Bioinformatics in Translational Medicine

Lu XIE (谢鹭)
Shanghai Center for Bioinformation Technology
（上海生物信息技术研究中心）

2014-06-19, The 12th Korean-Japan-China
Bioinformatics Training Course 2014, Jeju Island, Korea

谢鹭 （Lu Xie），Ph.D., M.D.
Translational Medicine Group, SCBIT
xielu@scbit.org， www.scbit.org

- ***Training***
  - 1985-2000 **Xiang Ya Medical College,** Central South University, M.D., Ph.D
  - -- 2000-2005 **Vanderbilt University,** USA, Postdoctoral Research Fellow

- ***Career***
  - 1997-2000 Xiang Ya Medical College, Central South Univ, Assitant Prof.
  - 2005-2009 Shanghai Center for Bioinformation Technolgoy, Assoc. Prof.
  - 2010-now **Shanghai Center for Bioinformation Technolgoy,** Professor.

- ***Research field***
  - **Proteomics Bioinformatics**: Proteogenomics, protein post-translational modification, signaling transduction network
  - **Cancer Bioinformatics**: tumor biomarker, clinical predictive models, transcriptional regulatory network

Outline

- Translational medicine definition
- Why bioinformatics in translational medicine
- Bioinformatics computational approaches
- Bioinformatics applications in Translational Medicine

- Work cases

Translational medicine

- **Translational medicine** is a discipline within biomedical and public health research that aims to improve the health of individuals and the community by “translating” findings into diagnostic tools, medicines, procedures, policies and education.

Wikipedia. Last updated on April 29, 2014

Traditional categorization

- Basic research (fundamental or pure research) is more speculative and takes a long time to be applied in any practical context. Basic research often leads to breakthroughs or paradigm-shifts in practice.
- Applied research is research that can have an impact in practice in a relatively short time but often represents an incremental improvement to current processes rather than delivering radical breakthroughs.

Translational research

- **Translational research** is engineering research that aims to make findings from basic science useful for practical applications that enhance human health and well-being. It is practiced in fields such as environmental and agricultural science, as well as the health, behavioral, and social sciences. Other terms: **translative research, translational science. (** Also: participative science and participatory action research.)

Wikipedia. Last updated on April 29, 2014

Translational medicine

- Translational medicine is a rapidly growing discipline in biomedical research and aims to expedite the discovery of new diagnostic tools and treatments by using a **multi-disciplinary, highly collaborative, "bench-to-bedside"** approach.

Translational medicine

- The National Institutes of Health (NIH): focus on cross-functional collaborations (between researchers and clinicians); **leveraging new technology and data analysis tools;** and increasing the speed at which new treatments reach patients.
- In December 2011, The National Center for Advancing Translational Science (NCATS) was established within the NIH: “transform the translational science process **so that new treatments and cures for disease can be delivered to patients faster.**”

Translational medicine

- Applying knowledge from basic science is a major stumbling block in science, partially due to the **compartmentalization** within science. Hence, translational research is seen as a key component to finding practical applications, especially within healthcare.
- Move from laboratory experiments **through clinical trials to point-of-care patient applications.**

Translational medicine

- Examples of failed translational research abound in the pharmaceutical industry, such as the failure of anti-aβ therapeutics in Alzheimer‘s disease. Other problems arise from the widespread **irreproducibility** in the translational research literature.

Why bioinformatics

- Translational research requires **a knowledge-driven ecosystem**, in which constituents **generate, contribute, manage and analyze data available from all parts of the landscape.** A continuous feedback loop to accelerate the translation of data into knowledge.
- Collaboration, **data sharing, data integration and standards** are important.
- Only by **seamlessly structuring** and integrating these data types will the complex and underlying causes and outcomes of illness be revealed, and effective prevention, early detection and personalized treatments be realized.

Why bioinformatics

- Translational research requires that **information and data flow from hospitals, clinics and study participants in an organized and structured format, to repositories and laboratories.**
- Meeting the increased operational requirements of larger studies, with ever **increasing specimen counts, larger and more complex systems biology data sets**, and government regulations.
- Most informatics systems today are inadequate to handle the tasks of complicated operations and contextually in data management and analysis.

Computational approaches

- Statistical inference (biomarker selection)

Classical Statistical Inference incorporates no prior information and assumes independent variables; the approach is used at all systems levels and underlies the primary tools, such as Student's *t* test. In contrast, *Bayesian Statistical Inference* does incorporate prior information as well as handle interdependent variables. The Bayesian "conditional probability" approach is used in genetic data analysis, clinical research and diagnostic medicine; complex Bayesian analyses use Markov Chain Monte Carlo (MCMC) computational methods.

Mary F. McGuire, et al. Computational Approaches for Translational Clinical Research in Disease Progression. J Investig Med. 2011 August ; 59(6): 893–903.

Common statistical software includes R (http://www.r-project.org/), Spotfire S+(http://spotfire.tibco.com/products/s-plus/statistical-analysis-software), SPSS (www.spss.com), and SAS (www.sas.com). OpenBUGS is open-source software for Bayesian analysis using MCMC methods (http://www.openbugs.info/w/).

Computational approaches

- **Class prediction: Mechanistic approaches (machine learning)**

An *artificial neural network* (ANN) is a mathematical model that mechanistically learns nonlinear patterns from a set of observations and then infers the optimal function that describes those observations. ANNs can train classifiers, approximate functions, filter and cluster data, and direct robotics.

SVM: support vector machine

Standard mathematical and statistical software including MATLAB, Mathematica, SPSS and SAS have built-in algorithms or add-on modules for neural network analysis and optimization.

Computational approaches

- **Interaction inference: Graphical approaches**

Cascades of molecular interactions can be represented as directed graphs and use computational methods from graph theory to explore pathways within the graph. G*raph theory* is supported by extensive computational methods from mathematics and computer science that are used for analysis of static and dynamic systems.

1. A *Bayesian network* is a probabilistic graphical model constructed as a directed acyclic graph (DAG) with nodes representing variables and edges representing the conditional dependencies between the nodes. process modeling and diagnostic reasoning. must satisfy the local Markov property - each node is conditionally independent of its non-descendents
2. *Generalized Bayesian Networks* (GBN), can model cyclic networks for use in translational systems biology.
3. software packages for Bayesian networks: Python library Pebl ; Petri Net Toolboxes (MATLAB)
4. *Pathway databases: IPA*

Computational approaches

- **Model inference:** Deterministic approaches

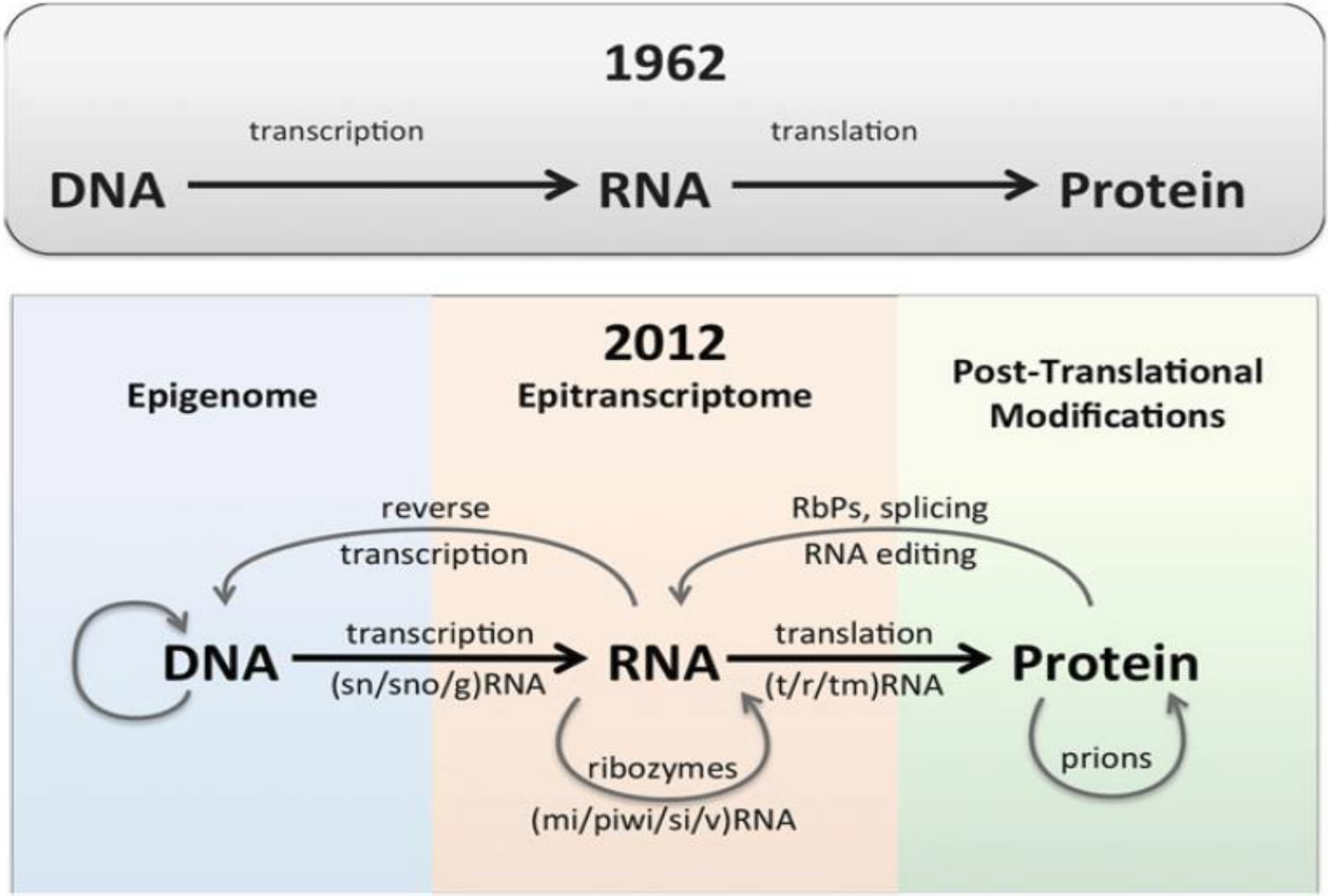
Deterministic approaches depend on initial states and chosen parameters.

Differential equations are the primary methods of deterministic dynamic analysis, and are mostly used at the molecular and cellular levels because they are computationally intensive at higher levels. Ordinary differential equations (ODEs) model dynamic changes in items, such as protein concentrations, over one independent variable whereas partial differential equations model simultaneous changes over two or more independent variables (MATLAB).

- *Matrix algebra* can be applied from molecular to organism levels. Stoichiometric matrices are used for flux-balance analysis (FBA) of metabolic biochemical reaction networks. Unlike differential equation approaches, FBA, the key assumptions are that the system is homeostatic with a balanced system of energy production and consumption and that the metabolites are “well stirred” so that Gillespie’s Algorithm can be used. generate signaling networks from sparse time series of observed data, has potential for analysis of signaling pathways in disease progression. (MATLAB. Code for Gillespie’s Algorithm: R)

Application: Biomarker Identification

- High dimensional data generation and challenge:
  - Genomics
  - Transcriptomics
  - Proteomics
  - Metabolomics
  - …

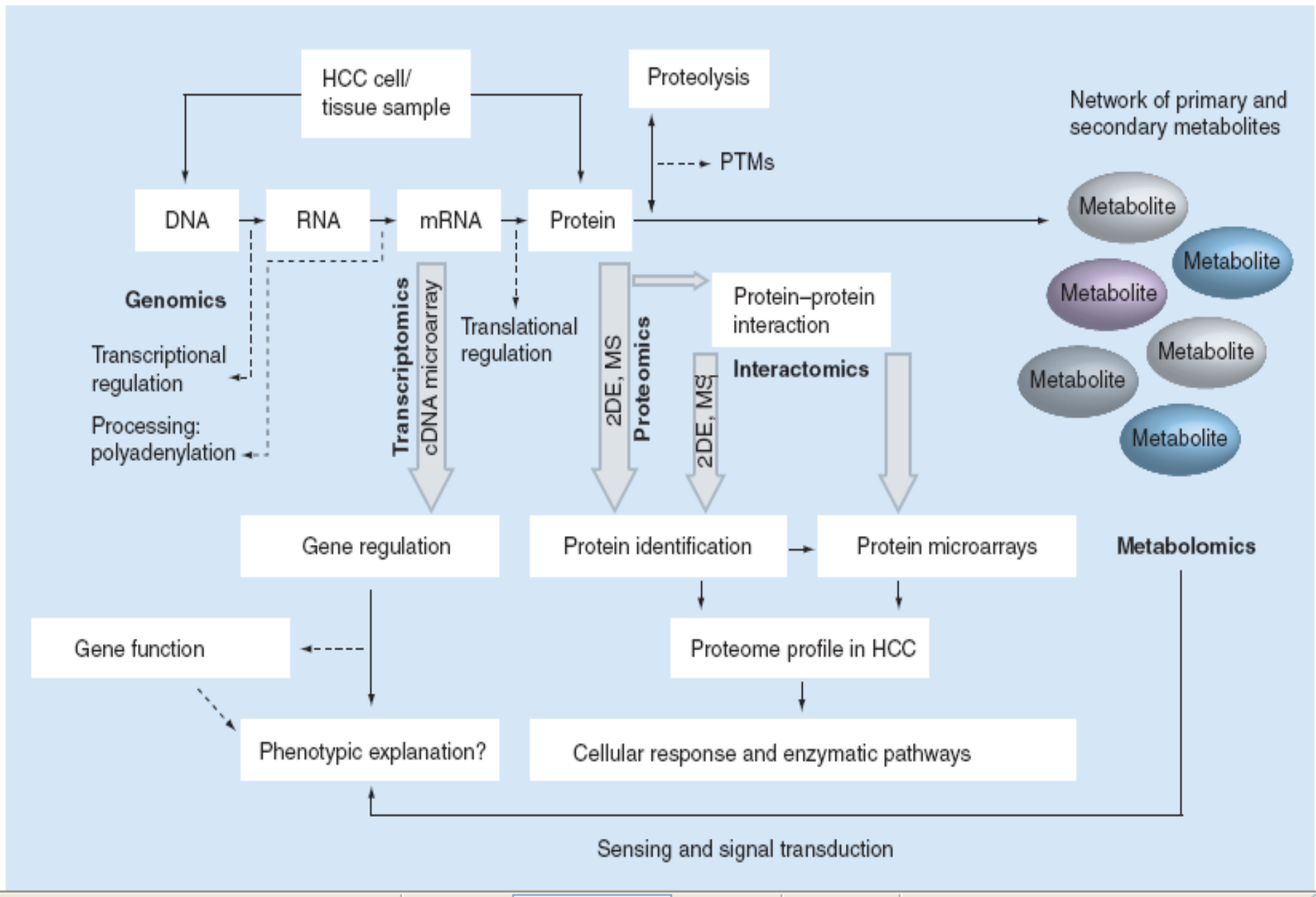


N. Maltsev et al. (eds.), *Systems Analysis of Human Multigene Disorders, Advances 15* in Experimental Medicine and Biology 799, DOI 10.1007/978-1-4614-8778-4_2,

High-dimensional Biology for complex diseases

- The term “high-dimensional biology”(HDB) is the integration of genomics, transcriptomics, proteomics and metabolomics in a biological sample with the goal of understanding the physiology or molecular mechanism of disease.

- **Personalized Therapy**

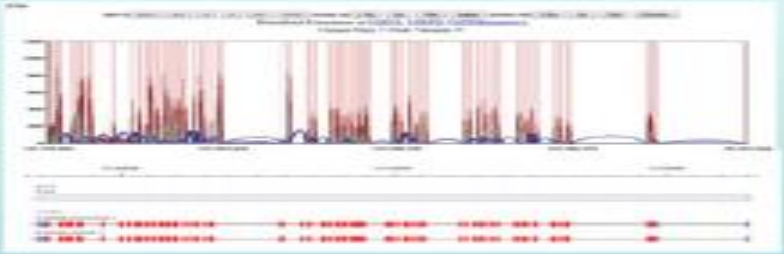


Perumal Vivekanandan and OM V Singh. ***Expert Rev Proteomics, 5(1): 45-60, 2008.***

Collect (1°) – Images to Reads

Sequences + Quality Values

Reduce (2°) – Reads to Annotated Variant List



Alignments Annotations

Compare (3°) – (*N*) Variant Lists to Gene Candidates

Comparisons Clustering

Discover (3°) – Gene Candidates to Clinical Utility

Ontologies Pathways Networks

Browse Verify

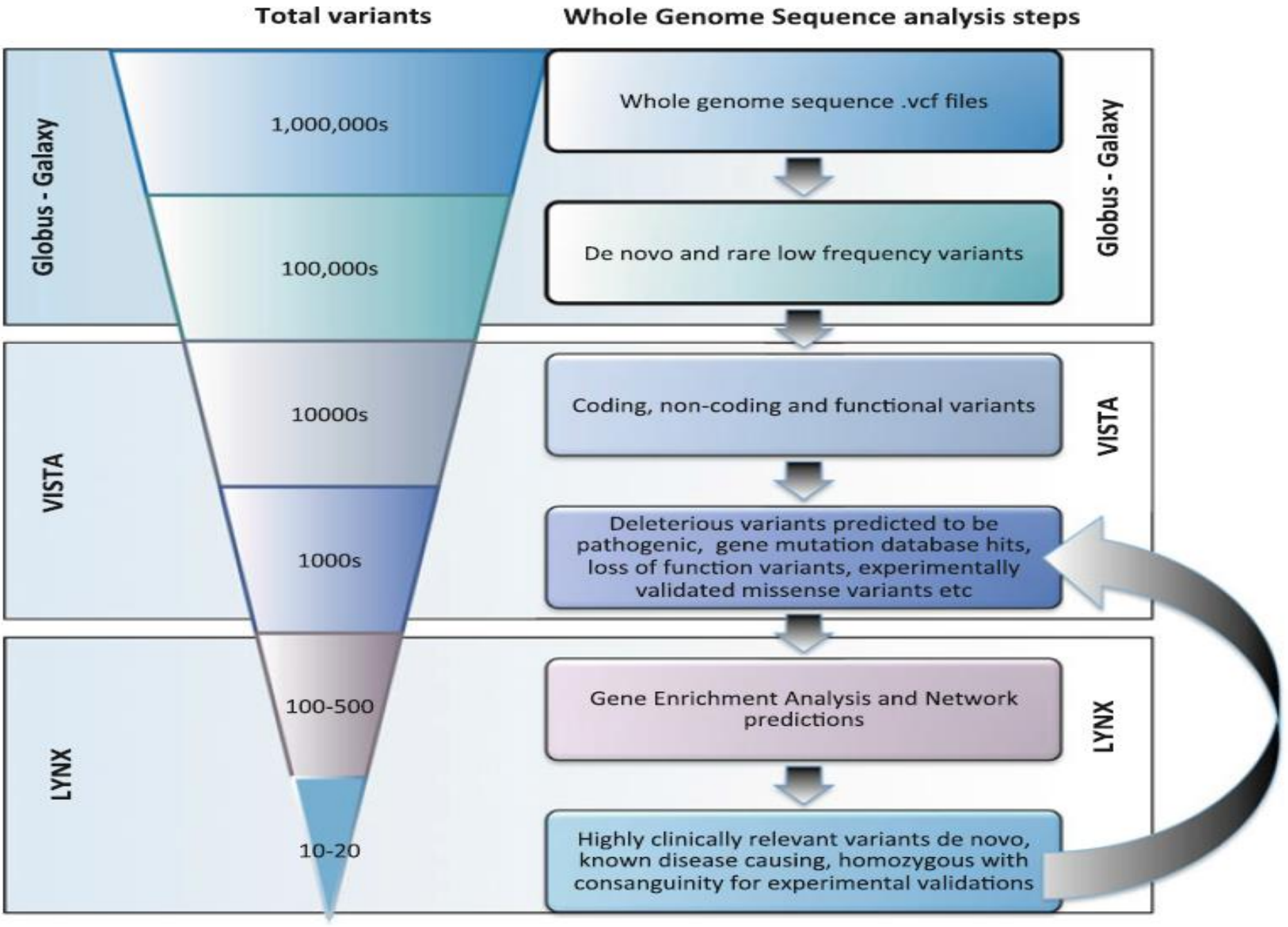
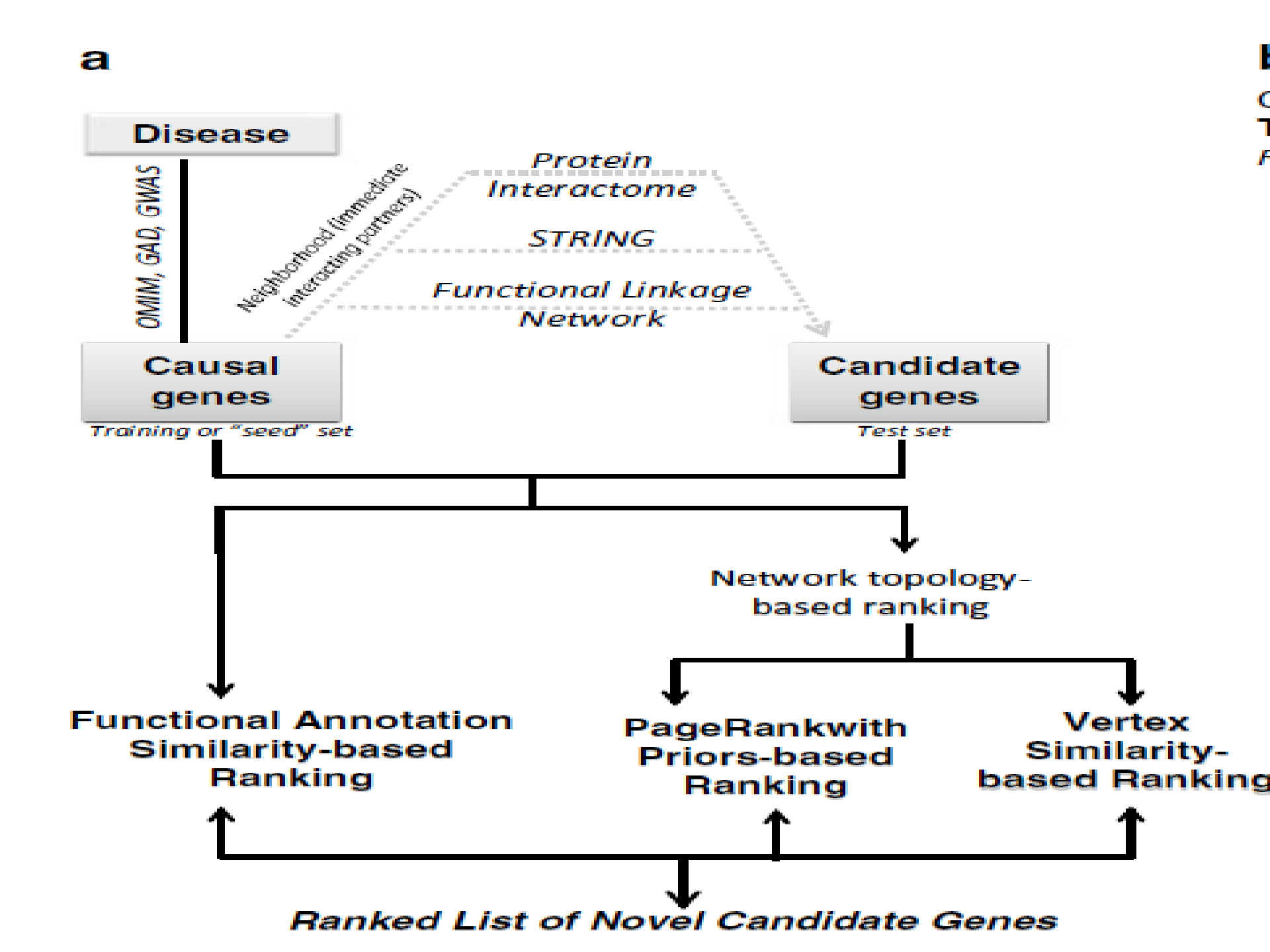
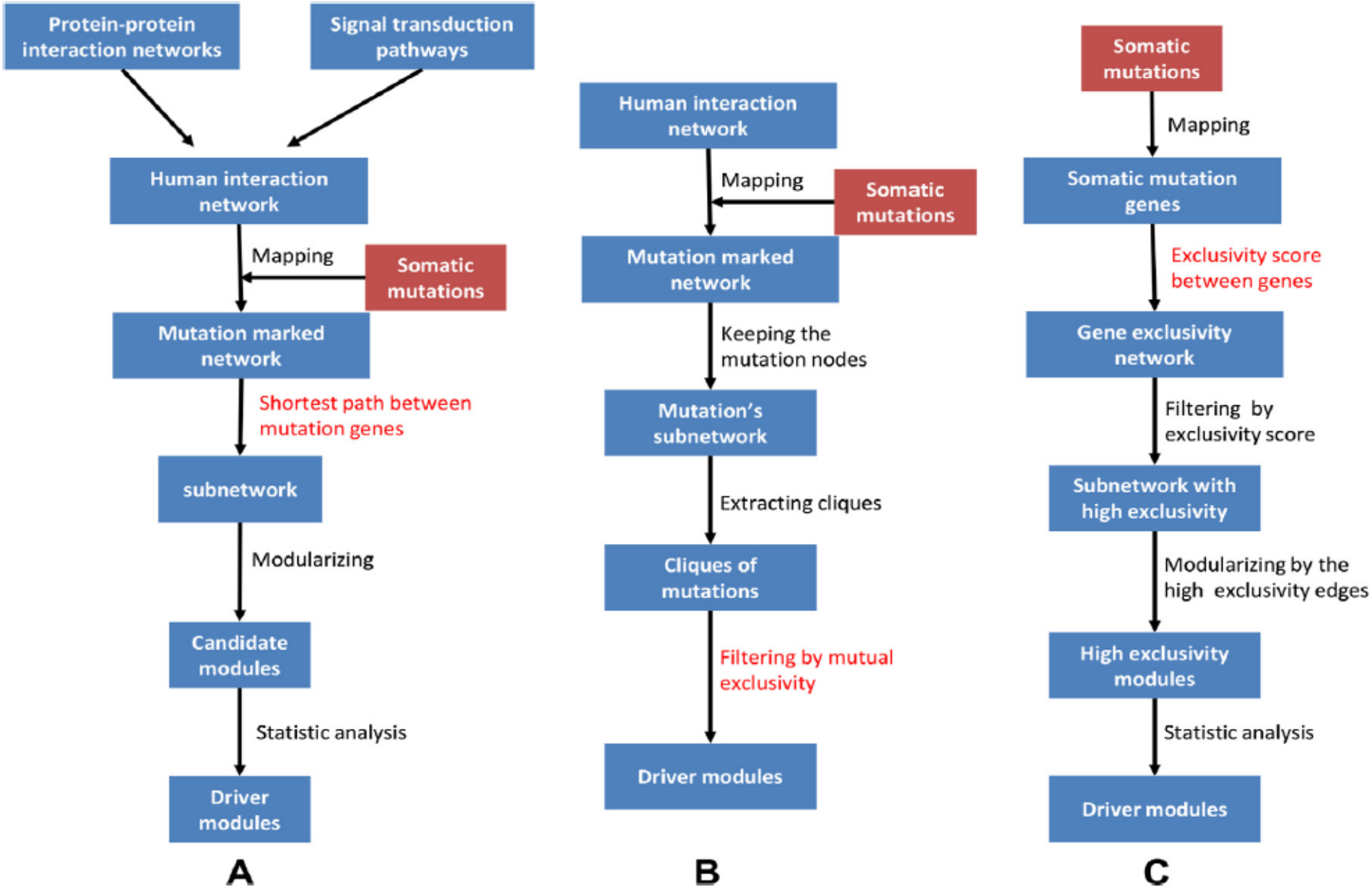


Fig. 3.2 Steps of whole genome sequence analysis


a
Disease
OMIM, GAD, GWAS
Neighborhood (immediate interacting partners)
Protein Interactome
STRING
Functional Linkage Network
Causal genes
Training or "seed" set
Candidate genes
Test set
Network topology-based ranking
Functional Annotation Similarity-based Ranking
PageRankwith Priors-based Ranking
Vertex Similarity-based Ranking
Ranked List of Novel Candidate Genes



Xiaoping Liu, et al. Whole-exome sequencing reveals recurrent somatic mutation networks in cancer. Cancer Letters 340 (2013) 270–276

Application: Class Prediction

Subclass

Treatment

Prognosis

…

Translational medicine -personalized medicine, individual therapy

- The goal of personalized treatment of patients is to get “the right drug to the right patient with the right dose”.

*Ulrich Kruse1, Marcus Bantscheff1, Gerard Drewes1, Carsten Hopf1,2**
Chemical and pathway proteomics: Powerful tools for oncology drug discovery and personalized health care. MCP Papers in Press. *Published on August 1, 2008 as Manuscript R800006-MCP200*

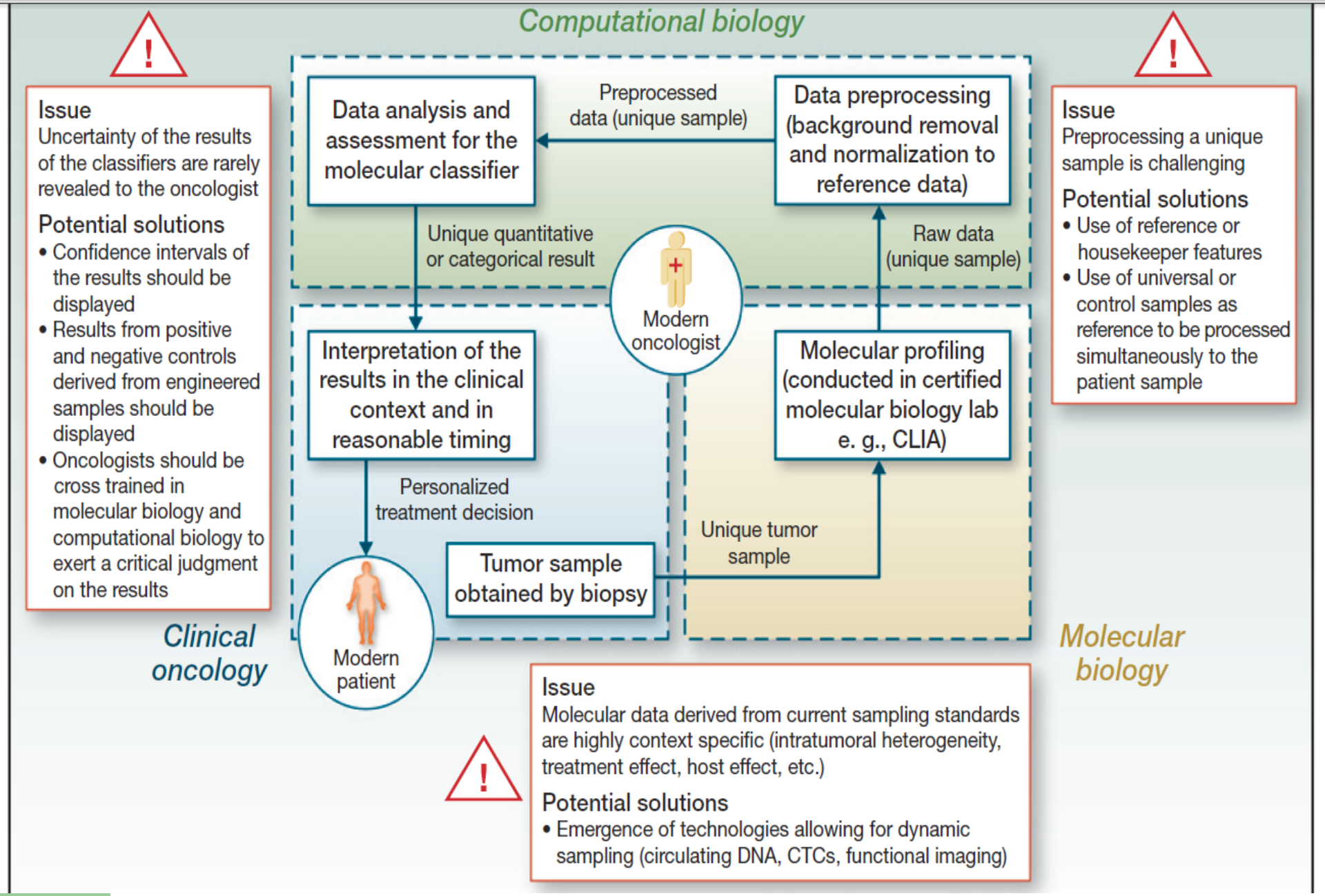
- precise knowledge of the molecular classes

Translational

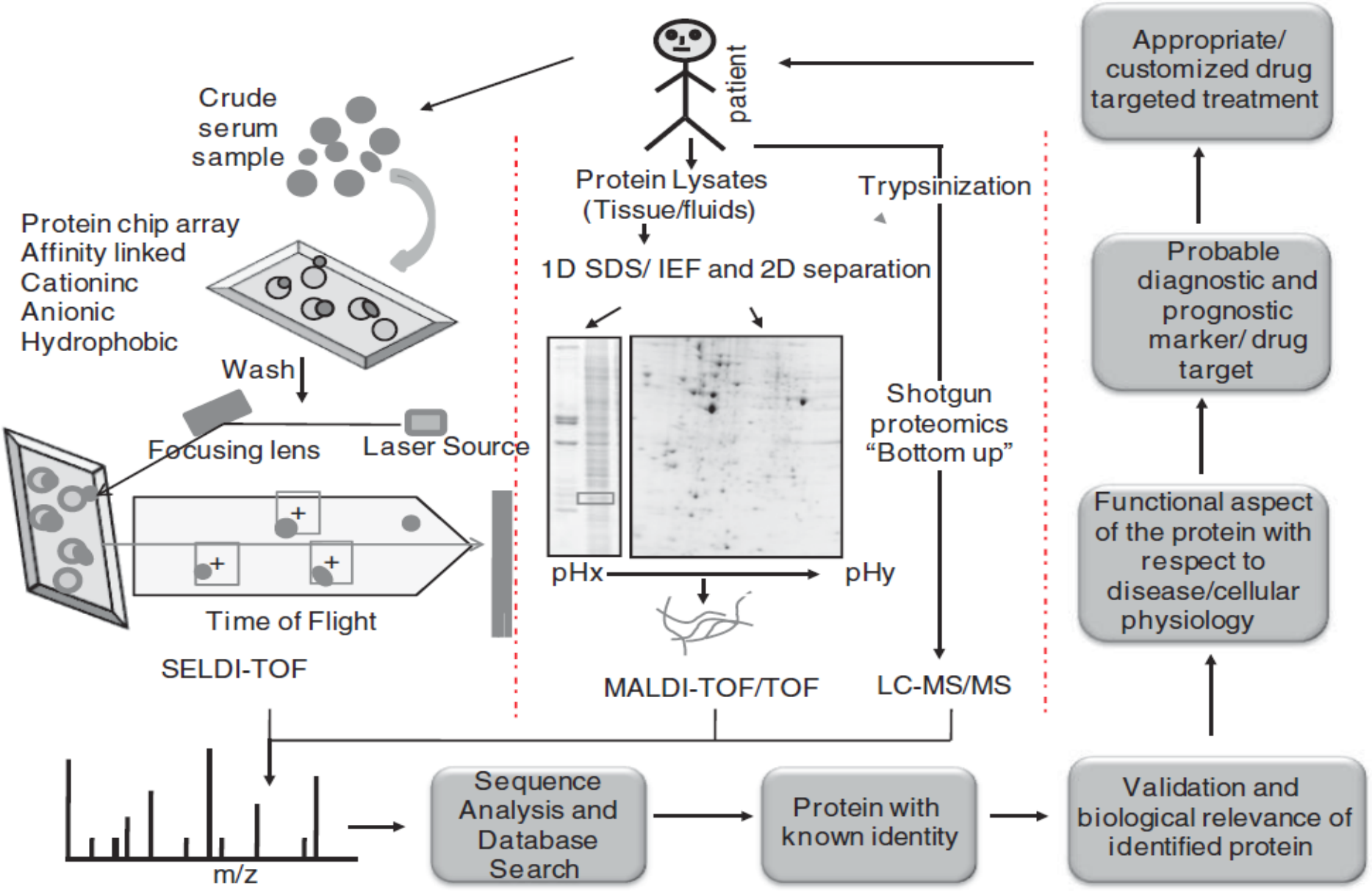
- effective blockade of specific pathways

- well designed clinical trials

The conjunction to implement personalized medicine (each individual patient classified to the right groups-the right treatment).



Charles Ferte, et al. Impact of Bioinformatic Procedures in the Development and Translation of High-Throughput Molecular Classifiers in Oncology. Clin Cancer Res; 19(16); 4315–25. 2013 AACR.



Isha Kapoor, et al. Proteomics approaches for myeloid leukemia drug discovery. Expert Opin. Drug Discov. (2012) 7(12):1165-1175

Customized predictive systems

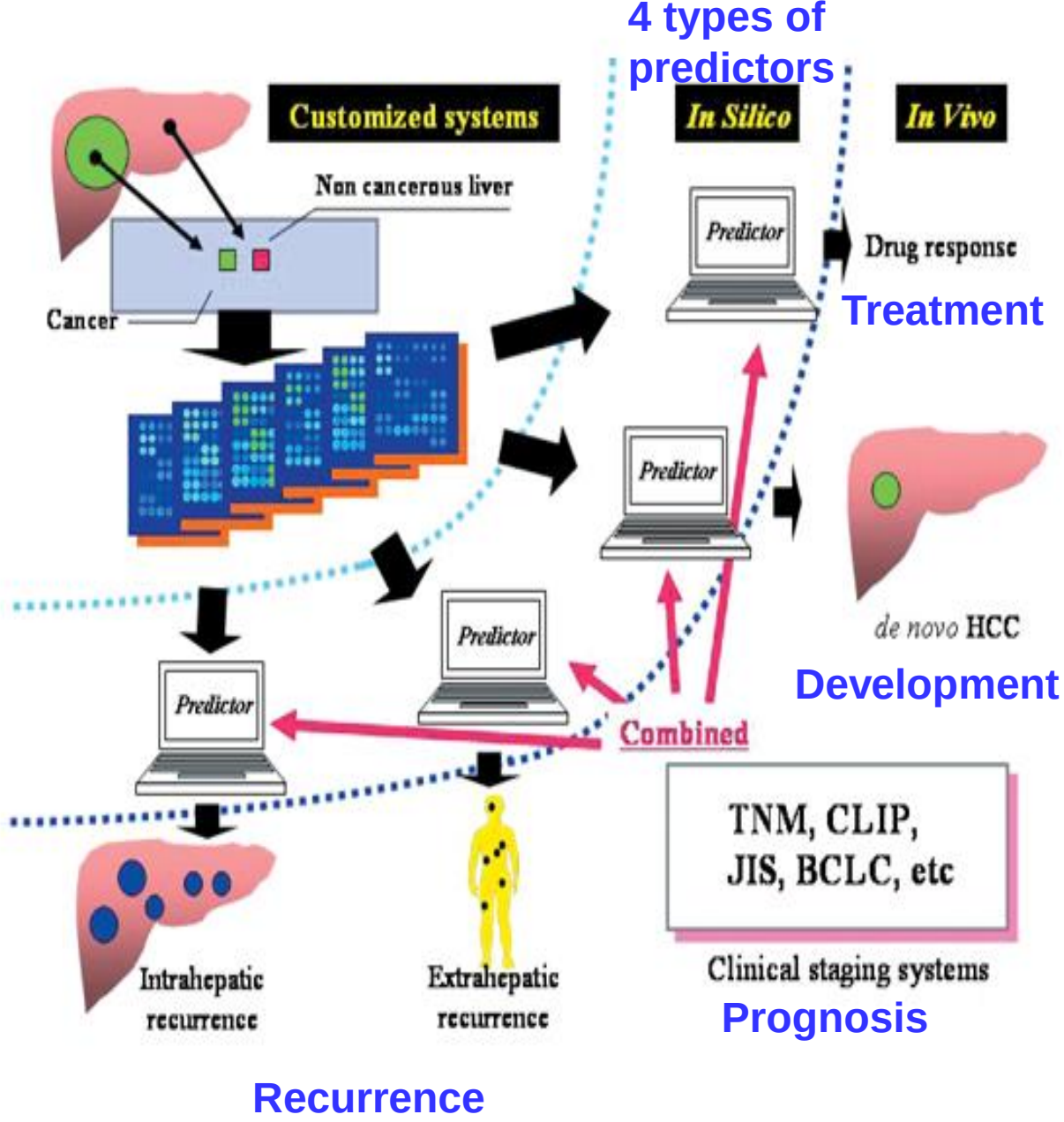


Fig. 1. Customized predictive systems for outcome of hepatocellular carcinoma patients. BCLC, Barcelona clinic liver cancer staging system; CLIP, Cancer of the Liver Italian Program; HCC, hepatocellular carcinoma; JIS, Japan integrated staging; TNM, tumor node metastasis.

Application: Drug Discovery

Drug target

Drug repositioning

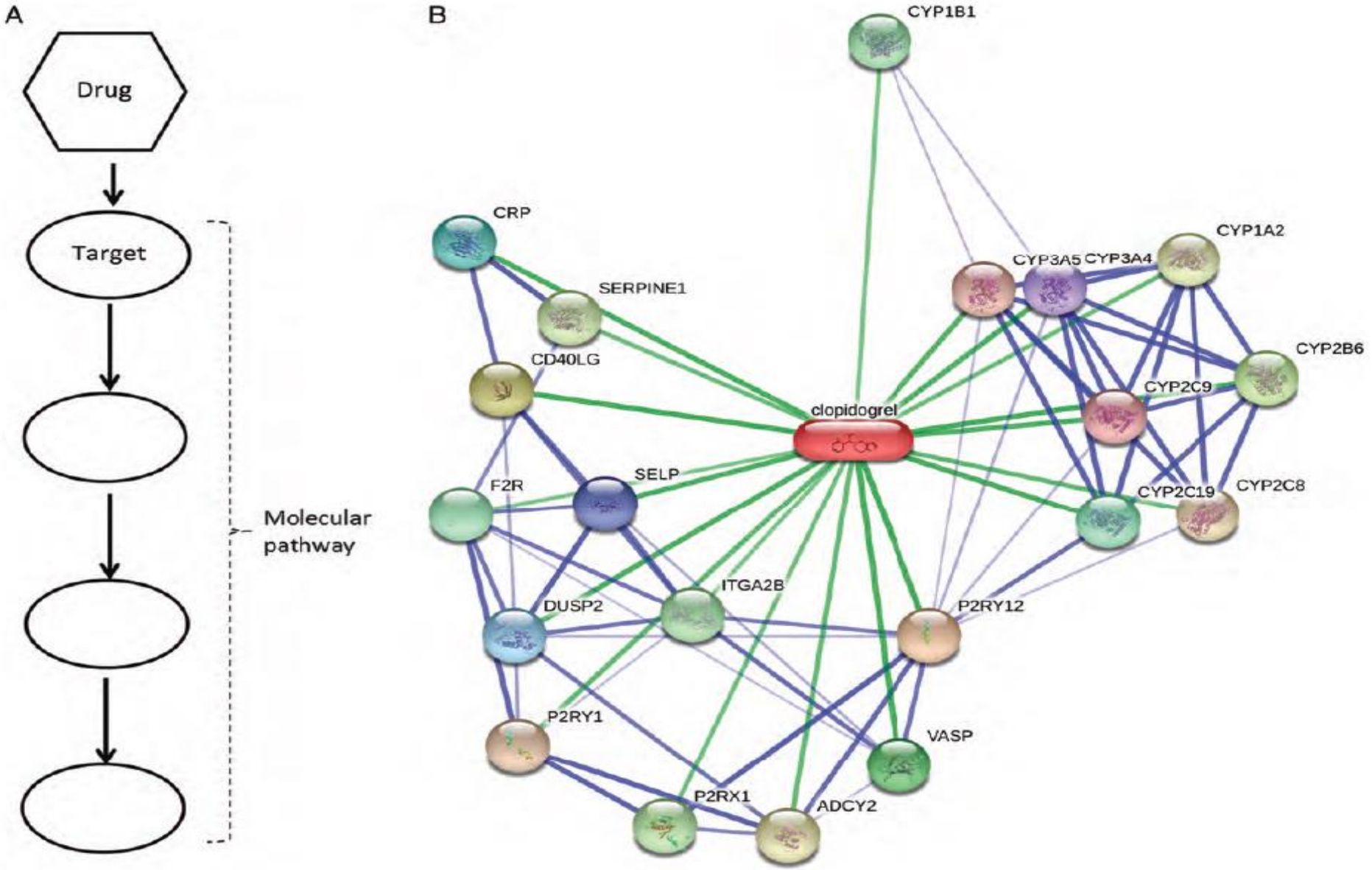
Network drug

Examples

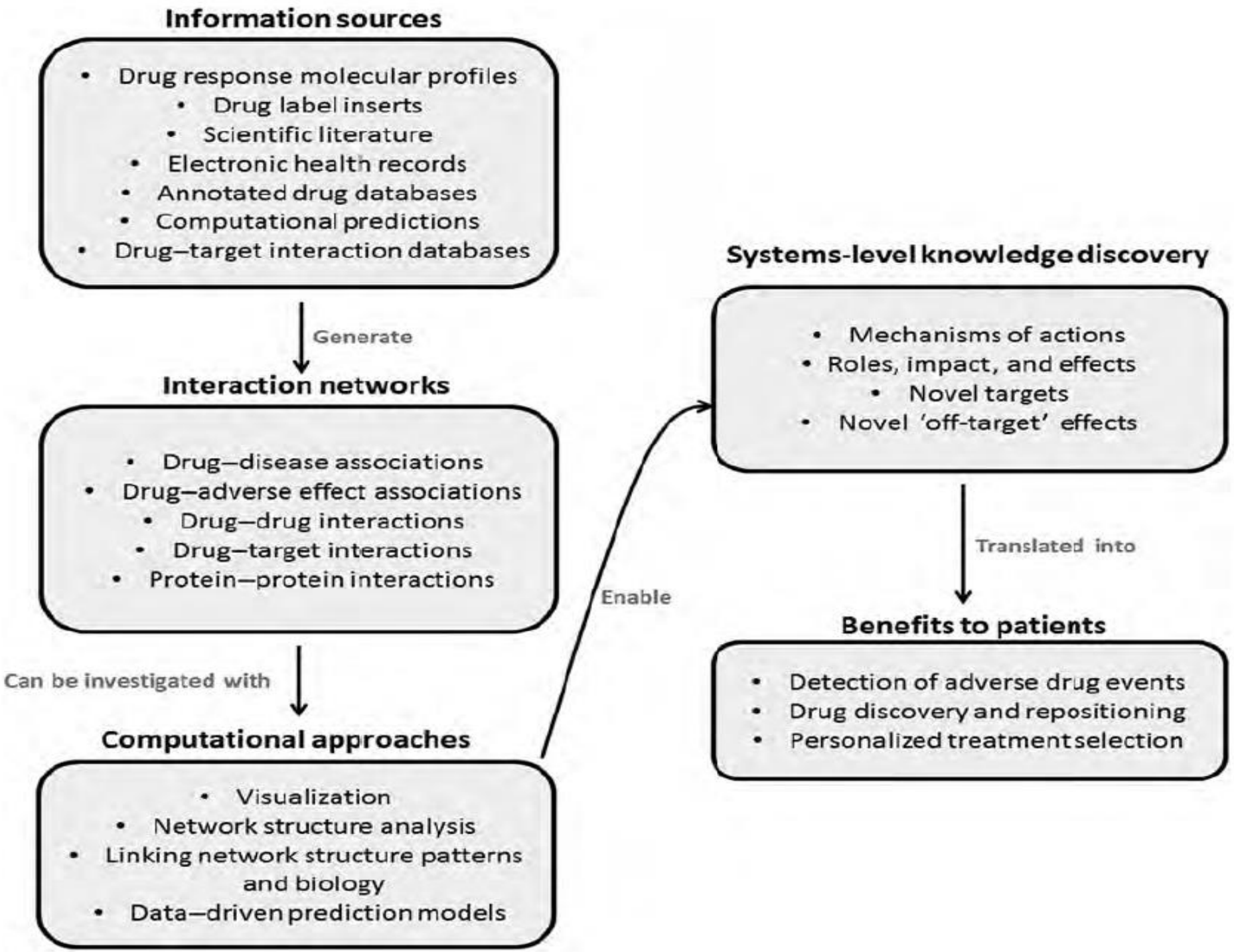
- Breast cancer:Brca1/2 mutations-prophylactic surgery；Her-2 overexpressing-Herceptin
- Non-small cell lung cancer：EGFR mutation positive-EGFR tyrosine kinase inhibitor-Iressa.

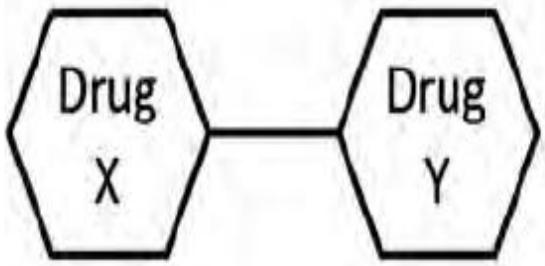
- Sorafenib (Nexavar): multi-tyrosine kinase (VEGFR2, PDGFR, c-Kit receptors, FLT-3） and -serine kinase (B-RAF, p38, Raf-1) inhibitor, The first multi-target drug approved by FDA.
- Anti-signaling transduction, anti-angiogenesis.
- Advanced renal cancer(and hepatocellular cancer, melanoma, tested on more than 20 malignancies).

- Cocktails of multi-target therapy



Francisco Azuaje. Drug interaction networks: an introduction to translational and clinical applications. Cardiovascular Research (2013) 97, 631–641

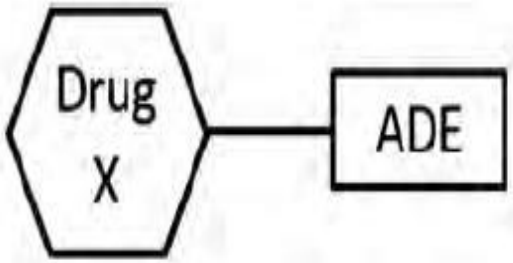

Information sources
Drug response molecular profiles
Drug label inserts
Scientific literature
Electronic health records
Annotated drug databases
Computational predictions
Drug–target interaction databases
Generate
Interaction networks
Drug–disease associations
Drug–adverse effect associations
Drug–drug interactions
Drug–target interactions
Protein–protein interactions
Can be investigated with
Computational approaches
Visualization
Network structure analysis
Linking network structure patterns and biology
Data–driven prediction models
Enable
Systems-level knowledge discovery
Mechanisms of actions
Roles, impact, and effects
Novel targets
Novel 'off-target' effects
Translated into
Benefits to patients
Detection of adverse drug events
Drug discovery and repositioning
Personalized treatment selection



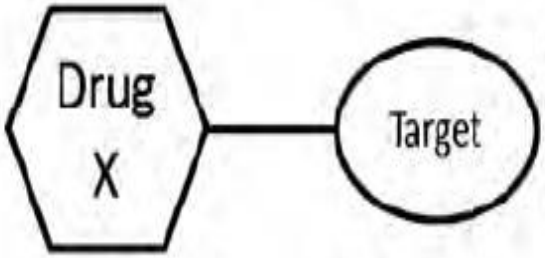
Known interaction

Drugs induce similar molecular profiles

Drugs have similar side effects

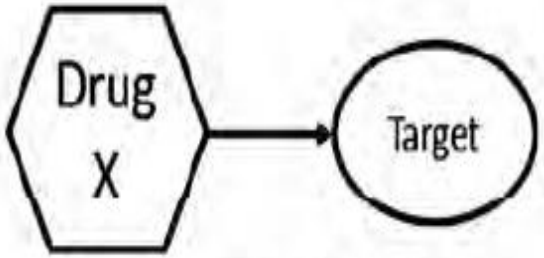


Drug is associated with specific adverse drug effect

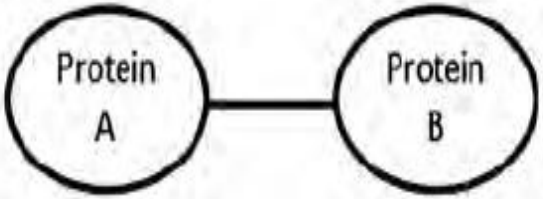


Physical interaction

Target activity is correlated with treatment

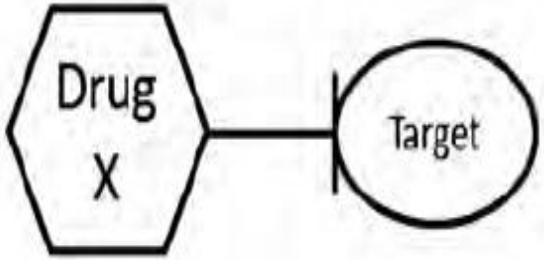


Drug activates target



Experimentally validated physical interaction

Computationally predicted association



Drug inhibits target

Table I Key information resources for the generation and analysis of drug–target interaction networks

Resource	Description	Website
cmap	Associations between drugs, genes and diseases	www.broadinstitute.org/cmap
DrugBank	Pharmacological information about drugs and their interactions	www.drugbank.ca
MANTRA	Drug–drug interactions derived from computational analysis of gene expression	Mantra.tigem.it
My-DTome	Drug–target interactions in human myocardial infarction	www.my-dtome.lu
PharmGKB	Impact of genomic variation on drug response, drug–target interaction information	www.pharmgkb.org
SIDER	Drug–ADE associations inferred from patient data	sideeffects.embl.de
STITCH	Drug–target interaction database	stitch.embl.de
STRING	Known and computationally predicted protein interactions	string-db.org
SuperTarget	Drug–target interaction database	insilico.charite.de/supertarget

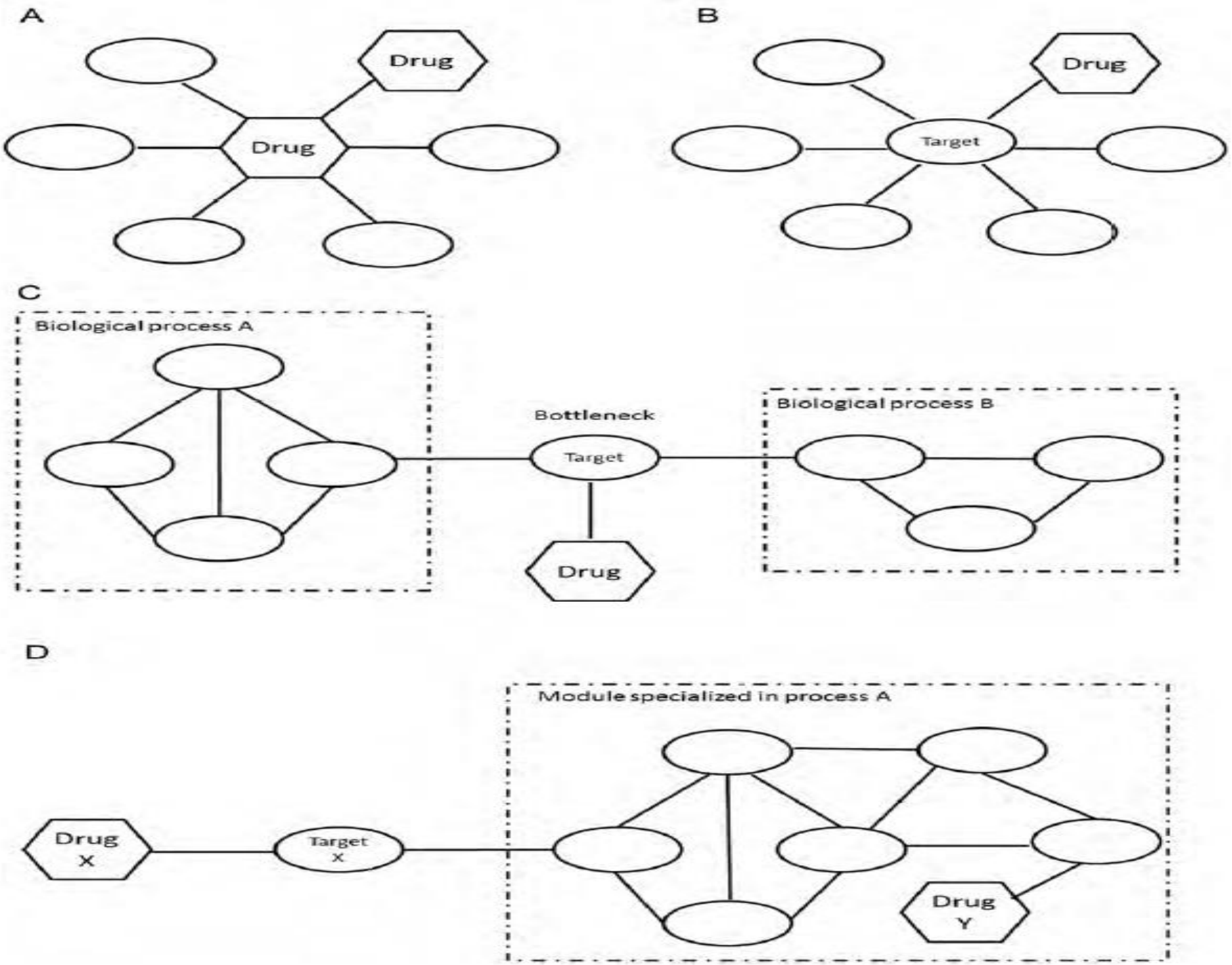

A
Drug
Drug
B
Drug
Target
C
Biological process A
Bottleneck
Target
Drug
Biological process B
D
Module specialized in process A
Drug
X
Target
X
Drug
Y

Table 5.1 Molecular subsets of NSCLC tumors with adenocarcinoma histology and clinically relevant mutations/translocations

Mutation/translocation	Lung cancer consortium	Sequist et al.
K-RAS mutation	25 %	24 %
EGFR mutation	23 %	13 %
ALK rearrangement	6 %	5 %
BRAF mutation	3 %	2 %
PIK3CA mutation	3 %	4 %
MET amplification	2 %	–
Her-2 mutation	1 %	< 1 %
MEK-1 mutation	0.4 %	–
N-RAS mutation	0.2 %	1 %
AKT-1 mutation	0 %	–
B-catenin mutation	–	2 %
IDH1 mutation	–	< 1 %

The second column depicts data from the Lung Cancer Mutation Consortium, which utilizes multiplexed assay for *K-RAS, EGFR, HER-2, BRAF, PIK3CA, AKT1*, *MEK1*, and *N-RAS* along with FISH for *ALK* rearrangement and *MET* amplification ($n = 830$, enrollment ongoing) [40]. The third column depicts data from 552 patients with NSCLC tested with multiplexed PCR-based assay (SNaPShot) along with FISH for *ALK* translocation [126]

N. Maltsev et al. (eds.), *Systems Analysis of Human Multigene Disorders, Advances 15* in Experimental Medicine and Biology 799, DOI 10.1007/978-1-4614-8778-4_2,

Table 5.3 Alterations in targetable oncogenic pathways in squamous-cell lung cancer

PI3K/AKT pathway alterations	Frequency
PTEN	15 %
PIK3CA	16 %
AKT1	<1 %
AKT2	4 %
AKT3	16 %
STK11	2 %
TSC1	3 %
TSC2	3 %
RTK/RAS pathway alterations	**Frequency**
EGFR	9 %
ERBB2	4 %
ERBB3	2 %
FGFR1	7 %
FGFR2	3 %
FGFR3	2 %
K-RAS	3 %
H-RAS	3 %
N-RAS	<1 %
RASA1	4 %
NF1	11 %
BRAF	4 %

Percentage of samples ($n = 178$) with alterations in the PI3K/RTK/RAS pathways, as obtained by the Cancer Genome Atlas Research Network, through whole-exome sequencing and whole-transcriptome expression profiles. Alterations are defined by somatic mutation, homozygous deletion, high-level, focal amplification, and in some cases by significant up- or downregulation of gene expression (AKT3, FGFR1, PTEN) [131]

Table 5.4 Molecular tumor board

Name	Age	PS	Stage	Histology	EGFR	ALK	K-RAS	ROS-1	MET	Other	Rx
JM	42	0	IIIB	NSCLC		+					crizotinib
RH	67	1	IIIA	SCCA						FGFR1	Phase I
GW	55	0	IV	NSCLC	+						erlotinib
EC	58	1	IV	NSCLC			+				Phase II
HS	63	2	E-S	SCLC							Pt/VP-16

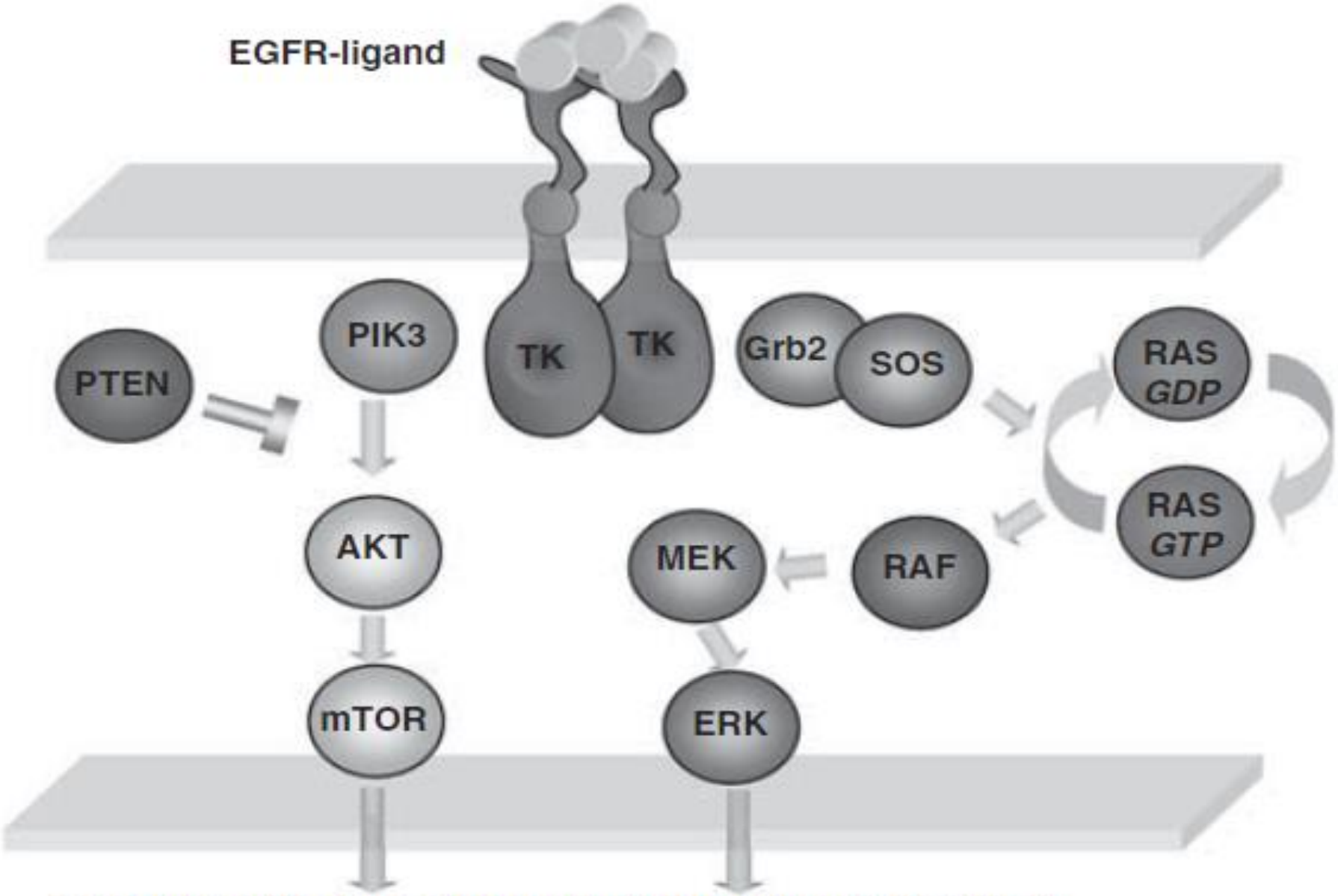
A sample representation of molecular tumor board from the thoracic oncology program at the University of Chicago

PS performance status, *E-S* extensive stage, *EGFR* activating epidermal growth factor receptor mutation, *ALK* anaplastic lymphoma kinase fusion, *ROS-1* reactive oxygen species-1 translocation, *FGFR1* fibroblast growth factor receptor 1 amplification, *Rx* proposed treatment, *Pt/VP-16* platinum-etoposide chemotherapy

Table 5.2 Mechanisms of acquired resistance to EGFR tyrosine kinase inhibitor-directed therapies

T790M and rare second-site mutations	60 %
Unknown—includes epithelial to mesenchymal transformation	30 %
Small-cell transformation	6 %
MET amplification	4 %

Percentages are based on aggregate data from the two largest re-biopsy series to date (Arcila et al., $n=99$, and Sequist et al., $n=37$) [53, 132, 133]. Small-cell transformation includes cases with histologic change to neuroendocrine differentiation. MET amplification represents cases without coexisting EGFR T790M



A Savonarola, et al. Pharmacogenetics and pharmacogenomics: role of mutational analysis in anti-cancer targeted therapy. The Pharmacogenomics Journal (2012) 12, 277 -- 286

Table 2. Mutation frequency of K-ras, BRAF, EGFR, c-KIT and PDGFRα genes obtained from overall pharmacogenetic studies in our laboratory

Disease	*Gene*	*Portion of gene analyzed*	*Frequency of mutation*
NSCLC	EGFR	Exons 18, 19, 20, 21	12.0%
mCRC	K-ras	Exons 1, 2	44.4%
	BRAF	Exon 15	3.9%
GIST	c-KIT	Exons 9, 11, 13, 17	34.2%
	PDGFRα	Exons 12, 14, 18	5.3%

Abbreviations: EGFR, epidermal growth factor receptor; GIST, gastrointestinal stromal tumor; mCRC, metastatic colorectal cancer; NSCLC, non-small cell lung cancer; PDGFRα, platelet-derived growth factor receptor α.

All molecular analyses are performed using direct sequence analysis (ABI 3130 Genetic Analyzer-Applied Biosystems) and all detected sequence variants are confirmed in duplicate experiments, using independently-extracted DNA samples.

Table 1. Most common pharmacogenomic biomarkers for selection of cancer therapy (http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm)

Clinical application	*PGx biomarker*	*Type of mutation*	*Therapeutic agent (brand)*	*Active ingredient*	*Original approval (FDA application no.)*	*Mutation-predicted therapeutic response*
NSCLC	EGFR	Tumor-activating *EGFR* mutations (exons 18–21)	Iressa (astrazeneca)	Gefitinib	5 May 2003 (NDA 021399)	*Response* to Gefitinib and Erlotinib in presence of *EGFR*-activating mutations *Resistance* to Gefitinib in presence of T790M mutation
			Tarceva (OSI Pharms)	Erlotinib Hydrochloride	18 November 2004 (NDA 021743)	
	ALK	*EML4–ALK* rearrangement (chr 2p23)	Xalcori (Pfizer)	Crizotinib	26 August 2011 (NDA 202570)	*Response* to Crizotinib in presence of *EML4–ALK* rearrangement *Resistance* to Crizotinib in presence of C1156Y and L1196M mutations
mCRC	K-ras	Tumor-activating *K-ras* mutations (mainly exon 1)	Erbitux (Imclone)	Cetuximab	12 February 2004 (BLA 125084)	*Response* to Cetuximab and Panitumumab in absence of *K-ras*-activating mutations
			Vectibix (Amgen)	Panitumumab	27 September 2006 (BLA 125147)	
GIST	PDGFRα	*PDGFRα* mutations (exons 12,14,18)	Gleevec (Novartis)	Imatinib Mesylate	18 April 2003 (NDA 021588)	*Good response* to Imatinib in presence of *c-KIT* mutations in exon 11 *Partial response* to Imatinib in presence of *c-KIT* mutations in exon 9 *Resistance* to Imatinib in presence of *c-KIT* mutations in exons 13 and 17, *PDGFRα* mutations (especially exon 18), acquisition of secondary mutations in the kinase domain
	c-KIT	Oncogenic *c-KIT* mutations (exons 9, 11, 13, 17)				
			Sutent (CPPI CV)	Sunitinib Malate	26 January 2006 (NDA 021938)	*Response* to Sunitinib in presence of *c-KIT* mutations in exon 9
Metastatic melanoma	BRAF	Mutation V600E	Zelboraf (Hoffmann La Roche)	Vemurafenib	17 August 2011 (NDA 202429)	*Response* to Vemurafenib in presence of *BRAF* mutation V600E *Resistance* to Vemurafenib in presence of reactivation of the MAPK pathway or PDGFRβ overexpression

Abbreviations: ALK, protein kinase B-alpha; EGFR, epidermal growth factor receptor; EML4, echinoderm microtubule-associated protein-like 4; FDA, Food and Drug Administration; GIST, gastrointestinal stromal tumor; K-ras, Kirsten murine sarcoma virus 2; MAPK, mitogen-activated protein kinase; mCRC, metastatic colorectal cancer; NSCLC, non-small cell lung cancer; PDGFRα, platelet-derived growth factor receptor α.

Work Cases

Biomarker Discovery

Prediction Modeling

Network Analysis

Biomarker resource

- Jun Wu, Qingchao Qiu, **Lu Xie,** Joseph Fullerton, Jian Yu, Yu Shyr, Alfred L George Jr and Yajun Yi. Web-based Interrogation of Gene Expression Signatures Using EXALT. ***BMC Bioinformatics,*** 2009, 10:420
- Ying He, Menghuan Zhang, Yuanhu Ju, Zhonghao Yu, Daqing Lv, Han Sun, Weilan Yuan, Fei He, Jianshe Zhang, Hong Li, Jing Li, Rui Wang-Sattler Yixue Li, Guoqing Zhang* and **Lu Xie***. dbDEPC 2.0: updated database of differentially expressed proteins in human cancers. ***Nucleic Acids Research.*** 2011, 1-8. doi:10.1093/nar/gkr936.

Individual biomarkers

- Zhuang G, Brantley-Sieders DM, Vaught D, Yu J, **Xie L,** Wells S, Jackson D, Muraoka-Cook R, Arteaga C, Chen J. Elevation of receptor tyrosine kinase EphA2 mediates resistance to trastuzumab therapy. ***Cancer Research***. 2010 Jan 1; 70(1):299-308.
- Zhong-Hua Tao, Jin-Liang Wan, Ling-Yao Zeng, **Lu Xie,** Hui-Chuan Sun, Lun-Xiu Qin, Lu Wang, Jian Zhou, Zheng-Gang Ren, Yi-Xue Li, Jia Fan, Wei-Zhong Wu. miR-612 Suppresses the Invasive-Metastatic Cascade in Hepatocellular Carcinom.***Journal of Experimental Medicine.*** 2013, 210(4):789-803. doi: 10.1084/jem.20120153.
- Han Sun, Xiaobin Xing, Jing Li, Fengli Zhou, Yunqin Chen, Ying He, Wei Li, Guangwu Wei, Xiao Chang, Jia Jia, Yixue Li*, **Lu Xie***. Identification of Gene Fusions from Human Lung Cancer Mass Spectrometry Data. ***BMC Genomics.*** 2013, 14(Suppl 8):S5) doi:10.1186/1471-2164-14-S8-S5.
- J Zhou, J Wu, B Li, D Liu, J Yu, X Yan, S Zheng, J Wang, L Zhang, L Zhang, F He, Q Li, A Chen, Y Zhang, X Zhao, Y Guan, X Zhao, J Yan, J Ni, M A Nobrega, B Löwenberg, R Delwel, P J M Valk, A Kumar, **L Xie,** D G Tenen, G Huang and Q-f Wang. PU.1 is essential for MLL leukemia partially via crosstalk with the MEIS/HOX pathway. ***Leukemia***. Advance online publication 21 January 2014; doi: 10.1038/leu.2013.384.

Work case 1: RNA level cancer biomarker

- **Web-based Interrogation of Gene Expression Signatures Using EXALT.** Jun Wu, Qingchao Qiu, Lu Xie…Yi Y BMC Bioinfomatics; 2009, 10:420

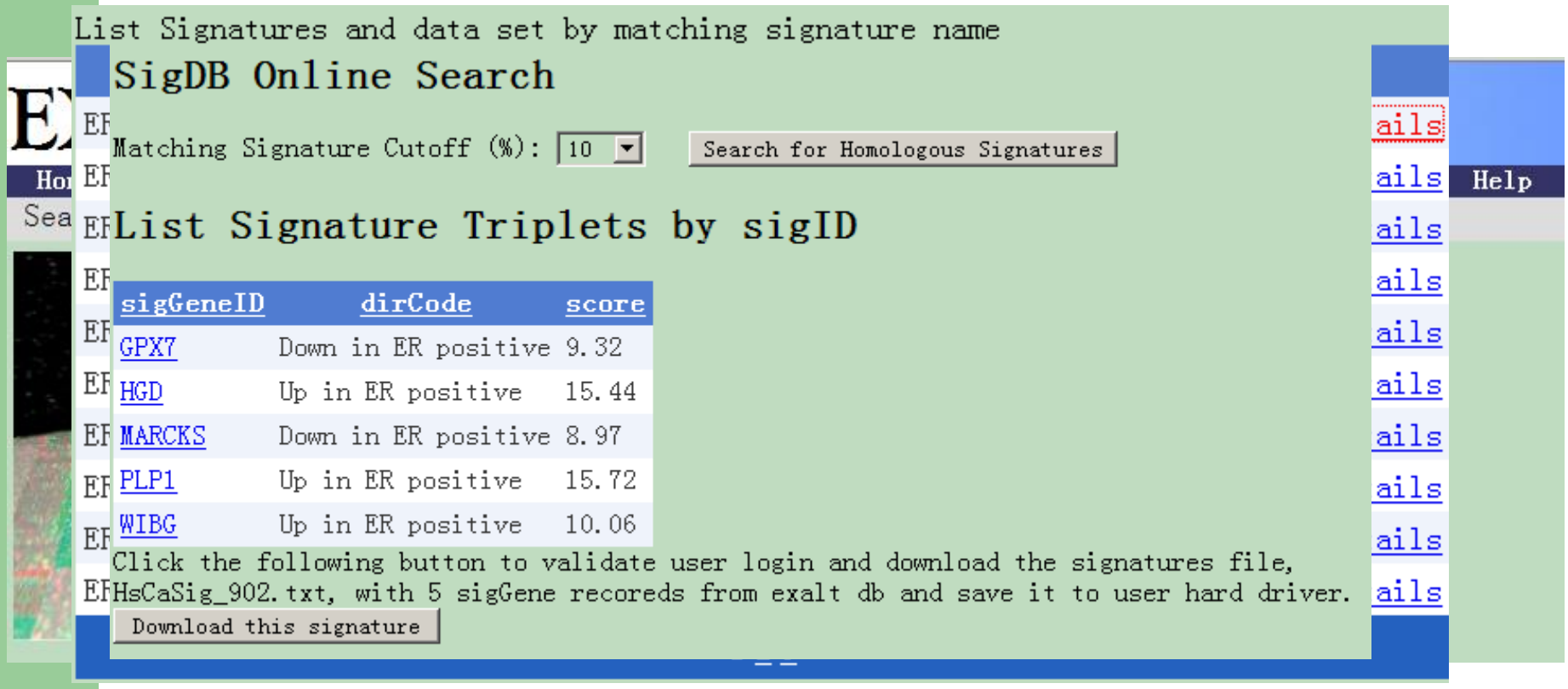
What Can EXALT Do?

- EXALT was developed to enable comparisons of microarray data across experimental platforms, different laboratories, and multiple species.

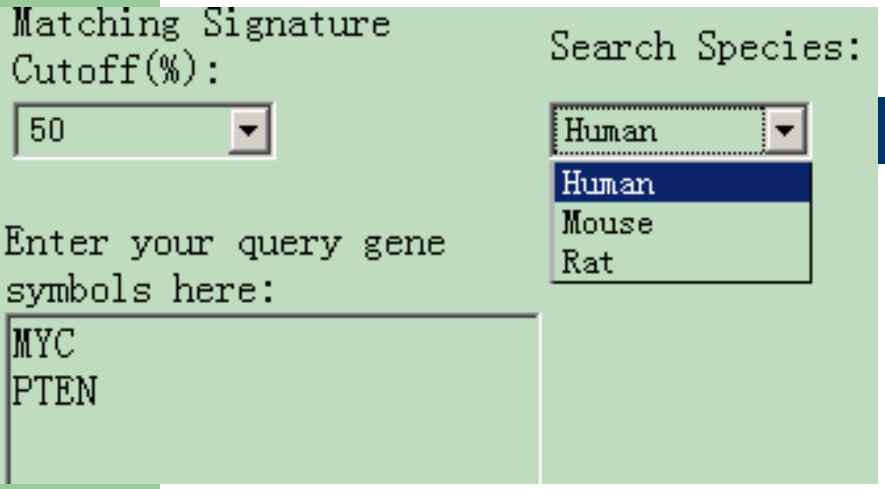
- EXALT allows investigators to use gene expression signatures (also referred to as gene sets) to query a large formatted collection of microarray results.

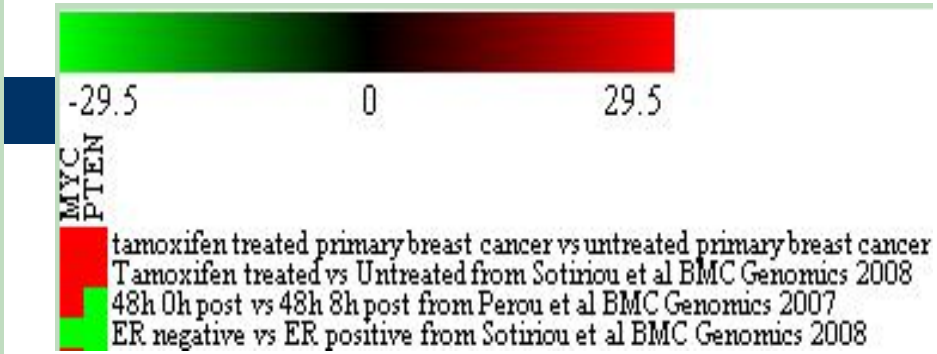
EXALT Now Supports:

GEO signature database and human cancer SigDB including almost 2000 datasets and more than 27,000 signatures. Here is an example of signature homology search:



Or you can upload a gene list to investigate whether these genes co-exist in signatures derived from other experiments





The colors in the heat map represents the direction of the differential gene expression within a given signature (red for up, green for down, and black for a missing match), and color intensities reflect the confidence levels of differential expression. From this analysis, profile suggests that MYC and PTEN both up-regulated in Tamoxifen treated breast cancer, and both down-regulated in ER- vs ER+ breast cancer.

- Compared with GeneAtlas, Oncomine or other current available tools, EXALT enables the identification of homologous data sets sharing similar expression profiles by comparing signatures, which can return more meaningful results.

- The search engine is time-effective and user-friendly to biologists and clinicians.

http://seq.mc.vanderbilt.edu/exalt/exaltHome.aspx

Wark case 2: Protein level cancer biomarker

Nucleic Acids Research Advance Access published November 16, 2011

Nucleic Acids Research, 2011, **1–8**
doi:10.1093/nar/gkr936

dbDEPC 2.0: updated database of differentially expressed proteins in human cancers

Ying He[1,2], Menghuan Zhang[2,3], Yuanhu Ju[2], Zhonghao Yu[4], Daqing Lv[2], Han Sun[1], Weilan Yuan[5], Fei He[2], Jianshe Zhang[2], Hong Li[1,2], Jing Li[3], Rui Wang-Sattler[4], Yixue Li[1,2], Guoqing Zhang[2,*] and Lu Xie[2,*]

[1]Key Laboratory of Systems Biology, Chinese Academy of Sciences, Shanghai 200031, [2]Shanghai Center for Bioinformation Technology, Shanghai 200235, [3]Department of Bioinformatics and Biostatistics, Shanghai Jiaotong University, Shanghai 200240, P. R. China, [4]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg 85764, Germany and [5]Biomedical Engineering for School of Life Sciences and Technology, Tongji University, Shanghai 200092, P. R. of China

dbDEPC

protein cancer MS experment profile network upload

http://lifecenter.sgst.cn/dbdepc/index.do

News

- dbDEPC 2.0 is released to public...
- Find association DEPC through a ...
- The profile function has a new ...
- Advanced function was achieved ...
- Datasets were expanded in dbDEPC...
- Text mining method facilitated ...
- Read more...

Statistics

- Protein: 4029
- Cancer (Subtype): 20 (18)
- MS Experiments: 331
- Literature: 241

Citation

Hong Li, Ying He, Guohui Ding, Chuan Wang, Lu Xie, and Yixue Li. dbDEPC: A Database of Differentially Expressed Proteins in Human

Welcome to dbDEPC 2.0

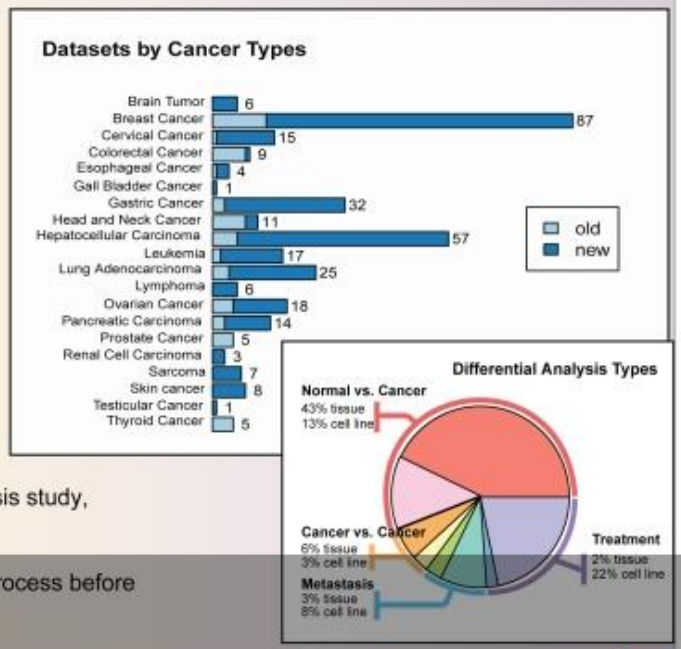
a database of Differentially Expressed Proteins in Human Cancer

Version 2 data update

331 MS experiments and 4029 DEPs

This version contains 4029 DEPs across 20 cancers with additional 18 subtypes, curated from 331 MS experiments.

Among the 4029 DEPs, 5% proteins were validated by low throughput assays, such as immunoblotting, western blotting analyais, etc.



Four main analysis types

According to the experimental design, datasets were classified into four categories, namely Normal vs. Cancer, Cancer vs. Cancer, Metastasis study, Treatment comparision.

Each dataset goes through a rigorous curation process before it is added to the dbDEPC.

1 2 3

DATABASE CONTENTS

- In the current release, dbDEPC contains information on **4092 differentially expressed proteins (DEPs)** in **20 different**

	dbDEP 1.0	**dbDEPC 2.0**
differentially expressed proteins (DEPs)	1803	4092
Cancer types	15	20 (18 subtype)
MS experiments	65	331
Publications	47	241

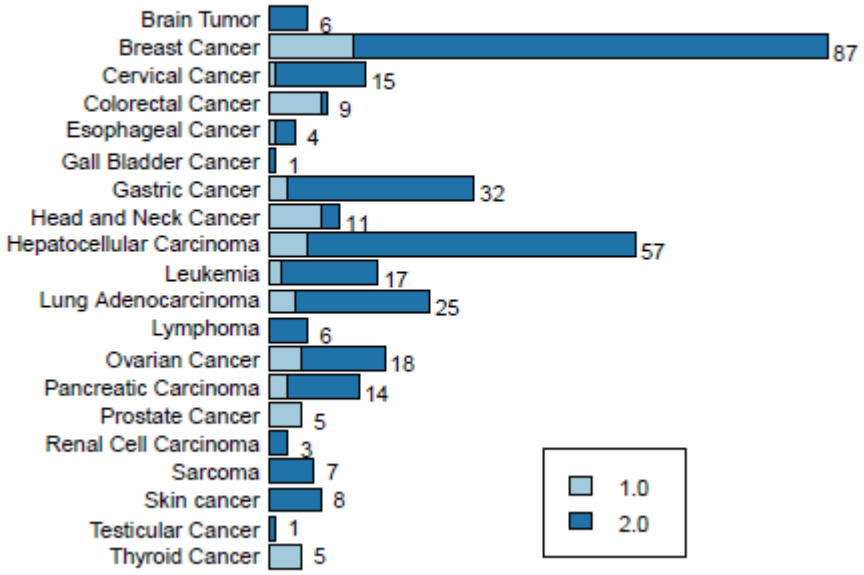
Cancer Type	Subtype *
lung adenocarcinoma	non-small cell lung carcinoma small cell lung carcinoma
hepatocellular carcinoma	hepatitis C virus hepatitis B virus
breast cancer	breast ductal carcinoma
pancreatic carcinoma	pancreatic ductal adenocarcinoma
leukemia	chronic myeloid leukemia chronic lymphocytic leukemia acute myeloid leukemia acute lymphoblastic leukemia
thyroid cancer	papillary thyroid carcinoma follicular thyroid carcinoma follicular thyroid adenoma
skin cancer *	melanoma non-melanoma
brain tumor *	neuroblastoma
head and neck cancer *	oral cancer oral premalignant lesions
gastric cancer	
colorectal cancer	
prostate cancer	
esophageal cancer	
cervical cancer	
ovarian cancer	
renal cell carcinoma *	
lymphoma *	
sarcoma *	
testicular cancer *	
gall bladder cancer *	

Table 1 Human cancer types in dbDEPC 2.0

* marked the new human cancer types and subtypes in dbDEPC 2.0.

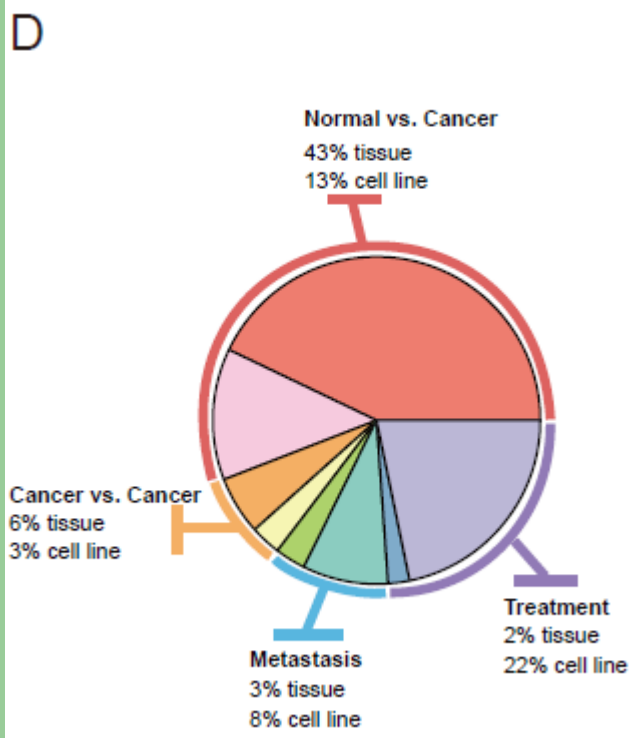
Data content additions

C



(C) Increasing number of experiment datasets in each cancer.

experimental descriptions



According to the experimental designs, all datasets could be categorized **into four types of studies:**

normal vs. cancer comparison: 56%
cancer vs. cancer comparison: 9%
cancer metastasis studies: 11%
treatment research:24%

A

DRAW A Cancer Profile Heatmap

Protein list

P16070
RCN1_HUMAN
IPI00879231
NP_058017.1
ENSP00000349259
GI:4504517
ARHGDIA

Experimental de

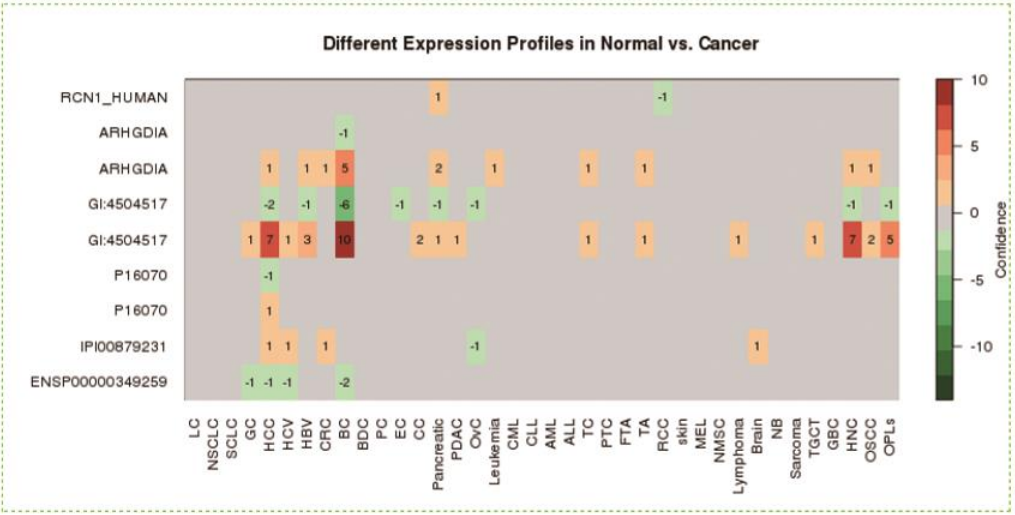
Normal vs Can
Metastasis
Treatment
Cancer vs Can

(Note: Please type in one protein per line.)

Profile | Reset | example

B

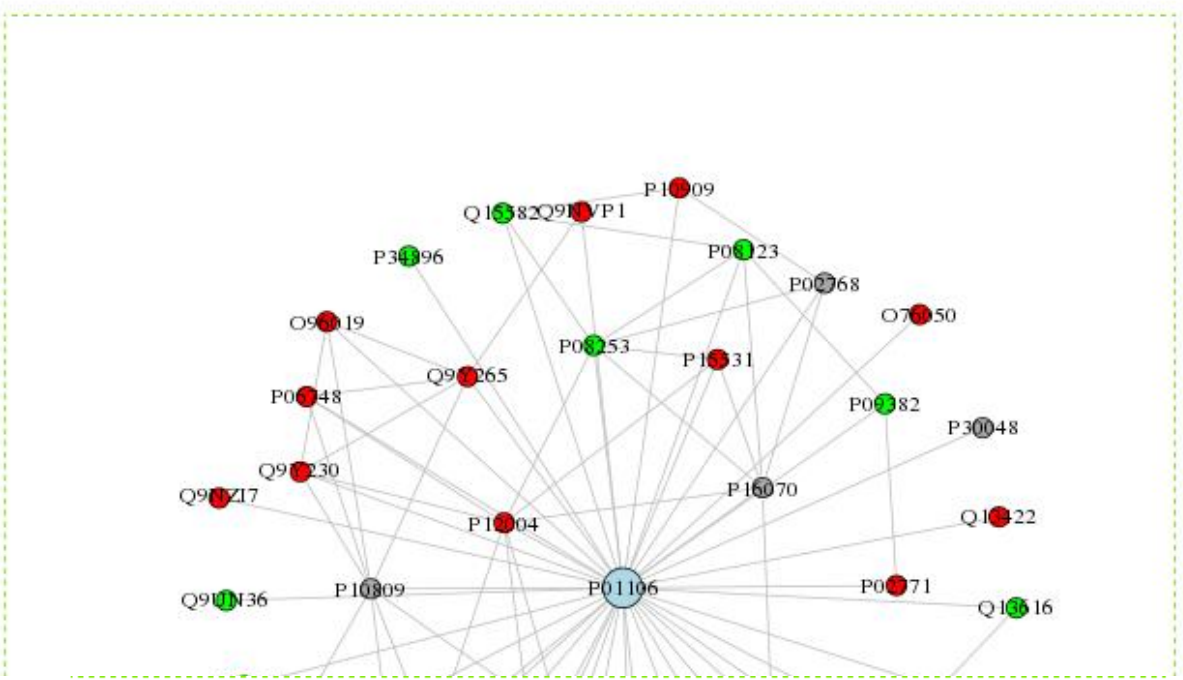
Profile Patterns in Cancers



All | Validated | Not validated

UniProt AC	Organism	Description	Cancer Name	Design	Diff	Ratio	ExpID
P04792	HUMAN	Heat shock protein beta-1	Hepatocellular Carcinoma	Normal vs. Cancer	Up	2.7	EXP00104
P04792	HUMAN	Heat shock protein beta-1	Hepatocellular Carcinoma	Normal vs. Cancer	Up	2.6	EXP00004
P04792	HUMAN	Heat shock protein beta-1	Breast Cancer	Normal vs. Cancer	Up	0.57	EXP00236
P04792	HUMAN	Heat shock protein beta-1	Hepatocellular Carcinoma	Normal vs. Cancer	Down	0.45	EXP00003
P04792	HUMAN	Heat shock protein beta-1	Breast Cancer	Normal vs. Cancer	Down	0.75	EXP00229
P04792	HUMAN	Heat shock protein beta-1	Hepatocellular Carcinoma (hepatitis C virus)	Normal vs. Cancer	Up	4.4	EXP00103
P04792	HUMAN	Heat shock protein beta-1	Breast Cancer	Normal vs. Cancer	Down	0.31	EXP00221
P04792	HUMAN	Heat shock protein beta-1	Hepatocellular Carcinoma (hepatitis B virus)	Normal vs. Cancer	Down	0.052	EXP00102
P04792	HUMAN	Heat shock protein beta-1	Breast Cancer	Normal vs. Cancer	Down	0.27	EXP00221
P04792	HUMAN	Heat shock protein beta-1	Breast Cancer	Normal vs. Cancer	Up	8	EXP00079

68 1 / 7 1 2 3 4 5 6 7 >> 10 1



We also provide a downloadable zip file containing two files, which can be imported by Cytoscape.

Associated DEPs list

- Query:MYC(P01106)
  - Lung Adenocarcinoma
    P22626 P48200 P15531 P12004 P30279 Q13145 Q13422 P06733
  - Lung Adenocarcinoma(Non-small cell lung carcinoma)
    P22626 P48200 P15531 P12004
  - Lung Adenocarcinoma(Small cell lung carcinoma)
  - Gastric Cancer
    P00338 P04792 P30048 P01833 P15531 P12004 P02768 Q15582 P09651 P10809 P09382 P08123 P08253 P06733 P04406 Q07021
  - Hepatocellular Carcinoma
    P22626 Q9NVP1 P02771 P10909 O96019 Q9NZI7 O76050 Q9Y265 P06748 P15531 P12004 P34932 P54868 P00338 P02786 Q13616 Q9UN36 Q92597 P30279 P34896 P02768 P16070 P04792 P30048 P11166 P06733 P04406 P10809
  - Hepatocellular Carcinoma(hepatitis C virus)
    P02768 P04792 P06748 P15531 P54868 P00338 P30048 P30279 P10809
  - Hepatocellular Carcinoma(hepatitis B virus)
    P02771 P10909 O96019 Q9NZI7 O76050 Q9Y265 P06748 P10809 P15531 P12004 P34932 P54868 Q13616 Q9UN36 P02768 P04792 P06733 P04406
  - Colorectal Cancer
    P06733 Q9Y230 P10809 P15531 Q13330 P02768 P25054

Sequence Variation

1	11	21	31	41
MSSSGTPDLP	VLLTDLKIQY	TKIFINNEWH	DSVSGKKFPV[A]	FNPATEEELC
51	61	71	81	91
QVEEGDKEDV	DKAVKAARQA	FQIGSPWRTM	DASERGRLLY	KLADLIERDR
101	111	121	131	141
LLLATMESMN	GGKLYSNA[S]YL	N[S]DLAG[R]CIKTL	RYCAGWADKI	QGRTIPIDGN
151	161	171	181	191
FFTYTRHEPI	GVCGQIIPWN	FPLVMLI[F]WKI	GPALSCGNTV	VVKPAEQTPL
201	211	221	231	241
TALHVASLIK	EAGFPPGVVN	IVPGYGPTAG	AAISSHMDID	KVAFTGSTEV
251	261	271	281	291
GKLIKEAAGK	SNLKRVTLEL	GGKSPCIVLA	DADLDNAVEF	AHHGVFYHQG
301	311	321	331	341
QCCIAASRIF	VEESIYDEFV	RRSVERAKKY	ILGNPLTPGV	TQGPQIDKEQ
351	361	371	381	391
YDKILDLIES	GKKEGAKLEC	GGGPWGNKGY	FVQPTVFSNV	TDEMRIAKEE
401	411	421	431	441
IFGPVQQIMK	FKSLDDVIKR	ANNTFYGLSA	GVFTKDIDKA	ITISSALQAG
451	461	471	481	491
TVWVNCYGVV	SAQCPFGGFK	MSGNGRELGE	YGFHEYTEVK	TVTVKISQKN
501				
S				

dbSNP cancer

- A possible **scenario** of how dbDEPC could benefit cancer studies.

st

metastasis

breast cancer

Experiment Results

Please select conditions to filter the experiment results.

Design: Metastasis

Cancer Name: Breast Cancer

Sample Type: please select

Sample Size: please select

Organism: please select

Mass-Spectrometry:

Search-Engine:

Quantification: please select

Filter

Please select experiments in the check box, and then view the differentially expressed proteins (DEPs); Or select two to three experiments to view the DEPs intersection among these experiments.

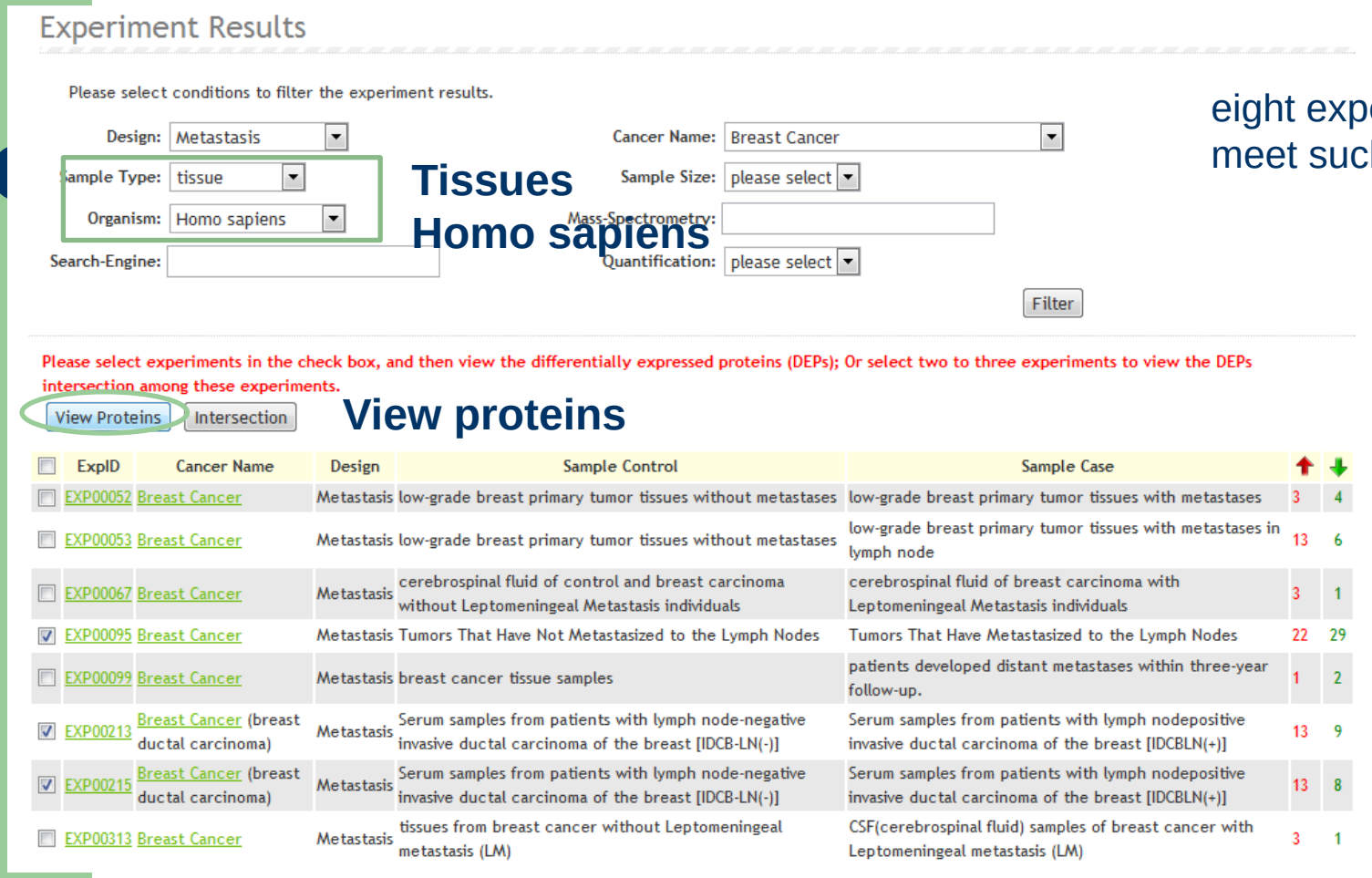
View Proteins | Intersection

ExpID	Cancer Name	Design	Sample Control	Sample Case	↑	↓
EXP00008	Breast Cancer	Metastasis	human MDA-MB-435 breast cancer cell line	highly metastatic variant of human MDA-MB-435 breast cancer cell line	10	0
EXP00037	Breast Cancer	Metastasis	HER-2/neu-negative breast cancer (MDA-MB-231 and MCF-10A cell line)	HER-2/neu-positive breast cancer (BT474 and SKBr3 cell line)	9	0
EXP00052	Breast Cancer	Metastasis	low-grade breast primary tumor tissues without metastases	low-grade breast primary tumor tissues with metastases	3	4
EXP00053	Breast Cancer	Metastasis	low-grade breast primary tumor tissues without metastases	low-grade breast primary tumor tissues with metastases in lymph node	13	6
EXP00063	Breast Cancer	Metastasis	67NR(The 67NR cancer cell line can form primary tumor but no tumor cells can be detected in any distant tissues including blood, lymph nodes and the lungs)	168FARN(cells of the 168FARN line disseminate from mammary fat pads and can be detected in lymph node but are rarely detectable in lung indicating that they are unable to accomplish extravasation effectively)	36	43
EXP00064	Breast Cancer	Metastasis	67NR(The 67NR cancer cell line can form primary tumor but no tumor cells can be detected in any distant tissues including blood, lymph nodes and the lungs)	4T07(cells of the 4T07 cells are able to spread to the lungs but cannot establish visible metastatic nodules)	40	58
EXP00065	Breast Cancer	Metastasis	67NR(The 67NR cancer cell line can form primary tumor but no tumor cells can be detected in any distant tissues including blood, lymph nodes and the lungs)	4T1(cells of the 4T1 lines are able to complete all steps of metastasis and form visible metastatic nodules in the lungs efficiently)	41	40
EXP00067	Breast Cancer	Metastasis	cerebrospinal fluid of control and breast carcinoma without Leptomeningeal Metastasis individuals	cerebrospinal fluid of breast carcinoma with Leptomeningeal Metastasis individuals	3	1

you can see 24 experiments

issues.

eight experiments would meet such criteria.

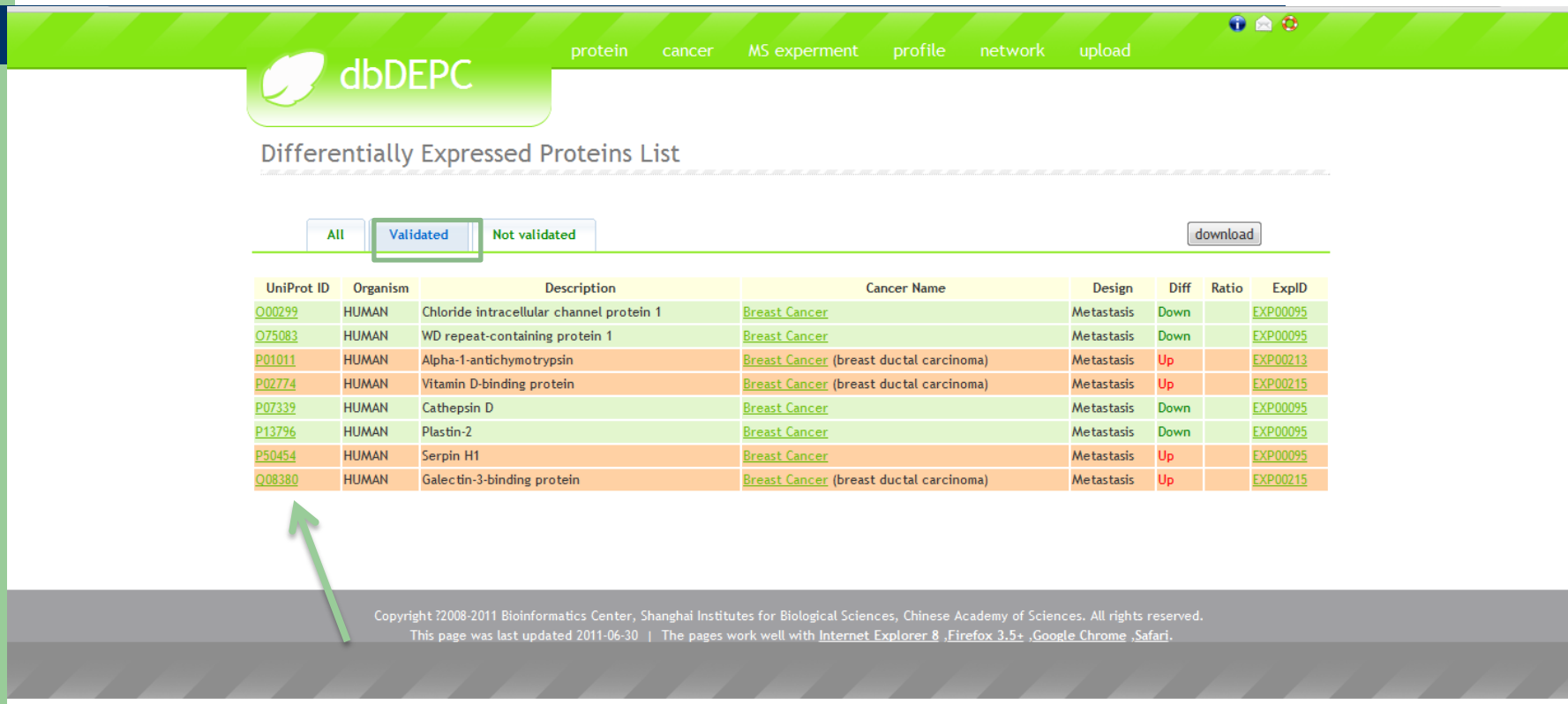


ExpID	Cancer Name	Design	Sample Control	Sample Case		
EXP00052	Breast Cancer	Metastasis	low-grade breast primary tumor tissues without metastases	low-grade breast primary tumor tissues with metastases	3	4
EXP00053	Breast Cancer	Metastasis	low-grade breast primary tumor tissues without metastases	low-grade breast primary tumor tissues with metastases in lymph node	13	6
EXP00067	Breast Cancer	Metastasis	cerebrospinal fluid of control and breast carcinoma without Leptomeningeal Metastasis individuals	cerebrospinal fluid of breast carcinoma with Leptomeningeal Metastasis individuals	3	1
EXP00095	Breast Cancer	Metastasis	Tumors That Have Not Metastasized to the Lymph Nodes	Tumors That Have Metastasized to the Lymph Nodes	22	29
EXP00099	Breast Cancer	Metastasis	breast cancer tissue samples	patients developed distant metastases within three-year follow-up.	1	2
EXP00213	Breast Cancer (breast ductal carcinoma)	Metastasis	Serum samples from patients with lymph node-negative invasive ductal carcinoma of the breast [IDCB-LN(-)]	Serum samples from patients with lymph nodepositive invasive ductal carcinoma of the breast [IDCBLN(+)]	13	9
EXP00215	Breast Cancer (breast ductal carcinoma)	Metastasis	Serum samples from patients with lymph node-negative invasive ductal carcinoma of the breast [IDCB-LN(-)]	Serum samples from patients with lymph nodepositive invasive ductal carcinoma of the breast [IDCBLN(+)]	13	8
EXP00313	Breast Cancer	Metastasis	tissues from breast cancer without Leptomeningeal metastasis (LM)	CSF(cerebrospinal fluid) samples of breast cancer with Leptomeningeal metastasis (LM)	3	1

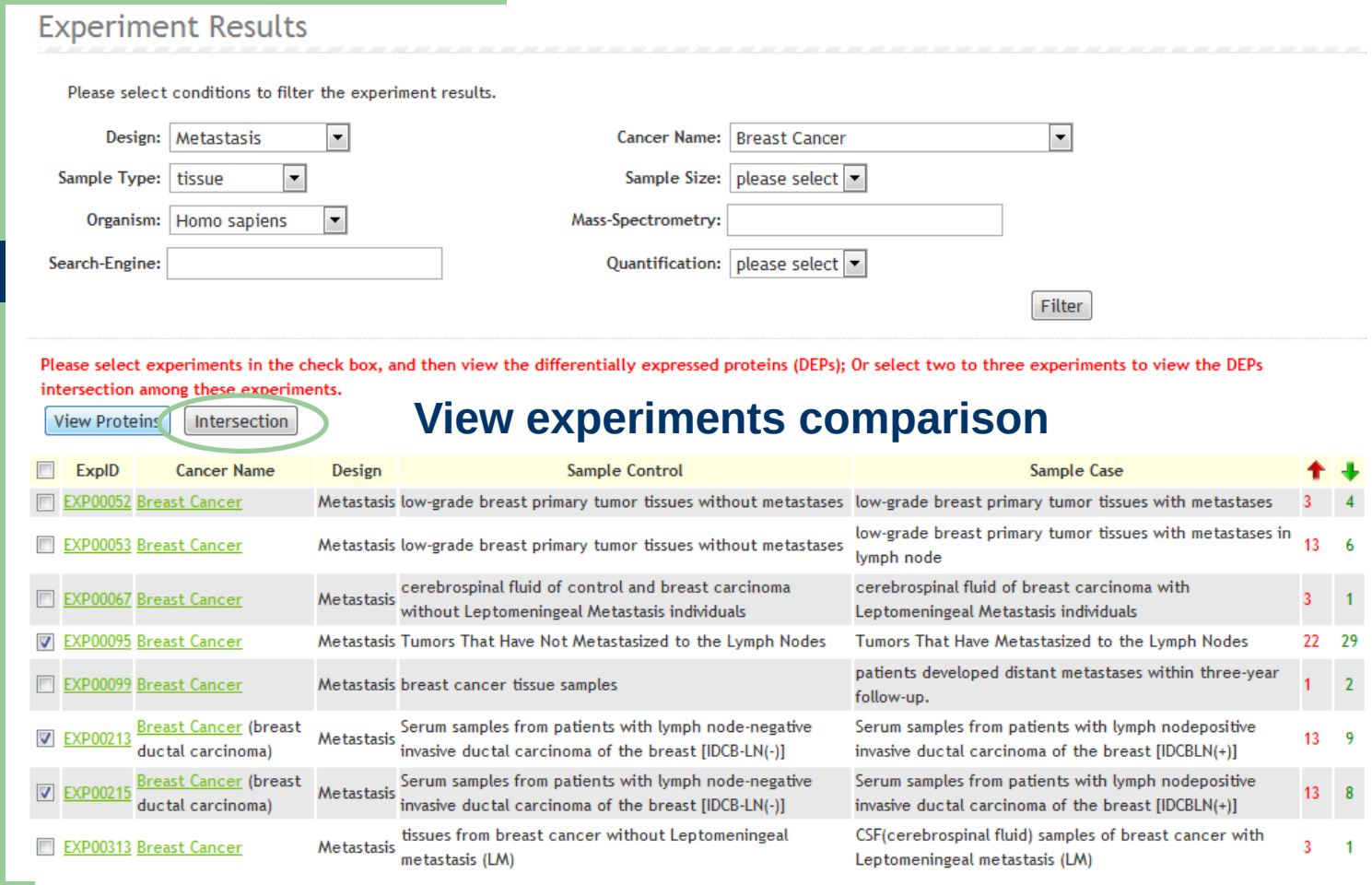
focusing on breast cancer concerning lymph node metastases, you select three experiments (EXP00095, EXP00213 and EXP00215)

•**“View Proteins” button**, all differentially expressed proteins identified in three experiments are provided as a downloadable tab separate text file, in this case, 94 proteins altogether.

•**“Validated” tag**, the user can see that only eight proteins were validated by traditional biochemical assays.

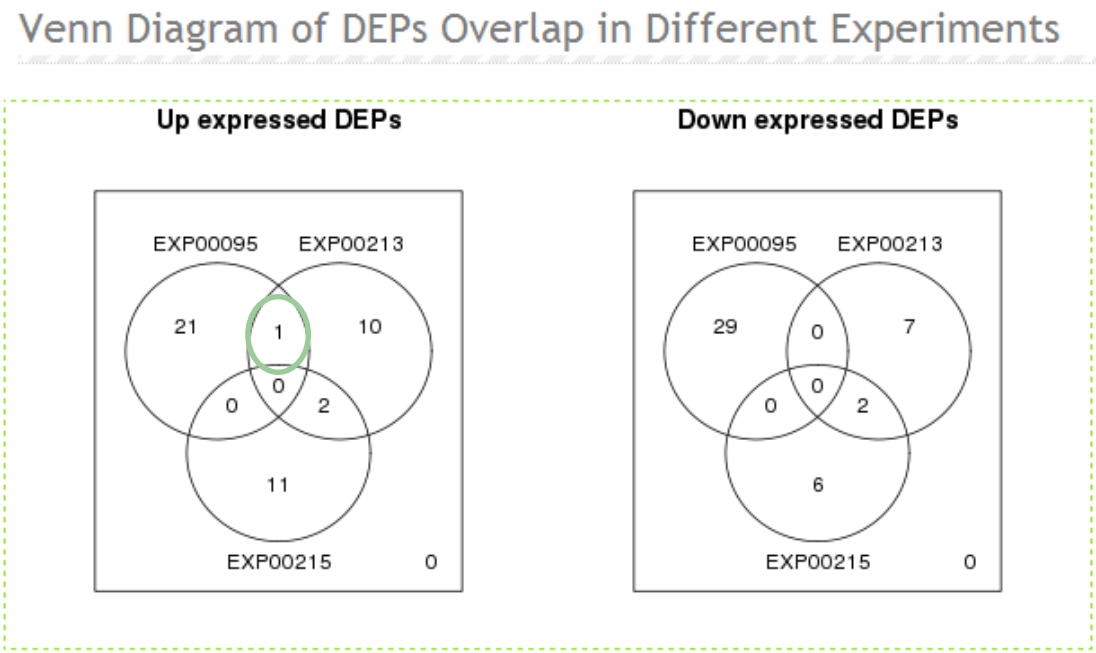


View detail protein annotation information



focusing on breast cancer concerning lymph node metastases, you select three experiments (EXP00095, EXP00213 and EXP00215)

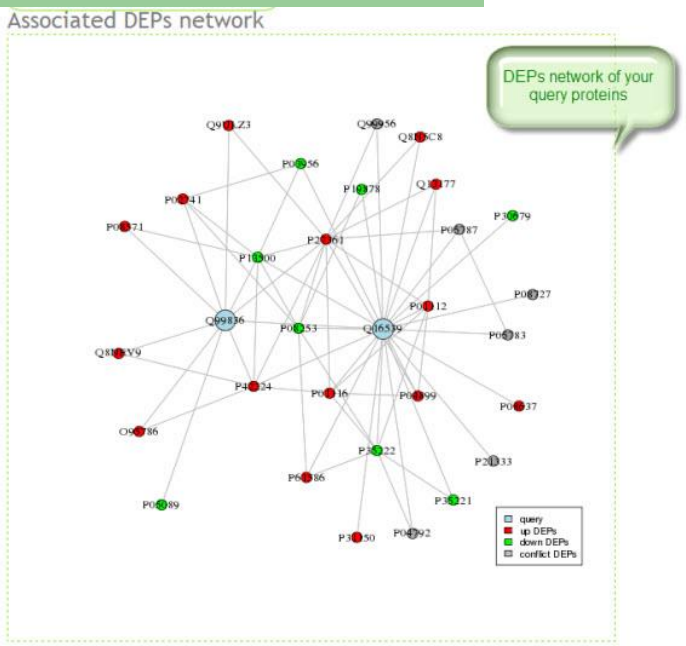
View experiments comparison



•the three experiments come up with two venn-diagrams showing **the intersection of up-regulated and down-regulated proteins** respectively in these experiments.

•Protein numbers that were identified by multiple experiments can be seen, such as, one up-regulated protein was identified by two studies (EXP00095 and EXP00213), etc.

DEPs association network tool



Citation by other papers

- major protein bioinformatics databases and resources that are relevant to comparative proteomics research.
- dbDEPC is a curated database that contains over 4000 protein entries, from 331 experiments across 20 types of human cancers. The database may be used to search for particular proteins to determine their range of expression and changes as may be related to genomic aberrations. Information is provided pertaining to experimental design, and tools are available for filtering and for network analysis.

WIREs Syst Biol Med 2012, 4:327–337. doi: 10.1002/wsbm.1169

Prediction Modeling

- Wenxi Li#, **Lu Xie#,** Xianghuo He#, Jinjun Li#, Kang Tu, Lin Wei , Jun Wu, Yong Guo, Xi Ma, Pingping Zhang, Zhimei Pan, Xin Hu, Yingjun Zhao, Haiyang Xie, Guoping Jiang, Taoyang Chen, Jianneng Wang, Shusen Zheng, Jing Cheng, Dafang Wan, Shengli Yang, Yixue Li, Jianren Gu. MicroRNA Signatures in Hepatocellular Carcinoma Diagnostic and Prognostic Implications. ***International Journal of Cancer,*** 2008, 123(7):1616-1622.
- Lin Wei, Baofeng Lian, Yuannv Zhang, Wei Li, Jianren Gu, Xianghuo He*, **Lu Xie***. Application of microRNA and mRNA expression profiling on prognostic biomarker discovery for hepatocellular carcinoma. ***BMC Genomics.*** 2014, **15**(Suppl 1):S13 doi:10.1186/1471-2164-15-S1-S13

- Tao Huang , Kang Tu, Yu Shyr, Chao-Chun Wei, **Lu Xie*** and Yi-Xue Li*. The prediction of interferon treatment effects based on time series microarray gene expression profiles. ***Journal of Translational Medicine,*** 2008, 6:44.
- Cai Y-D*, Huang T, Feng K-Y, Hu L, **Xie L*.** (2010) A Unified 35-Gene Signature for both Subtype Classification and Survival Prediction in Diffuse Large B-Cell Lymphomas. ***PLoS ONE*** 2010, 5(9): e12726. doi:10.1371/journal.pone.001272

Work case 3: Prognosis prediction

- microRNAs profiles of 78 matched cancer/noncanerous liver tissues from HCC patients and 10 normal liver tissues. 69 miRNAs were differentially expressed between hepatocellular carcinoma (HCC) and corresponding noncancerous liver tissues (N). The set of differentially expressed miRNAs could distinctly classify HCC, N and normal liver tissues (NL).

- some of these differentially expressed miRNAs were related to the clinical factors of HCC patients.
- high expression of hsa-miR-125b was correlated with good survival of HCC patients. overexpression of miR-125b in HCC cell line could suppress cell growth and phosporylation of Akt.

TABLE I – A SET OF 10 miRNAs WHICH COULD DISCRIMINATE HCC *VERSUS* NONCANCEROUS LIVER (N) VERSUS NORMAL LIVER (NL) ($p < 0.05$)

miRNAs	Chromosome location	Host gene	Log ratios of miRNAs expression			Relevance to cancer
			HCC *vs.* N	HCC *vs.* NL	N *vs.* NL	
hsa-miR-101	1p3.1	Intergenic	−0.958	−2.236	−1.278	LC, CLL, BC, HCC
hsa-miR-148a	7p15.2	Intergenic	−0.483	−1.363	−0.880	PC,BC
hsa-miR-181a	9q33.3	NR6A1	0.438	1.314	0.876	PG
hsa-miR-214	1q24.3	DNM3	−0.383	0.787	1.170	
hsa-miR-221	Xp11.3	Intergenic	1.570	2.704	1.133	CLL, CC, PAC, SC, PG, PTC, PC, BC
hsa-miR-222	Xp11.3	Intergenic	1.405	2.389	0.984	PG, HCC, PTC, PC, BC, CLL
hsa-miR-25	7q22.1	MCM7	0.941	1.764	0.823	PTC, PC, PAC, SC, PG, PC, BC
hsa-miR-29c	1q32.2	Intergenic	−0.694	−1.894	−1.200	CLL, PTC, PC, BC
hsa-miR-34a	1p36.22	Intergenic	0.480	1.480	1.001	CLL, CC
hsa-miR-424	xq26.3	OTTHUMT00000058372	−0.472	−1.817	−1.345	

BC, breast cancer; CC, colorectal cancer; CLL, chronic lymphocytic leukemia; HCC, hepatocellular carcinoma; LC, lung cancer; PA, pituitary adenomas; PAC, pancreatic cancer; PC, prostate cancer; PG, primary glioblastoma; PTC, papillary thyroid carcinoma; SC, stomach cancer.

10 miRNAs to discriminate HCC vs Non-tumor vs Normal liver

TABLE II – MICRORNAs CORRELATED WITH DIFFERENT CLINICO-PATHOLOGICAL FEATURES OF HCC PATIENTS

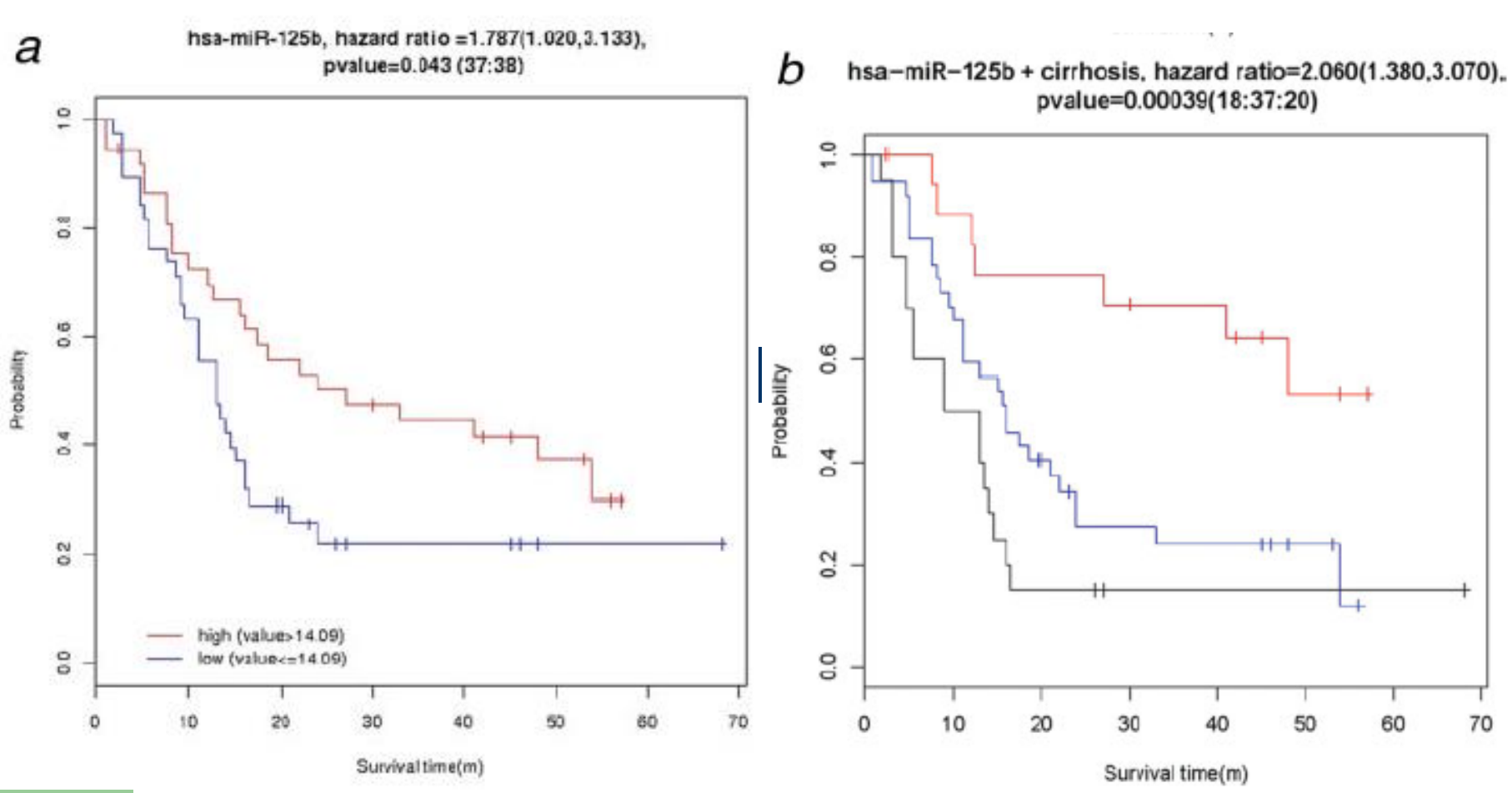
Expression of miRNAs	Metastasis			Cirrhosis			HBV		
	No	Yes	*p* value	No	Yes	*p* value	Negative	Positive	*p* value
hsa-miR-106a	Low	High	0.006						
hsa-miR-106b	Low	High	0.000						
hsa-miR-17-5p	Low	High	0.007						
hsa-miR-194	Low	High	0.000						
hsa-miR-25	Low	High	0.000						
hsa-miR-296	Low	High	0.004	Low	High	0.014			
hsa-miR-30d	Low	High	0.000						
hsa-miR-92	Low	High	0.003						
hsa-miR-93	Low	High	0.000						
hsa-miR-99b	Low	High	0.020						
hsa-miR-202	High	Low	0.013						
hsa-miR-424	High	Low	0.042						
hsa-miR-516-3p	High	Low	0.023						
hsa-miR-98	High	Low	0.000						
hsa-miR-122a	High	Low	0.044				Low	High	0.002
hsa-miR-210							High	Low	0.035
hsa-miR-103				High	Low	0.004			

miRNA expression correlated with clinical features

TABLE III – MULTIVARIATE COX HAZARD REGRESSION ANALYSIS FOR PROGNOSTIC FACTORS

Variate	Subset	Hazard ratio (95% confidence interval)	*p* value
Sex	Male/female	1.373 (0.614–3.071)	0.440
Age	≥50/<50	0.951 (0.486–1.863)	0.880
AFP	>400/<400µg/l	1.085 (0.599–1.965)	0.790
Cirrhosis	Yes/no	2.821 (1.409–5.647)	0.003
HBsAg	Positive/negative	0.772 (0.330–1.799)	0.550
Alcohol abuse	Yes/no	0.481 (0.207–1.118)	0.089
hsa-miR-125b	High/low	0.562 (0.307–1.027)	0.061

Prognostic clinical features



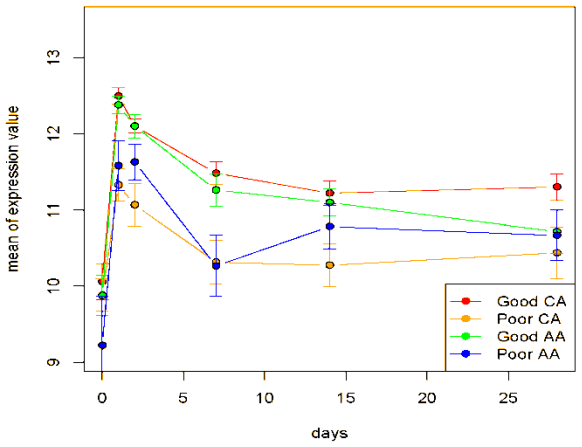
Prognosis by single miRNA
Prognosis by single miRNA + cirrhosis

Work case 4: treatment prediction

- Tao Huang , Kang Tu, Yu Shyr, Chao-Chun Wei, Lu Xie* and Yi-Xue Li*. The prediction of interferon treatment effects based on time series microarray gene expression profiles. **Journal of Translational Medicine,** 2008, 6:44.

- Gene expression profiles—subtypes of patients—response to therapy—choice of therapy

- Time series gene expression profiles—prediction of therapy response



Personalized Medicine

Data Source

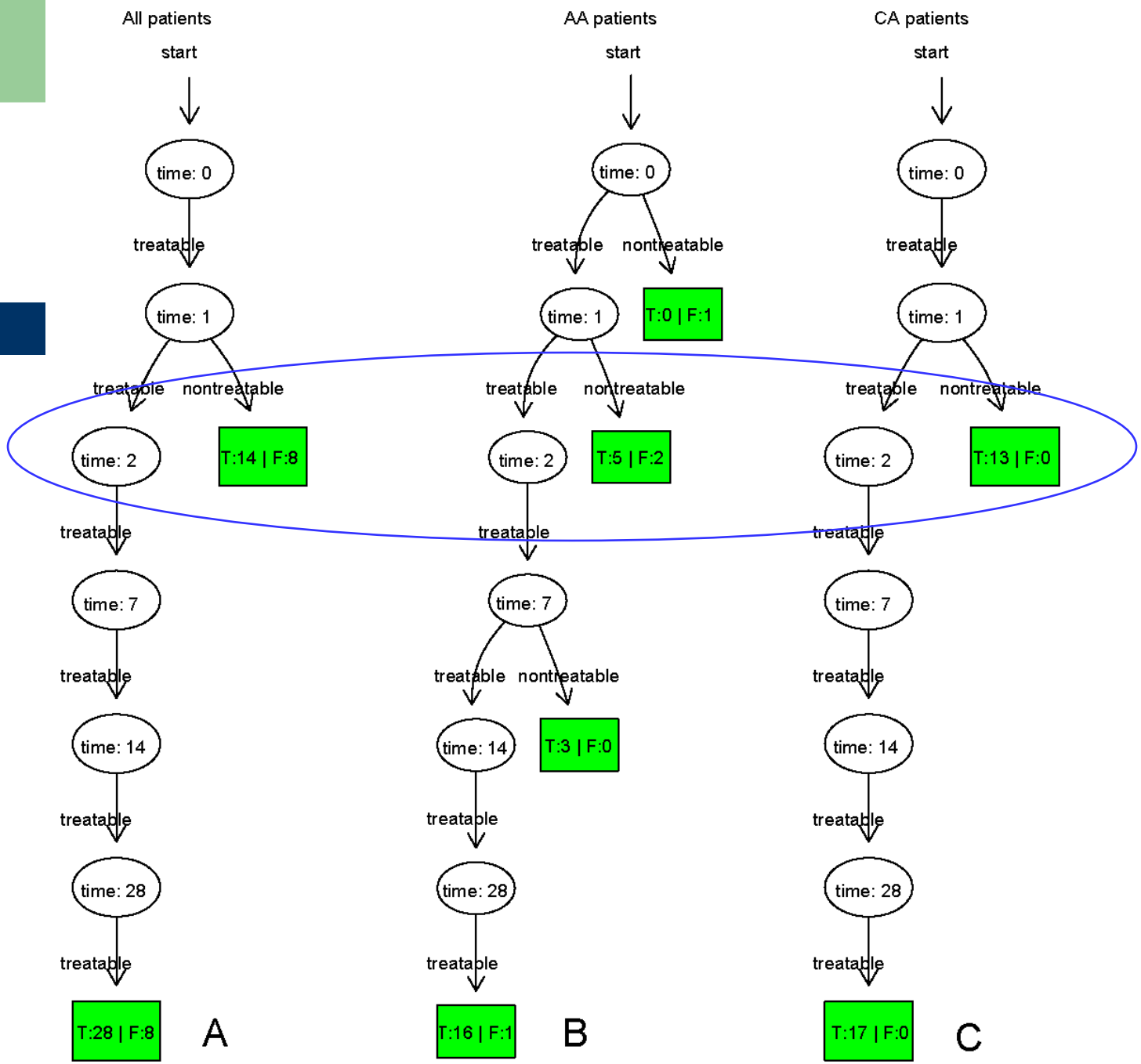
- In Milton W. Taylor's study, 33 African-American (AA) and 36 Caucasian American (CA) patients with chronic HCV genotype 1 infection received pegylated interferon and ribavirin therapy.

- HG-U133A GeneChip was used to analyze the global gene expression in peripheral blood mononuclear cells (PBMC) of all the patients on days 0 (pretreatment), 1, 2, 7, 14, and 28 (GSE7123).
- HCV RNA levels were assessed by a quantitative PCR-based assay on day 0 (pretreatment) and 28.

JOURNAL OF VIROLOGY, Apr. 2007, p. 3391–3401

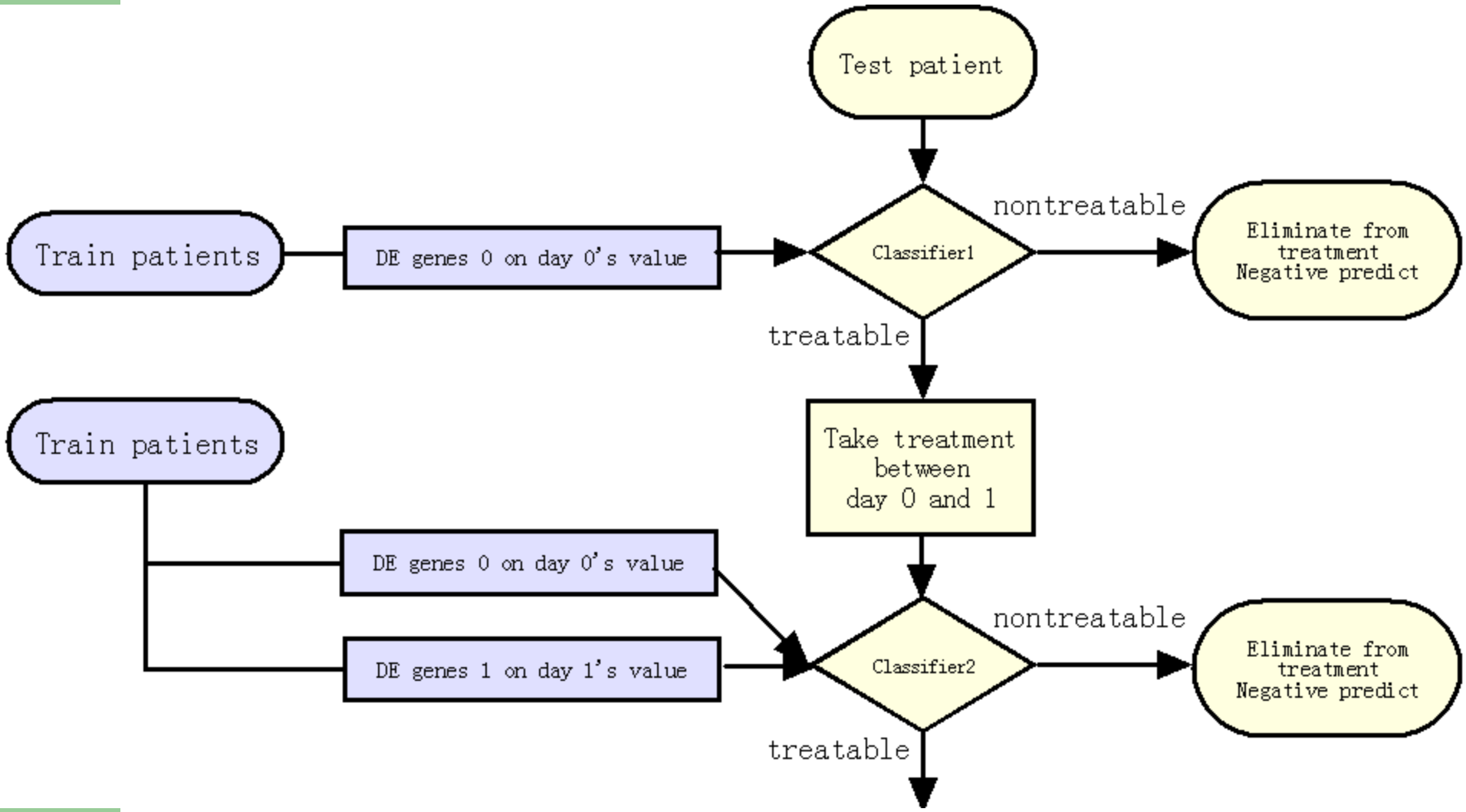
Time-dependent diagnostic model

- The main idea of our model is to fully utilize gene expression profiles before and during treatment to predict the final treatment outcome.


All patients
start
time: 0
treatable
time: 1
treatable
nontreatable
time: 2
T:14 | F:8
treatable
time: 7
treatable
time: 14
treatable
time: 28
treatable
T:28 | F:8
A
AA patients
start
time: 0
treatable
nontreatable
time: 1
T:0 | F:1
treatable
nontreatable
time: 2
T:5 | F:2
treatable
time: 7
treatable
nontreatable
time: 14
T:3 | F:0
treatable
time: 28
treatable
T:16 | F:1
B
CA patients
start
time: 0
treatable
time: 1
treatable
nontreatable
time: 2
T:13 | F:0
treatable
time: 7
treatable
time: 14
treatable
time: 28
treatable
T:17 | F:0
C

Classifier : C4.5

Simplified predictor: Day 1 classifier



The Leave-one-out cross-validation results of all patients, AA patients and CA patients on day 1

所有病人　　黑人　　白人

		All patients		AA patients		CA patients	
		Predicted Good	Predicted poor	Predicted Good	Predicted poor	Predicted Good	Predicted poor
2×2 Table	Actual Good	28	8	16	3	17	0
	Actual poor	8	14	4	5	0	13
Accuracy		72.4%		75%		100%	

疗效预测准确率

* Good means good response, poor means poor response.

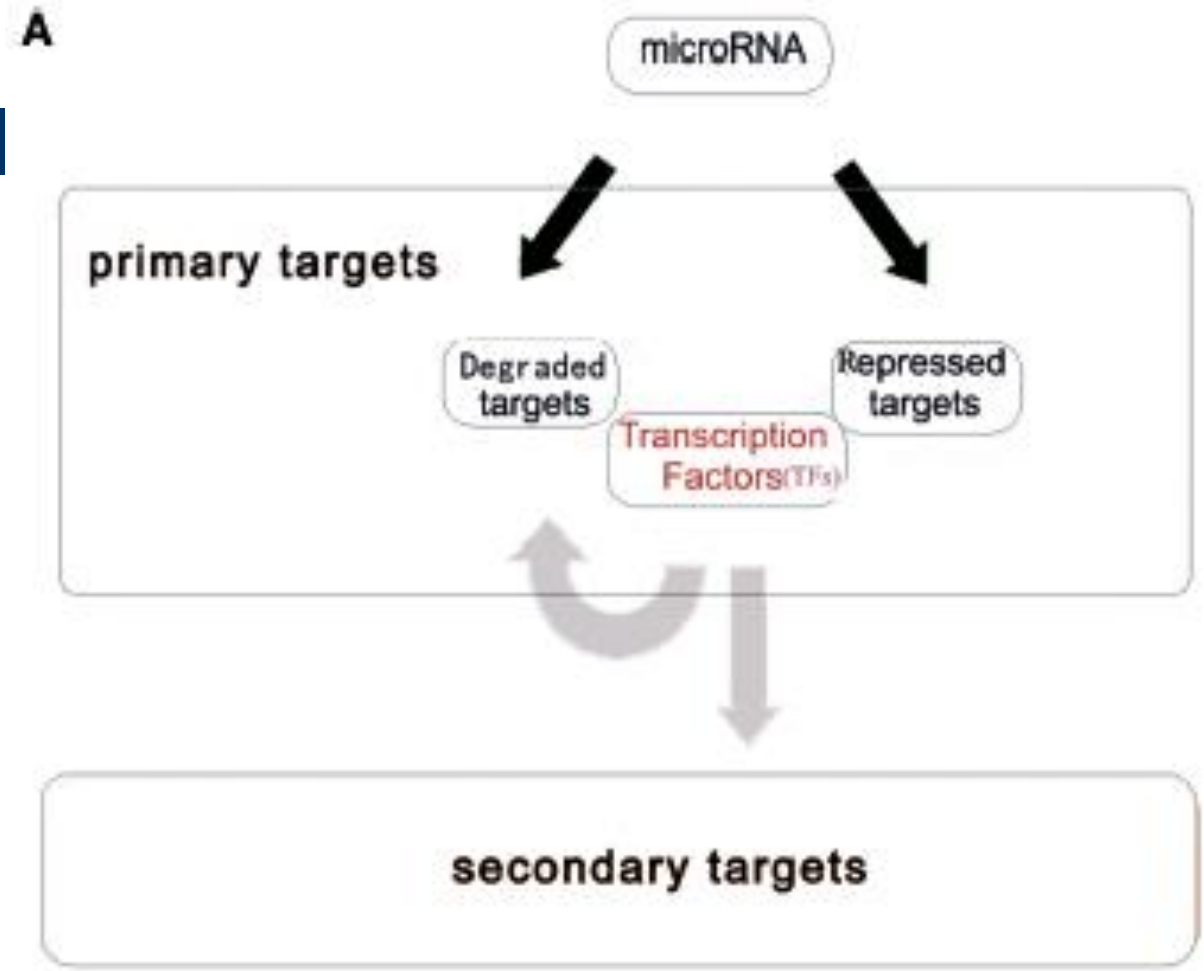
Function Module: network analysis

- Kang Tu, Hui Yu, Youjia Hua, Yuan-Yuan Li, Lei Liu*, **Lu Xie***, Yixue Li*. Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms. ***Nucleic Acids Research,*** 2009, 37 (18): 5969-5980
- Hui Yu, Kang Tu, Yi-Jie Wang, Jun-Zhe Mao, **Lu Xie***, Yuan-Yuan Li*, Yi-Xue Li*. Combinatorial Network of Transcriptional Regulation and microRNA Regulation in Human Cancer. ***BMC Systems Biology.*** 2012, **6**:61 doi:10.1186/1752-0509-6-61.
- Lingyao Zeng, Jian Yu, Tao Huang, Huliang Jia, Qiongzhu Dong, Fei He, Weilan Yuan, Lunxiu Qin, Yixue Li*, **Lu Xie***. Differential combinatorial regulatory network analysis related to venous metastasis of hepatocellular carcinoma. ***BMC Genomics.*** 2012, 13(Suppl 8):S14. http://www.biomedcentral.com/1471-2164/13/S8/S14.

- Tao Huang, Lei Liu, Qi Liu, Guohui Ding, Eugene Tan, Zhidong Tu, Yixue Li, Hongyue Dai*, **Lu Xie*.** The role of hepatitis C virus in the dynamic protein interaction networks of hepatocellular cirrhosis and carcinoma. ***International Journal of Computational Biology and Drug Design*** 2011, 4(1): 5-1
- Weilan Yuan, Tao Huang, Jian Yu, Lingyao Zeng, Baofeng Lian, Qinwen He, Yixue Li, Xiaoyan Zhang*, Fengli Zhou*, **Lu Xie*.** Comparative analysis of viral protein interaction networks in Hepatitis B Virus and Hepatitis C Virus infected HCC. ***Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics***. Available online 14 June 2013 http://dx.doi.org/10.1016/j.bbapap.2013.06.002

Work case 5: regulatory network

- Kang Tu, Hui Yu, Youjia Hua, Yuan-Yuan Li, Lei Liu, Lu Xie*, Yixue Li*. Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms. **Nucleic Acids Research, 2009,** doi:10.1093/nar/gkp638.


A
microRNA
primary targets
Degraded targets
Repressed targets
Transcription Factors(TFs)
secondary targets

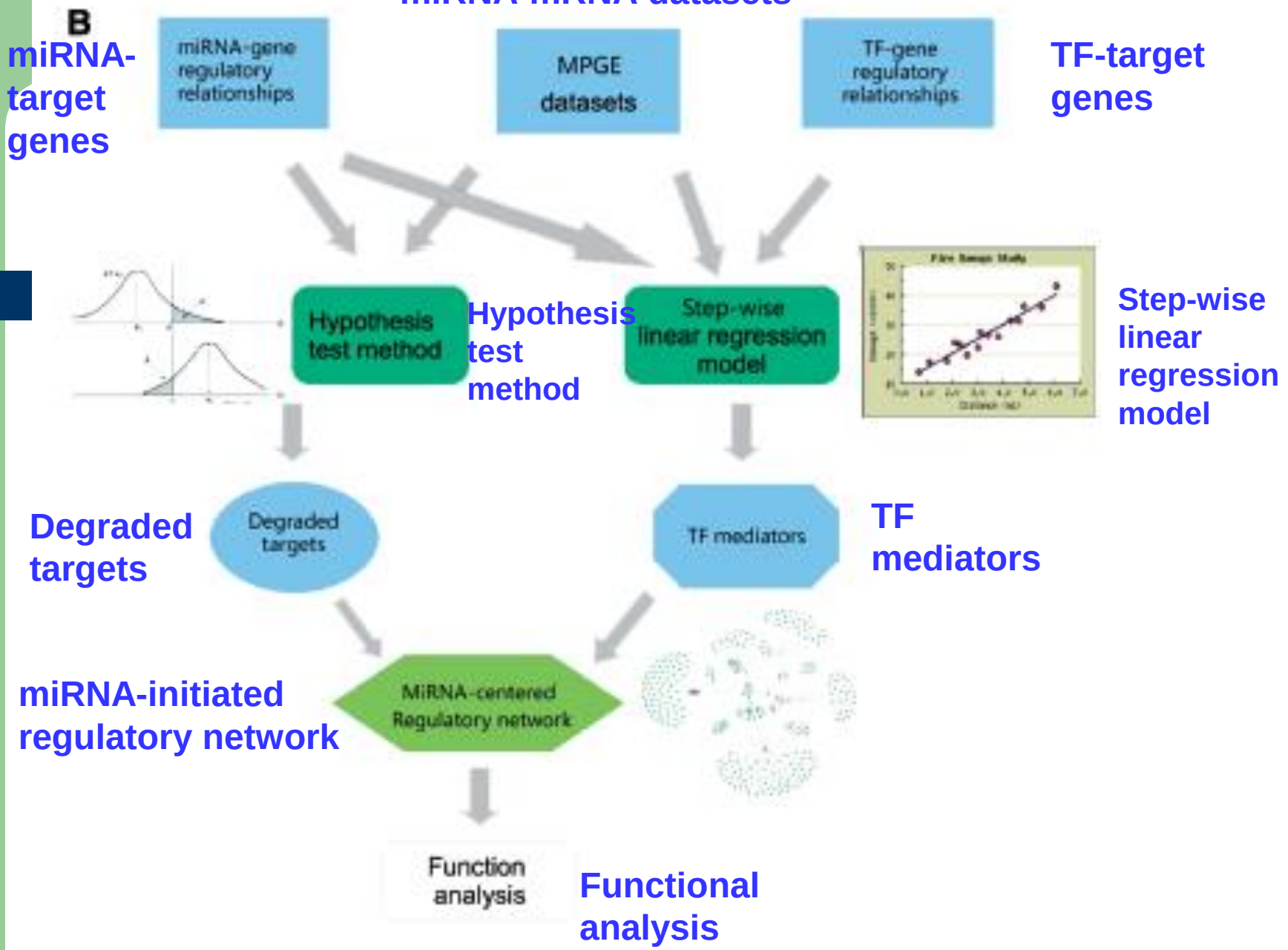
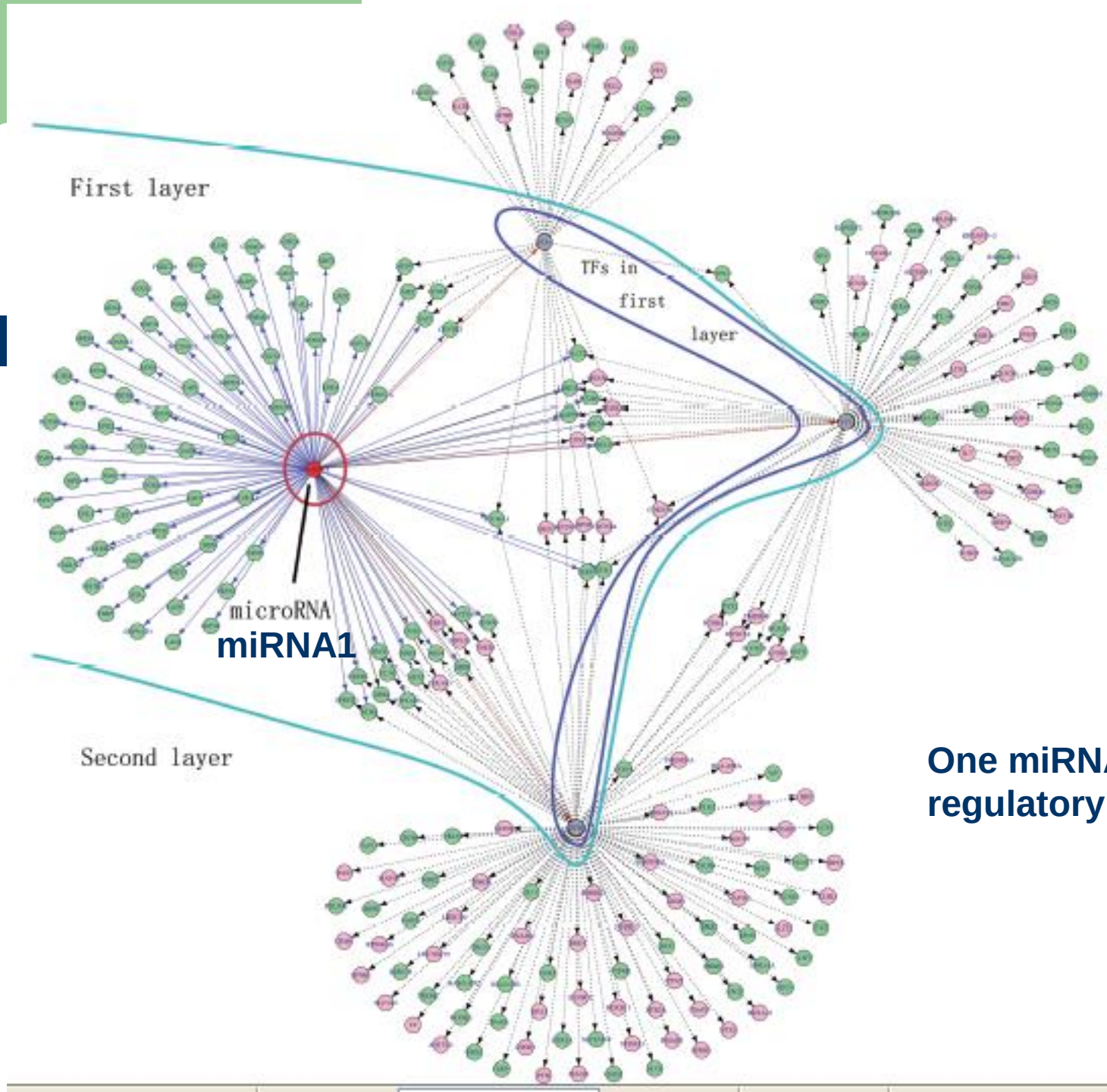

B
miRNA-mRNA datasets
miRNA-target genes
miRNA-gene regulatory relationships
MPGE datasets
TF-gene regulatory relationships
TF-target genes
Hypothesis test method
Hypothesis test method
Step-wise linear regression model
Step-wise linear regression model
Degraded targets
Degraded targets
TF mediators
TF mediators
miRNA-initiated regulatory network
MiRNA-centered Regulatory network
Function analysis
Functional analysis

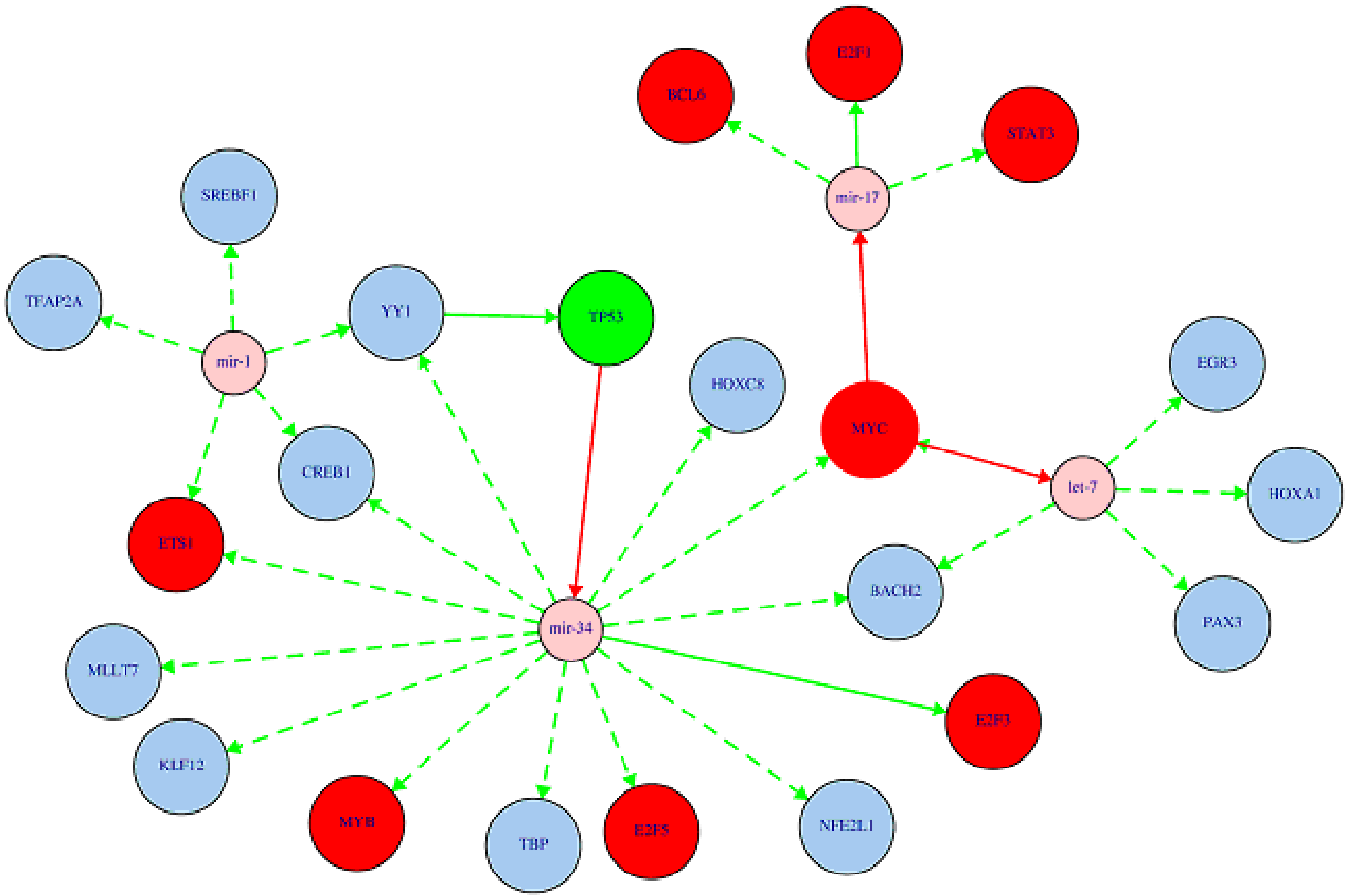
Table 1. Targets of miRNA-induced degradation and TF mediators of miRNA-triggered regulation, summarized for each of 53 MPGE datasets

Dataset group	miRNA	Cell line	Time point	K–S test *P*-value	De-graded targets	TF mediators	Shuffling *P*-value
GDS1858	miR-1	HeLa	12	1.2*e*–15	91	ETS1, CREB1, YY1	0
			24	4.7*e*–15	107	TFAP2A, CREB1, YY1, SREBF1	0
	miR-124		12	1.8*e*–12	109	GLI3*	0.04
			24	6.5*e*–22	132	MLLT7, NKX6.1	0.03
	miR-373		12	9.7*e*–05	17	MYCN	0.03
			24	4.0*e*–4	12	NFYA, TAL1, TFAP4*, KLF12	0
GDS2657	miR-124	HepG2					
						NR3C1*, BACH2, IRF1	
			24	1.3*e*–73	366	AHR*, RREB1	0
			32	6.2*e*–64	329	AHR*, SP1*, EGR1, RELA*, RREB1, NR3C1*, SP2	0
			72	2.2*e*–59	292	CREB1, SP1, ETS1, MLLT7, SP2	0
			120	1.1*e*–19	144	AHR*, SP1*, MLLT7	0
GSE6474	let-7a3	A549	Not known	1.1*e*–2	1	PAX3, HOXA1, BACH2, EGR3, MYC	0.02
GSE6838	let-7c	HCT116 Dicer−/− #2	24	1.5*e*–54	211	MYC	0.05
	miR-103		10	5.7*e*–08	82	MEF2A	0.08
			24	4.4*e*–22	77	NFATC3, MEF2A	0.04
	miR-106		6	3.4*e*–55	234	–	–
			10	8.0*e*–34	158	FOXJ2*	0.05
			24	6.0*e*–51	246	EGR2	0.07
	miR-107		10	3.2*e*–06	1	FOXJ2*	0.04
			24	7.8*e*–15	0	–	–
	miR-15a		6	1.3*e*–13	0	HOXC8, TBP, POU3F2, FOXC1	0.01
			10	8.3*e*–51	224	–	–
			14	7.2*e*–25	0	POU3F2	0.02
			24	3.9*e*–36	78	WT1	0.06
	miR-15b		10	2.0*e*–26	0	FOXC1	0.06
			24	4.6*e*–37	69	FOXC1	0.07
	miR-16		6	2.3*e*–23	85	–	–
			10	2.6*e*–29	181	–	–
			14	5.9*e*–40	0	BACH2	0.11
			24	1.1*e*–20	46	–	–

53 miRNA-perturbed mRNA expression profiling datasets

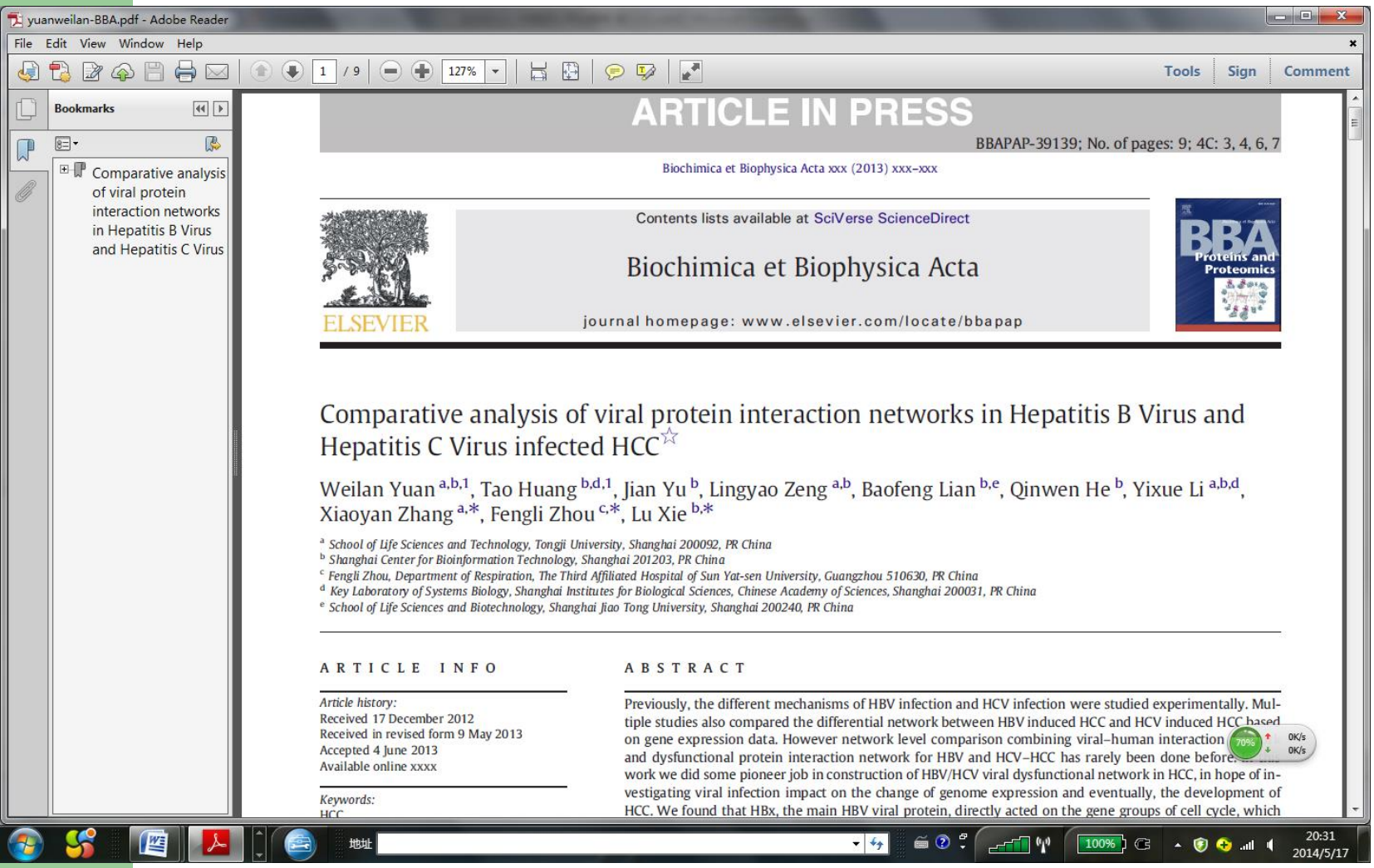


One miRNA initiated regulatory network

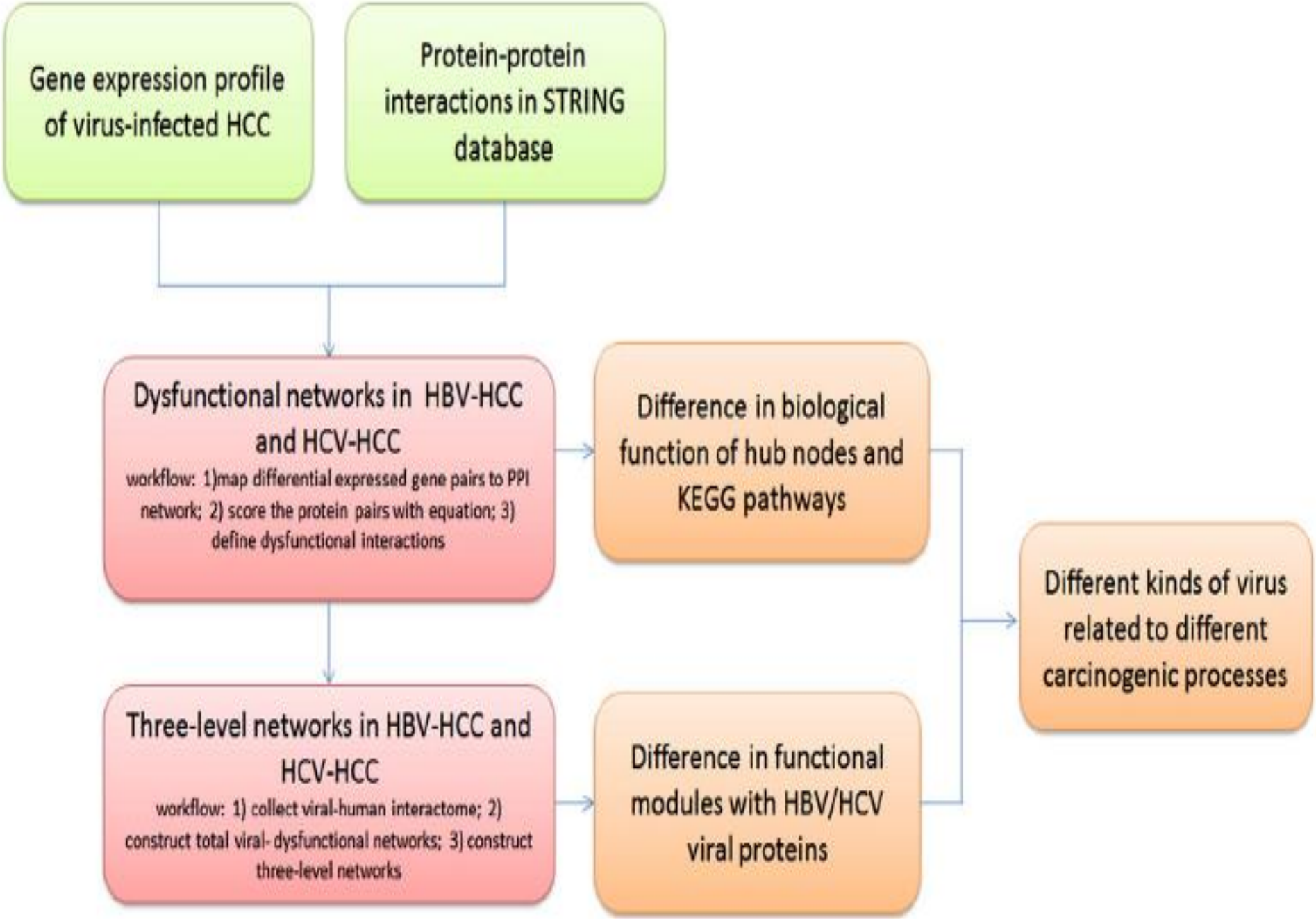


the regulatory cascades triggered by the miR-34, let-7, miR-17 and miR-1 families may be linked together to form a tumor-related regulatory network.

Work case 6: protein interaction network



This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications.


Gene expression profile of virus-infected HCC
Protein-protein interactions in STRING database
Dysfunctional networks in HBV-HCC and HCV-HCC
workflow: 1)map differential expressed gene pairs to PPI network; 2) score the protein pairs with equation; 3) define dysfunctional interactions
Difference in biological function of hub nodes and KEGG pathways
Three-level networks in HBV-HCC and HCV-HCC
workflow: 1) collect viral-human interactome; 2) construct total viral- dysfunctional networks; 3) construct three-level networks
Difference in functional modules with HBV/HCV viral proteins
Different kinds of virus related to different carcinogenic processes

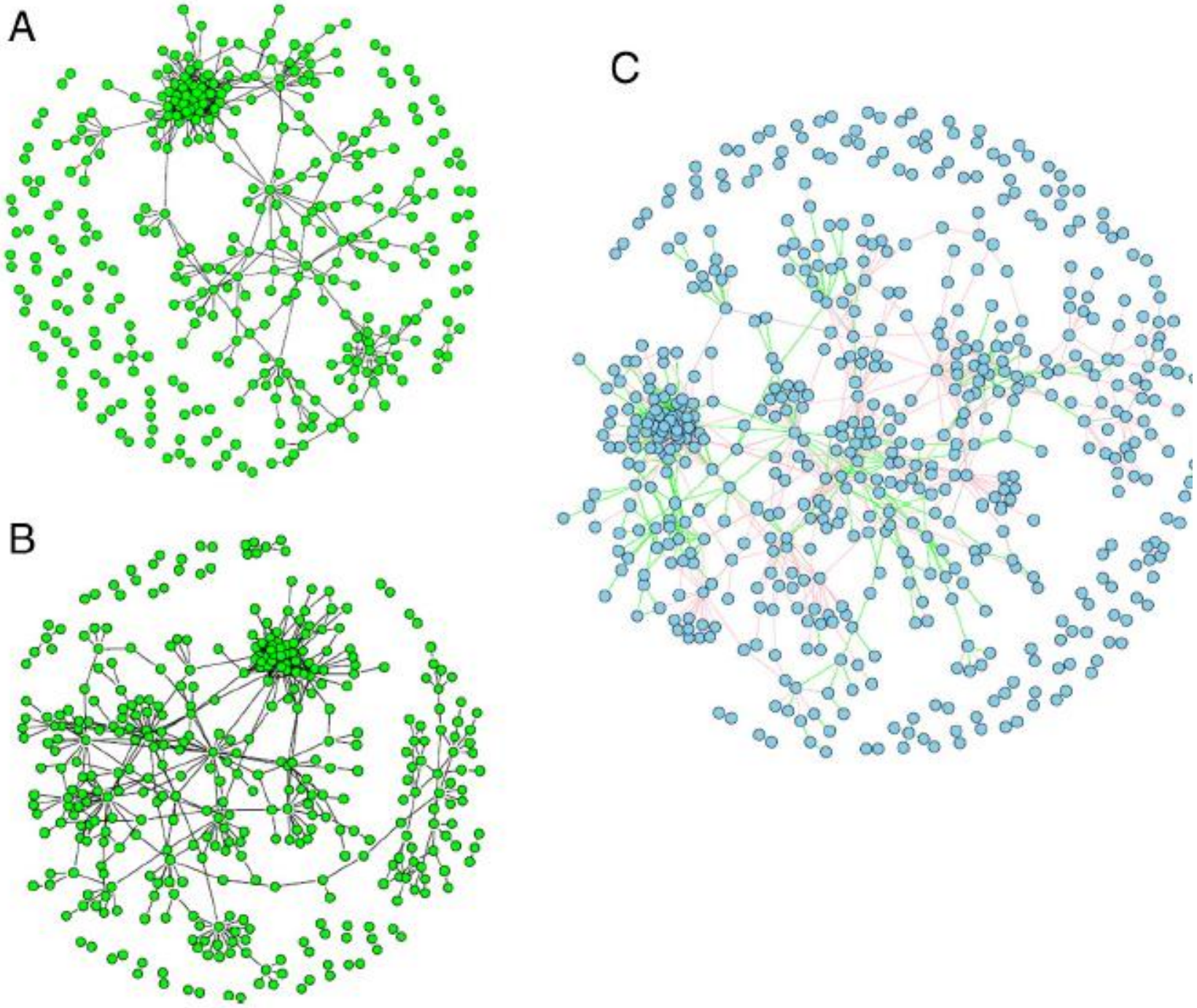
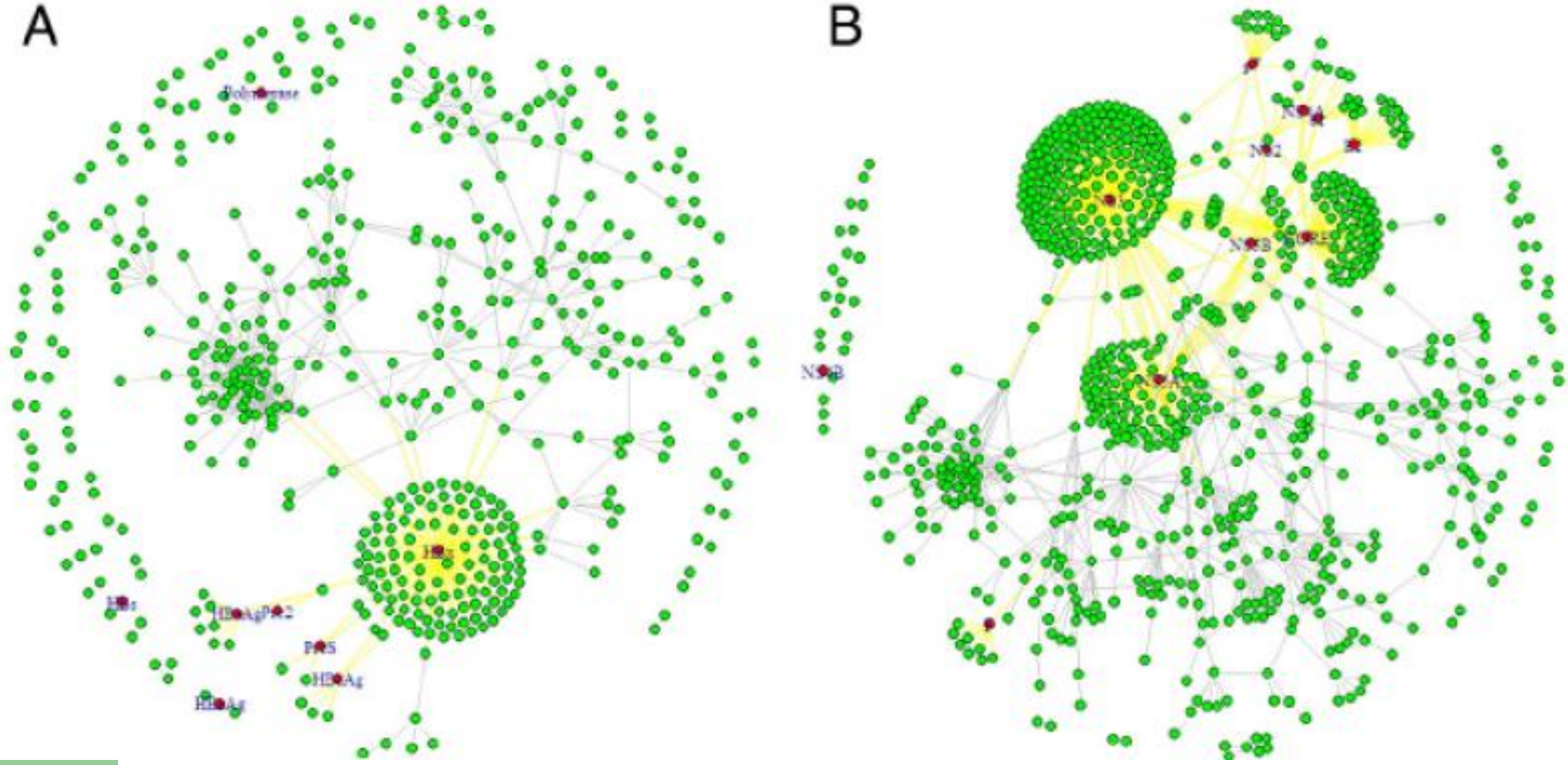
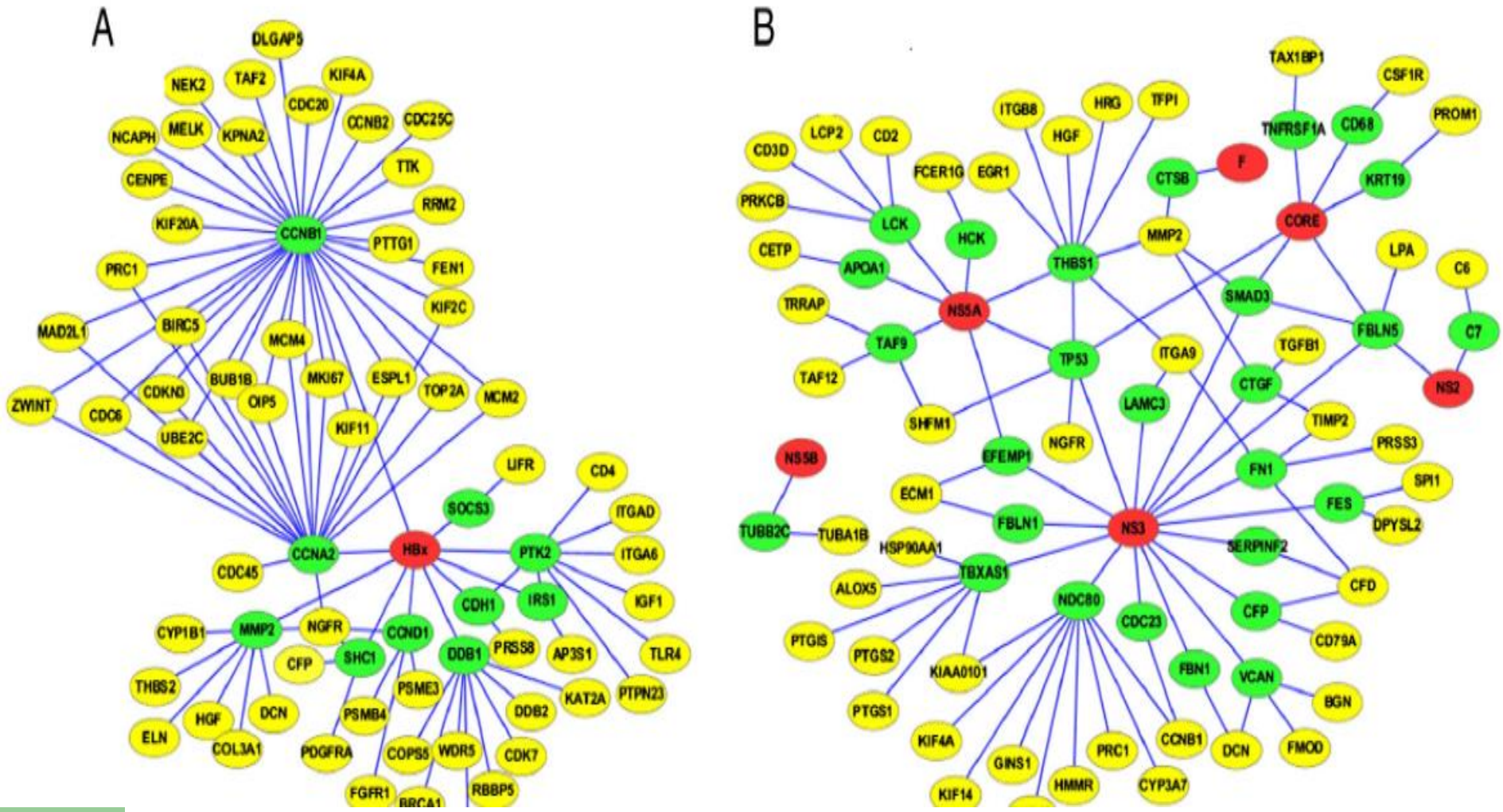

A
B
C

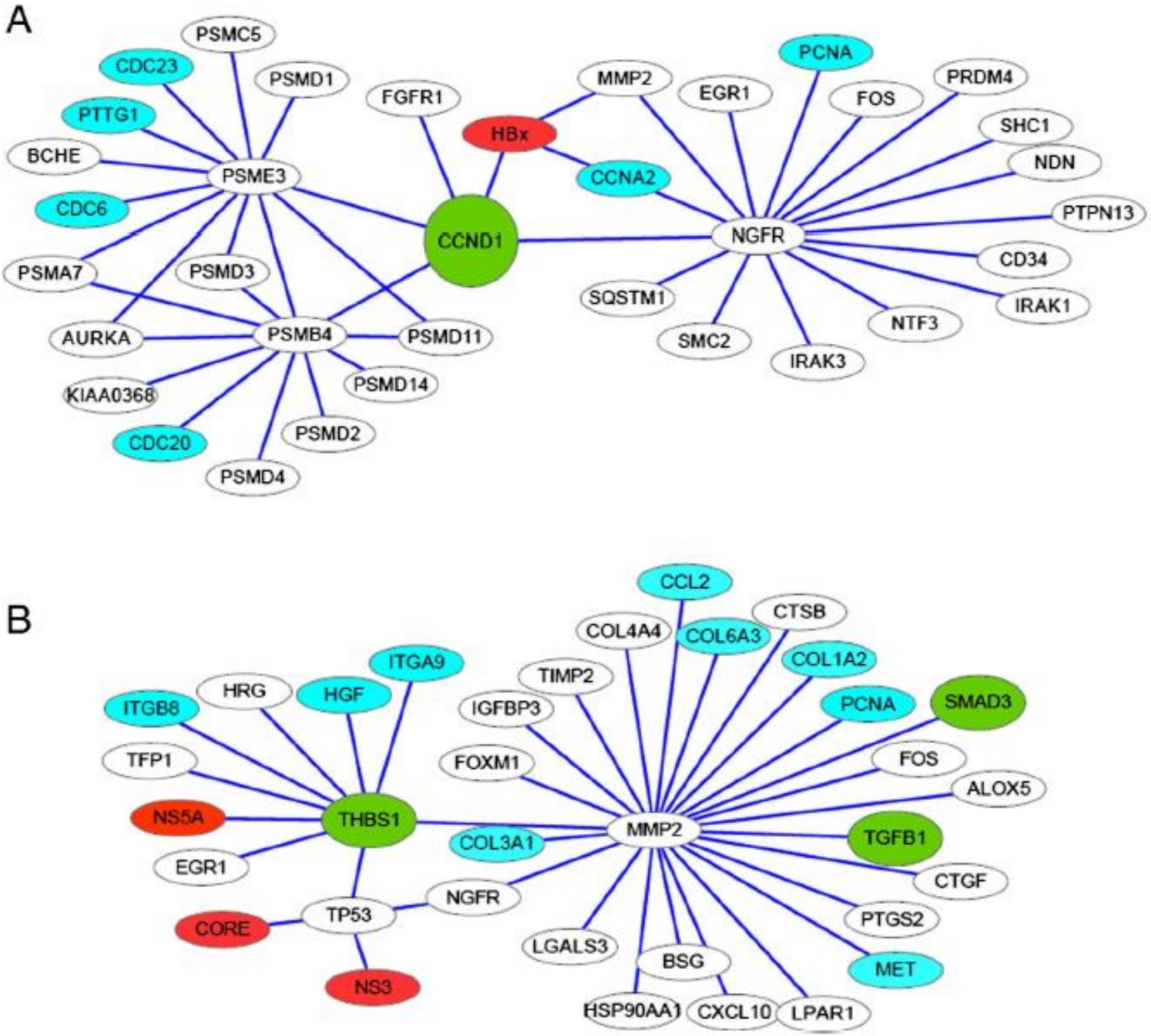
Table 2

Characteristics of the dysfunctional network of HBV/HCV-infected HCC.

	Nodes	Edges	The KEGG enrichment of dysfunctional nodes	Hub nodes
Dysfunctional network of HBV-infected HCC	359	625	Cell cycle, DNA replication	MKI67, BIRC5, CCNB1, CDKN3, MELK, PRC1
Dysfunctional network of HCV-infected HCC	391	663	Focal adhesion, ECM–receptor interaction, chemokine signaling pathway, complement and coagulation cascades, arachidonic acid metabolism	TOP2A, KIF20A, CCNB1, MMP2, CCL19, KIF4A, ITGA9


A
B


A
DLGAP5
NEK2
TAF2
KIF4A
CDC20
NCAPH
MELK
KPNA2
CCNB2
CDC25C
CENPE
TTK
KIF20A
CCNB1
RRM2
PTTG1
PRC1
FEN1
KIF2C
MAD2L1
BIRC5
MCM4
ZWINT
CDKN3
BUB1B
MKI67
ESPL1
TOP2A
MCM2
CDC6
OIP5
UBE2C
KIF11
LIFR
CD4
SOCS3
ITGAD
CCNA2
HBx
PTK2
ITGA6
CDC45
CDH1
IRS1
IGF1
CYP1B1
MMP2
NGFR
CCND1
DDB1
PRSS8
AP3S1
TLR4
CFP
SHC1
PSME3
KAT2A
PTPN23
THBS2
PSMB4
DDB2
HGF
DCN
ELN
COL3A1
PDGFRA
COPS5
WDR5
CDK7
FGFR1
RBBP5
B
TAX1BP1
CSF1R
ITGB8
HRG
TFPI
TNFRSF1A
CD68
PROM1
LCP2
CD2
HGF
CD3D
F
CTSB
KRT19
FCER1G
EGR1
PRKCB
LCK
CORE
HCK
MMP2
CETP
THBS1
LPA
C6
APOA1
SMAD3
TRRAP
NS5A
FBLN5
C7
TAF9
TP53
ITGA9
TGFB1
NS2
TAF12
CTGF
LAMC3
TIMP2
SHFM1
NGFR
PRSS3
NS5B
EFEMP1
FN1
SPI1
ECM1
FES
TUBB2C
TUBA1B
FBLN1
NS3
DPYSL2
HSP90AA1
SERPINF2
TBXAS1
CFD
ALOX5
NDC80
CFP
CDC23
PTGIS
CD79A
PTGS2
KIAA0101
FBN1
VCAN
PTGS1
BGN
KIF4A
CCNB1
DCN
FMOD
GINS1
PRC1
KIF14
HMMR
CYP3A7


A
PSMC5
CDC23
PSMD1
PTTG1
BCHE
PSME3
CDC6
PSMA7
PSMD3
AURKA
PSMB4
PSMD11
KIAA0368
PSMD14
CDC20
PSMD2
PSMD4
FGFR1
HBx
CCND1
MMP2
EGR1
PCNA
FOS
PRDM4
SHC1
NDN
CCNA2
PTPN13
NGFR
CD34
IRAK1
SQSTM1
NTF3
SMC2
IRAK3
B
ITGA9
HRG
HGF
ITGB8
TFP1
NS5A
THBS1
EGR1
CORE
TP53
NS3
NGFR
CCL2
CTSB
COL4A4
COL6A3
COL1A2
TIMP2
IGFBP3
PCNA
SMAD3
FOXM1
FOS
ALOX5
COL3A1
MMP2
TGFB1
CTGF
PTGS2
LGALS3
BSG
MET
HSP90AA1
CXCL10
LPAR1

Thank you for your attention