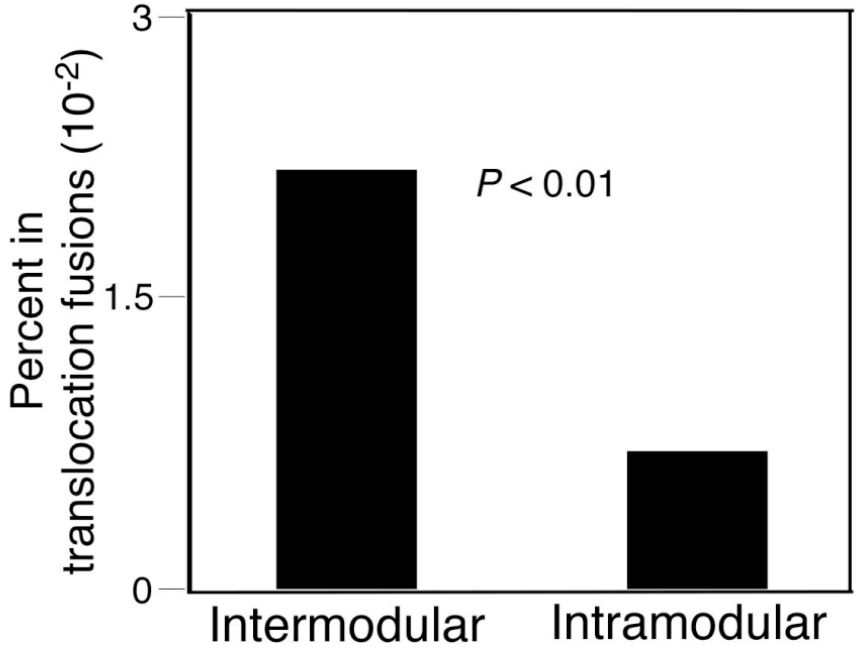
3. Characterization of hubs in cancer

Genes that encode for intermodular hubs are more frequently associated with *entries* related with cancer in OMIM database



3. Characterization of hubs in cancer

Genes that encode for intermodular hubs are more frequently associated with translocations/fusions in OMIM database



3. Characterization of hubs in cancer

The results obtained from OMIM suggest that intermodular hubs are more important in cancer comparing to intramodular hubs

Question: are alterations in modularity present in PPI networks in cancer?

We need data from patients with cancer!

La prognosi (dal greco: pro-, "prima" + gnòsis, "conoscere, sapere") è un giudizio di previsione sul probabile andamento della malattia. Viene formulata dal medico una volta fatta la diagnosi, prendendo in considerazione le condizioni del malato, le possibilità terapeutiche, le possibili complicazioni o le condizioni ambientali.

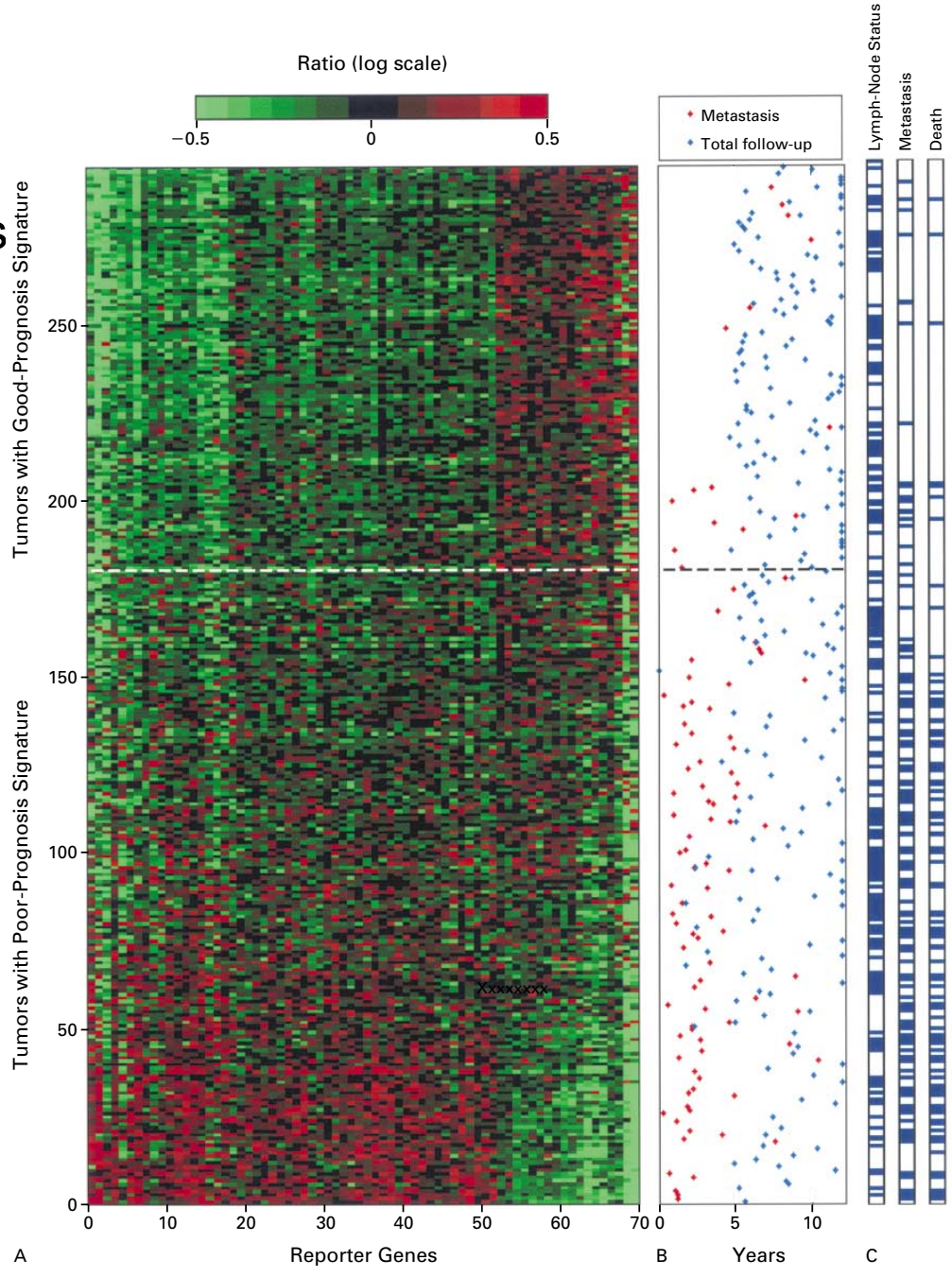
Prognosis: prediction of the probable course and outcome of a disease.

3. Characterization of hubs in cancer

Pattern of Expression of Genes Used to Determine the Prognosis and Clinical Characteristics of 295 Patients with Breast Cancer
(van de Vijver et al., 2002)

(studio precedente effettuato su 295 pazienti con cancro al seno che identifica i geni che predicono la sopravvivenza (esito favorevole o sfavorevole)

Nel panel B viene mostrato il tempo fino alla prima metastasi (in rosso) e il tempo del follow up (in blu)
Nel panel C sono evidenziati i pazienti con metastasi linfonodale o in altri organi (a distanza) o deceduti



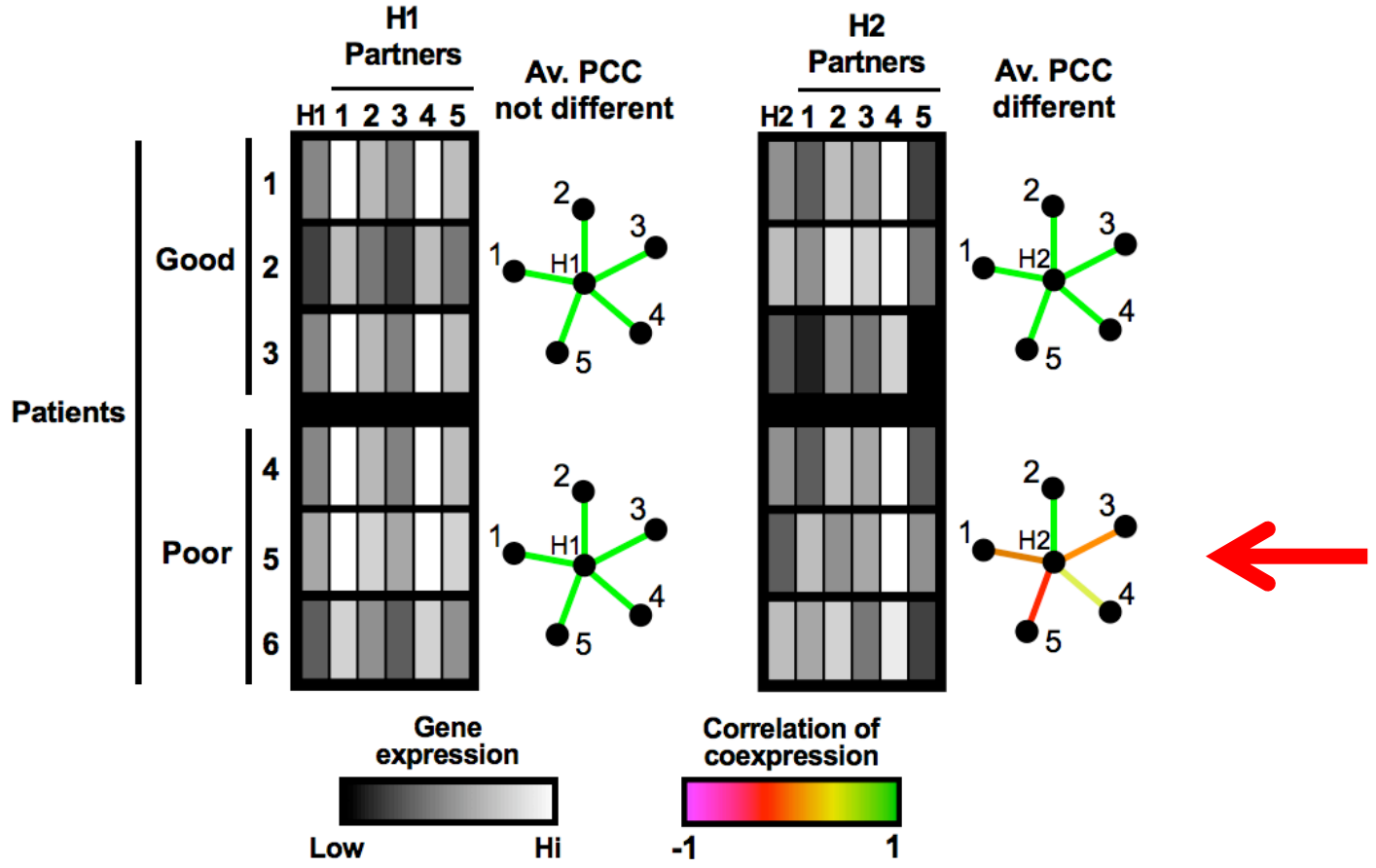
Explanation of previous slide

Pattern of Expression of Genes Used to Determine the Prognosis and Clinical Characteristics of 295 Patients with Breast Cancer.
Panel A shows the pattern of expression of the 70 marker genes (also referred to as prognosis-classifier genes) in a series of 295 consecutive patients with breast carcinomas. Each row represents the prognostic profile of the 70 marker genes for one tumor, and each column represents the relative level of expression of one gene. The tumors are numbered from 1 to 295 on the y axis, and the genes are numbered from 1 to 70 on the x axis. Red indicates a high level of expression of messenger RNA (mRNA) in the tumor, as compared with the reference level of mRNA, and green indicates a low level of expression. The dotted line is the previously determined threshold between a good-prognosis signature and a poor-prognosis signature. Tumors are rank-ordered according to their correlation with the previously determined average profile in tumors from patients with a good prognosis. Panel B shows the time in years to distant metastases as a first event for those in whom this occurred, and the total duration of follow-up for all other patients. Panel C shows the lymph-node status (blue marks indicate lymph-node–positive disease, and white lymph-node–negative disease), the number of patients with distant metastases as a first event (blue marks), and the number of patients who died (blue marks).

3. Characterization of hubs in cancer

Mean PCCs (co-expresion) between hubs and their interactors in patients with good *versus* poor prognosis

Example of analysis



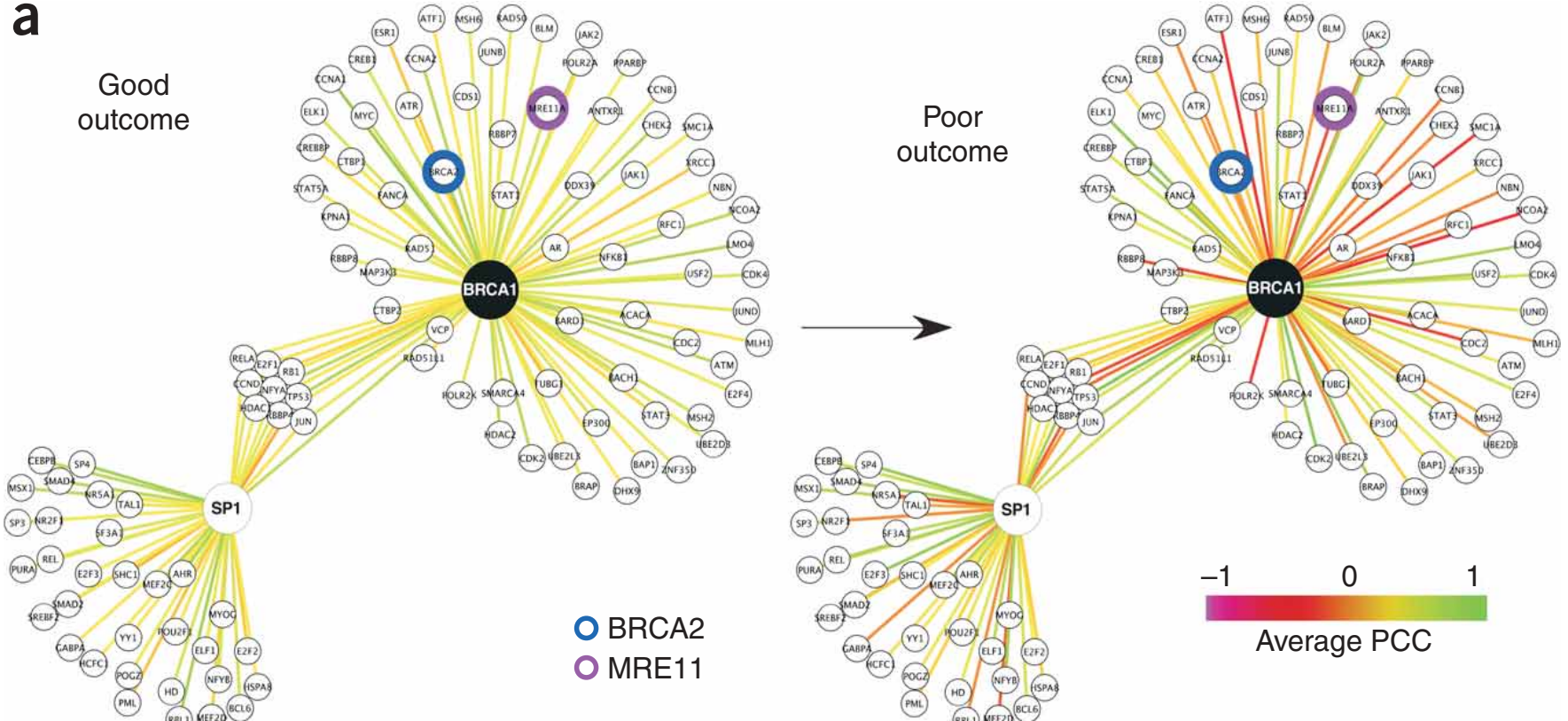
Explanation of previous slide

Schematic of dynamic network modularity associated with breast cancer outcome. Hypothetical gene expression patterns are shown for two hubs (H1 and H2), each with 5 partners, as indicated. Relative expression for these 12 genes in 6 hypothetical patients, 3 with good prognosis and 3 with poor prognosis is shown (grey scale). The correlation of expression of each partner and its hub in the patient groups is shown by edge colour according to the coloured gradient. In this example, H2 shows a difference in average PCC as a function of disease outcome, whereas H1 does not.

3. Characterization of hubs in cancer

<u>The analysis revealed 256 hubs that displayed altered PCC as a function of disease outcome.</u>

Example: the BRCA1 hub is strongly coexpressed in patients with good prognosis (green edges), whereas is less coexpressed in patients with poor prognosis (many red edges)



Of the BRCA1 partners highly correlated in good outcome tumors, both MRE11 and BRCA2 were notable as they are members of the BRCA1-associated genome surveillance complex (BASC) and are misregulated in poor prognosis breast cancer.

The results suggest that disorganization of the BASC by loss of coordinated co-expression of components is associated with poor outcome.

BRCA1 = Breast cancer type 1 susceptibility protein (BRCA1)
E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and **plays a central role in DNA repair by facilitating cellular responses to DNA damage**. **The E3 ubiquitin-protein ligase activity is required for its tumor suppressor function.** The **BRCA1-BARD1 heterodimer coordinates a diverse range of cellular pathways such as DNA damage repair, ubiquitination and transcriptional regulation to maintain genomic stability.** Regulates centrosomal microtubule nucleation. **Required for normal cell cycle progression from G2 to mitosis**.
BRCA1 partners such as both MRE11 and BRCA2 are members of the BRCA1-associated genome surveillance complex (BASC).

MRE11 = Double-strand break repair protein MRE11A
Alternative name: Meiotic recombination 11 homolog 1
Component of the MRN complex, which **plays a central role in double-strand break repair**, **DNA recombination**, **maintenance of telomere integrity and meiosis**.

BARD-1 (BRCA1-associated RING domain protein 1) is a protein that in humans is encoded by the BARD1 gene. The BARD1-BRCA1 interaction is essential for BRCA1 stability. Mutations in the BARD1 protein that affect its structure appear in many breast, ovarian, and uterine cancers, suggesting the mutations disable BARD1's tumor suppressor function.

BRCA1 and *BRCA2* are human genes that produce tumor suppressor proteins. These proteins help repair damaged DNA ensuring the stability of the genetic material. If mutated, cells are more likely to develop additional genetic alterations that can lead to cancer.

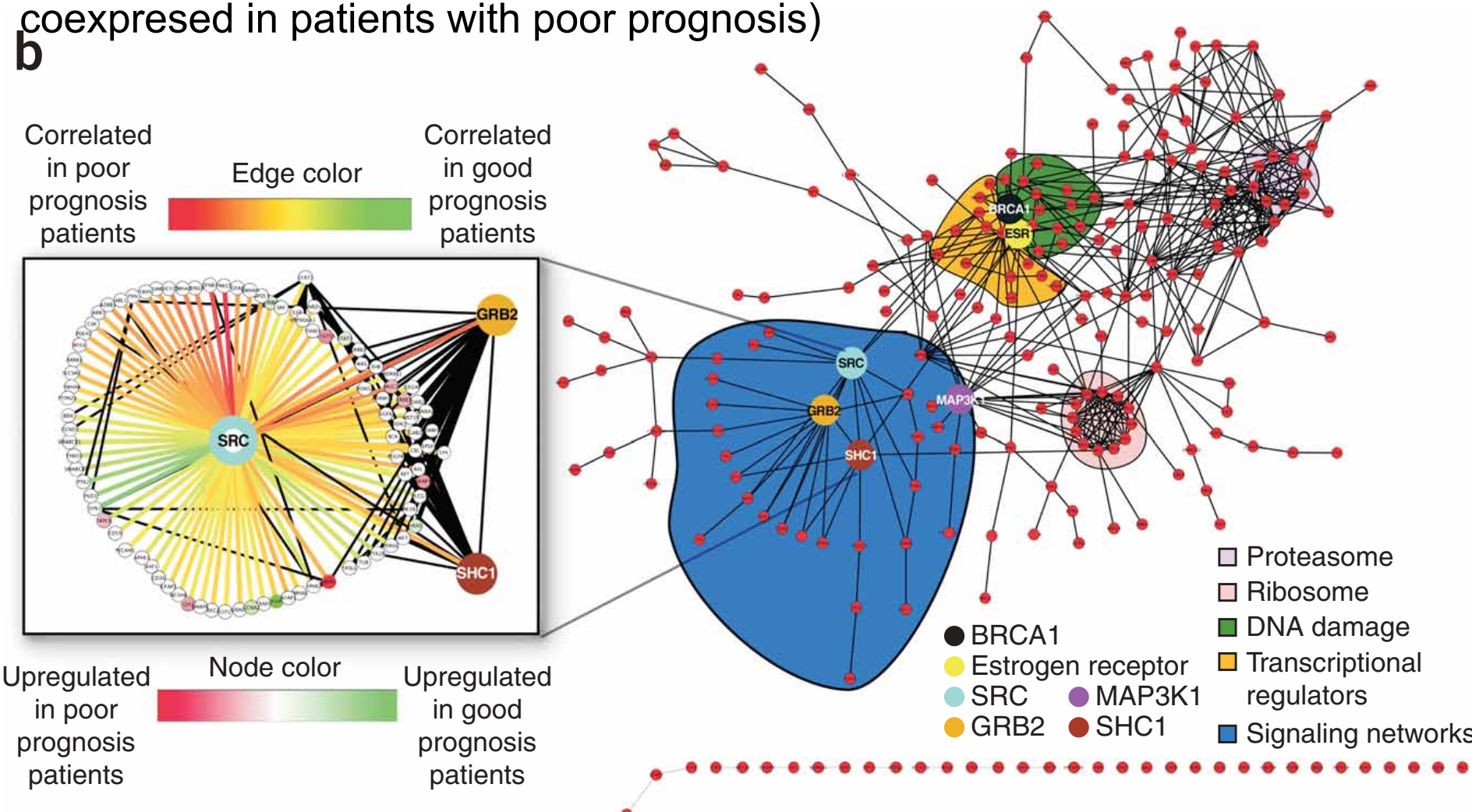
Mutations in BRCA1 and BRCA2 increase the risk of female **breast and ovarian cancers**.
Men with BRCA2 mutations are at increased risk of **breast cancer and prostate cancer**.

Both men and women with harmful BRCA1 or BRCA2 mutations are at increased risk of **pancreatic cancer**.

3. Characterization of hubs in cancer

Differences in the dynamic properties of cancer networks

252 hubs had altered co-expression in patients with poor prognosis and they form an interconnected network (In red –right- nodes differently coexpresed in patients with poor prognosis)



SHC1 (Src homology 2 domain containing) transforming protein 1.
Signaling adapter protein that **couples activated growth factor receptors to signaling pathways** (couples activated receptor tyrosine kinases to Ras pathway)

SRC = v-**src** sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)
Alternative names: tyrosine-protein kinase **SRC;** Proto-oncogene c-**Src**
Src kinases are key upstream mediators of PI3-K and MAPK signaling pathways, and have been shown to have important **roles in cell proliferation, migration and survival**.

Growth factor receptor-bound protein 2 (**Grb2)** is an **adaptor protein** involved in signal transduction/cell communication.

The protein encoded by this gene binds receptors such as the epidermal growth factor receptor. Grb2 function is involved in developmental processes in various organisms and **transformation and proliferation of various cell types.** Targeted gene disruption of Grb2 in mouse is lethal at an early embryonic stage.

Grb2 is best known for its ability to link the epidermal growth factor receptor tyrosine kinase to the activation of Ras and its downstream kinases, ERK1,2.

3. Characterization of hubs in cancer

Differences in the dynamic properties of cancer networks:

The analysis of the hubs with altered PCC shows that:

- The 252 hubs with altered co-expression form an interconnected network

- The majority of hubs (77%) show a significant change in co-expression with their neighbors (thus of connectivity) in patients with poor outcome

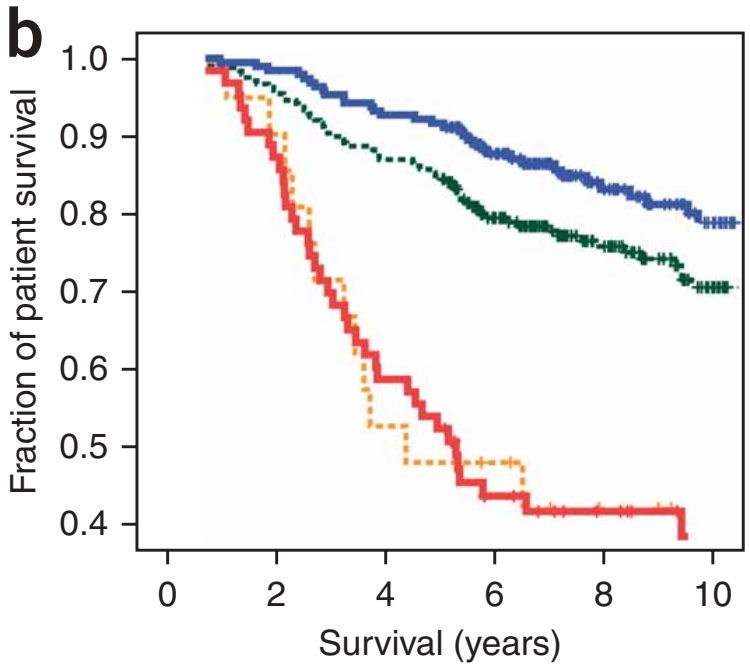
The data show a network dynamic change.
Ex: SRC (inset).

- The network of 252 hubs include molecules such as MAP3K1, GRB2, SHC, SRC, BRCA1, MRE11, estrogen receptor 1 (ESR1) that were incriminated in the pathogenesis of cancer.

4. Prediction of cancer evolution through dynamic properties of PPI networks

First step: computation of the relative expression of hubs with each of their interacting partners, determined for which hubs the relative expression differed significantly between patients who survived *versus* those who died from disease

Second step: application of "**affinity propagation clustering**" algoritm to clusterize patients with similar data (L'algoritmo ha permesso di assegnare un valore di probabilità di prognosi in base alla signature molecolare).



Full model = network probabilities integrated with clinical data: age, tumor stage and tumor grade using a logistic regression model

In statistics, **logistic regression** (sometimes called the logistic model or logit model) **is used for prediction of the probability of occurrence of an event** by fitting data to a logit function- logistic curve. It is a generalized **linear model** used for binomial regression. Like many forms of regression analysis, **it makes use of several predictor variables that may be either numerical or categorical.** For example, the probability that a person has a heart attack within a specified time period might be predicted from knowledge of the person's age, sex and body mass index. **Logistic regression is used extensively in the medical and social sciences fields**, as well as marketing applications such as prediction of a customer's propensity to purchase a product or cease a subscription.

The affinity clustering algoritm

REPORTS

16 FEBRUARY 2007 VOL 315 **SCIENCE** www.sciencemag.org

Clustering by Passing Messages Between Data Points

Brendan J. Frey* and Delbert Dueck

Clustering data by identifying a subset of representative examples is important for processing sensory signals and detecting patterns in data. Such "exemplars" can be found by randomly choosing an initial subset of data points and then iteratively refining it, but this works well only if that initial choice is close to a good solution. We devised a method called "affinity propagation," which takes as input measures of similarity between pairs of data points. Real-valued messages are exchanged between data points until a high-quality set of exemplars and corresponding clusters gradually emerges. We used affinity propagation to cluster images of faces, detect genes in microarray data, identify representative sentences in this manuscript, and identify cities that are efficiently accessed by airline travel. Affinity propagation found clusters with much lower error than other methods, and it did so in less than one-hundredth the amount of time.

5. Conclusions

1. PPI networks have modular architecture.

2. In cancer there is alteration of PPI network modularity: a dynamic change (co-expression) and a “rewiring” of the network

3. Changes in modularity may have prognostic relevance on disease outcome and may improve the predictive value of existing clinical indicators.

4. Multi-modal therapies that target hubs with altered modularity may be effective in patients with cancer.

P4 medicine

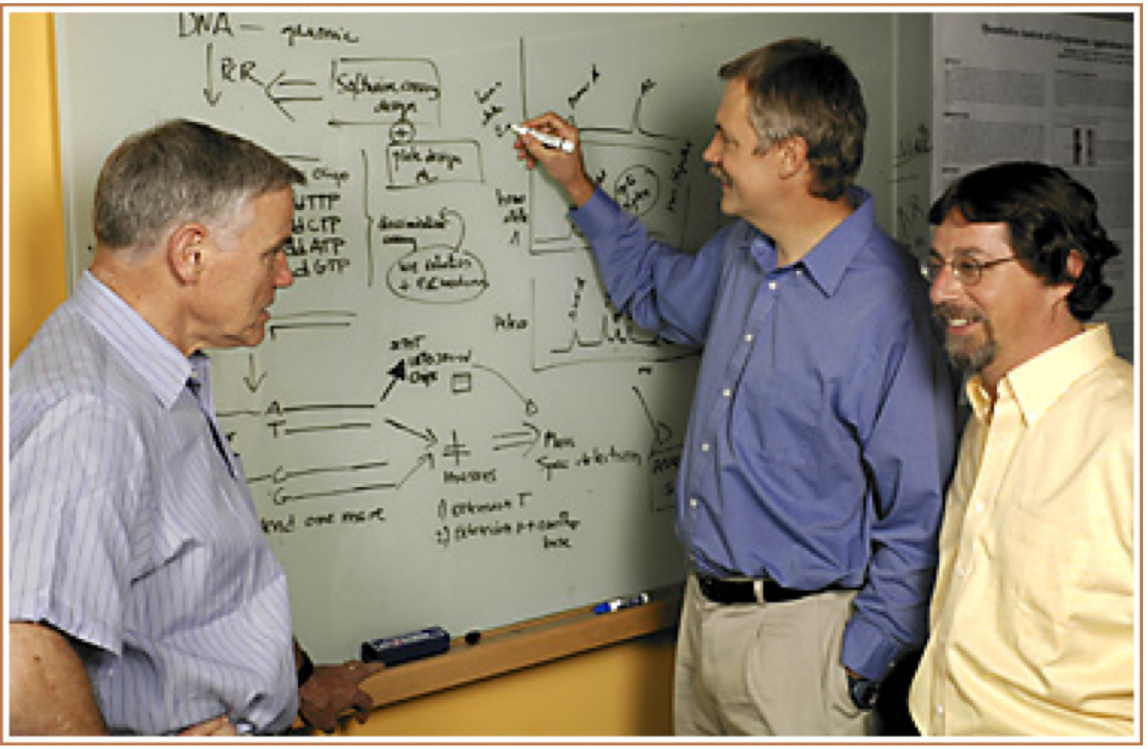
The Institute for Systems Biology, Seattle

"Everyone realizes that biology now requires a multi-science approach and some institutions are attempting to create centers where different disciplines can collaborate on complex biological problems. However, ISB is the only institution that I know of that was founded on and dedicated to the principle of bringing together biologists, mathematicians and engineers, computer scientists and physicists, in an interactive and collaborative environment."

Lee Hartwell, M.D, Ph.D.
Nobel Laureate
President and Director, Fred Hutchinson Cancer Research Center
Seattle, Washington

The Institute for Systems Biology, Seattle

The Institute for Systems Biology, Seattle



Leroy Hood

Ruedi Aebersold

Alan Aderem

"Studying the interactions and interplay of many levels of biological information, systems biology will enable us not only to cure complex diseases but also to predict an individual's health and extend the human body's natural lifespan by preventing diseases. The new era of predictive, preventive, and personalized medicine—made possible by systems biology—represents a profound shift in the practice of medicine and will reach into many corners of our lives."

Leroy Hood, Ph.D, M.D.
President
Institute for Systems Biology

Why systems medicine?

1. Today medicine is largely reactive.

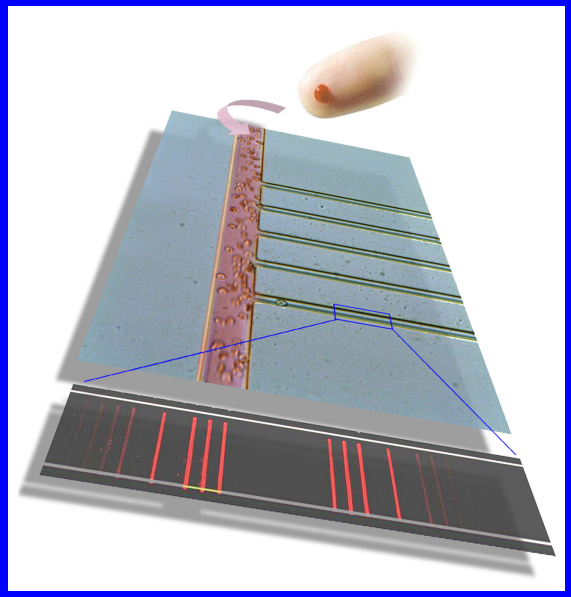
2. Identification of biomarkers allowing diseases to be detected and treated much earlier than is possible today.

3. Generation and analysis of "big data" sets

4. P4 medicine

Predictive, Personalized, Preventive and Participatory (P4) Medicine

- **Driven by systems approaches to disease, new measurement (nanotechnology) and visualization technologies and powerful new computational tools, P4 medicine will emerge over the next 10-20 years**



Predictive, Personalized, Preventive and Participatory (P4) Medicine

- **Its two major objectives are to:**

1. **quantify wellness**

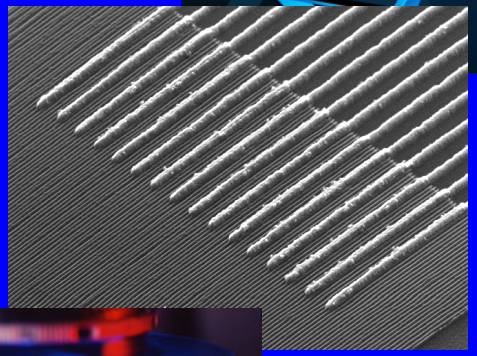
2. **demystify disease**

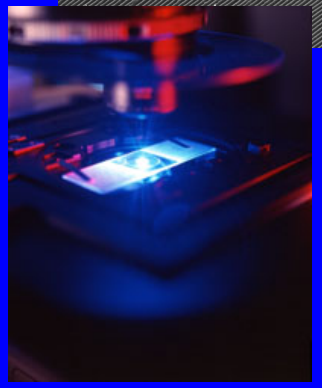
Which are the technologies that will transform systems or P4 medicine?

- High throughput DNA sequencing for individual human genome (for less than 1,000$ in the next years)
- Targeted MRM (multiple reaction monitoring) mass spectrometry for discovery, validation and typing (initially) of blood fingerprints
- Microfluidic protein chip to measure blood organ-specific protein fingerprints and type millions of individuals and assess their key biological networks
- Single-cell analyses--deciphering the interplay of the digital genome and the environment
- In vivo and in vitro molecular imaging to assess disease distribution and follow therapy

P4 Medicine

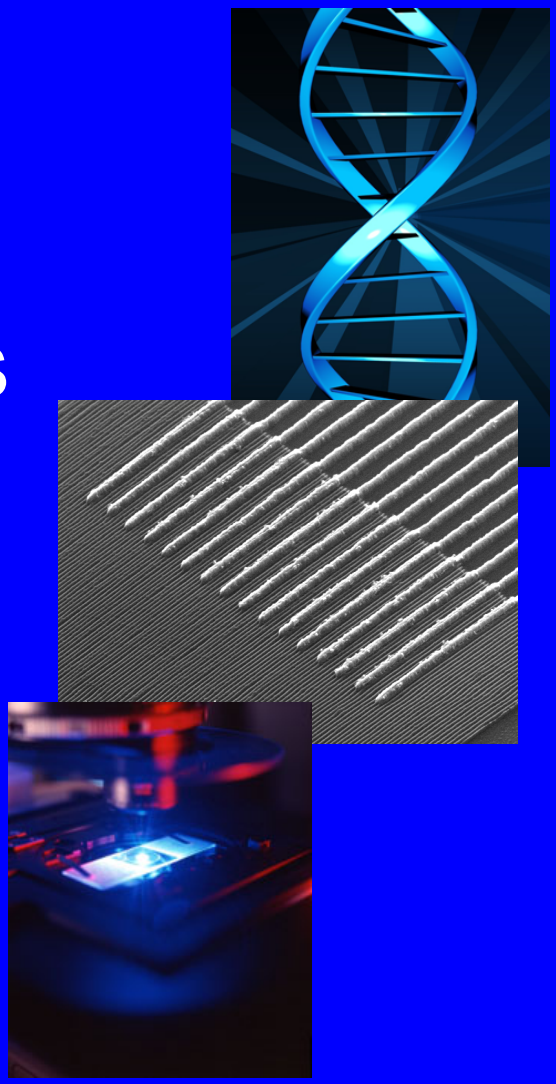
- Predictive:
  - Probabilistic health history--DNA sequence
  - Biannual multi-parameter blood protein measurements
  - In vivo molecular imaging





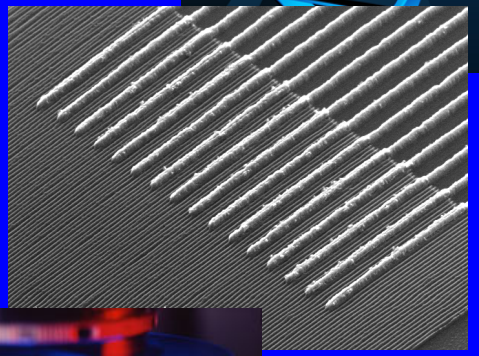
P4 Medicine

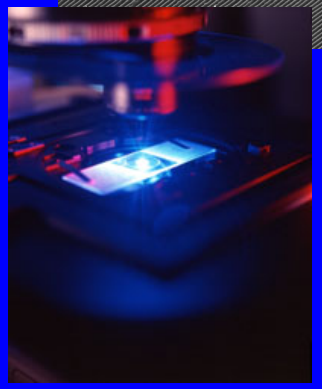
- Preventive:
  - Design of therapeutic and preventive drugs via systems approaches
  - Systems approaches to vaccines will transform prevention of infectious diseases
  - Transition to wellness assessment (from reaction to prevention)



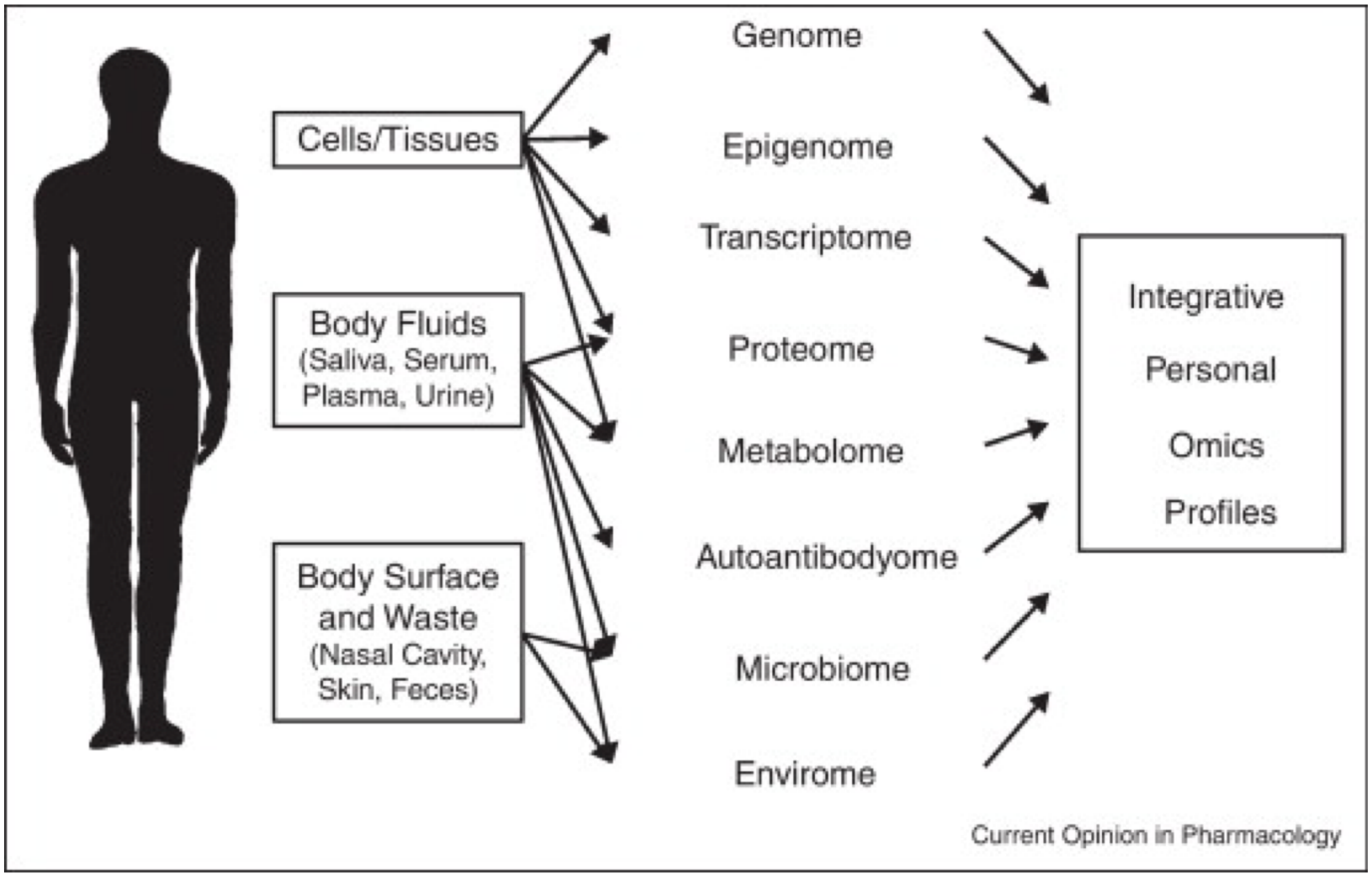
P4 Medicine

- Personalized:
  - Unique individual human genetic variation mandates *individual treatment*
  - Billions of data points on each individual
  - Personalized treatment





Systems biology: personalized medicine for the future?

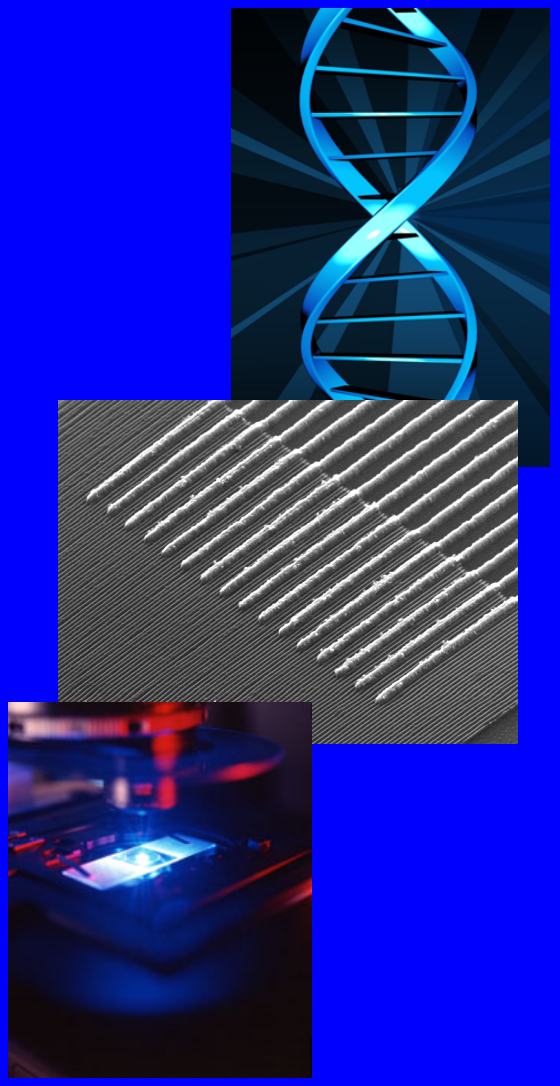


Integrative Personal Omics Profile (iPOP) analysis. Various types of systems data can be generated and integrated with the iPOP analysis. Note that this approach is highly modular and can be tailored to meet specific needs of different studies.

Curr. Opin. Pharmacol. 2012

P4 Medicine

- Participatory:
  - Patient understands and participates in medical choices
  - Patient increasing will make choices with doctor intervention



Scientific wellness embodies
P4 MEDICINE:

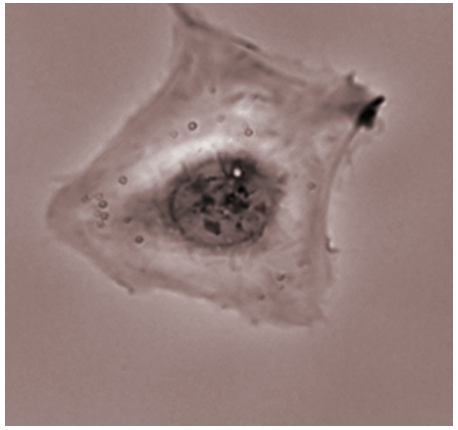
https://www.systemsbiology.org/research/p4-medicine/

Digitalization of Biology and Medicine Will Transform Medicine

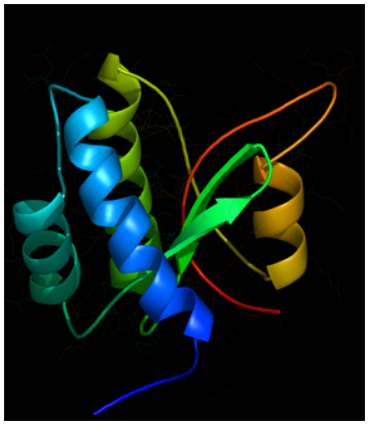
- Analysis of single molecules, single cells and single individuals
- A revolution that will transform medicine even more than digitalization transformed information technologies and communications
- Digitization of medicine will lead to dramatically lower healthcare costs

Single individual

Single cell



Single molecule

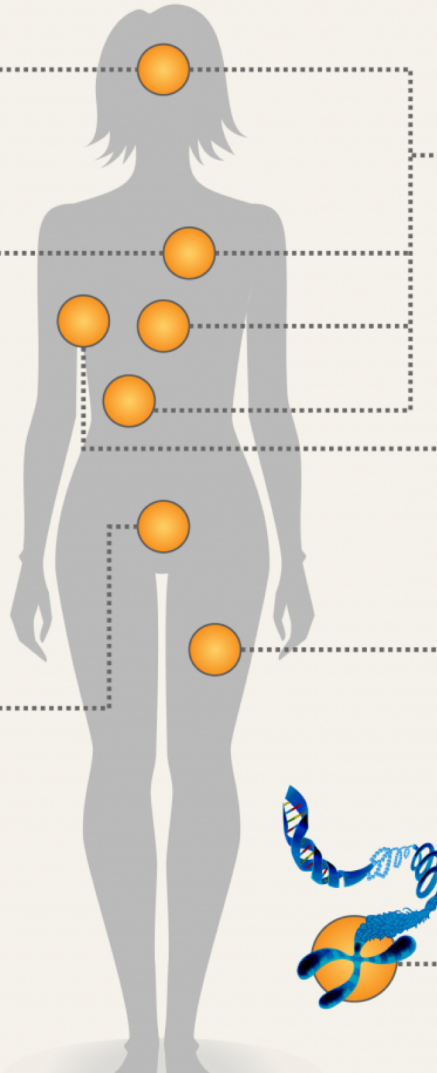


ISB: SCIENTIFIC WELLNESS

100K
100K WELLNESS PROJECT
Institute for Systems Biology

AN EXAMINED LIFE

The longitudinal study collected data at daily and three-month intervals, and allowed personalized interventions -- such as changes in diet -- as the study proceeded.



BRAIN
- **What's measured:** Sleep patterns
- **Frequency:** Daily
- **Method:** Wrist sensor

HEART
- Pulse, physical-activity level
- Daily
- Wrist sensor

COLON
- Microbiome ecology
- Every three months
- Stool sample

LIVER, LUNGS, BRAIN & HEART
- 100 proteins to track organ health
- Every three months
- Blood sample

LYMPHATIC SYSTEM
- Immune-cell activity
- Every three months
- Blood sample

INSULIN SENSITIVITY
- Blood glucose
- Every three months
- Blood sample

CHROMOSOMES
- Whole-genome sequence
- At enrollment
- Blood sensor

Institute for Systems Biology
Revolutionizing Science. Enhancing Life.

A wellness study of 108 individuals using personal, dense, dynamic data clouds

Nathan D Price[1,2,6,7], Andrew T Magis[2,6], John C Earls[2,6], Gustavo Glusman[1], Roie Levy[1], Christopher Lausted[1], Daniel T McDonald[1,5], Ulrike Kusebauch[1], Christopher L Moss[1], Yong Zhou[1], Shizhen Qin[1], Robert L Moritz[1], Kristin Brogaard[2], Gilbert S Omenn[1,3], Jennifer C Lovejoy[1,2] & Leroy Hood[1,4,7]

Personal data for 108 individuals were collected during a 9-month period, including whole genome sequences; clinical tests, metabolomes, proteomes, and microbiomes at three time points; and daily activity tracking. Using all of these data, we generated a correlation network that revealed communities of related analytes associated with physiology and disease. Connectivity within analyte communities enabled the identification of known and candidate biomarkers (e.g., gamma-glutamyltyrosine was densely interconnected with clinical analytes for cardiometabolic disease). We calculated polygenic scores from genome-wide association studies (GWAS) for 127 traits and diseases, and used these to discover molecular correlates of polygenic risk (e.g., genetic risk for inflammatory bowel disease was negatively correlated with plasma cystine). Finally, behavioral coaching informed by personal data helped participants to improve clinical biomarkers. Our results show that measurement of personal data clouds over time can improve our understanding of health and disease, including early transitions to disease states.

NATURE BIOTECHNOLOGY VOLUME 35 NUMBER 8 AUGUST 2017

"An increased scale of personal, dense, dynamic data clouds in future holds the potential to improve our understanding of scientific wellness and delineate early warning signs for human diseases. «

Personal Genome project

Personal Genome Project: PersonalGenomes.org (/) PersonalGenomes.org Participate Global Network Donate (/organization/donate)

Sharing Personal Genomes

The Personal Genome Project was founded in 2005 and is dedicated to creating public genome, health, and trait data. Sharing data is critical to scientific progress, but has been hampered by traditional research practices—our approach is to invite willing participants to publicly share their personal data for the greater good.

Learn more about the PGP > (/organization/pgp)

Participation

Donating your genome and health data to science is a great way to enable advances in understanding human genetics, biology, and health. We seek volunteers willing to donate diverse personal information to become a public resource.

Learn about participating > (/organization/pgp-sign-up)

Open Data

Open data is a critical component of the scientific method, but genomes are both identifiable and predictive. As a result, many studies choose to withhold data from participants and restrict access to researchers. The PGP's public data is a common ground to collaborate and improve our understanding of genomes.

Use PGP data > (/organization/data)

Global Network

The pilot group for the Personal Genome Project has been based at Harvard, but we are a global group, with projects starting around the world.

Meet our PGP groups » (/organization/network)

http://www.personalgenomes.org

The field of cancer research has markedly benefited from WGS/WES.
*whole genome sequencing (WGS), whole exome sequencing (WES)

Cancer genomes include breast cancer, ovarian cancer, small-cell lung cancer, melanoma, chronic lymphocytic leukemia, Sonic-Hedgehog medulloblastoma, pediatric glioblastoma , and hepatocellular carcinoma, etc. In addition to bulk cancer sequencing, single-cell level cancer exomes have also been examined with WES.

When compared to normal tissues, these efforts identified somatic mutations for the specific cancer genomes as well as molecular markers for cancer subtyping, which may provide potential targets and guides for personalized cancer treatment.

International Cancer Genome Consortium

Enter keywords Search

Home | Cancer Genome Projects | Committees and Working Groups | Policies and Guidelines | Media

ICGC Cancer Genome Projects

Committed projects to date: 71

Sort by: Project

Biliary tract cancer – Singapore	Bladder Cancer – United States	Bladder cancer – China
Blood Cancer – United States	Blood cancer – China	Blood cancer – South Korea
Blood cancer – United States	Bone Cancer – United Kingdom	Brain Cancer – Canada
Brain Cancer – United States	Brain cancer – China	Brain cancer – United States
Breast Cancer – European Union / United Kingdom	Breast Cancer – France	Breast Cancer – Mexico
Breast Cancer – United Kingdom	Breast Cancer – United States	Breast cancer – China
Breast cancer – South Korea	Cervical Cancer – United States	Chronic Lymphocytic Leukemia – Spain
Chronic Myeloid Disorders – United Kingdom	Colon Cancer – United States	Colorectal cancer – China
Endocrine Tissues Cancer	Endometrial Cancer – United States	Esophageal Cancer – United Kingdom

ICGC Goal: To obtain a **comprehensive** description of **genomic, transcriptomic and epigenomic changes** in **50 different tumor types and/or subtypes** which are of clinical and societal importance across the globe.

Read more »

Launch Data Portal »

Apply for Access to Controlled Data »

Announcements

15/May/2014 - The ICGC Data Coordination Center (DCC) is pleased to announce ICGC data portal data release 16 (http://dcc.icgc.org).

ICGC data release 16 in total comprises data from 11,633 cancer genomes.

Updates

Currently, the ICGC has received commitments from funding organizations in Asia, Australia, Europe, North America and South America for 74 project teams in 17 jurisdictions to study over 25,000 tumor genomes. Projects that are currently funded are examining tumors affecting: the biliary tract, bladder, blood, bone, brain, breast, cervix, colon, eye, head and neck, kidney, liver, lung, nasopharynx, oral cavity, ovary, pancreas, prostate, rectum, skin, soft tissues, stomach, thyroid and uterus. The genomic analyses of tumors conducted by ICGC members in Australia (ovarian and pancreatic cancer), Canada (pancreatic, pediatric brain and prostate cancer), China (bladder, esophageal, gastric and renal cancer), European Union/France (renal cancer), France (liver cancer), Germany (blood, brain and prostate cancer), India (oral cancer), Japan (liver cancer), Saudi Arabia

National Cancer Institute | National Human Genome Research Institute

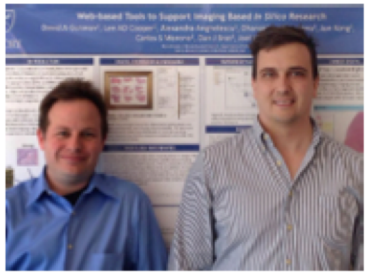
The Cancer Genome Atlas

Understanding genomics to improve cancer care

Launch Data Portal | Contact Us | For the Media

Search | Search

Home | About Cancer Genomics | Cancers Selected for Study | Research Highlights | Publications | News and Events | About TCGA



The Cancer Digital Slide Archive

Dr. David Gutman and Dr. Lee Cooper developed The Cancer Digital Slide Archive (CDSA), a web platform for accessing pathology slide images of TCGA samples. Find out how they did it and how to use the CDSA website in this Case Study.

Learn More

 **Cancer Digital Slide Archive Case Study**

 Batch Effects Case Study

 Cancers Selected for Study

 About TCGA

Leadership Update

April 2014

TCGA's Pan-Cancer Efforts and Expansion to Include Whole Genome Sequence

Carolyn Hutter, Ph.D., Program Director of NHGRI's Division of Genomic Medicine, discusses the expansion of TCGA's Pan-Cancer efforts to include the Pan-Cancer Analysis of Whole Genomes (PAWG) project.

View All

Perspectives

April 2014

TCGA and Its Vital Role in Understanding How Germline Variation Informs the Landscape of Somatic Alterations in Cancer

Dr. Stephen Chanock, M.D., Director of the Division of Cancer Epidemiology & Genetics at the NCI, discusses how TCGA provides a strong foundation for understanding key biological alterations in cancer.

View All

TCGA in Action

March 2014

CASE STUDY: The Cancer Digital Slide Archive: A Web Platform for Accessing TCGA Data

Dr. David Gutman and Dr. Lee Cooper developed The Cancer Digital Slide Archive (CDSA), a web platform for accessing pathology slide images of TCGA samples. Find out how they did it and how to use the CDSA website in this Case Study.

February 2014

CASE STUDY: Using Forensics to Untangle Batch Effects in TCGA Data

Rehan Akbani, Ph.D., and colleagues at the University of Texas MD Anderson Cancer Center developed a tool called MBatch to detect, diagnose, and correct batch effects in TCGA data. Read more about batch effects in this Case Study.

More Stories

News Releases and Announcements

April 30, 2014

The ICGC-TCGA DREAM Somatic Mutation Calling (SMC) Challenge

The SMC Challenge is an open, crowd-sourced initiative to identify the most accurate and robust methods for detecting cancer-associated mutations.

March 12, 2014

The Center for Cancer Genomics Data Portal Launches

The new Data Portal for the Center for Cancer Genomics (CCG) will serve as the access point for data generated by CCG programs, collaborations or member offices.

View All

Launch Data Portal

The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA.

Questions About Cancer

Visit www.cancer.gov

Call 1-800-4-CANCER

Use LiveHelp Online Chat

Multimedia Library

- Images
- Videos and Animations
- Podcasts
- Interactive

Stay Connected

- Sign up for email updates
- RSS newsfeeds
- Twitter
- YouTube

Molecular Systems Biology 7; Article number 496; doi:10.1038/msb.2011.26
Citation: ***Molecular Systems Biology* 7:496**

www.molecularsystemsbiology.com

molecular systems biology

PREDICT: a method for inferring novel drug indications with application to personalized medicine

Assaf Gottlieb[1], Gideon Y Stein[2,3], Eytan Ruppin[1,2] and Roded Sharan[1,*]

[1] The Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel, [2] Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel and [3] Department of Internal Medicine 'B', Beilinson Hospital, Rabin Medical Center, Petah-Tikva, Israel

* Corresponding author. The Blavatnik School of Computer Science, Tel-Aviv University, Haim-Levanon, Tel-Aviv 69978, Israel.
Tel.: +972 3 640 7139; Fax: +972 3 640 9357; E-mail: roded@post.tau.ac.il

Received 12.1.11; accepted 12.4.11

Inferring potential drug indications, for either novel or approved drugs, is a key step in drug development. Previous computational methods in this domain have focused on either drug repositioning or matching drug and disease gene expression profiles. Here, we present a novel method for the large-scale prediction of drug indications (PREDICT) that can handle both approved drugs and novel molecules. Our method is based on the observation that similar drugs are indicated for similar diseases, and utilizes multiple drug–drug and disease–disease similarity measures for the prediction task. On cross-validation, it obtains high specificity and sensitivity (AUC=0.9) in predicting drug indications, surpassing existing methods. We validate our predictions by their overlap with drug indications that are currently under clinical trials, and by their agreement with tissue-specific expression information on the drug targets. We further show that disease-specific genetic signatures can be used to accurately predict drug indications for new diseases (AUC=0.92). This lays the computational foundation for future personalized drug treatments, where gene expression signatures from individual patients would replace the disease-specific signatures.

Molecular Systems Biology **7**:496; published online 7 June 2011; doi:10.1038/msb.2011.26
Subject Categories: bioinformatics; molecular biology of disease
Keywords: drug indication prediction; drug repositioning; drug repurposing; machine learning; personalized medicine

FINE CORSO BIOLOGIA DEI SISTEMI

“Systems Biology” and biomedical applications

1) Diseases– introduction
2) Networks in biomedicine – introduction
3) Application: The human diseasome
4) Application: Comorbidity
5) Innate immunity: introduction and applications
6) Inflammation: introduction and applications
7) Tumors: introduction and applications
8) P4 medicine

Definitions

Cross-validation (**statistics**) = a technique for estimating the performance of a predictive model.
Cross-validation, sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set.

Cross-validation (**analytical chemistry**) = the practice of confirming an experimental finding by repeating the experiment using an independent assay technique