Системная биология: перспективы и проблемы

Илья Серебрийский
Fox Chase Cancer Center

Системная биология в моде!..

SBI The Systems Biology Institute

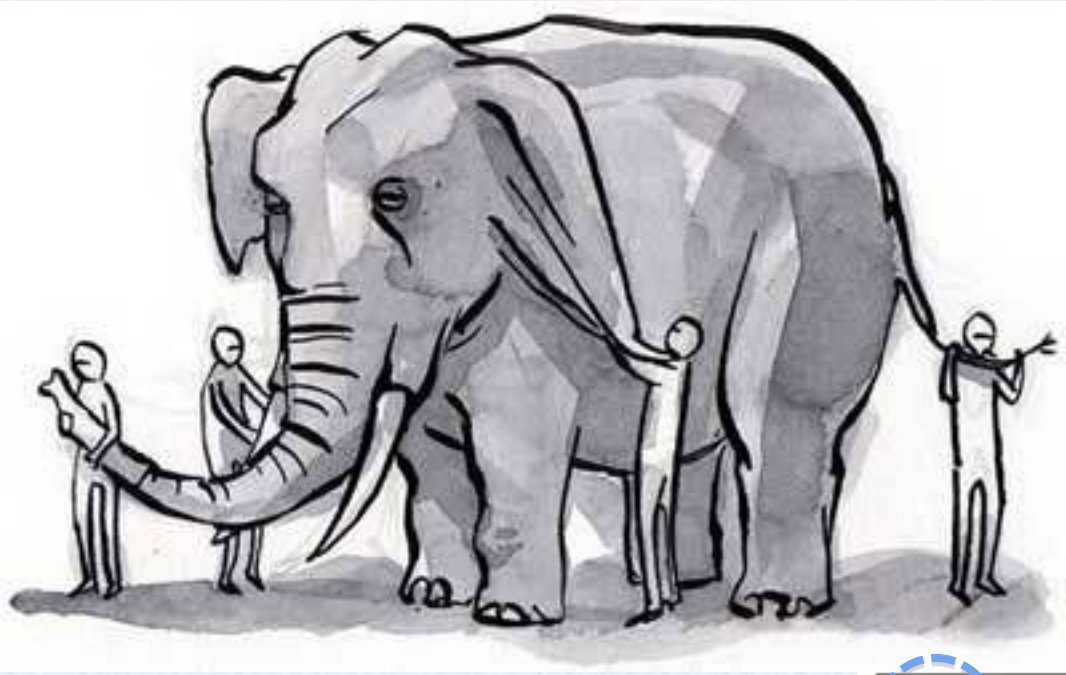
5 лет: ~50публикации

~ 50 публикаций в год

ООО «Институт системной биологии», г. Новосибирск, Россия

3 сотрудника

... но системная биология была всегда!



В.И. Вернадский

M. Á. Medina

«целое больше, чем сумма его частей»

все части и процессы целого взаимовлияют и взаимообуславливают друг друга

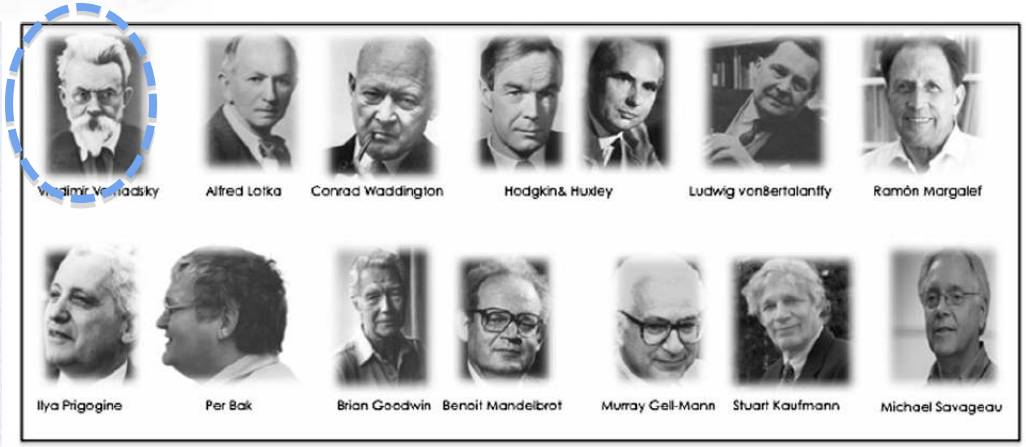


Fig. 1 Some relevant scientists with outstanding systemic contributions to the understanding of complex biological problems before the emergence of modern systems biology

некоторые

Системная биология: перспективы и проблемы

Мои основные задачи на эти лекции:

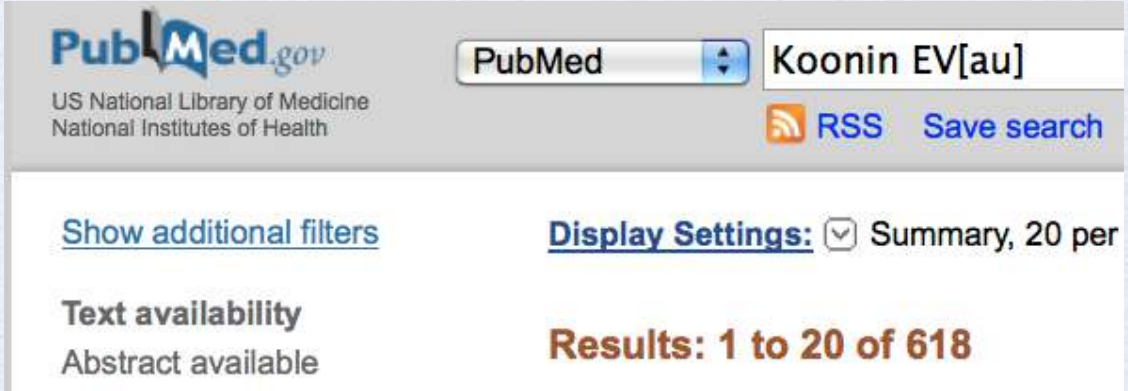
- постараться дать общее представление о системных свойствах биологических объектов
- вкратце описать основные составляющие системной биологии
- выборочно осветить некоторые достижения
- привлечь ваше внимание к общедоступным ресурсам
- предостеречь об основных проблемах в системной биологии и рассказать попытках их разрешения

А как насчёт определения «системной биологии»?

NCBI Evolutionary Genomics Research Group

Евгений Кунин:

Вообще говоря, что такое системная биология, не знает никто. Оглядываясь назад, кажется, что такого термина, наверное, лучше было бы вообще не придумывать.
... Но его придумали, и он уже укоренился.



Ну хотя бы определение «системы»?

WIKIPEDIA
Die freie Enzyklopädie

Der Begriff **System** (von griechisch σύστημα, „das Gebilde, Zusammengestellte, Verbundene“) bezeichnet allgemein eine Gesamtheit von Elementen, die so aufeinander bezogen bzw. miteinander verbunden sind und in einer Weise interagieren, dass sie als eine aufgaben-, sinn- oder zweckgebundene Einheit angesehen werden können.

WIKIPÉDIA
L'encyclopédie libre

Un **système** est un ensemble d'éléments interagissant entre eux selon certains principes ou règles.
Un système est déterminé par :
la nature de ses éléments constitutifs ;
les interactions entre ces derniers ;
sa frontière, c'est-à-dire le critère … déterminant si une entité appartient au système ou fait au contraire partie de son environnement).

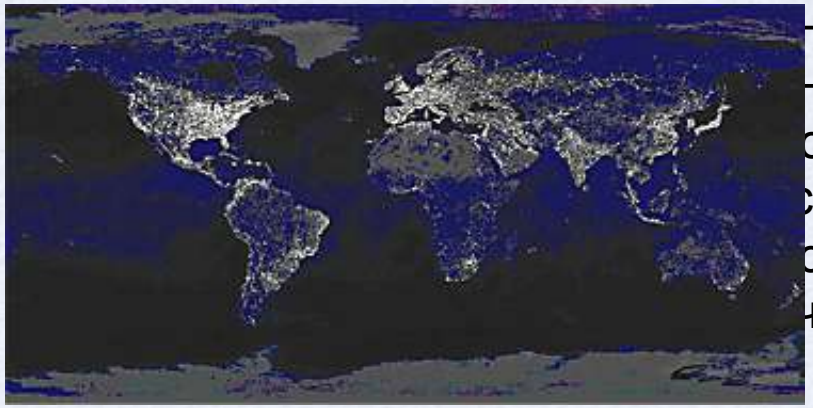
WIKIPEDIA
The Free Encyclopedia

•A **system** is a set of interacting or interdependent components forming an integrated whole or a set of elements (often called *'components'*) and relationships which are different from relationships of the set or its elements to other elements or sets.

Ну хотя бы определение «системы»?

Свойства, общие для всех систем

•Целостность — система есть абстрактная сущность, обладающая целостностью и определенная в своих границах.
Целостность системы подразумевает, что в некотором существенном аспекте «сила» или «ценность» связей элементов *внутри системы* выше, чем сила или ценность связей элементов системы с элементами *внешних систем* или *среды*.



- появление у системы свойств, не нципиальная несводимость оставляющих её компонентов стемы превосходят сумму стей; общая производительность чше, чем у простой суммы

•Иерархичность — каждый компонент системы может рассматриваться как система; сама система также может рассматриваться как элемент некоторой надсистемы (суперсистемы).

Свойства биологических систем и основные задачи их всестороннего исследования (1)

Некоторые наиболее важные свойства:

- *открытые системы*, находятся в постоянном взаимообмене информацией и материей с окружающей средой
- такой обмен приводит к постоянным изменениям – это *динамические системы*
- *имеют иерархическую структуру* и многоуровневые взаимодействия и взаимозависимости между компонентами

Свойства биологических систем и основные задачи их всестороннего исследования (2)

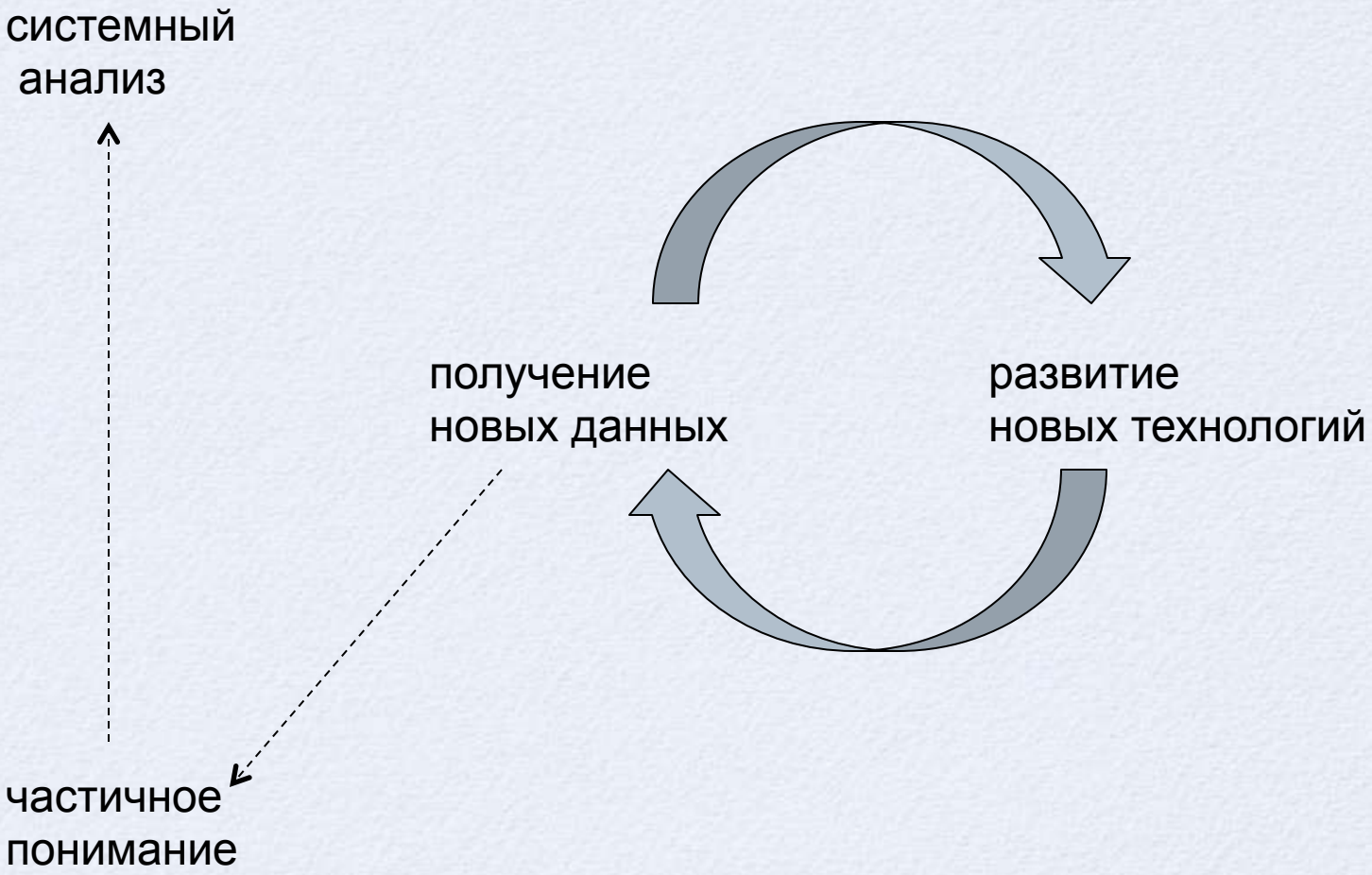
Важнейшие задачи:

- изучение структуры системы
- понимание динамики системы
- определение правил саморегуляции системы

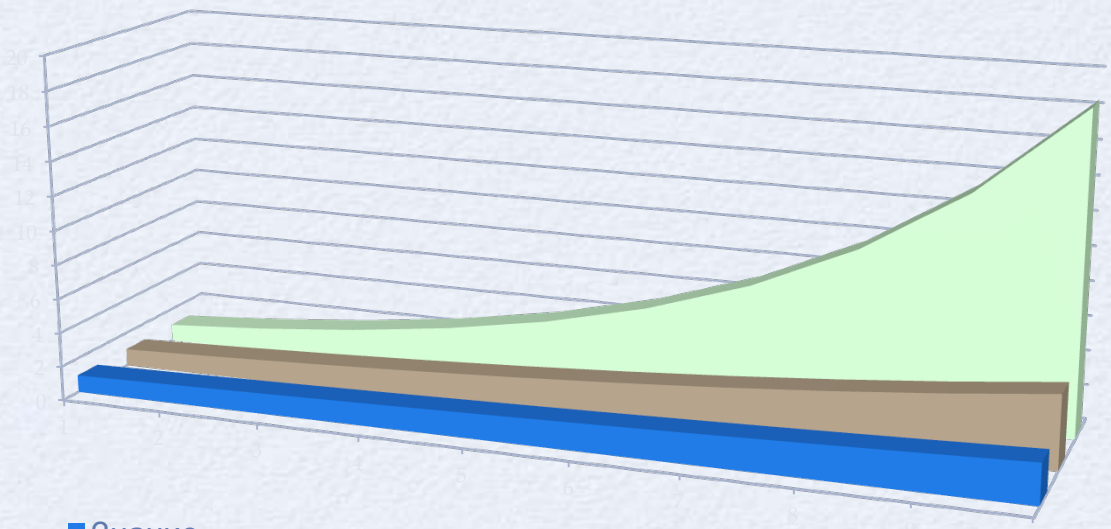
Способы решения:

- планомерное и систематическое воздействие
- измерение ответа системы на разных уровнях
- интеграция данных
- построение ***моделей***, позволяющих предсказывать поведение системы

Этапы становления системной биологии (1)



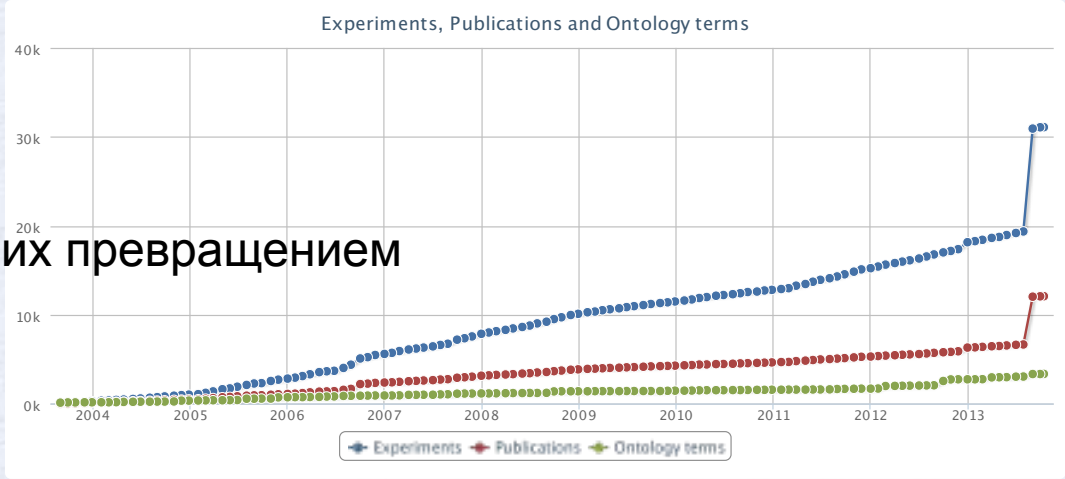
Девятый вал данных...



Знание

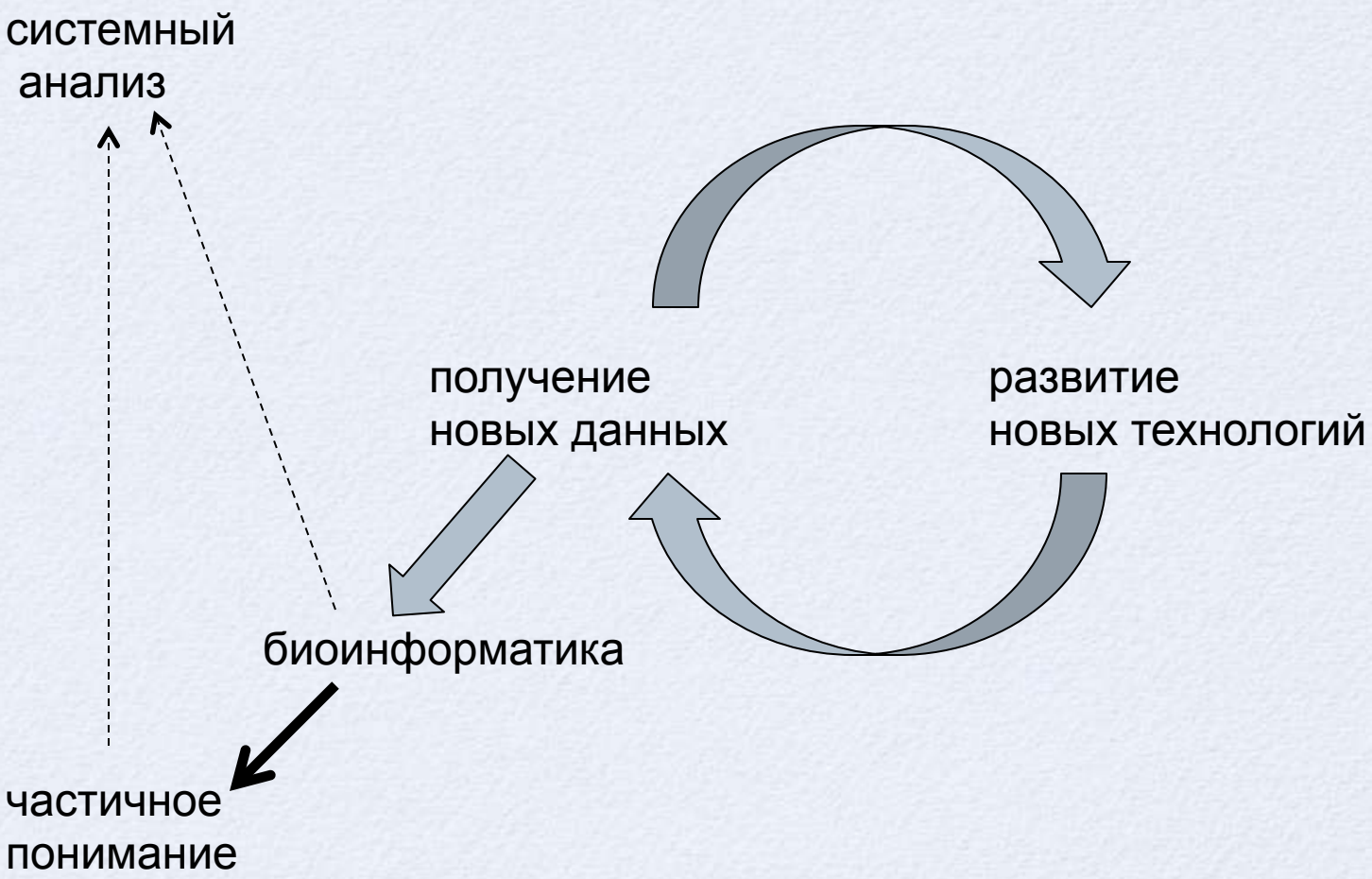
Информация

Данные

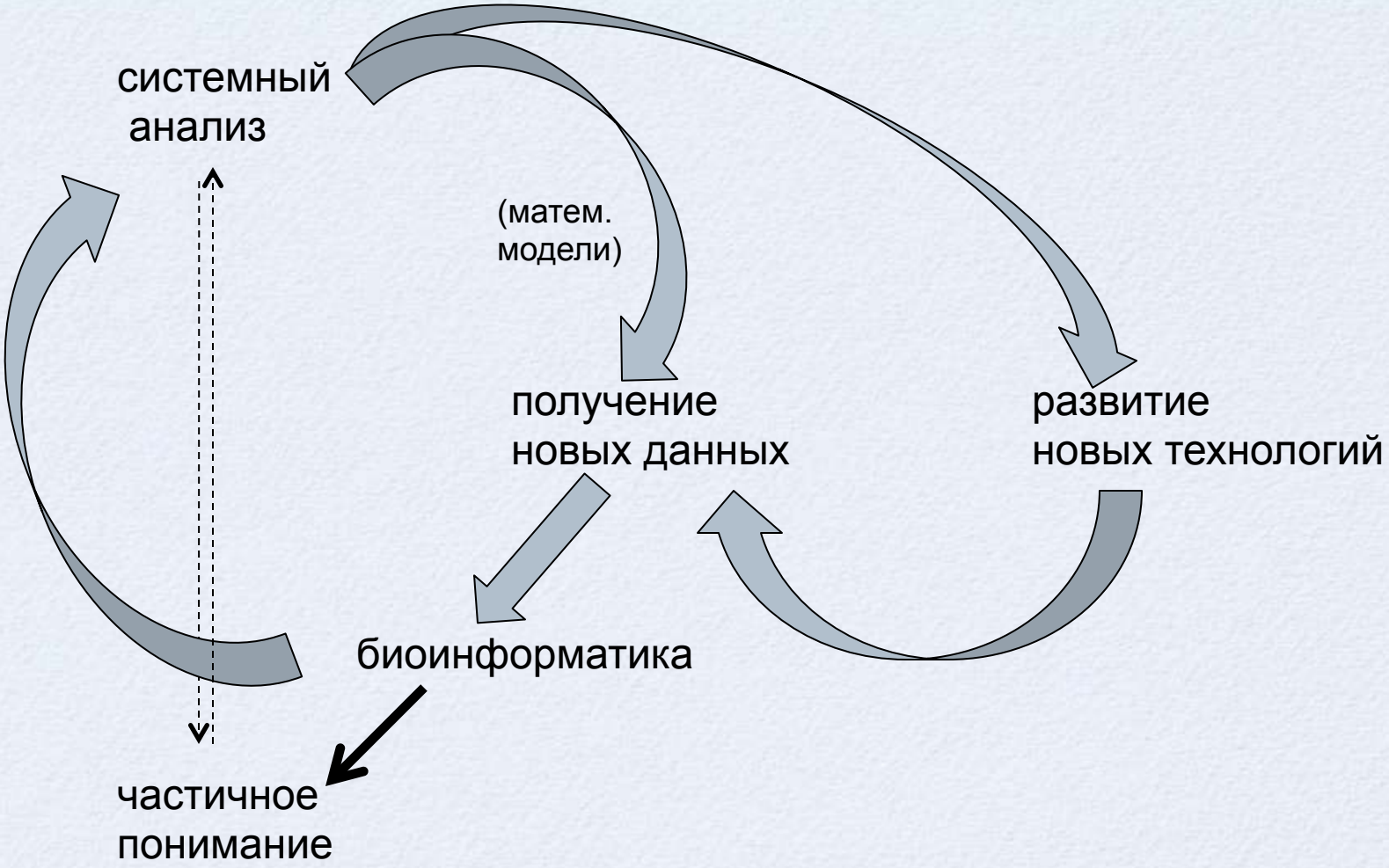


Разрыв между наработкой данных и их превращением в знания продолжает увеличиваться

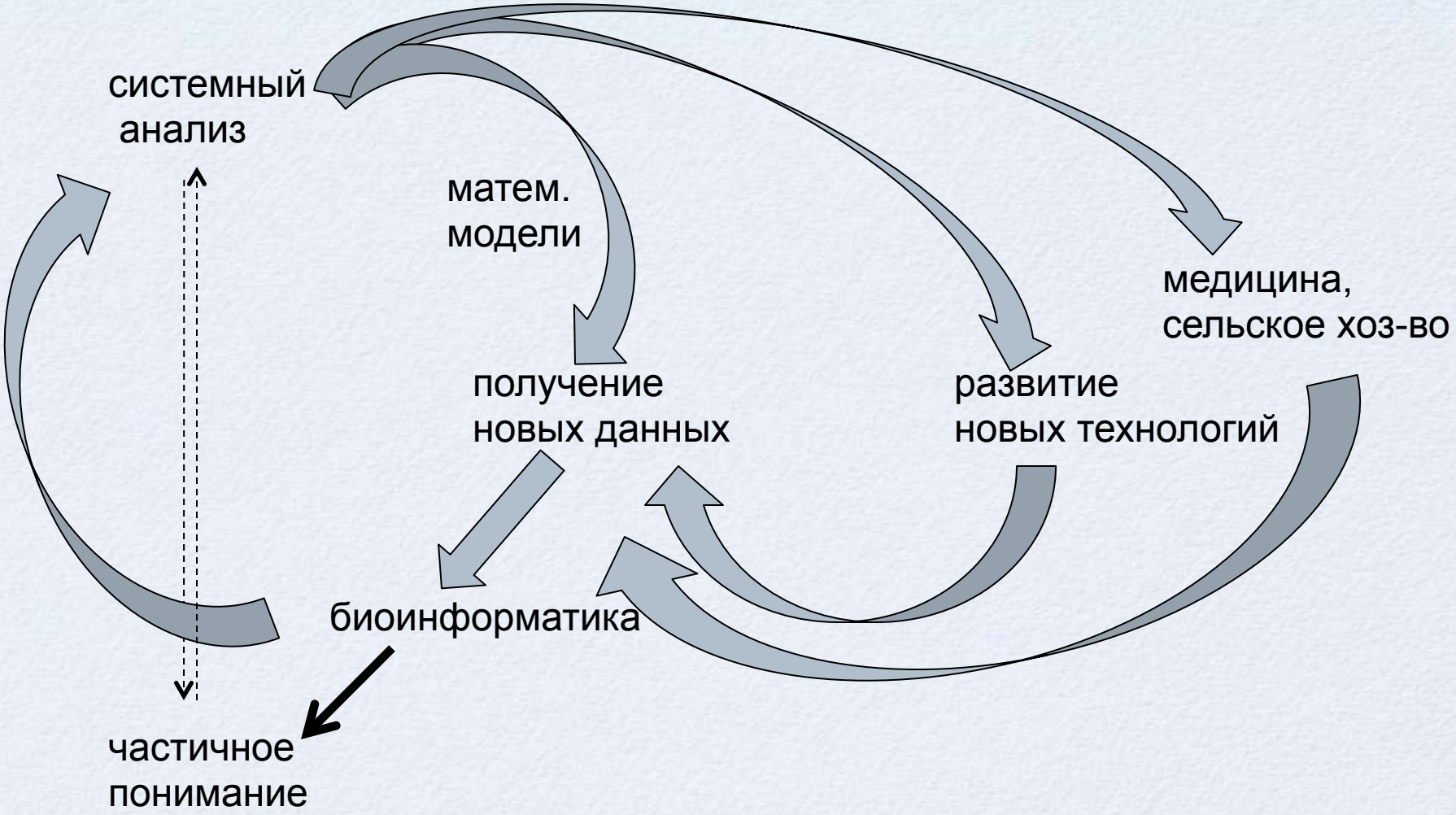
Этапы становления системной биологии (2)



Этапы становления системной биологии (3)



Этапы становления системной биологии (4)



Этапы становления системной биологии (2)

Евгений Кунин:

За словами "системная биология" обычно стоит идея о том, что нужно изучать организм, клетку или клетку с вирусом как некую систему в целом. Но как именно это сделать, никто на самом деле не понимает. Таким образом, реально под этот зонтик попадают широкомасштабные (high throughput) исследования чего-либо и комбинации таких исследований...

Пока существующая Системная Биология подразумевает не столько исследования каких-то биологических систем, сколько подробное описание их частей. Хотя, конечно, на основании полученных таким образом данных уже можно строить какие-то модели, но это другой вопрос.

системный анализ

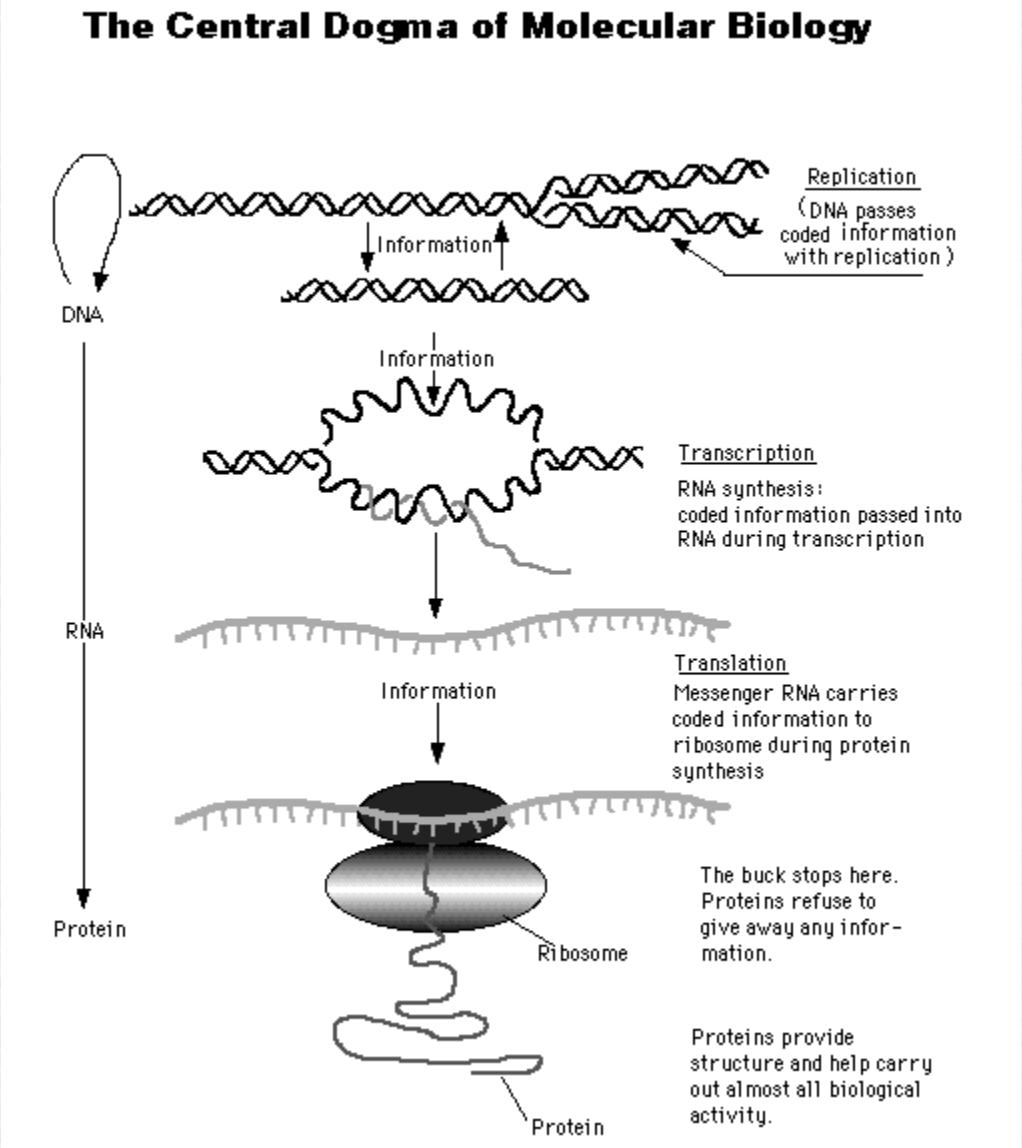
получение новых данных

биоинформатика

частичное понимание

“-omics” sciences

Основные источники информации в системной биологии



Основные источники информации в системной биологии

- **Геномика** – основная задача заключается в определении и общем изучении нуклеотидных последовательностей РНК и ДНК на уровне геномов
- **Транскриптомика** – идентификация всех РНК в клетке, (кодирующих или некодирующих), определение количества каждой индивидуальной РНК, определение закономерностей экспрессии
- **Протеомика** – определение всех белковых компонентов на уровне клетки, их функционального состояния и динамики
- **Метаболомика** – описывает биохимическое состояние клетки, определяя и измеряя концентрации так называемых малых молекул, метаболитов

Эпигеномика, гликомика, микробиомика, резистомика...

только ленивый не прибавляет «омика» к названию своей науки...

Badomics generator

Your badomics word is:

Ciliomics

Smith et al (2012) Sequencing the Ciliome reveals insights into evolution and the environment. Nat Biotechnol 30(7):213-222

DISCLAIMER: This webpage is intended to be funny. It is not meant to offend. It is not aimed at any inidvidual, group or institution. If this page causes offence in any way then please e-mail me ASAP at mick.watson@roslin.ed.ac.uk and I will take it down.

Our Sponsors

ROSLIN
ROSLIN FOUNDATION
BBSRC
Technology Strategy Board
Driving Innovation
EUROPE & SCOTLAND

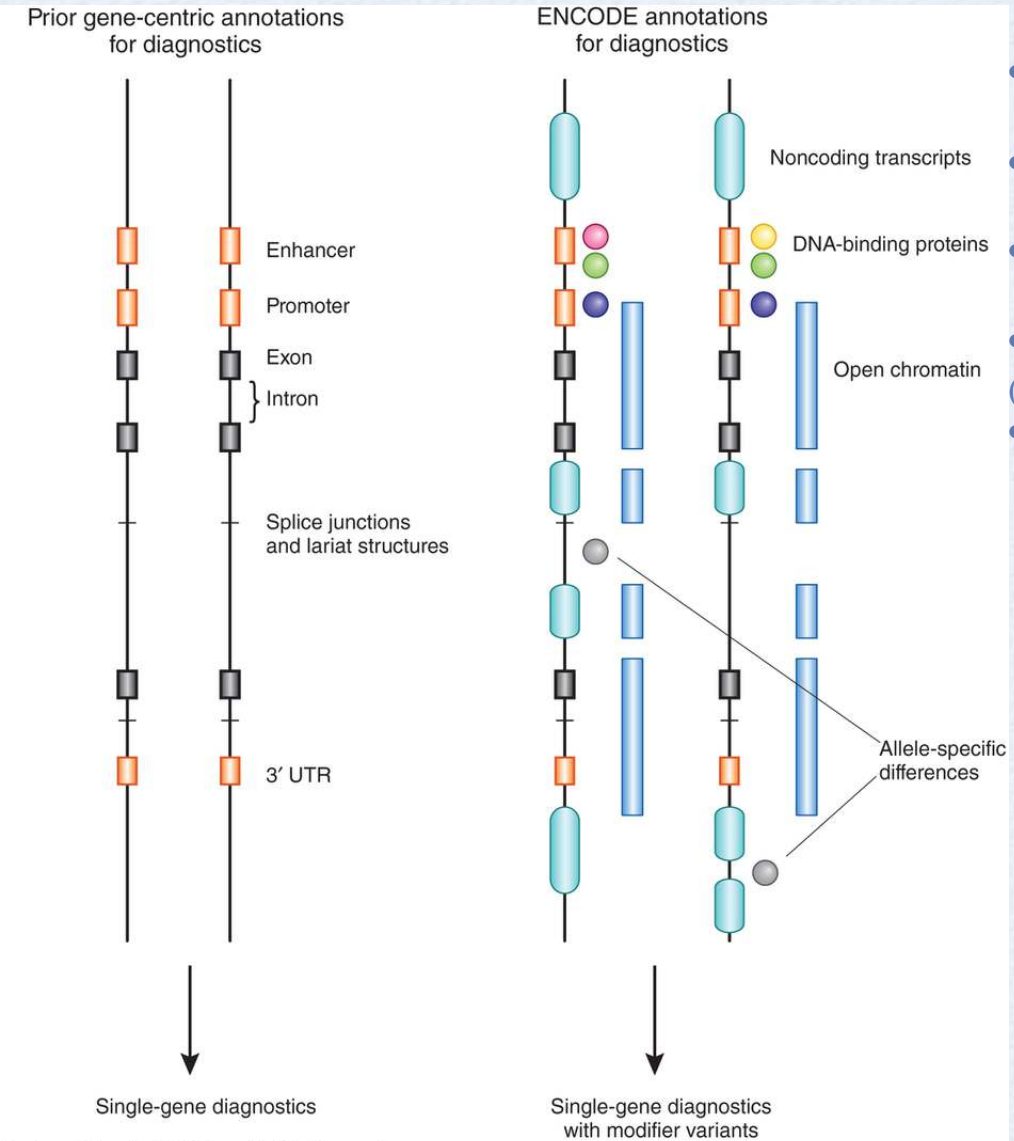
Геномика

nature International weekly journal of science
Nature 489, 57–74 (06 September 2012)

An integrated encyclopedia of DNA elements in the human genome

- The Encyclopedia of DNA Elements (ENCODE) project aims to delineate all functional elements encoded in the human genome
- Operationally, a functional element is defined as a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA) or displays a reproducible biochemical signature (protein binding, or a specific chromatin structure).

An integrated encyclopedia of DNA elements in the human genome

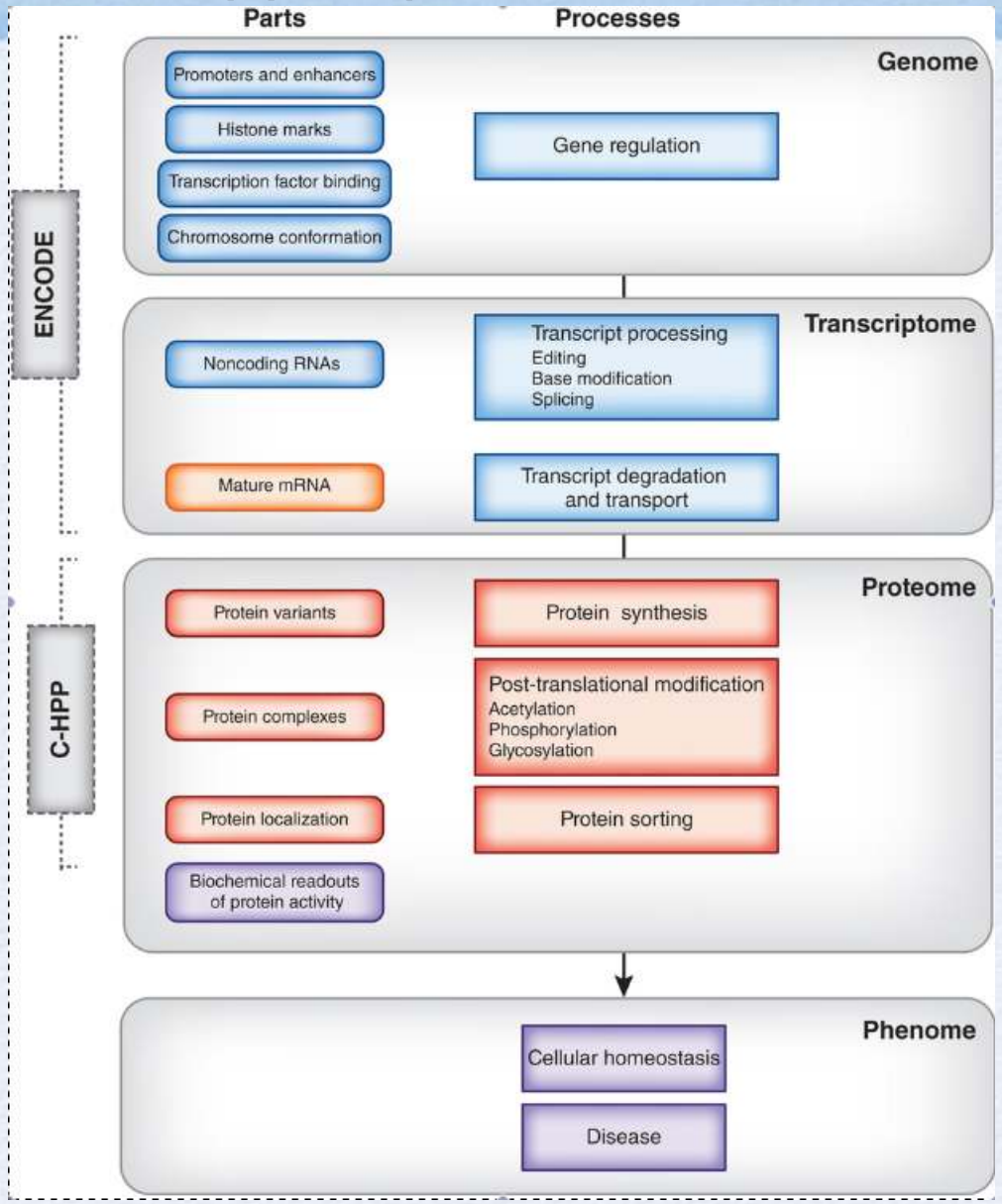


The elements mapped (and approaches used)

- RNA transcribed regions (RNA-seq, CAGE, RNA-PET)
- protein-coding regions (mass spectrometry)
- transcription-factor-binding sites (ChIP-seq and DNase-seq)
- chromatin structure (DNase-seq, FAIRE-seq, histone ChIP-seq and MNase-seq),
- DNA methylation sites (RRBS assay)

The vast majority (80.4%) of the human genome participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type. Much of the genome lies close to a regulatory event: 95% of the genome lies within 8 kilobases (kb) of a DNA–protein interaction (as assayed by bound ChIP-seq motifs or DNase I footprints), and 99% is within 1.7 kb of at least one of the biochemical events measured by ENCODE.

Геномика – взаимодействие с другими дисциплинами системной биологии



Эпигеномика, микробиомика, резистомика...

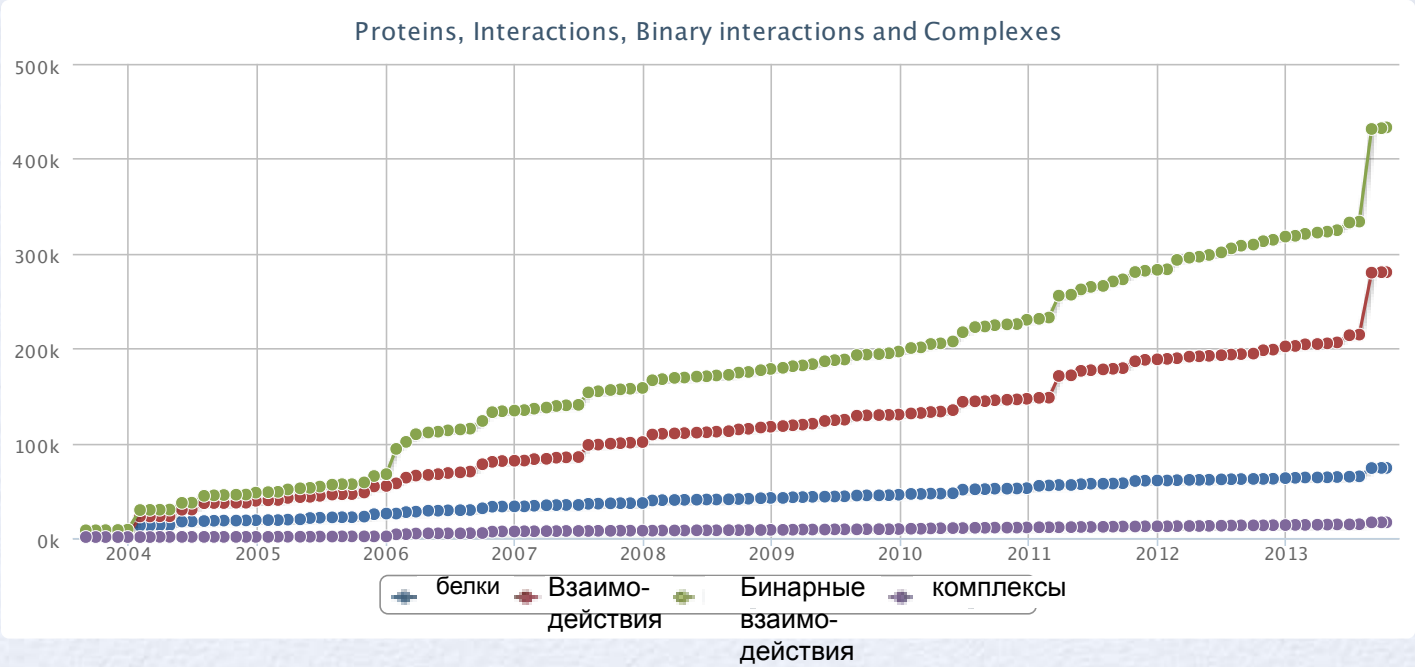
1000 Genomes
A Deep Catalog of Human Genetic Variation

华大基因 BGI

Million Plant & Animal Genomes Project
Million Human Genomes Project
Million Micro-ecosystem Genomes Project

Интерактомика – важнейший компонент системной биологии

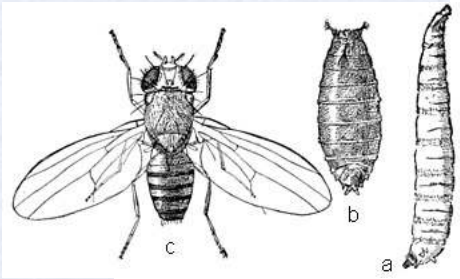
- **Интерактомика** – основная задача заключается в определении взаимодействий молекул (белков, ДНК, РНК) на уровне клетки, а также в изучении динамики этих взаимодействий
- позволяет получить первое приближение к модели – карты сетей взаимодействия

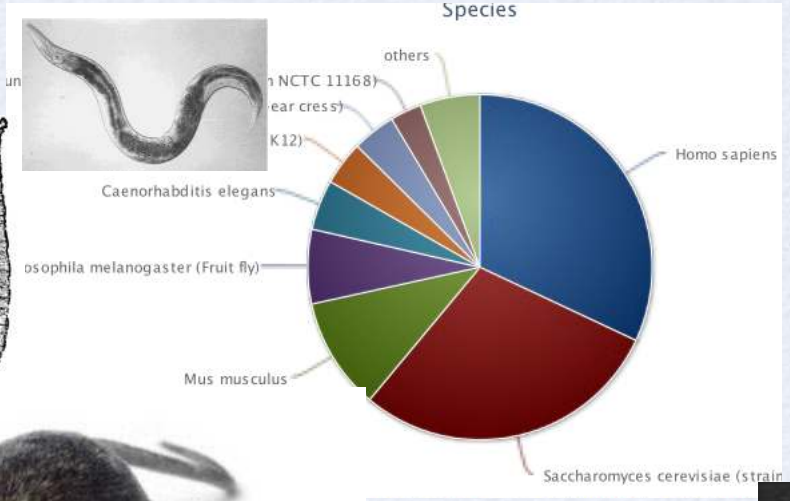


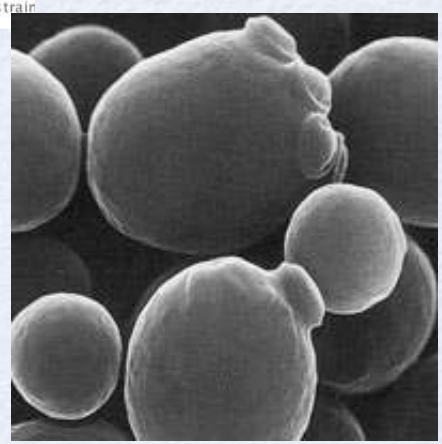
Общее количество белок-белковых взаимодействий в базах данных растёт намного быстрее общего количества известных белков

Интерактомика – важнейший компонент системной биологии

Большая часть белок-белковых взаимодействий в базах данных получена на основе исследований модельных организмов

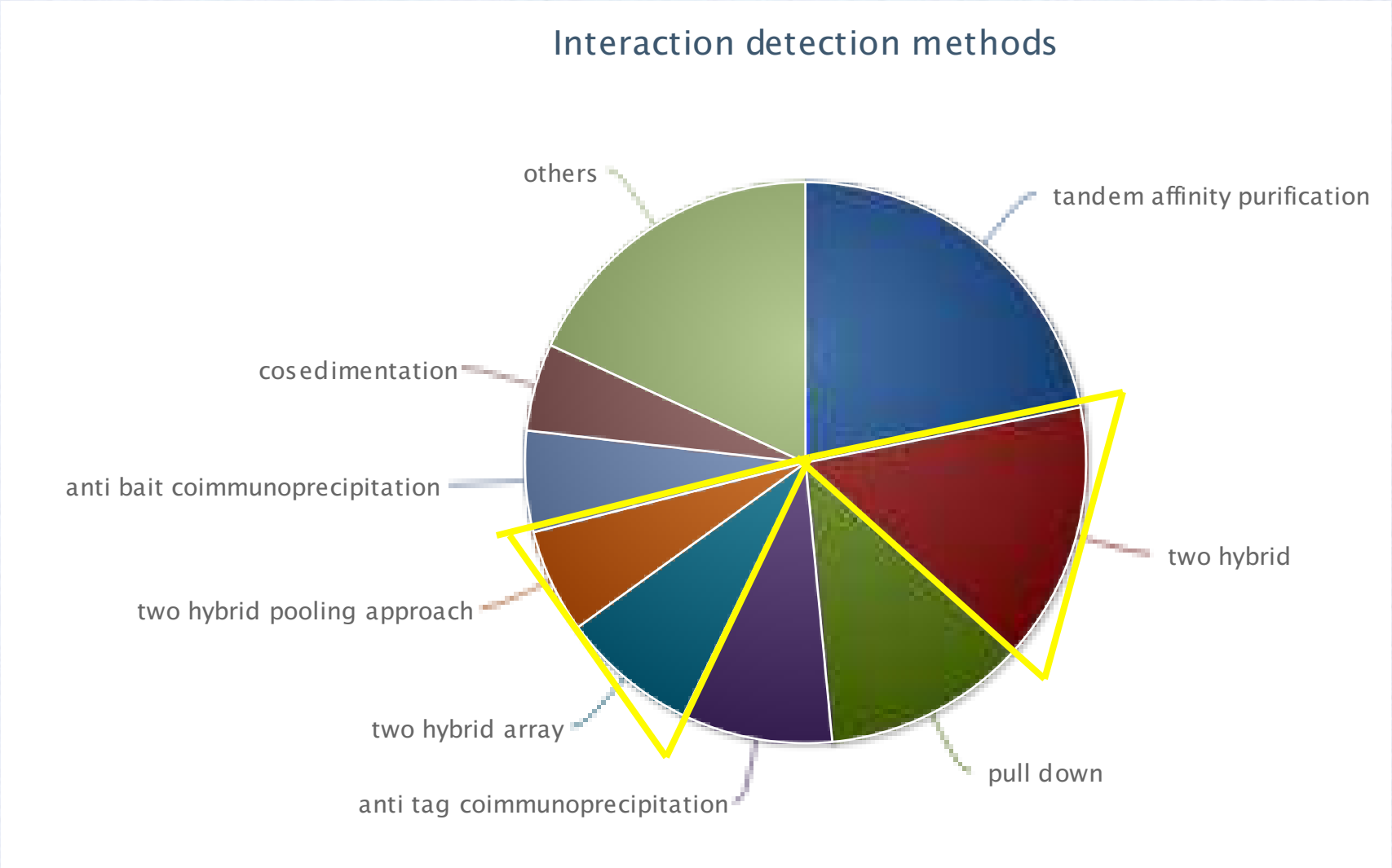




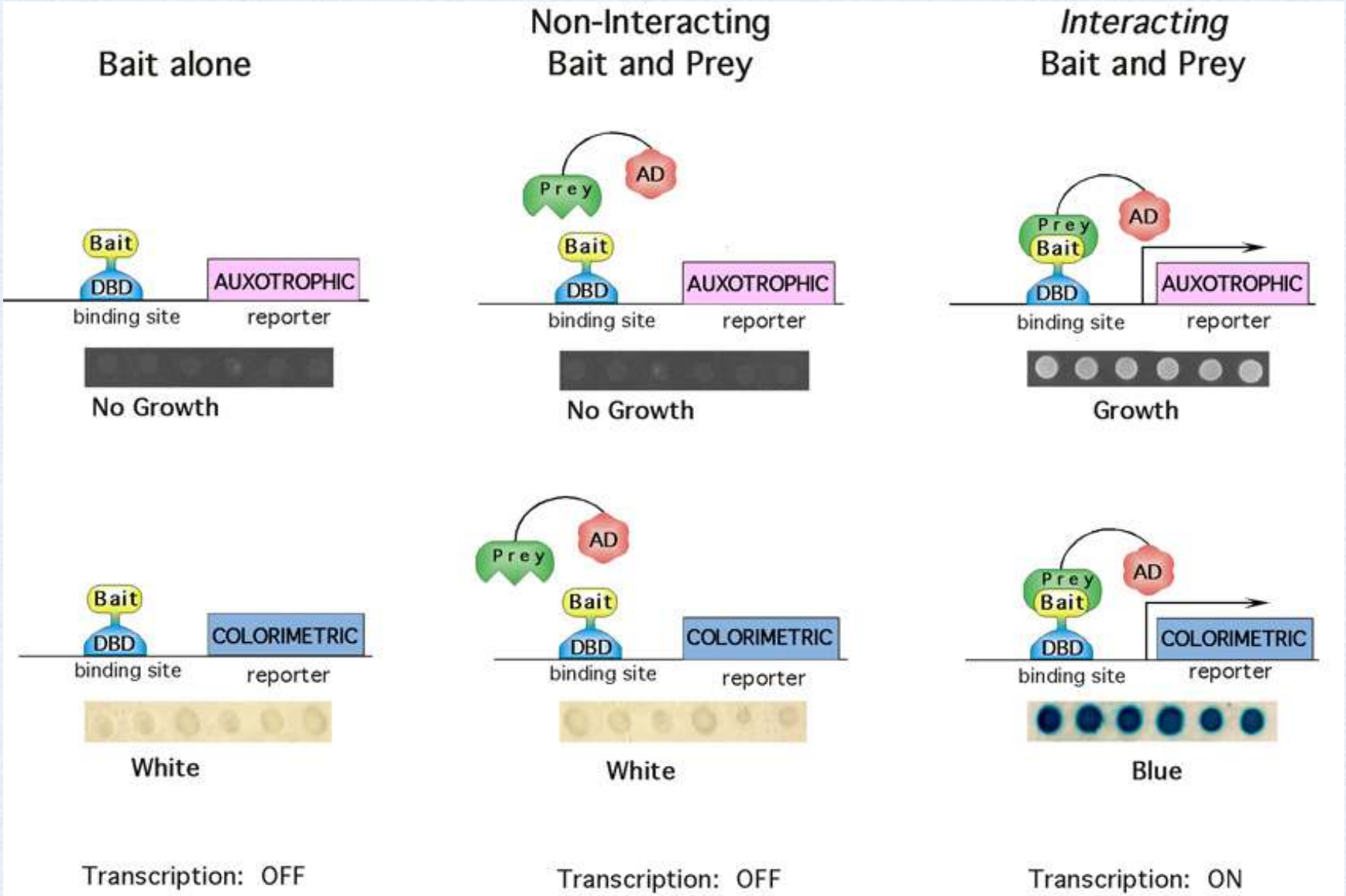


“ **interolog**” - conserved interaction between a pair of homologous proteins

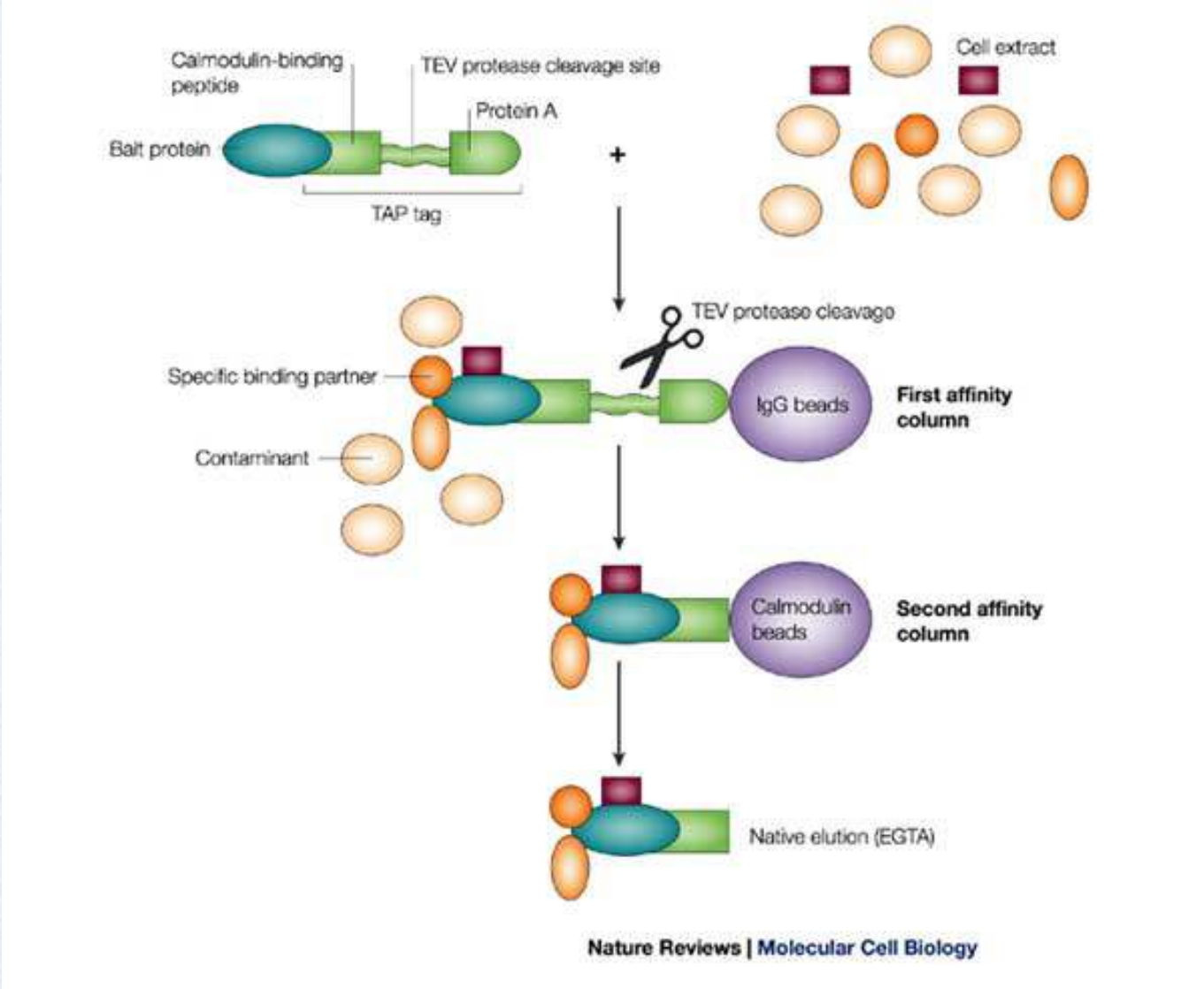
Методы определения белок-белковых взаимодействий



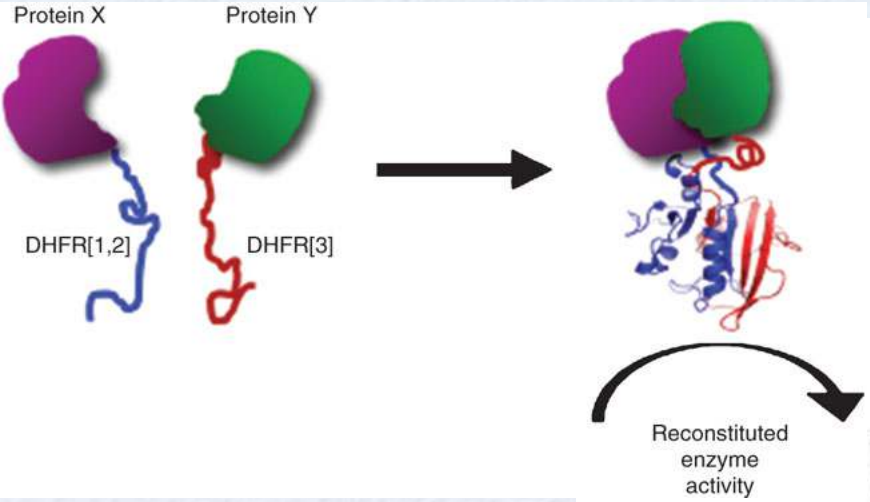
Yeast Two-hybrid

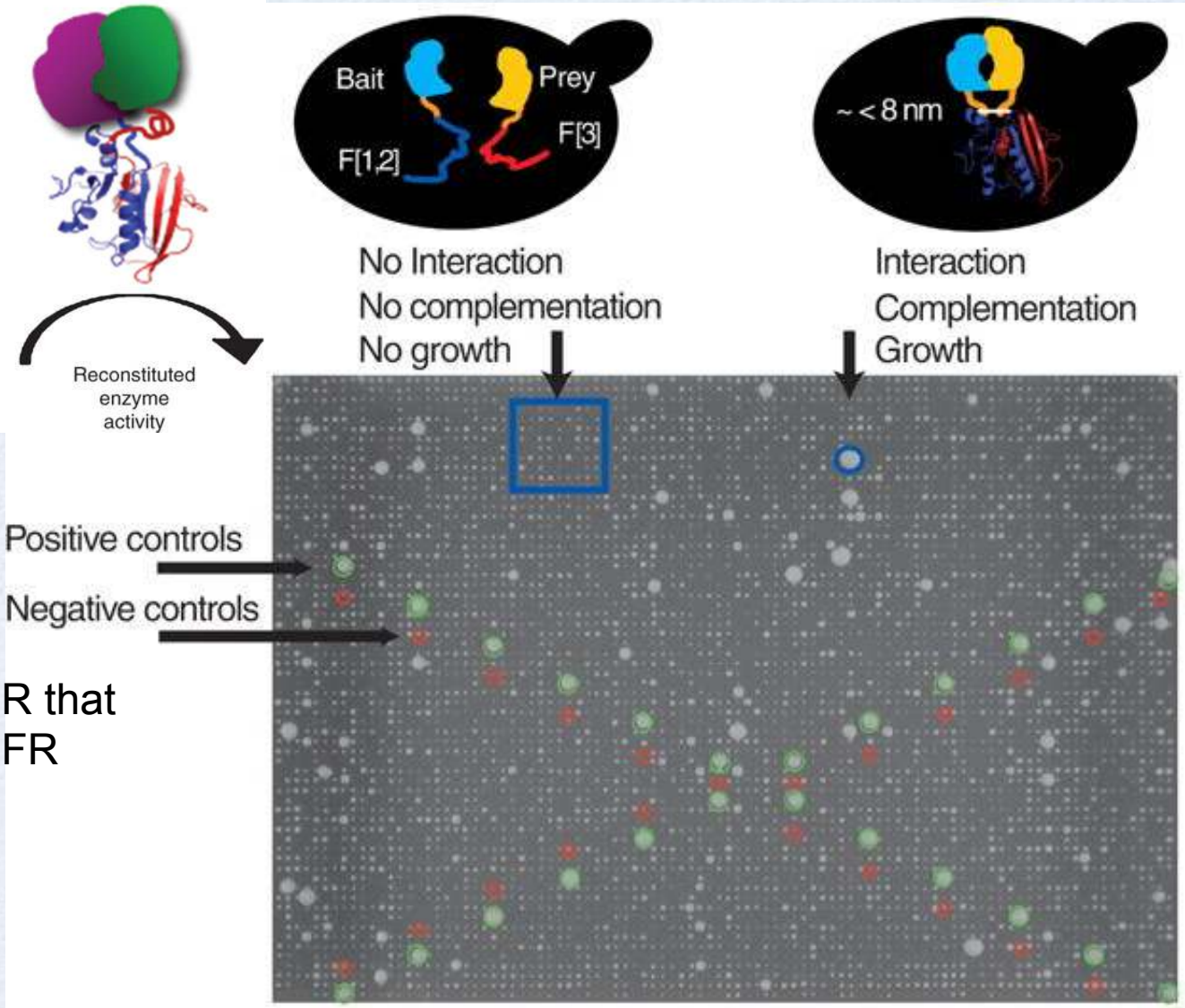


Tandem affinity purification



Protein fragment complementation assay

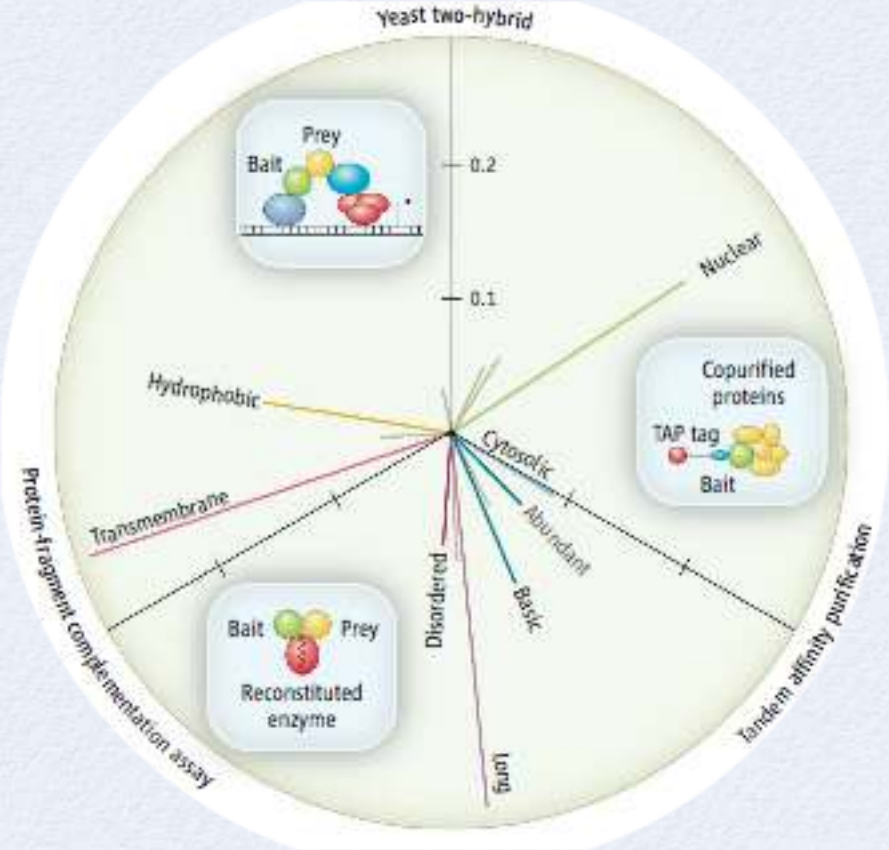




Using mutant of mDHFR that is insensitive to the DHFR inhibitor methotrexate

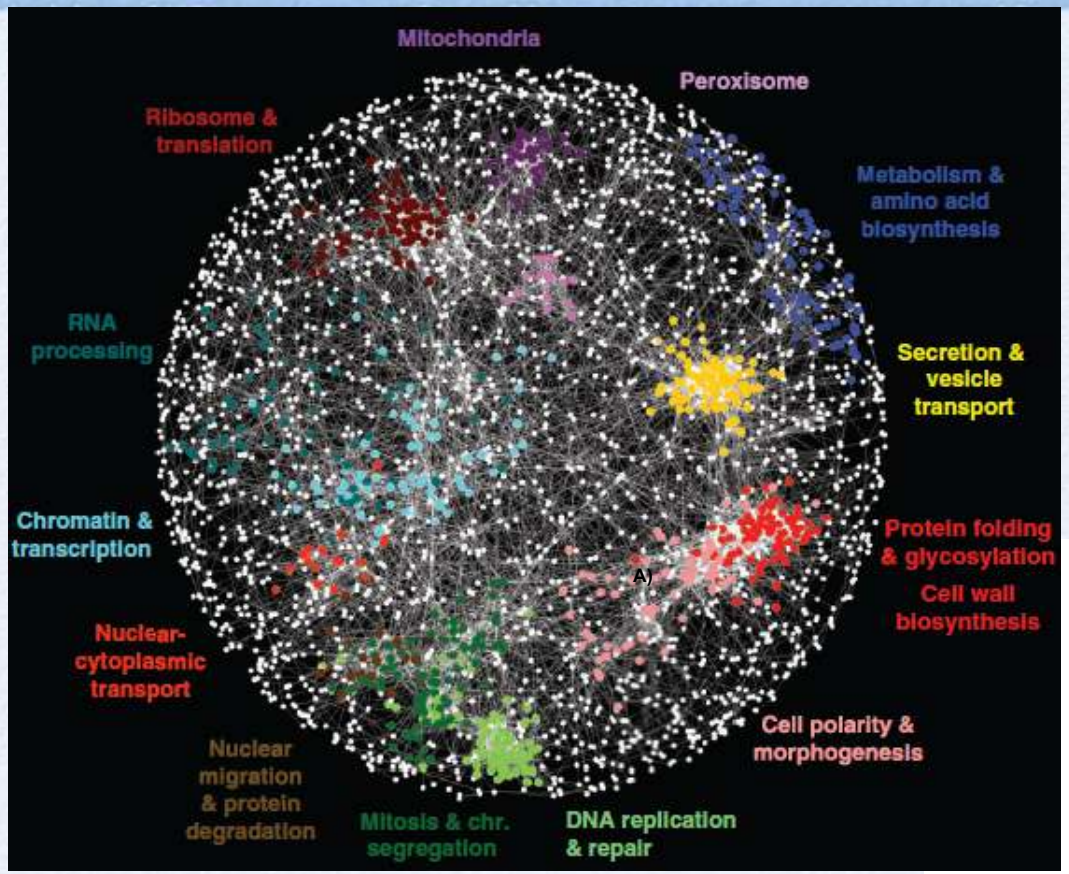
Методы определения белок-белковых взаимодействий

Своя специфика у каждого из них...

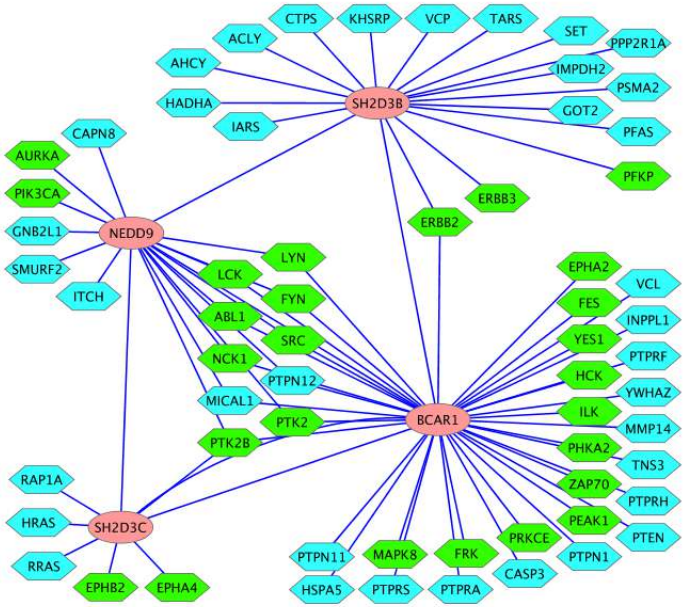


The three methods shown for detecting protein interactions function in fundamentally different ways and hence have different physiochemical constraints. The length of each line represents the strength of the bias.

Сети взаимодействий позволяют получить дополнительную информацию



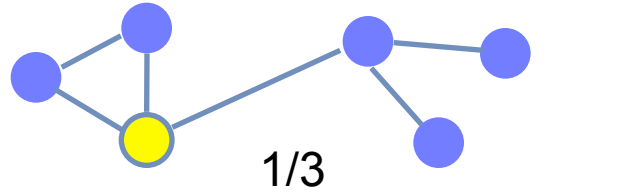
Глобальная сеть генетических взаимодействий в дрожжах

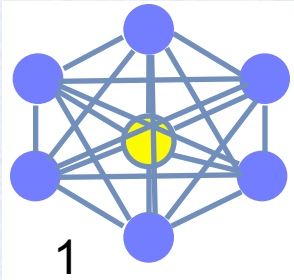


Упрощённая сеть физических взаимодействий, построенная вокруг изучаемых генов

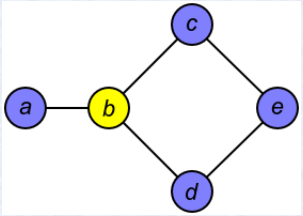
Топологические характеристики сети описываются многими параметрами

Коэффициент кластеризации



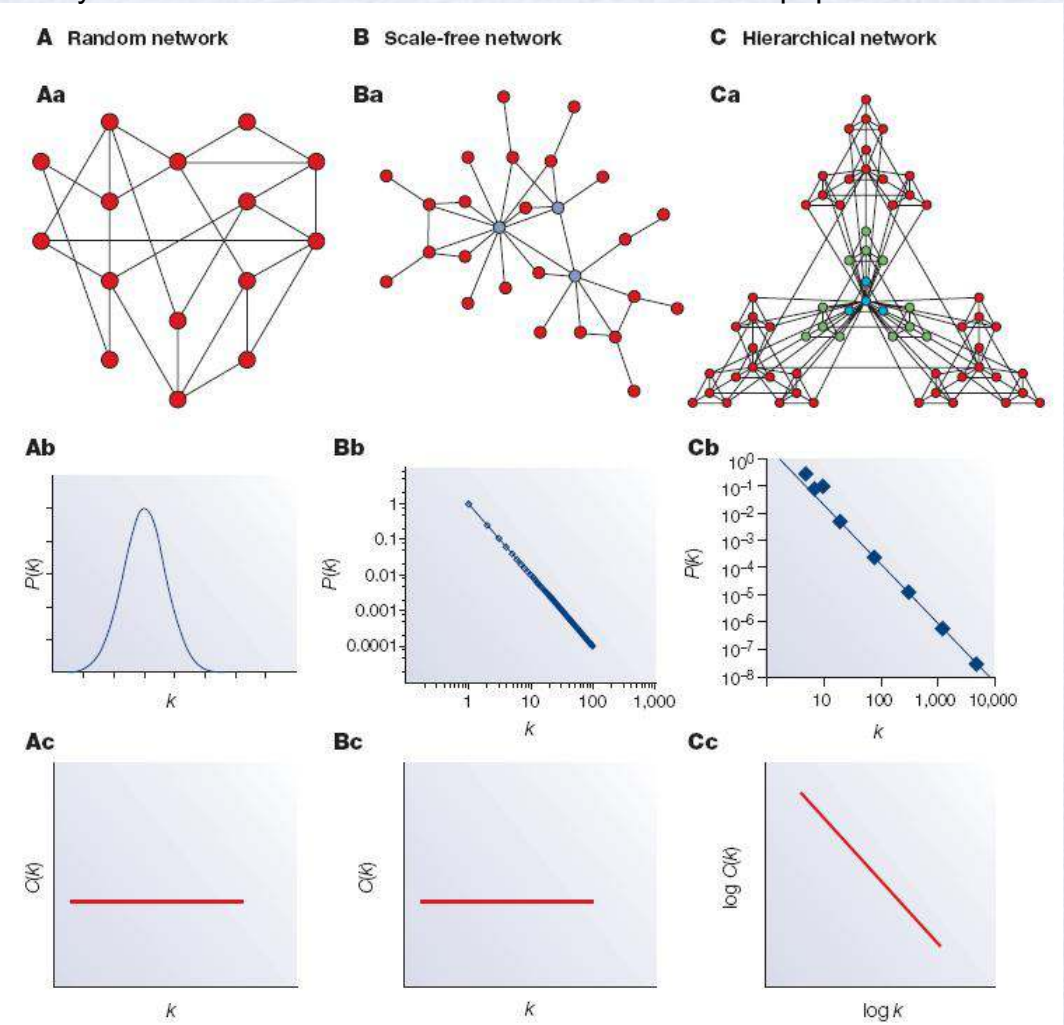


Промежуточная центральность

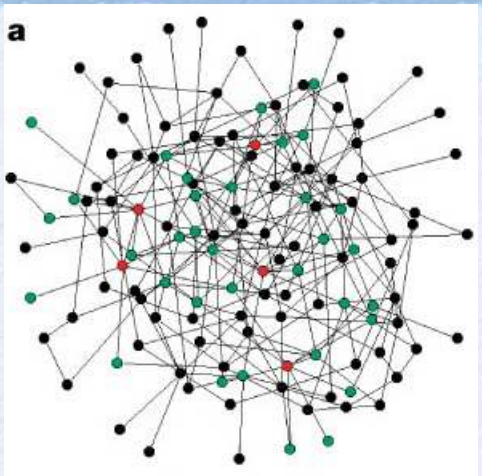


$$Cb(b) = ((\sigma_{ac}(b) / \sigma_{ac}) + (\sigma_{ad}(b) / \sigma_{ad}) + (\sigma_{ae}(b) / \sigma_{ae}) + (\sigma_{cd}(b) / \sigma_{cd}) + (\sigma_{ce}(b) / \sigma_{ce}) + (\sigma_{de}(b) / \sigma_{de})) / 6 = ((1 / 1) + (1 / 1) + (2 / 2) + (1 / 2) + 0 + 0) / 6 = 3.5 / 6 \approx 0.583$$

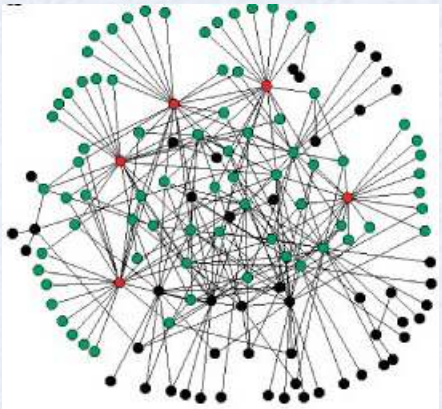
Случайная сеть | Безмасштабная сеть | Иерархическая сеть



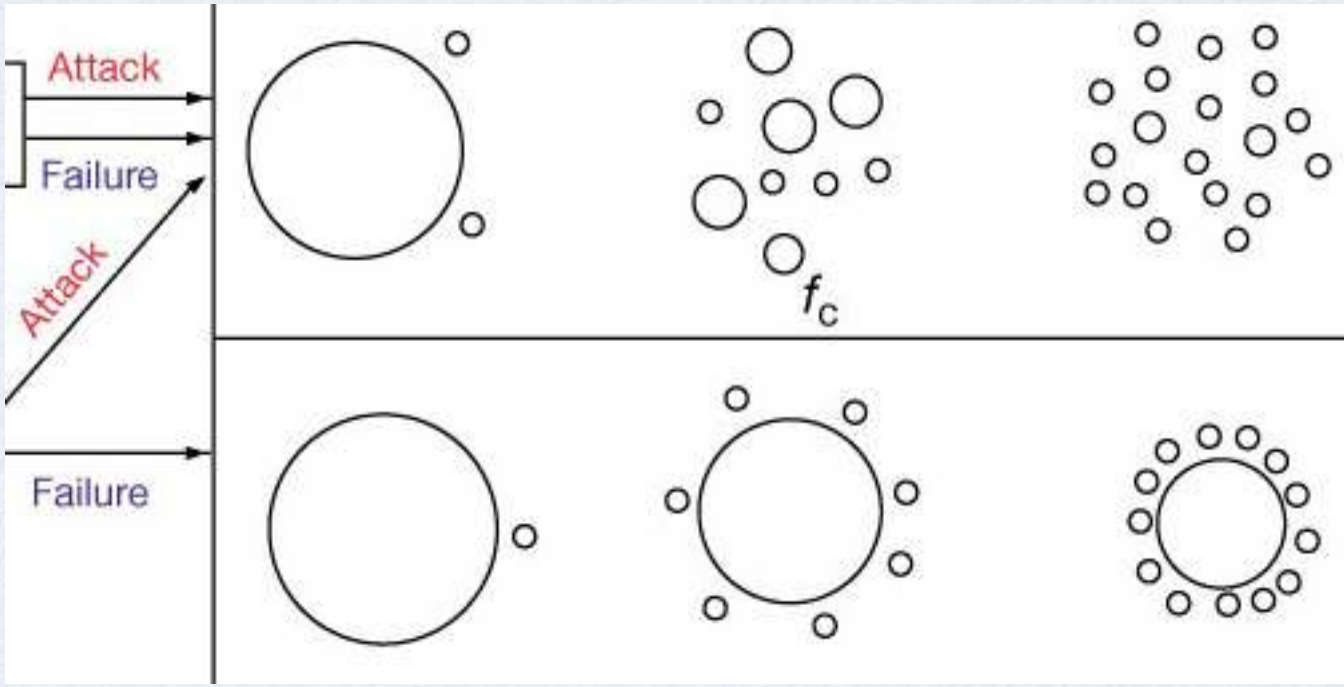
Топологические характеристики определяют многие свойства сети



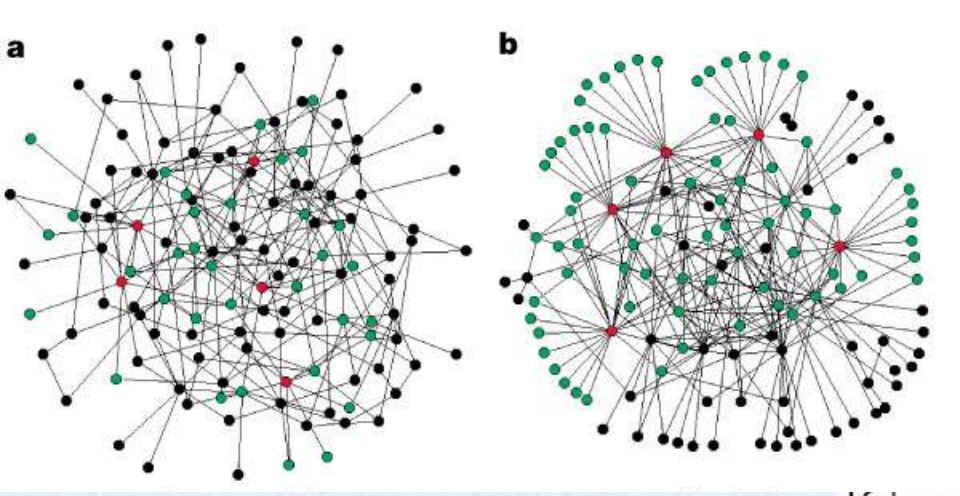
Случайная сеть



Безмасштабная сеть



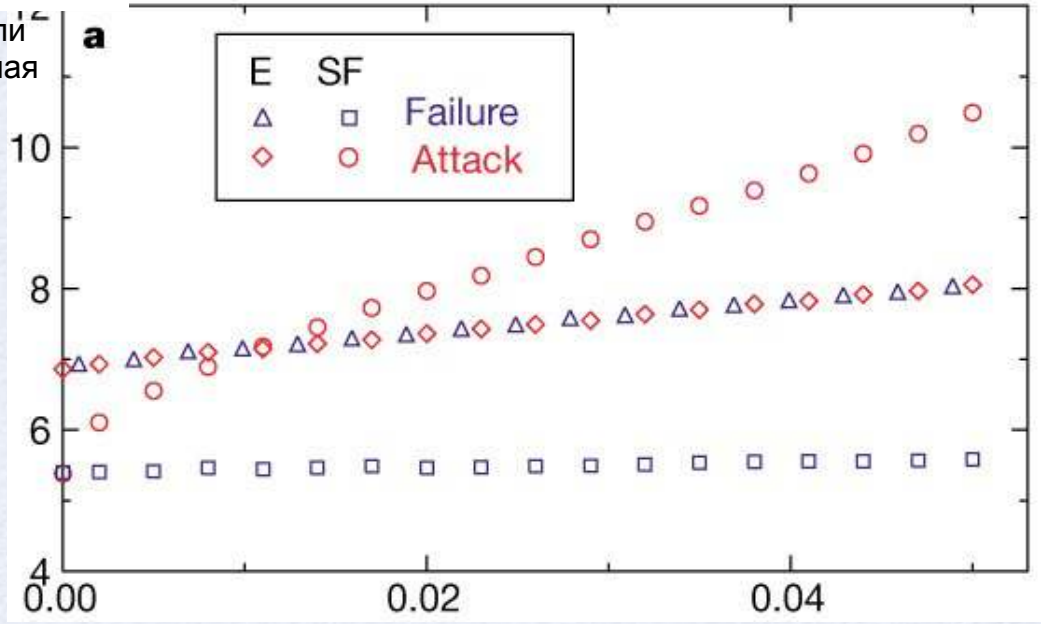
Топологические характеристики определяют многие свойства сети



Случайная сеть

Безмасштабная сеть или масштабно-инвариантная сеть

- Скорость синхронизации сигнала
- Устойчивость к разрушению

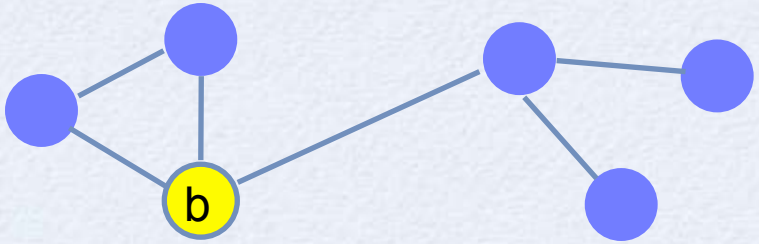


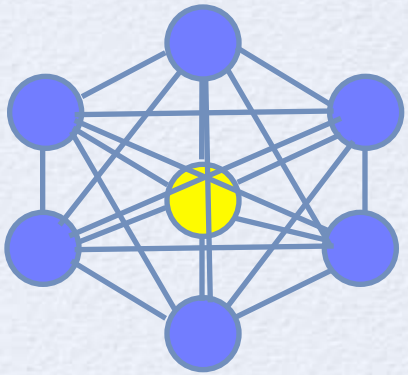
Топологические характеристики определяют многие свойства сети

Для любой вершины можно определить *степень скопления* (clustering coefficient) как показатель взаимосвязанности её соседей.

Clustering coefficient Коэффициент кластеризации

The clustering coefficient is a ratio N / M, where N is the number of edges between the neighbors of n, and M is the maximum number of edges that could possibly exist between the neighbors of n. The clustering coefficient of a node is always a number between 0 and 1.





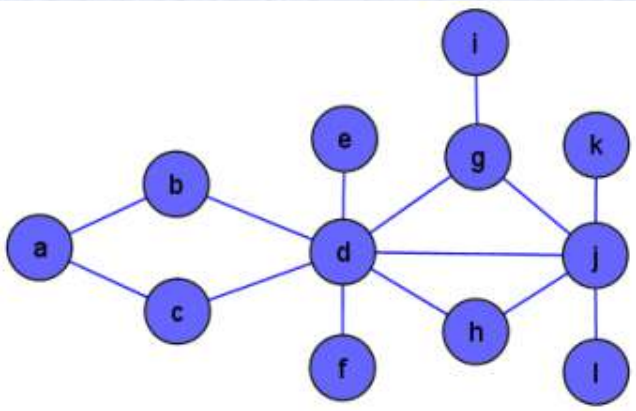
the clustering coefficient $C_b = 1/3$

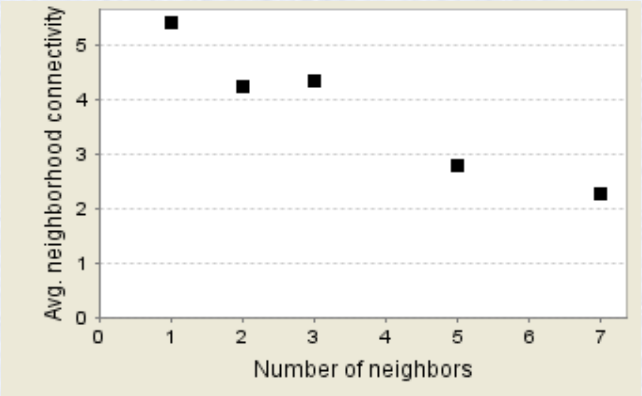
The average clustering coefficient distribution gives the average of the clustering coefficients for all nodes n with k neighbors for k = 2,…. It was used to identify a modular organization of metabolic networks

Топологические характеристики определяют многие свойства сети

Neighborhood connectivity Связность окружения

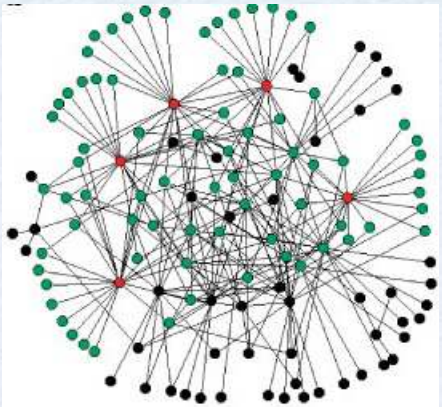
The connectivity of a node is the number of its neighbors. The neighborhood connectivity of a node n is defined as the average connectivity of all neighbors of n . The neighborhood connectivity distribution gives the average of the neighborhood connectivities of all nodes n with k neighbors for k = 0,1,….





If the neighborhood connectivity distribution is a decreasing function in k, edges between low connected and highly connected nodes prevail in the network

Топологические характеристики определяют многие свойства сети



Промежуточная центральность

Betweenness centrality

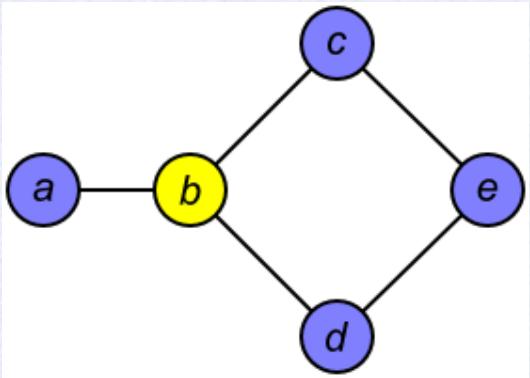
The betweenness centrality Cb(n) of a node n is computed as follows:

$C_b(n) = \sum_{s \neq n \neq t} (\sigma_{st}(n) / \sigma_{st})$.

It is always a number between 0 and 1.

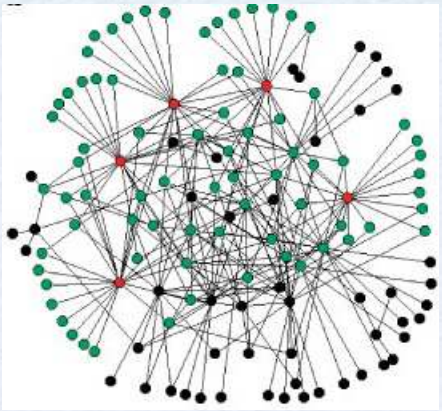
$$Cb(b) = ((\sigma_{ac}(b) / \sigma_{ac}) + (\sigma_{ad}(b) / \sigma_{ad}) + (\sigma_{ae}(b) / \sigma_{ae}) + (\sigma_{cd}(b) / \sigma_{cd}) + (\sigma_{ce}(b) / \sigma_{ce}) + (\sigma_{de}(b) / \sigma_{de})) / 6 =$$

$$((1 / 1) + (1 / 1) + (2 / 2) + (1 / 2) + 0 + 0) / 6 = 3.5 / 6 \approx 0.583$$



The betweenness centrality of a node reflects the amount of control that this node exerts over the interactions of other nodes in the network . This measure is higher for nodes that join communities (dense subnetworks), rather than nodes which just lie inside a community.

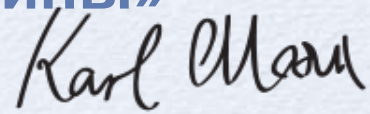
Топологические характеристики определяют многие свойства сети



Какой параметр важнее?

Промежуточная центральность?
Связность окружения?

« Практика - критерий истины»



Когда Маркс говорит, что критерий истины есть практика, то он выражает этим, прежде всего, именно точку зрения относительности истины.
С изменением содержания практики людей изменяется и их истина. То, что было истиною в пределах практики более узкой, перестает быть ею в практике более широкой, А для В.Ильина «критерий практики», это нечто вроде такого экзамена, после которого истине выдается окончательный аттестат: выдержала несколько веков, была безвредна — отлично, истина признается «объективной», вечной и т.д.; не выдержала — заблуждение, и тоже объективное, вечное и пр.

В.И. Ленин "Материализм и эмпириокритицизм"

Топологические характеристики определяют многие свойства сети ?

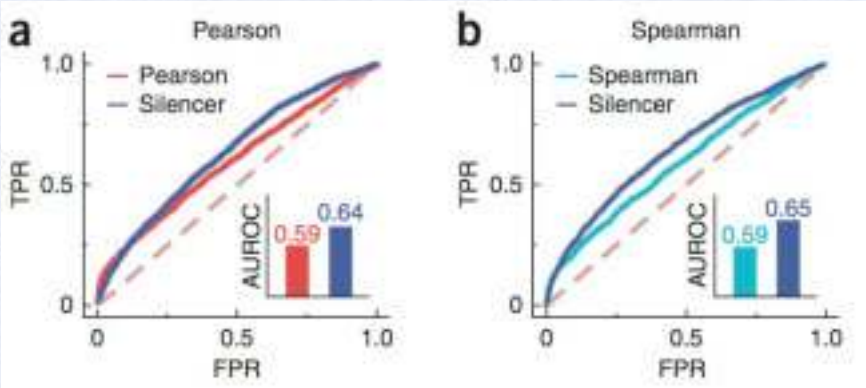
nature biotechnology

Home | Current issue | News & comment | Research | Archive ▾ | Authors & referees

Network link prediction by global silencing of indirect correlations

Baruch Barzel & Albert-László Barabási

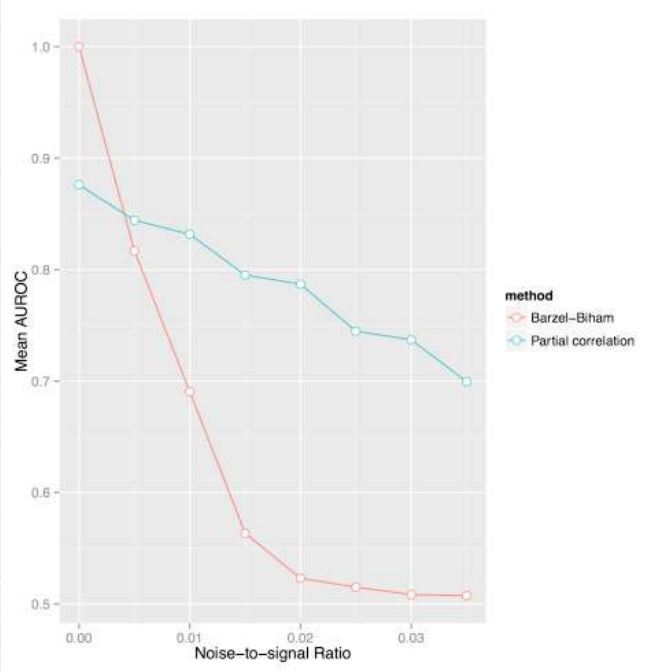
Nature Biotechnology 31, 720–725 (2013)



Bits of DNA

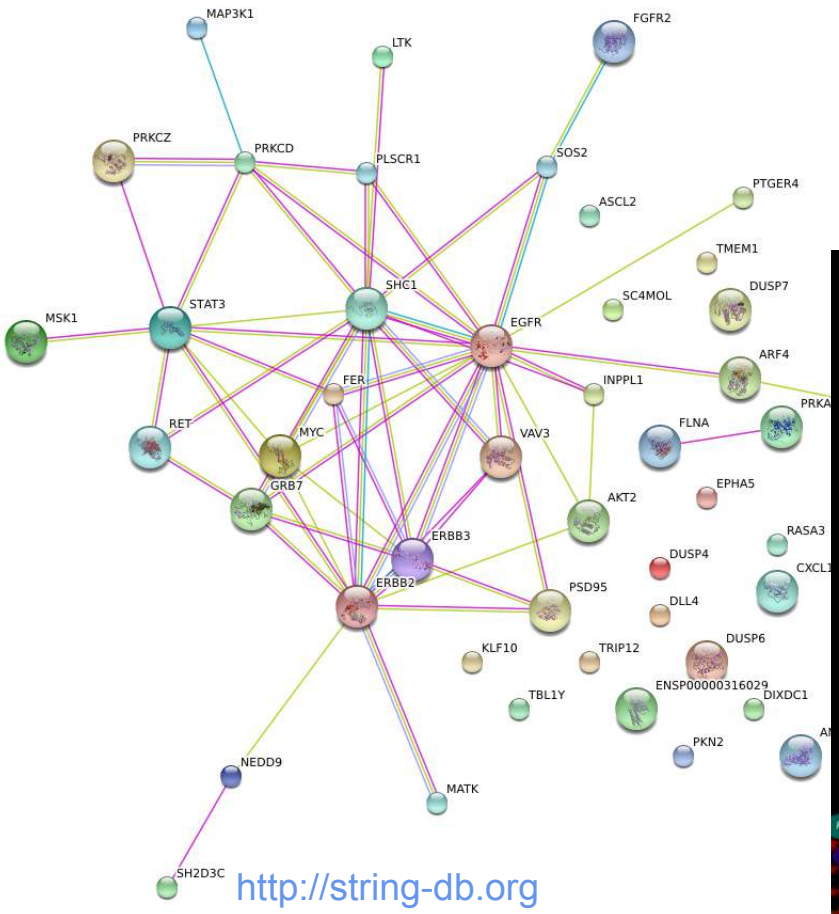
Reviews and commentary on computational biology

The network nonsense of Albert-László Barabási



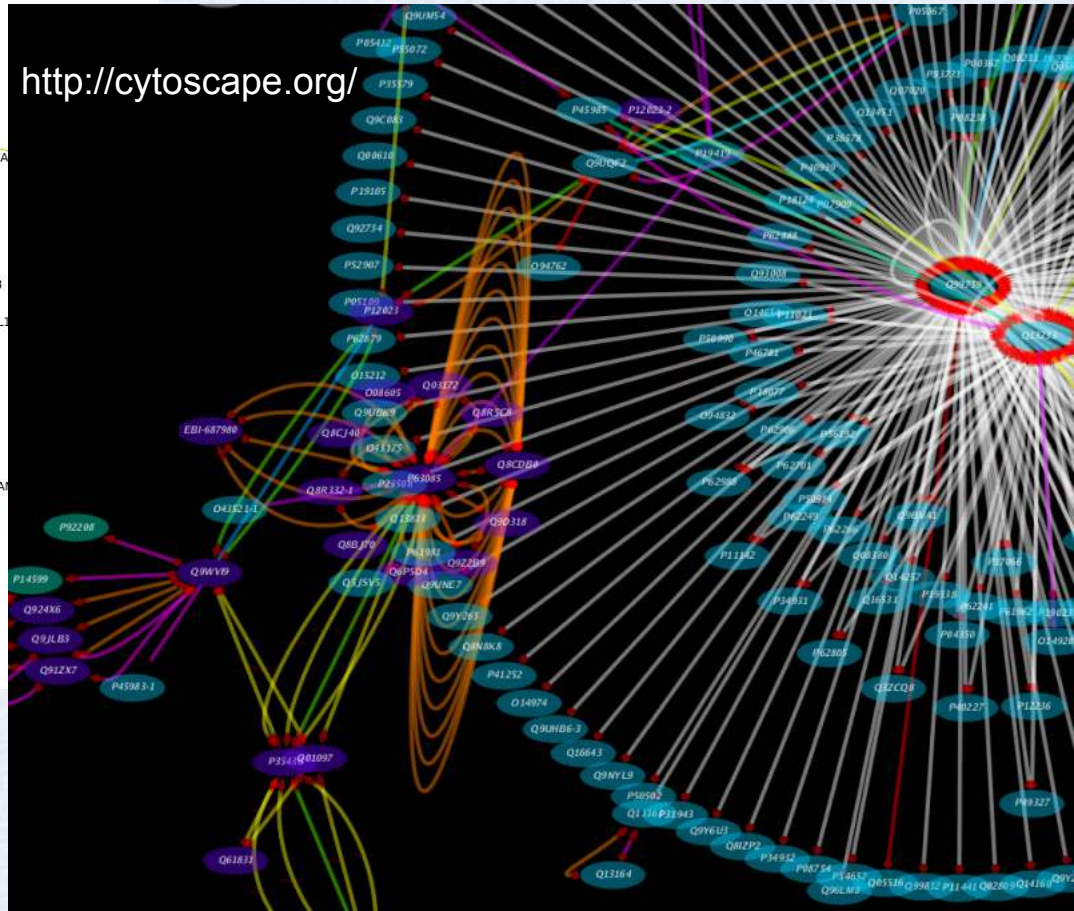
http://liorpachter.wordpress.com/2014/02/10/the-network-nonsense-of-albert-laszlo-barabasi/

Источники информации для построения и визуализации сетей



http://string-db.org

- Метапоисковые системы – для быстрого поиска и визуализации
- Индивидуальные базы данных для более тщательного поиска



http://cytoscape.org/

- Специализированные программы визуализации

Построение моделей – важнейшая задача системной биологии

- отсутствие универсального, формализированного языка для описания биологических процессов тормозит обмен информацией

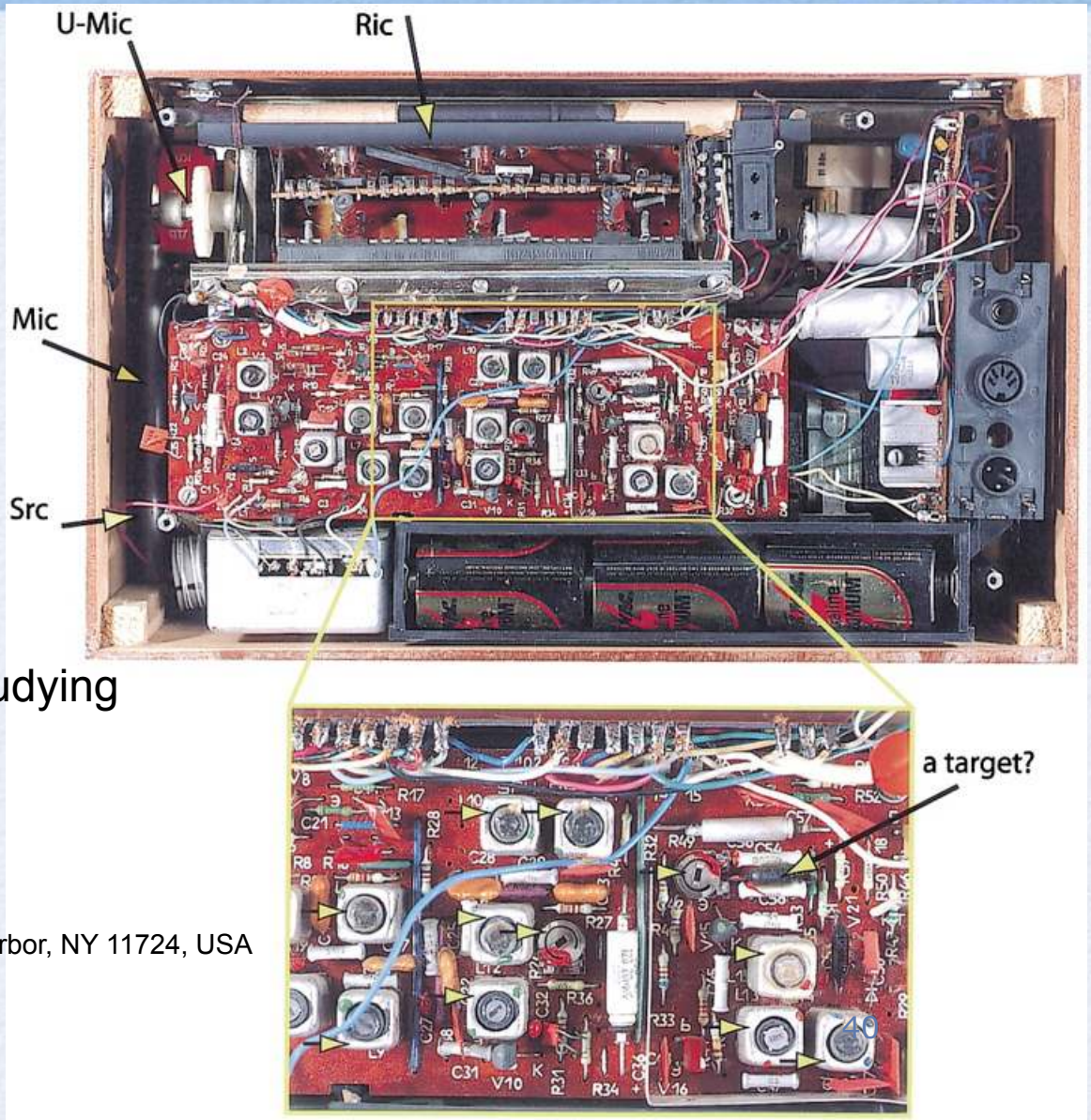
Impact Factor 7.35

Oncogene. 2006 Aug 7;25(34):4633-46.
Glycolysis inhibition for anticancer treatment.
Pelicano H, Martin DS, Xu RH, Huang P.

The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

For instance, mitochondrial hexokinase activity ***seems*** to be required for the growth factor-induced cell survival, and Akt (protein kinase B) signaling ***appears to promote*** the association of hexokinase with VDAC

Может ли биолог починить радиоприёмник?



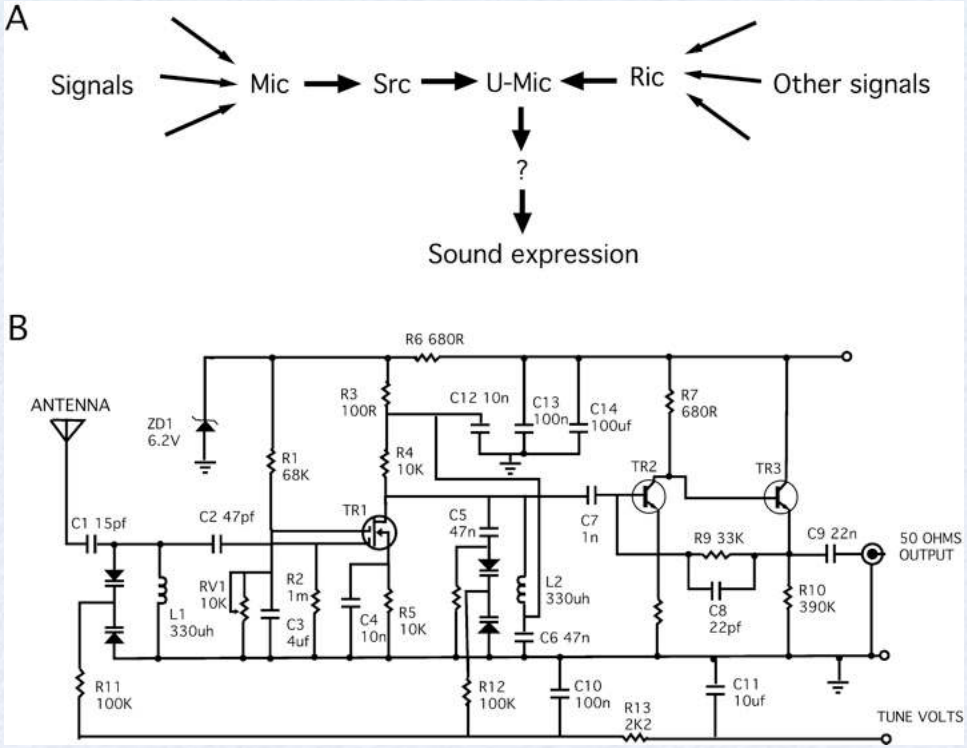
Can a biologist fix a radio?
—Or, what I learned while studying apoptosis

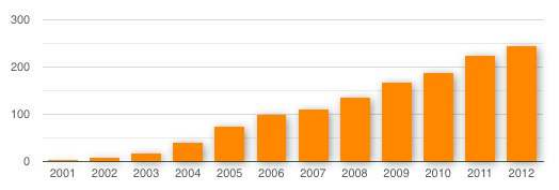
Yuri Lazebnik,

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

CANCER CELL : SEPTEMBER 2002

Построение моделей – важнейшая задача системной биологии

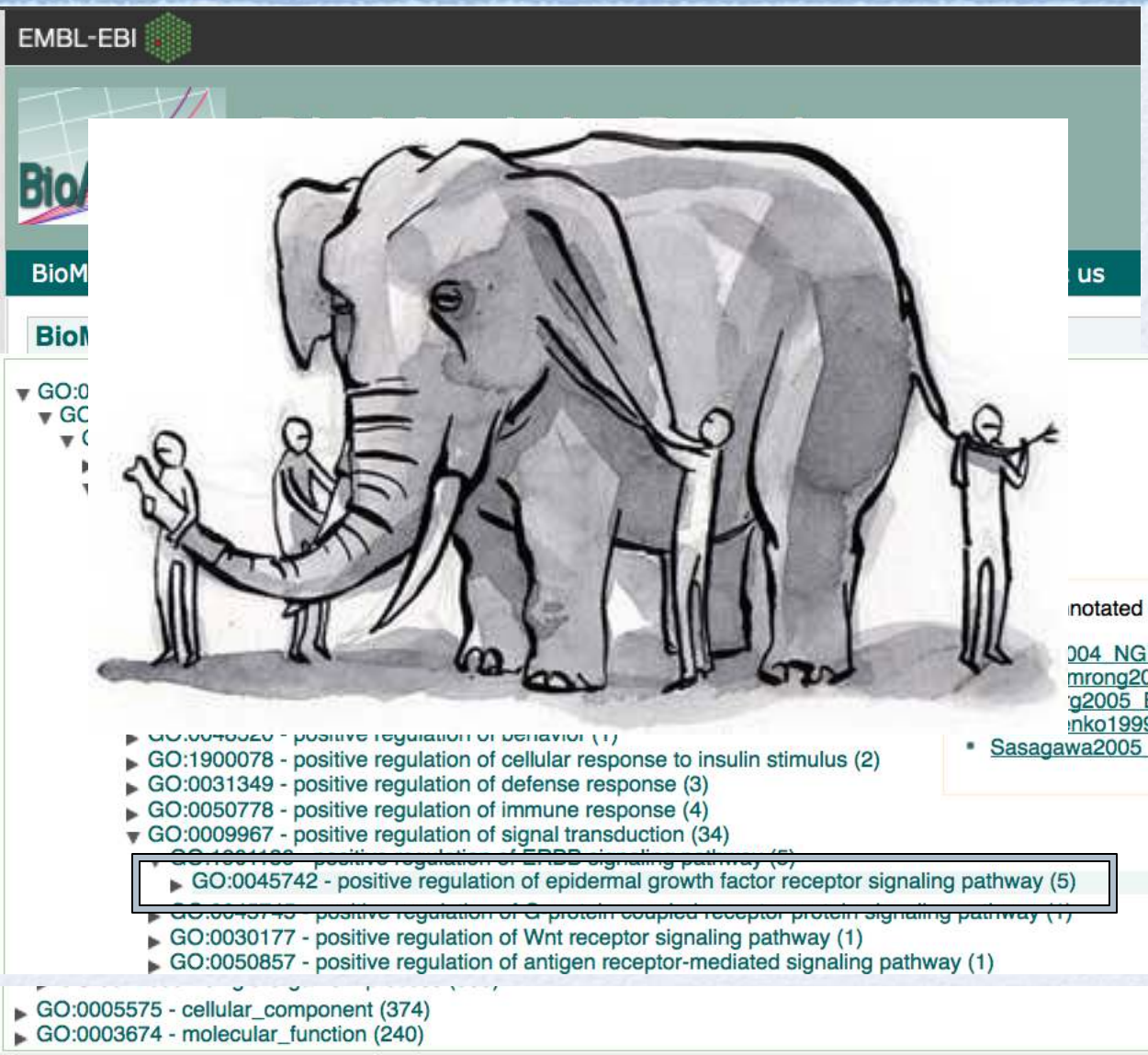




SBML.org The Systems Biology Markup Language

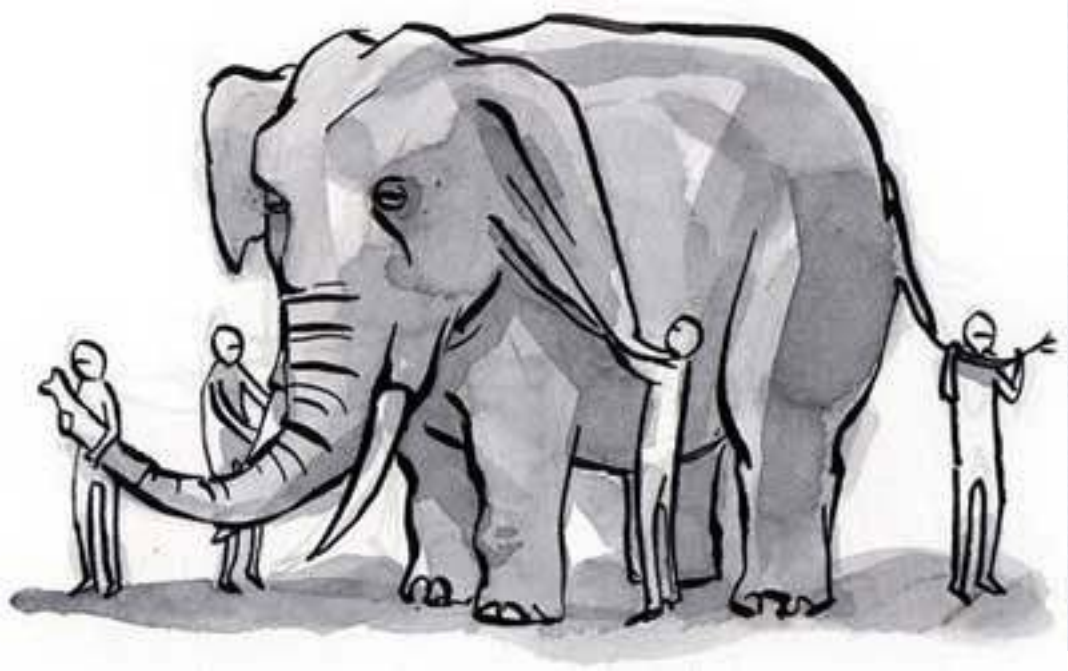
- SMBL – формат представления моделей, пригодный для описания путей передачи сигнала, метаболических путей, регуляции генов, биохимических реакций и т.д.
- Используется в более чем 250 компьютерных программах

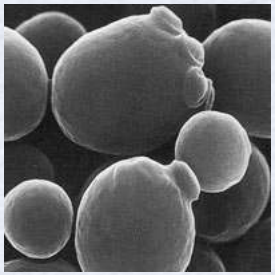
Существуют модели множества отдельных процессов...

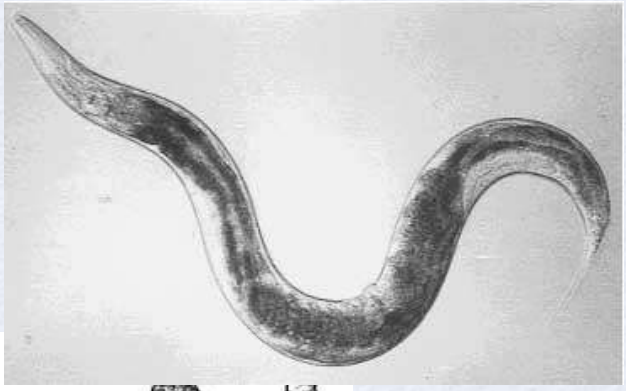


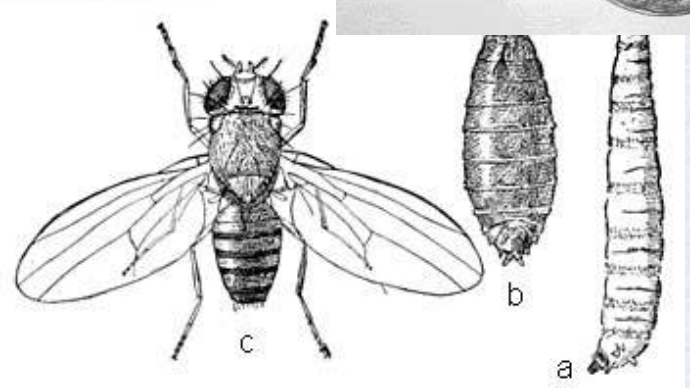
- Circadian Rhythms
- Cardiovascular Circulation
- . . .

... и даже отдельных организмов !

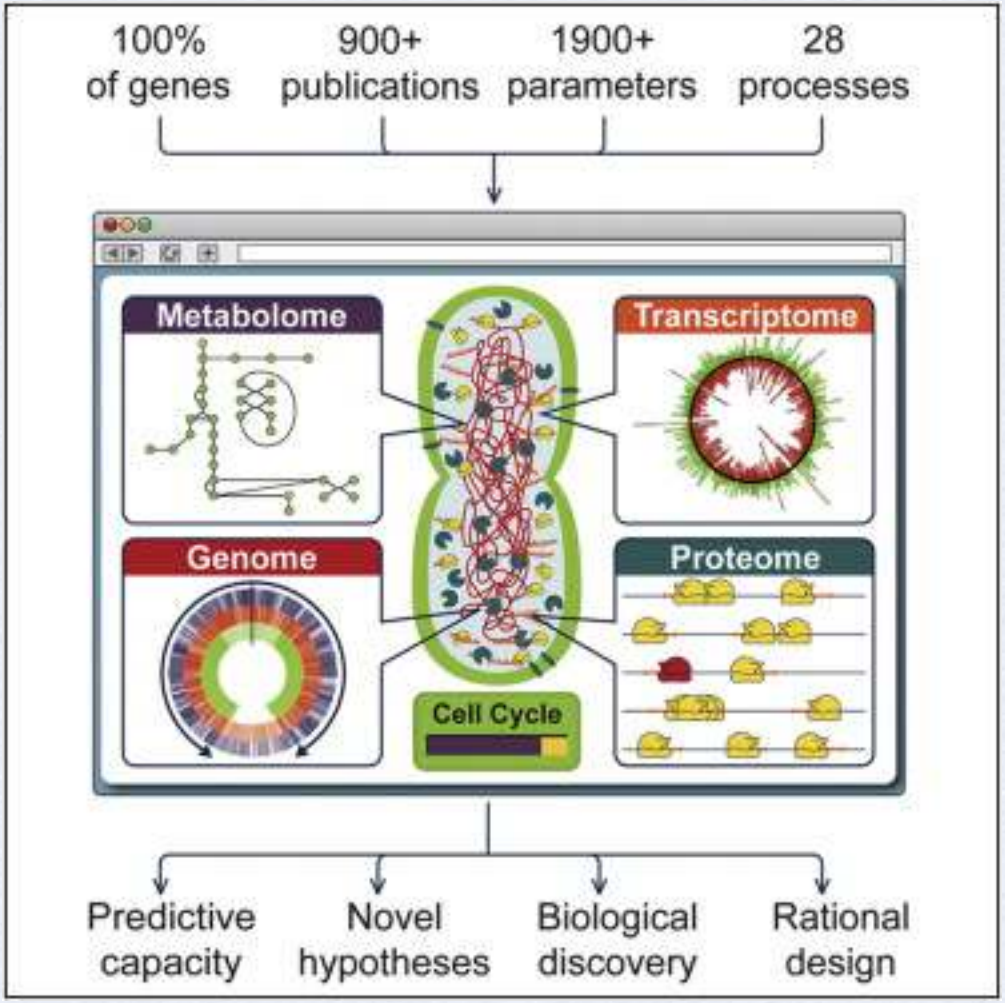








Первая модель живой клетки



Mycoplasma genitalium

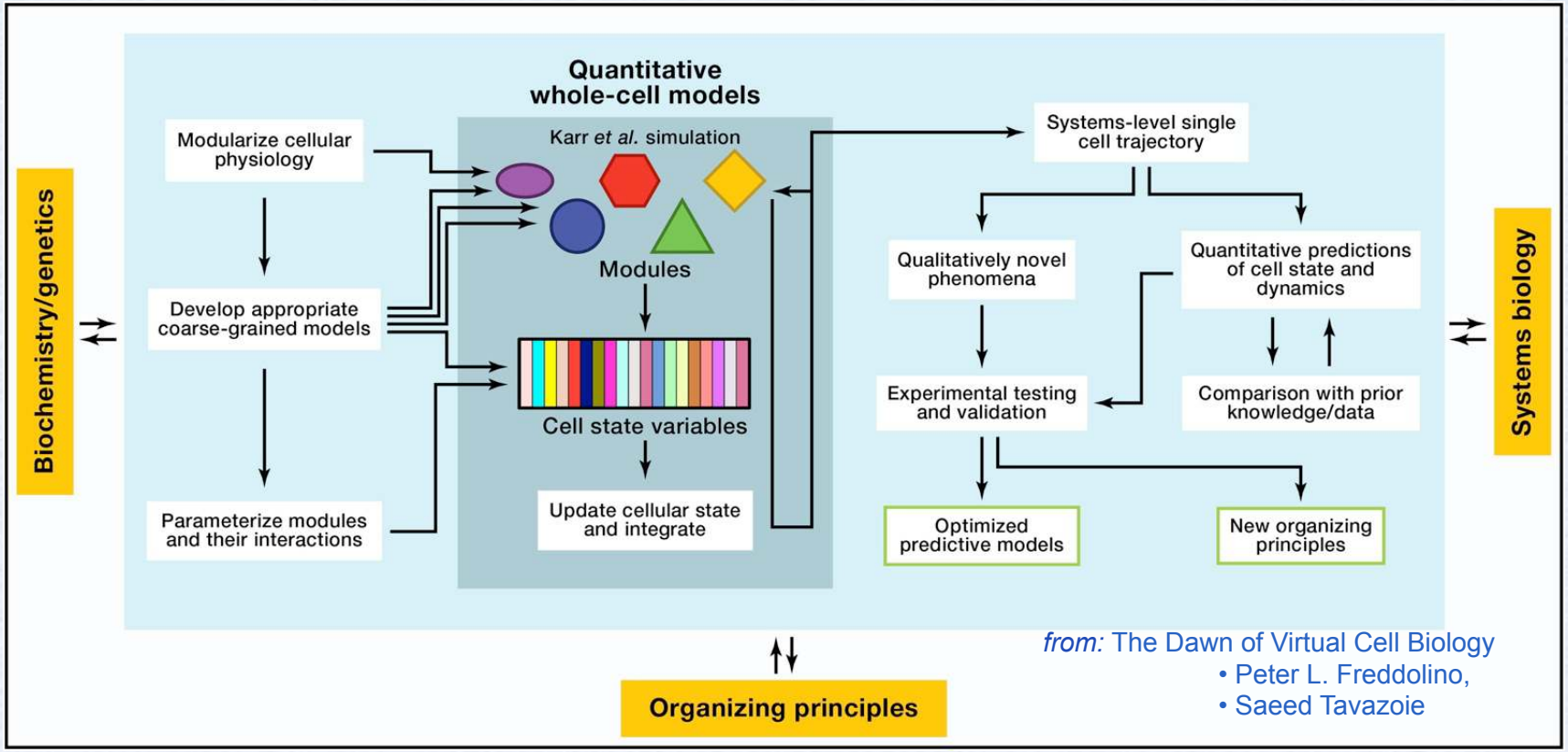
- всего 525 генов

Предложенная новая гипотеза: хотя в целом циклы деления идут с одинаковой скоростью, разные его стадии могут протекать с разной скоростью.

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW. A Whole-Cell Computational Model Predicts Phenotype from Genotype. Cell 150, 389-401 (2012)

Do we currently know enough to attempt a mesoscale simulation that will produce nontrivial results?

The only way to address this question is to produce whole-cell models that generate testable predictions.



from: The Dawn of Virtual Cell Biology
- Peter L. Freddolino,
- Saeed Tavazoie

Е. Кунин: Я не желаю никого обижать, но, с моей точки зрения, пока это - детский конструктор. Утверждается, что такие модели необходимы для понимания того, как это все работает. Но если вы не можете предсказуемым образом ими манипулировать, то это пустые слова. Чтобы научиться делать такие модели, все-таки должно пройти еще несколько лет.

Вклад системной биологии в развитие молекулярной медицины

Основные задачи:

- построение глобальных сетей генетических взаимодействий
- идентификация биомаркеров заболевания на основе изучения и моделирования биохимических путей и путей передачи сигнала
- идентификация генов, изменения в которых приводят к заболеваниям

Некоторые интересные и общедоступные ресурсы (относящиеся к онкологии)

- базы данных клеточных линий
- базы данных опухолевых образцов

International Cancer Genome Consortium

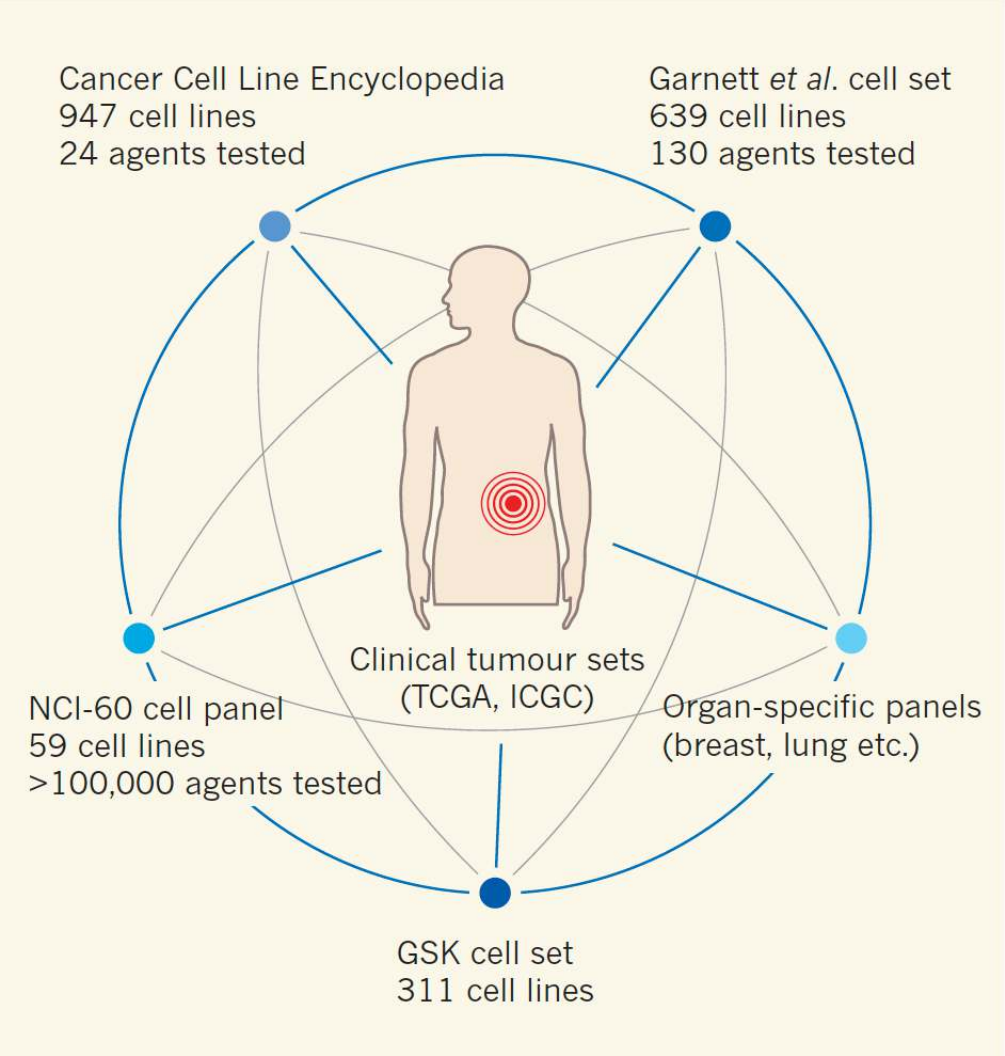
The Cancer Genome Atlas Data Portal

Understanding genomics to improve cancer care

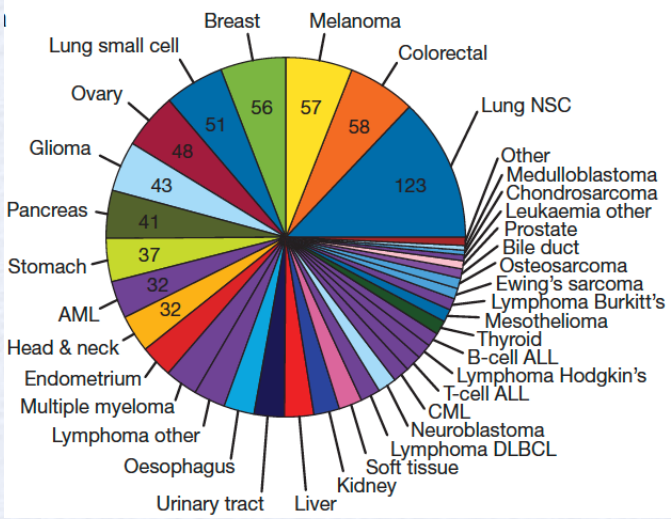
COSMIC

Catalogue of somatic mutations in cancer

Клеточные линии как средство изучения рака



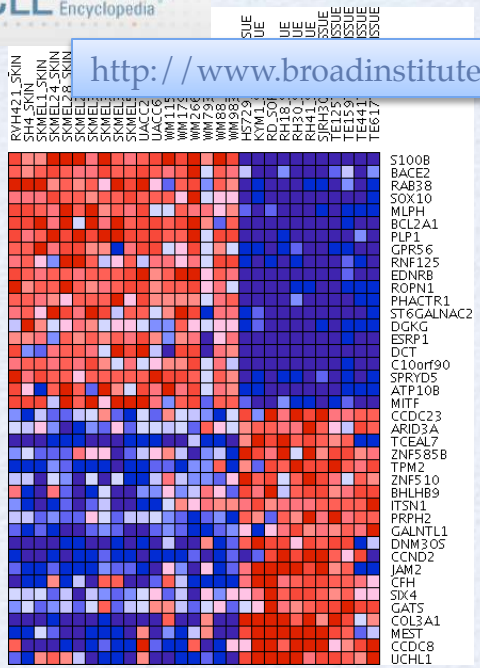
- уровни экспрессии генов
- хромосомные аномалии
- мутации в ключевых генах
- устойчивость к лекарствам



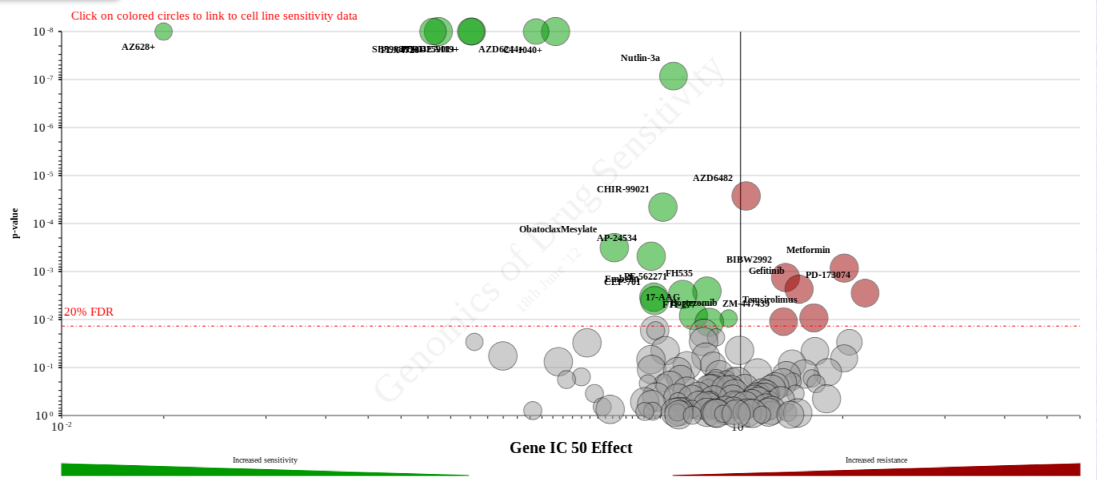
Клеточные линии как инструмент для изучения злокачественных опухолей

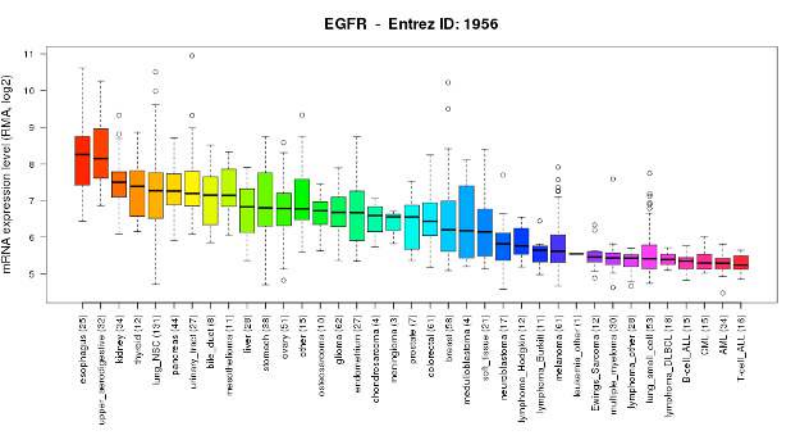
CCLE Cancer Cell Line Encyclopedia

http://www.broadinstitute.org/ccle/home



http://www.cancerrxgene.org/





- уровни экспрессии генов
- различия в уровнях экспрессии между тканями
- гены с похожими профилями экспрессии
- влияние мутации на устойчивость к лекарствам

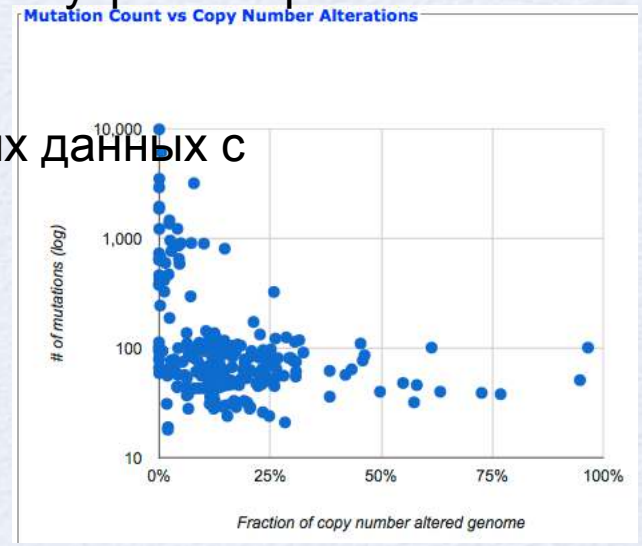
Memorial Sloan-Kettering Cancer Center

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

HOME | DATA SETS | WEB API | R/MATLAB | TUTORIALS | FAQ | NEWS | ABOUT

- Даёт доступ к данным более 13 000 образцов опухолей
- Позволяет провести анализ и визуализацию данных широкомасштабных исследований опухолевых тканей
- Данные включают в себя геномные перестройки, уровни эскпрессии генов на уровне РНК и белка, мутации
- Позволяет оценить взаимозависимость внутри набора генов
- Даёт возможность проследить связь этих данных с клиническим исходом



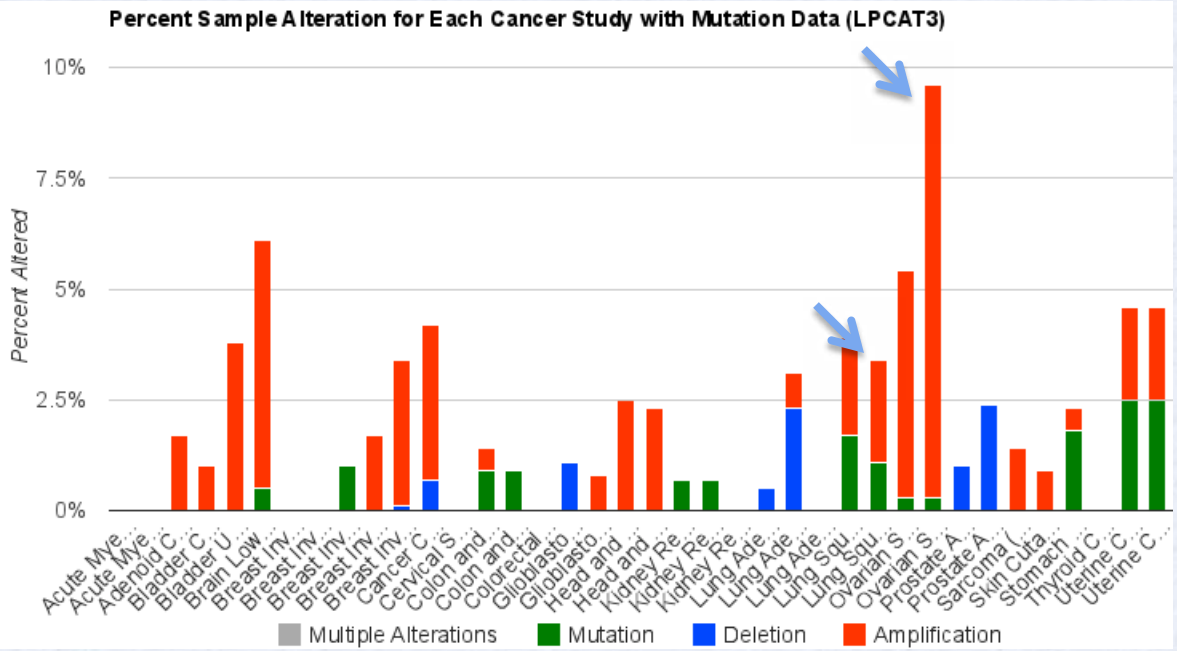
Memorial Sloan-Kettering Cancer Center

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

HOME | DATA SETS | WEB API | R/MATLAB | TUTORIALS | FAQ | NEWS | ABOUT

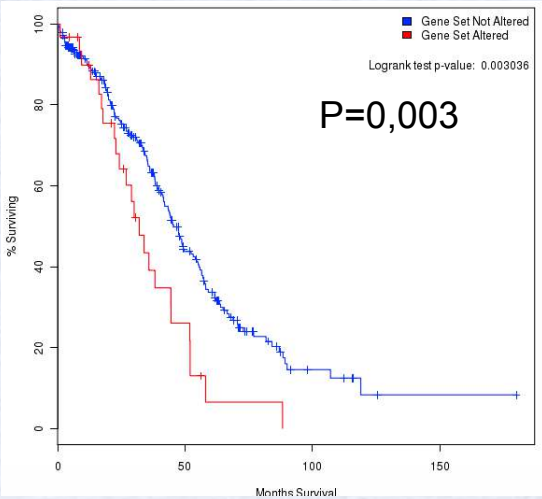
Исследование одного гена



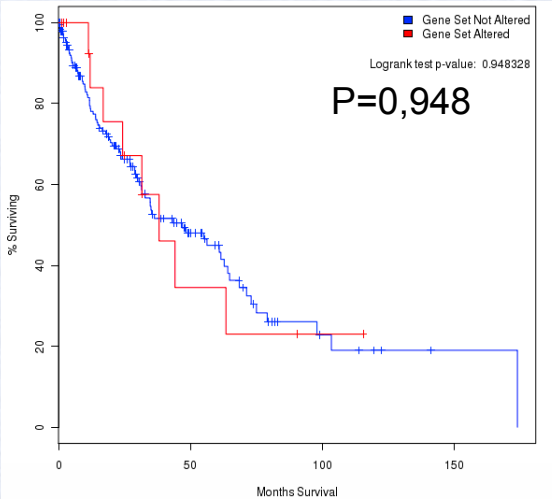
Сравнение геномных характеристик по разным типам опухолей

- значительно более часто амплифицирован в раке лёгких, чем в раке яичников
- можно сравнить кривые выживаемости больных при повышенной экспрессии этого гена

Рак яичников (311)



Рак лёгких (387)

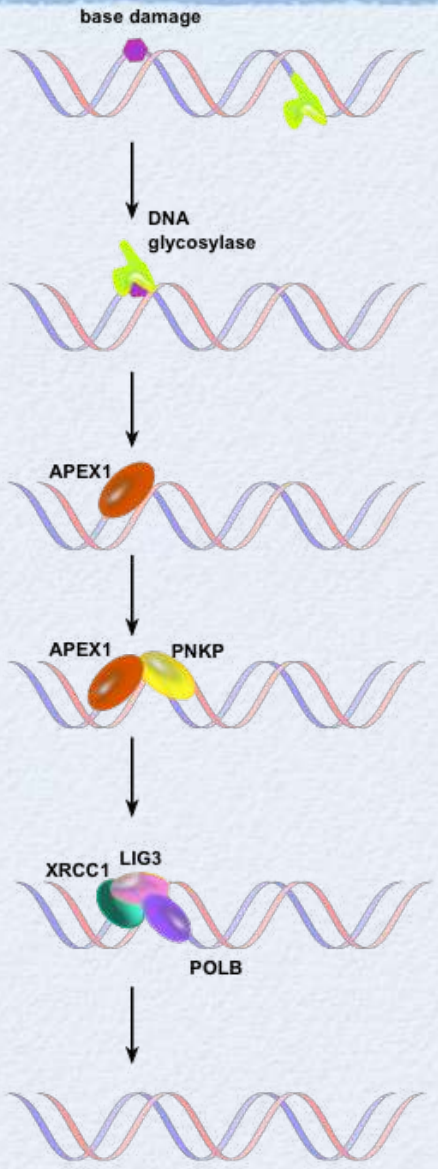


Memorial Sloan-Kettering Cancer Center

cBioPortal for Cancer Genomics

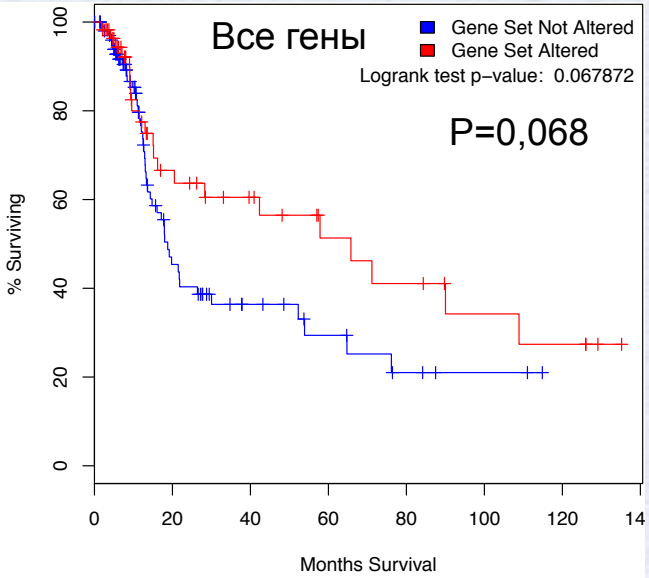
Visualize, analyze, discover.

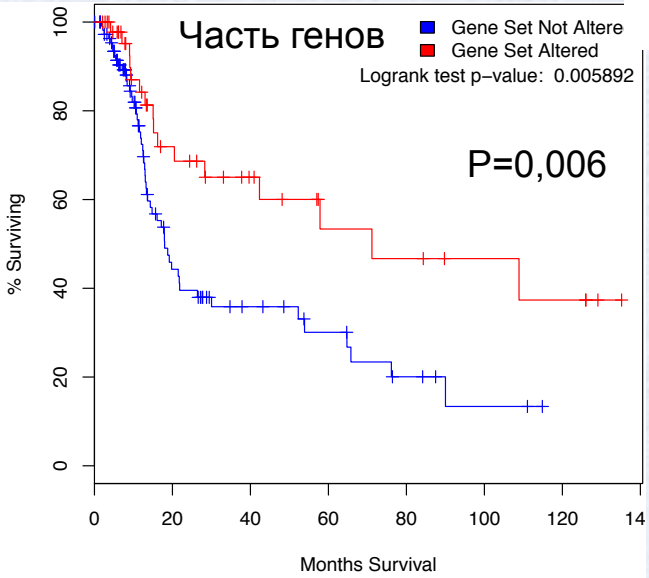
HOME | DATA SETS | WEB API | R/MATLAB | TUTORIALS | FAQ | NEWS | ABOUT



Исследование функционально связанной группы генов

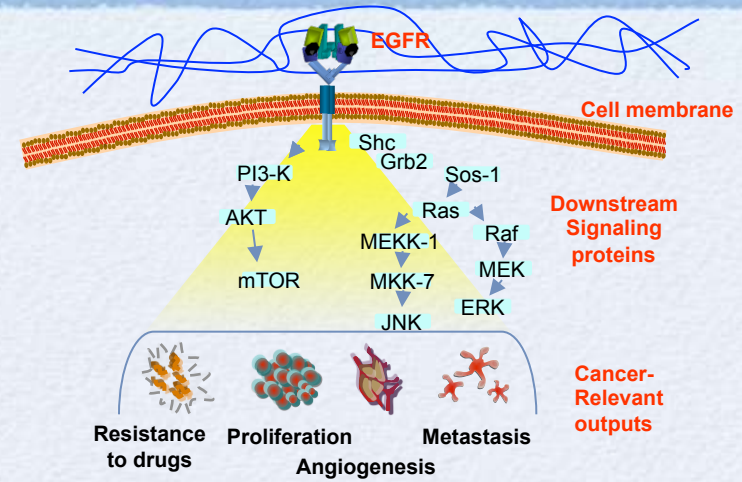
- Выбран тип опухоли, где часто применяются ДНК-повреждающие лекарства
- Проверены все пути репарации ДНК





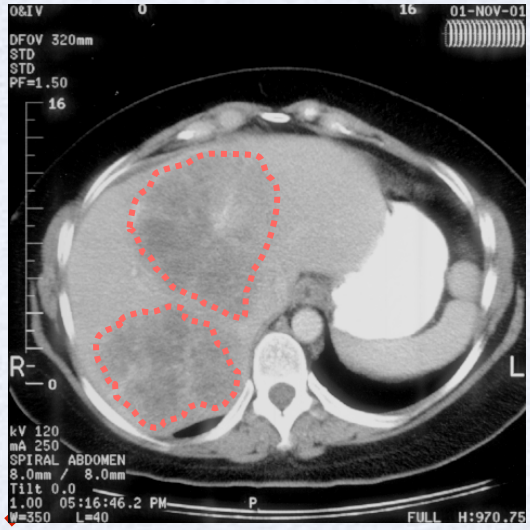
- определён набор кандидатов в биомаркеры ответа (необходима дальнейшая проверка)

Наши собственные исследования с применением системной биологии

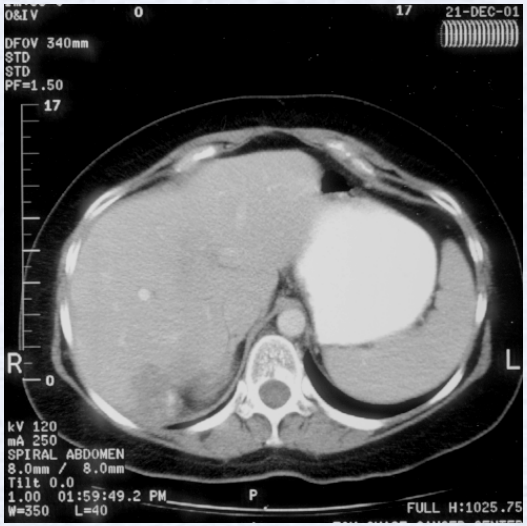


Year ~2000 EGFR signaling

Best case: one patient, EGFR-targeted therapy



Before treatment: metastases in liver

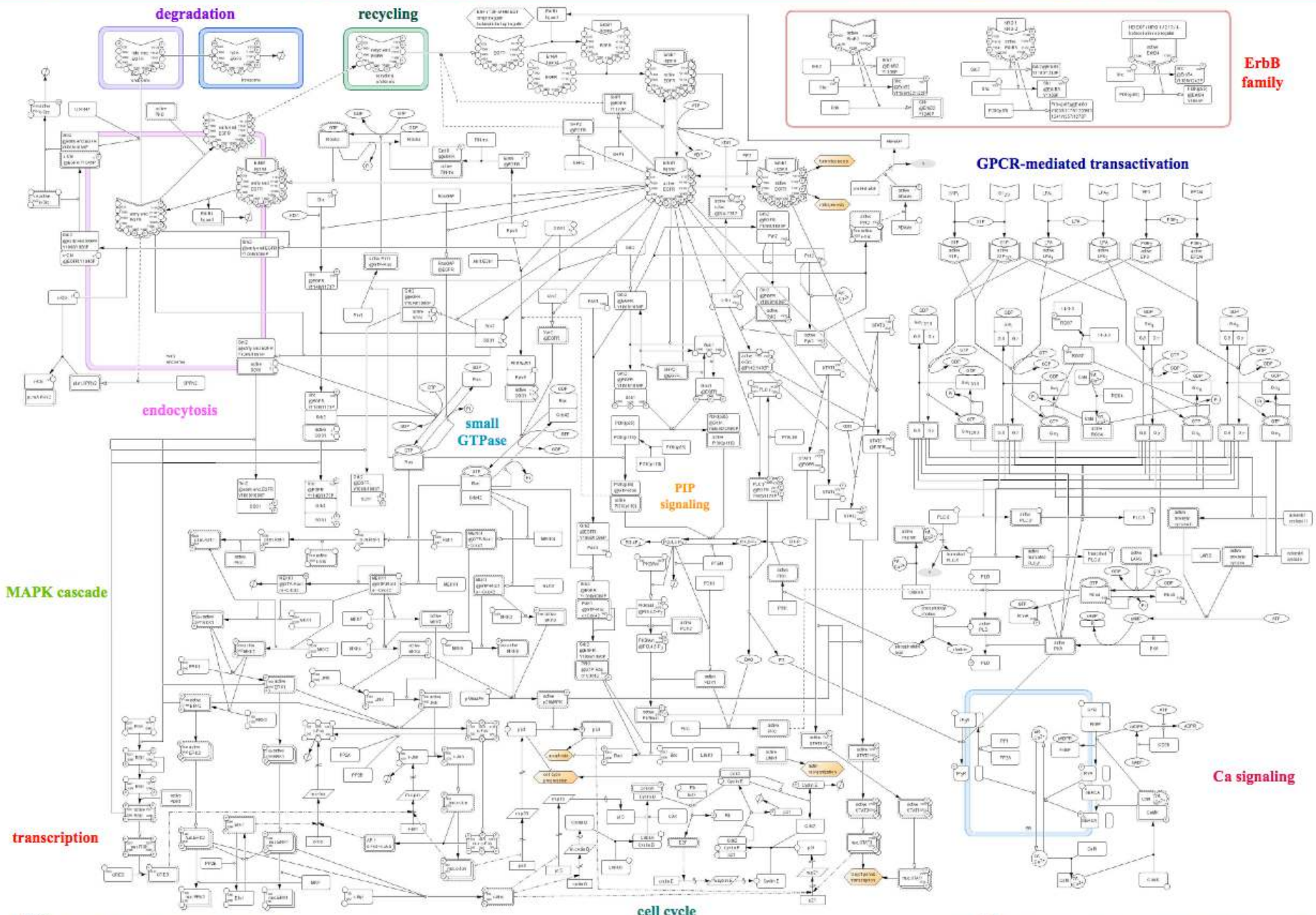


6 weeks: Metastases regressed

8 months: No regrowth

Epidermal Growth Factor Receptor Pathway Map

Kanae Oda (1,2,3), Yukiko Matsuoka (4), Hiroaki Kitano (1,2,5)
(1) The Systems Biology Institute, (2) Department of Fundamental Science and Technology, Keio University, (3) Department Molecular Epidemiology, Medical Research Institute, Tokyo Medical and Dental University, (4) ERATO-SORST Micro Symbiotic Systems Project, Japan Science and Technology Agency, (5) Sony Computer Science Laboratories, Inc.



Concept: escape pathways will cluster near incapacitated targets

Synthetic lethality:
a binary event leading to network failure

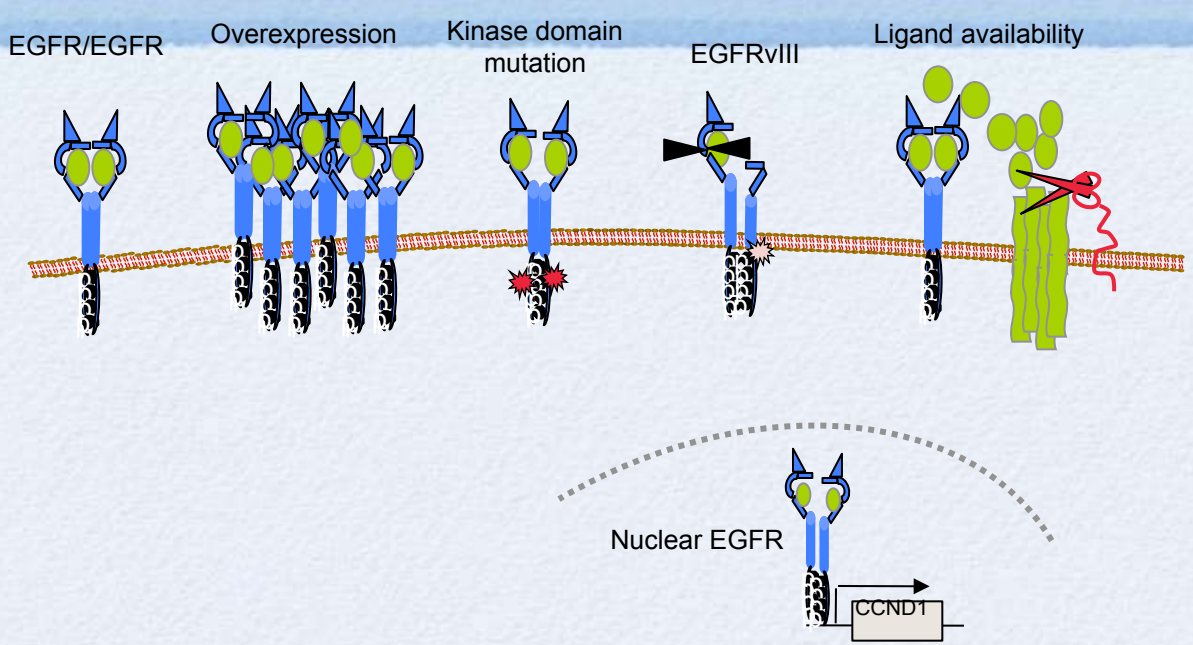
Motivating hypothesis for our work:
if we can effectively identify targets likely to yield synthetic lethality, we will reduce cost and improve outcomes in combination therapy

"Large-scale mapping of genetic interactions among nonessential yeast genes showed that **synthetic interactions are highly biased toward genes that have related functions**". Tong et al, Science 303, 808–813, 2004.

"Perhaps the most notable property of the essential genetic network (based on synthetic lethal analysis) is its density…our results indicate that **essential genes are highly connected hubs on the genetic interaction network**". Daverwala et al, Nat Genet 37:1147, 2005

"Thus, **genetic profiles of members of PPI pairs tend to correlate** better, not only to their interaction partners within the same species, but also to the orthologs of their interaction partner in an evolutionarily distant organism". Roguev et al Science 322:405, 2008

our own success story:

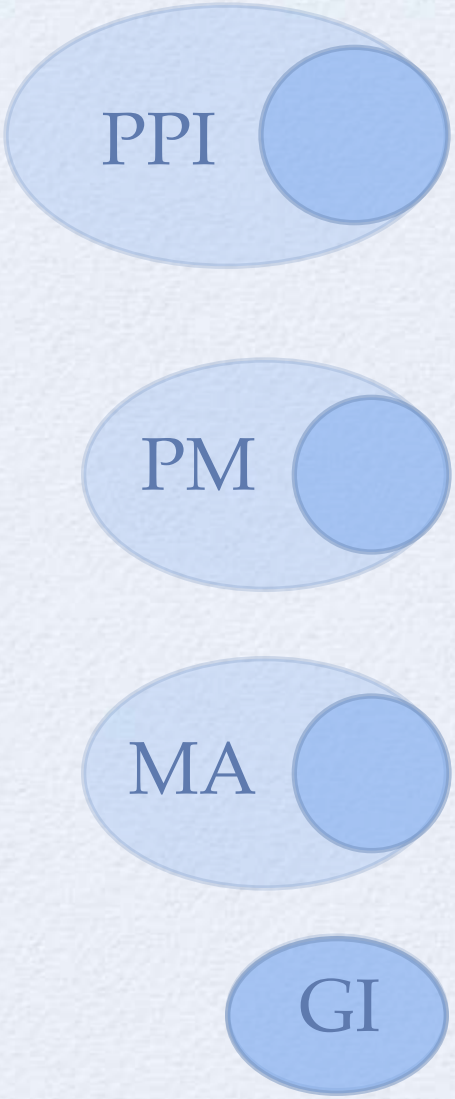


Antibodies
cetuximab
panitumumab

Small molecule
erlotinib

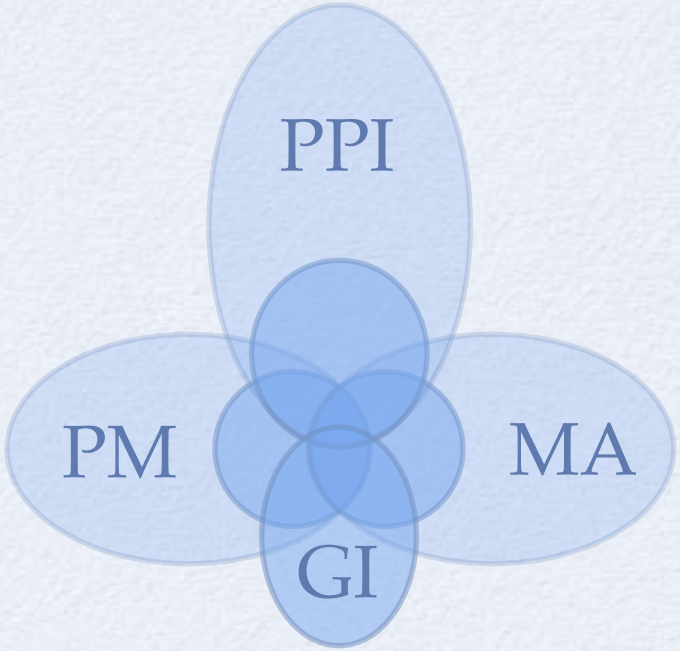
- Hypothesis: If we develop a network of proteins defined by close physical or functional linkage to a therapeutic target of interest, this network should be rich in proteins that regulate response to therapeutic inhibition of the target
- Such "focused" library would be optimal for screening in search for sensitizer genes

Building focused library

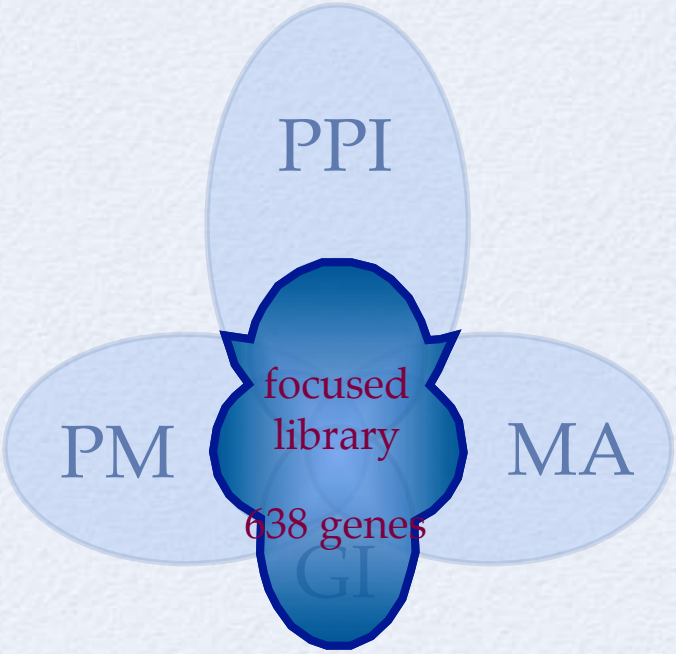


- Protein-protein interactions
- Pathway maps
- Microarray Data
- Genetic interactions in model orgainsms

Building focused library

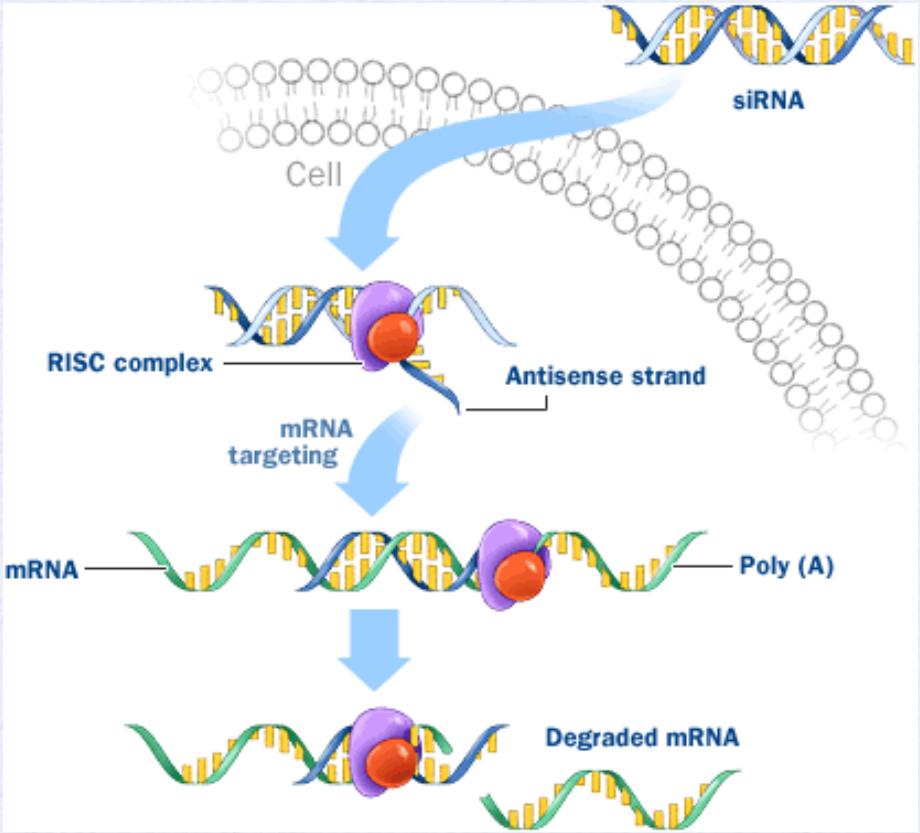


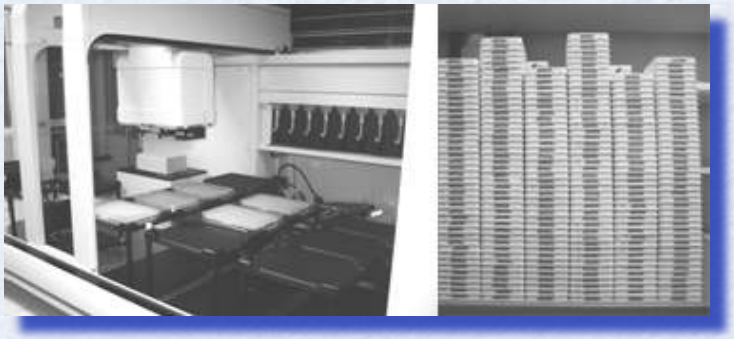
Building focused library

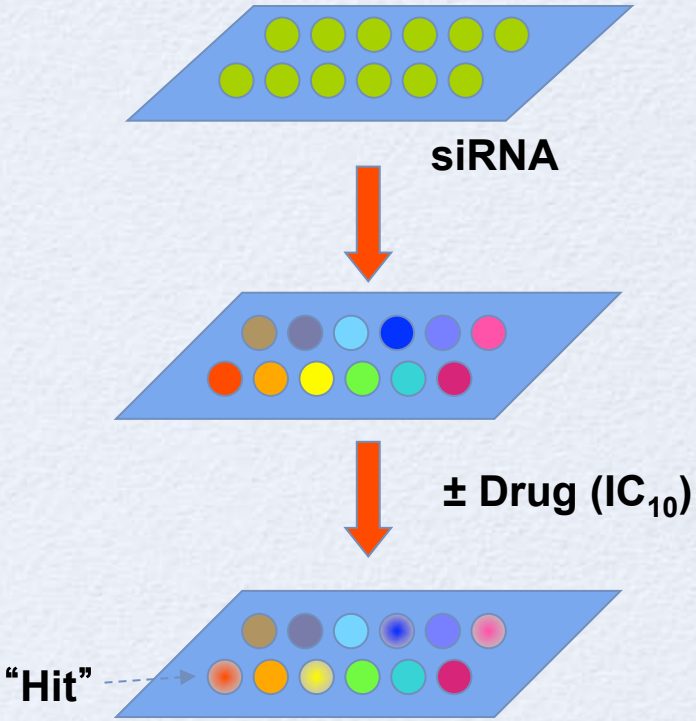


Combining the “cores” plus the intersections of less reliable sources

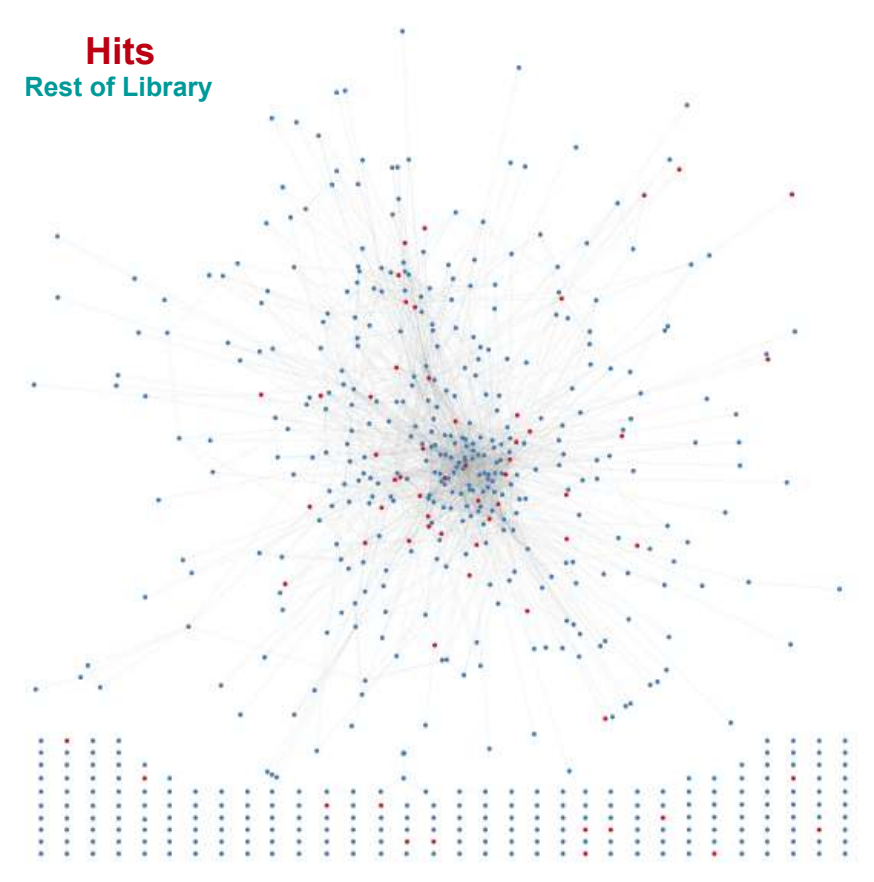
Used siRNA to screen the library in search for synergistic targets

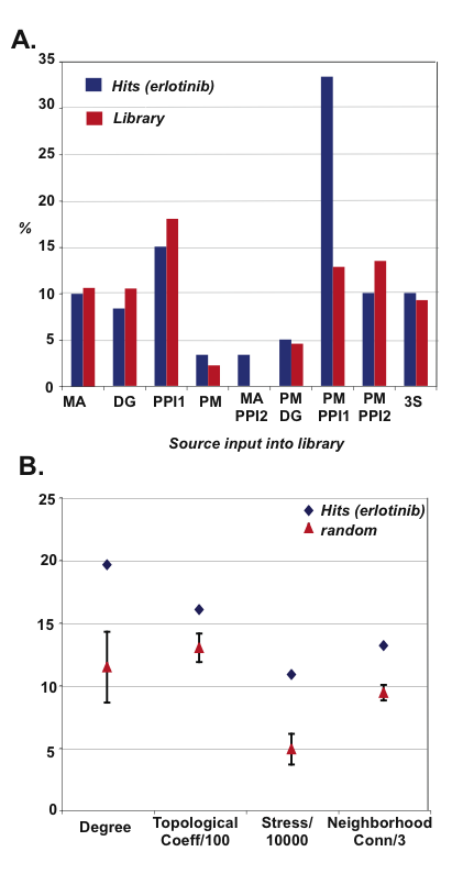




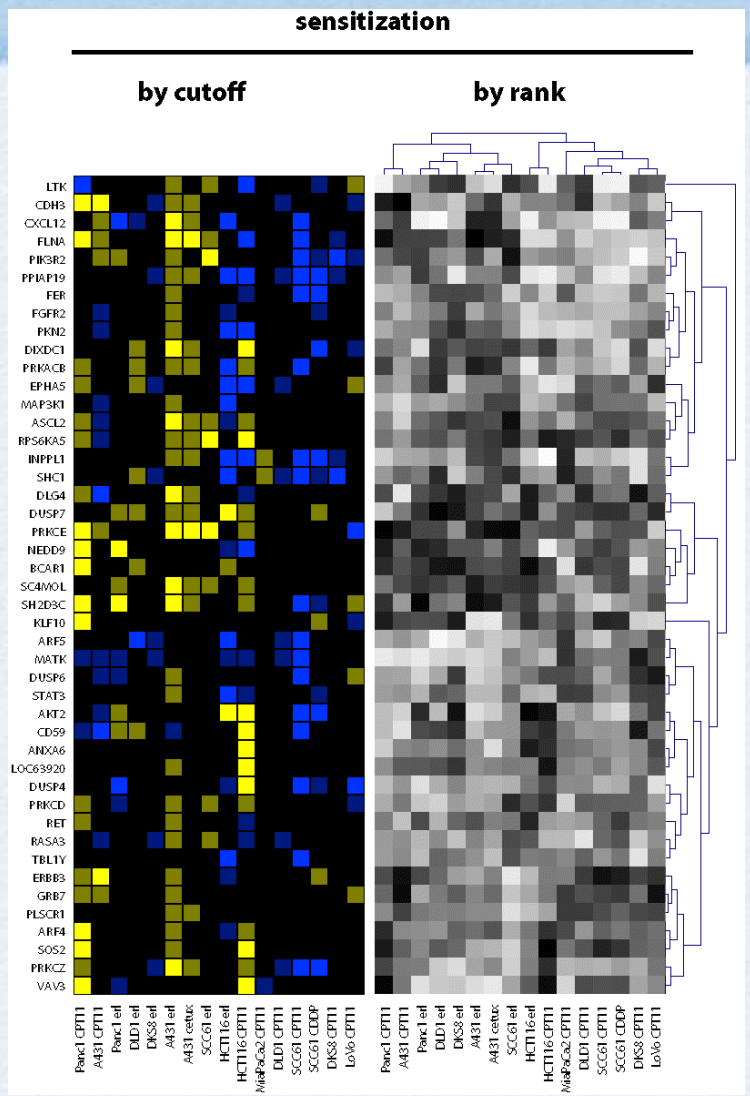


Results of screening library: erlotinib-sensitizing hits concentrate among first order PPIs of EGFR "core"





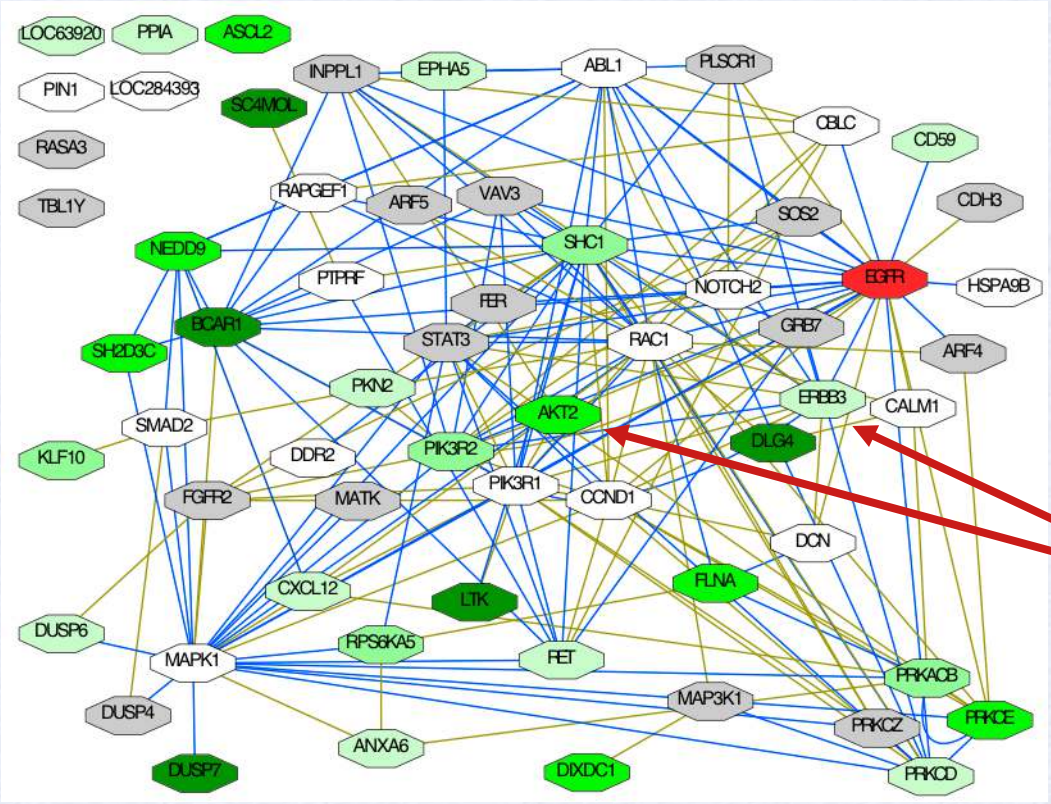
Hit activity profiles



--After repeated re-screening in different cell lines and with different drugs, we developed a list of hits with varying desirable properties, including some:

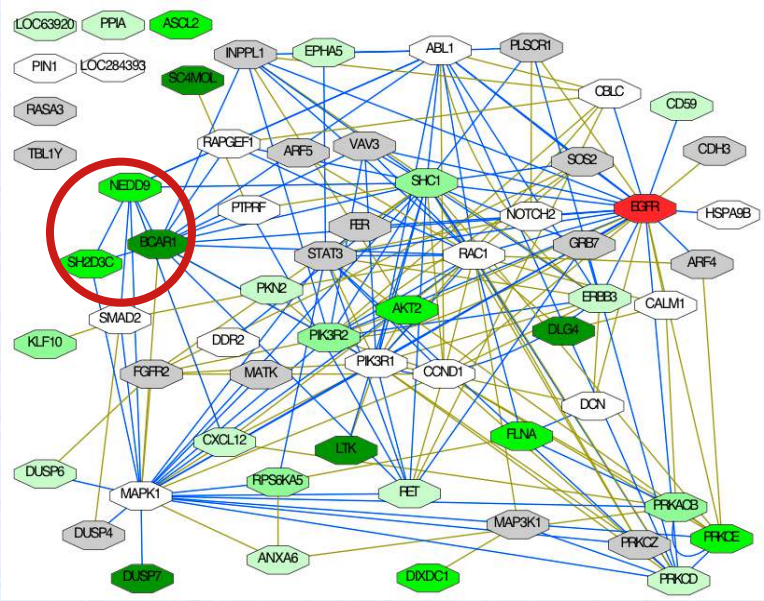
1. specific for EGFR-targeting agents
2. active in multiple cell lines
3. selectively inducing apoptosis (not cytostasis) in a drug-dependent manner

Hit network



Includes targets of drug development…

Focus on BCAR1-SH2D3C-NEDD9 cluster



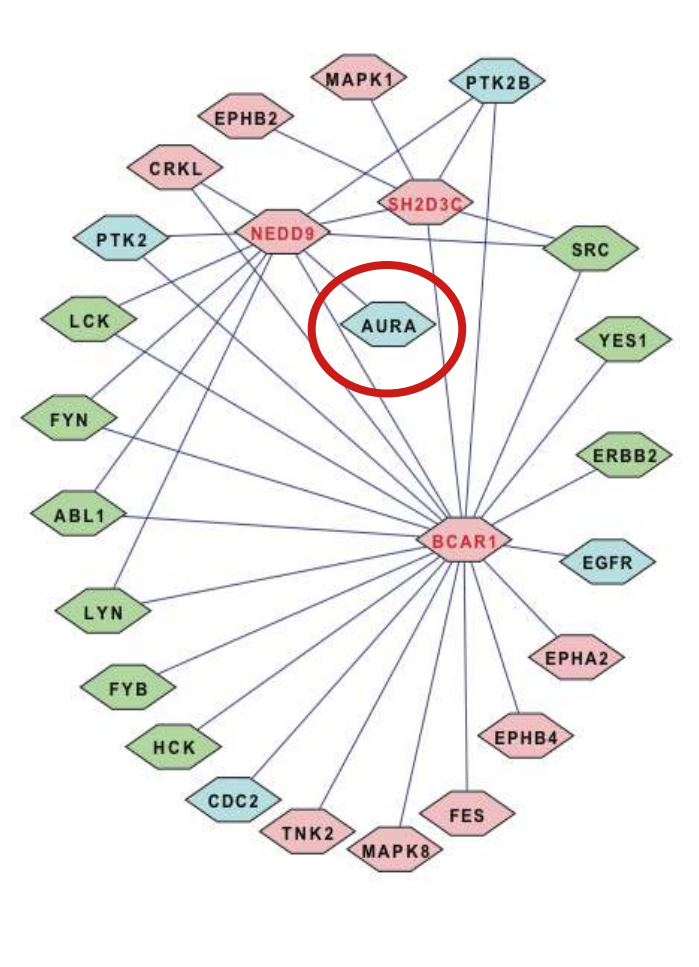
Relevant biology:

Scaffolding proteins with numerous partners

Interesting, cancer-relevant biology

BUT: not catalytic, not obviously druggable

Extrapolate to Aurora-A



Hypothesis

catalytic targets closely linked to this cluster are well positioned for synthetic lethality

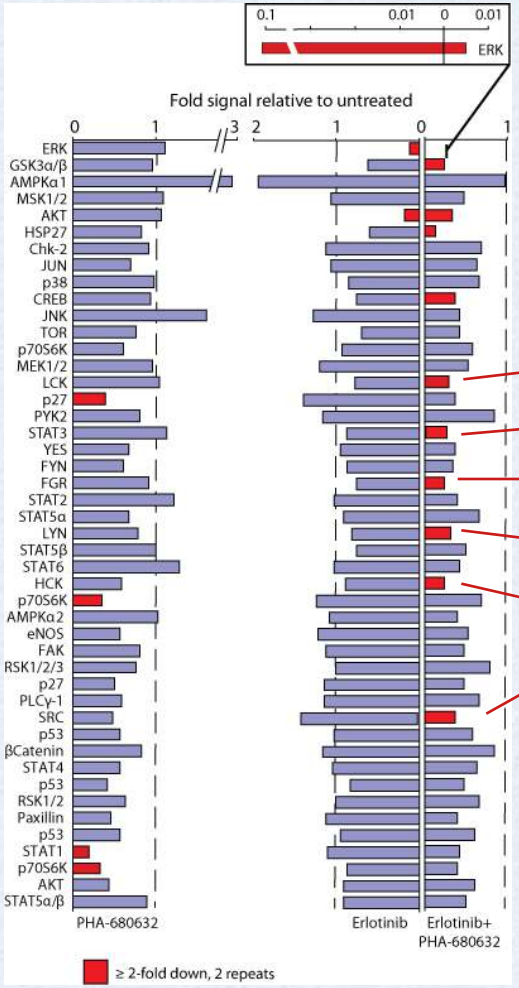
Observation

Aurora-A and a cluster of Src-family kinases are closely linked to NEDD9, SH2D3C, BCAR1

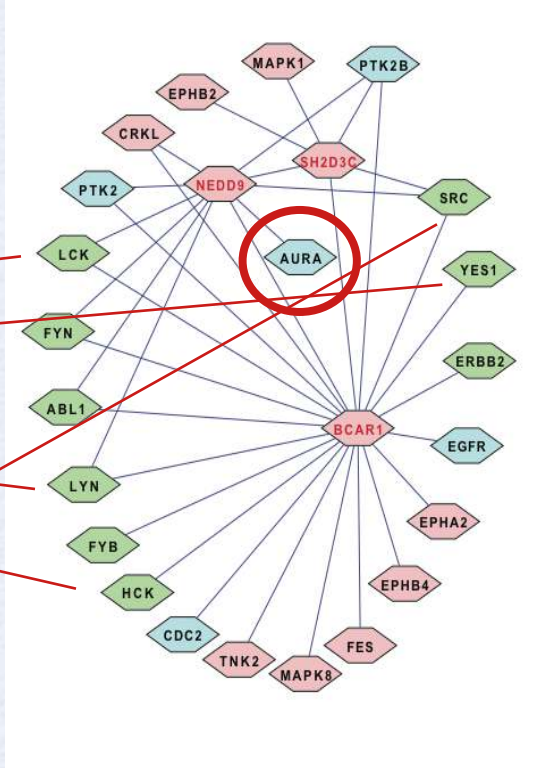
Test:

combine existing inhibitors of Aurora-A with those of EGFR (as surrogate for targeting NEDD9, SH2D3C, BCAR1…)

Emphasizing connectivity, dual inhibition of Aurora-A and EGFR inhibits cluster of Src family kinases



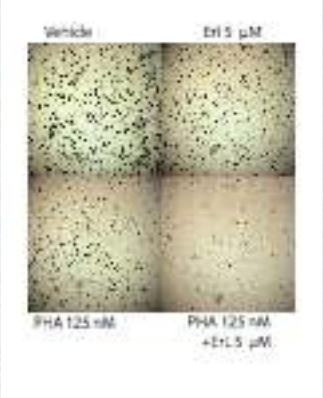
(SRC)



Already evidence that dual targeting of Src and EGFR is beneficial:
Epidermal growth factor receptor cooperates with Src family kinases in acquired resistance to cetuximab. Wheeler et al, Cancer Biol Ther 2009, etc…

Synergy, EGFR-Aurora-A inhibition

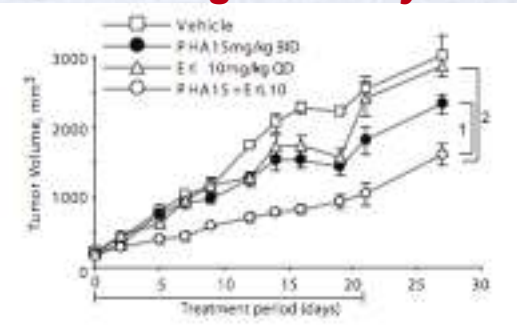
also in soft agar



Chou-Talalay analysis: multiple combinations of EGFR and AurA inhibitors result in synergy

Cell line	Inhibitors		Molar Ratio	Coefficient of Interaction (average ± std. dev.) ED50	ED75	ED90
HCT116	erlotinib	PHA-680632	1:50	0.80 ± 0.04	0.28 ± 0.18	0.14 ± 0.13
			1:100	0.58 ± 0.16	0.31 ± 0.12	0.21 ± 0.06
	cetuximab	PHA-680632	1:16	0.46 ± 0.27	0.62 ± 0.05	1.07 ± 0.78
			1:33	0.31 ± 0.10	0.45 ± 0.06	0.73 ± 0.39
	erlotinib	C1368	1:25	0.64 ± 0.12	0.71 ± 0.17	0.89 ± 0.56
	cetuximab	C1368	1:8	0.42 ± 0.18	0.31 ± 0.10	0.24 ± 0.06
A431	erlotinib	PHA-680632	1:5	0.46 ± 0.11	0.37 ± 0.17	0.34 ± 0.23

and xenograft analysis

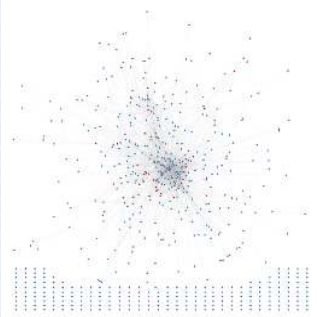


ClinicalTrials.gov: NCT01471964
A Phase I/II Trial to Assess Safety and Tolerability of MLN8237, In Combination With Erlotinib to Treat Non-Small Cell Lung Cancer

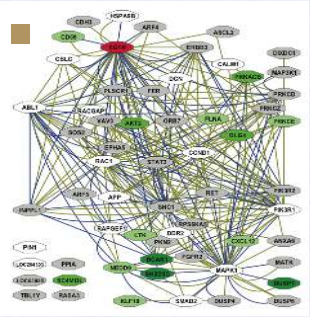
ClinicalTrials.gov: NCT01540682
Phase I Study of MLN8237 in Combination With Cetuximab and Definitive Radiation in Patients With Locoregionally Advanced Squamous Cell Carcinoma of the Head and Neck

Summary

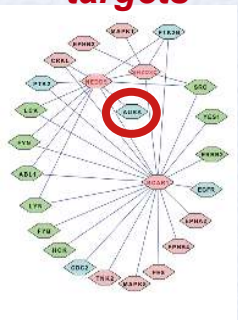
Network



Sub-network structure



Re-expanded towards drug targets



Biomarkers

Biology

3 Way Synergy