

Борисов Николай Михайлович

доктор технических наук
Chief Scientist, ООО ПОНКЦ и Pathway Pharmaceuticals, Ltd
Chief Biomathematician, Insilico Medicine, Inc
В.н.с. НИЦ «Курчатовский институт»

Математические основы системно- биоинформатической платформы OncoFinder



ВАЖНЕЙШАЯ ПРОБЛЕМА ОНКОЛОГИИ

Нельзя лечить всех пациентов по единой схеме



Одна из главных проблем онкологии - индивидуализация подхода к терапии и выбор оптимального для каждого пациента курса лечения



При одинаковой картине появления и развития заболевания, индивидуальная реакция пациентов на проводимые курсы лечения может различаться диаметрально противоположно

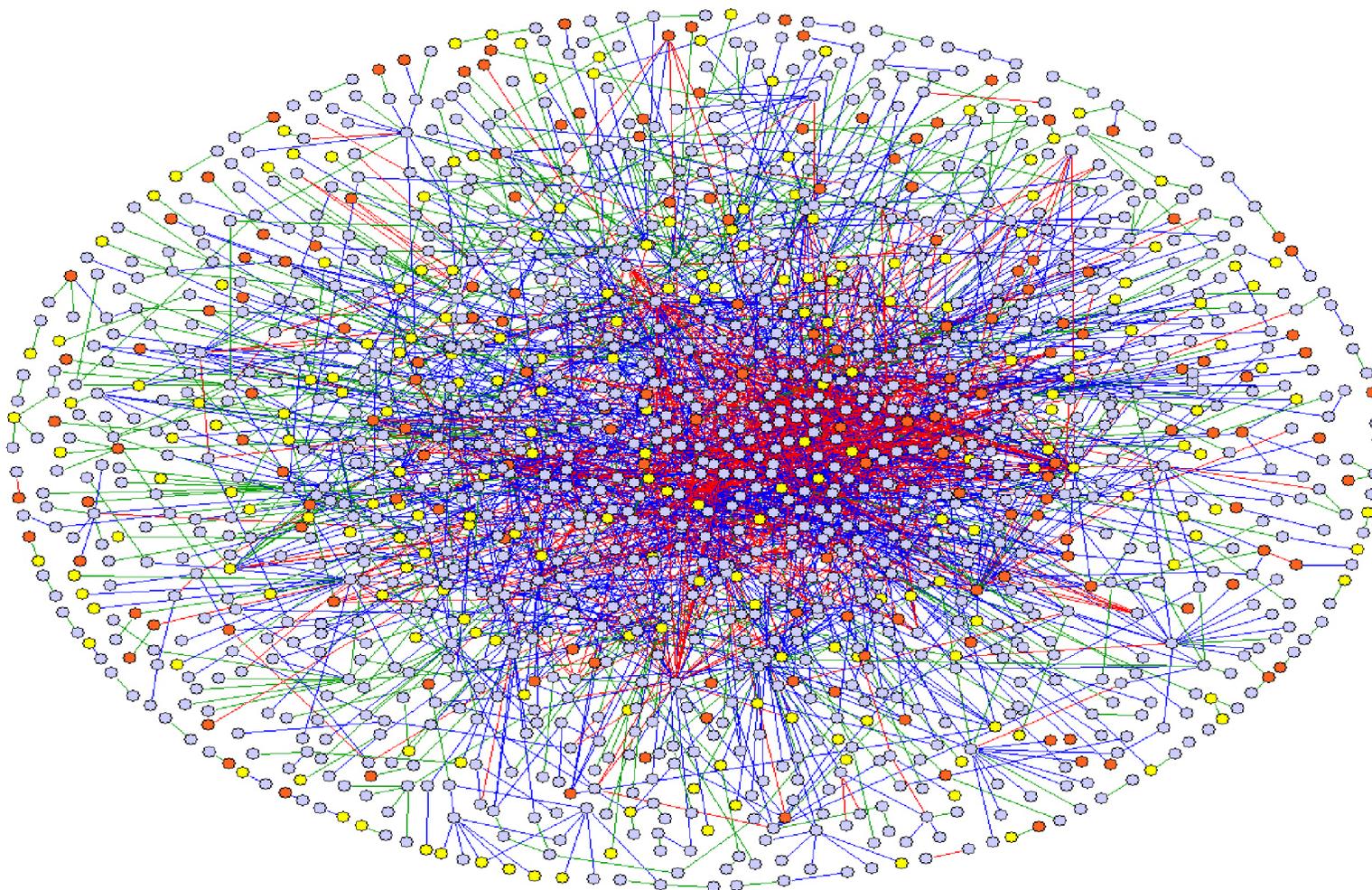


Если лечить всех пациентов по единой схеме – около трети всех больных обрекаются на смерть только из-за упущенного времени и побочных токсических эффектов

Причины развития рака во многом индивидуальны

Поэтому подходы к лечению обязательно должны учитывать индивидуальные особенности возникновения и протекания болезни

ИНТЕРАКТОМ ЧЕЛОВЕКА



Таргетные противоопухолевые препараты: восторг биологов и фармацевтов и скептицизм клиницистов

Жуков Н.В., Тюляндин С.А. Биохимия, 2008, т. 73, №5, с. 751-768

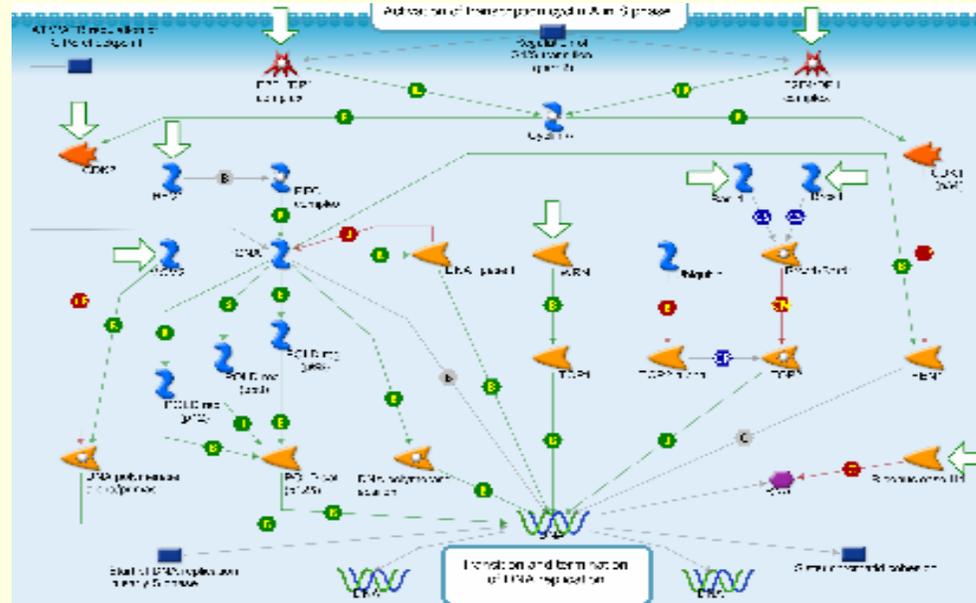
Надежды на мабы и нибы, представлявшие «магическими пулями», не оправдались. Клиническая эффективность этих методов терапии очень часто оказывается низкой.

Причина такого положения вполне понята с точки зрения системной биологии клеточных сигналов. Если в ходе рутинных ИГХ-исследований определяют уровень экспрессии считанных онкогенов, то в настоящий момент хорошо исследованы десятки про- и антимитотических путей, каждый из которых содержит десятки, если не сотни, белков и генов-участников



Клиническую эффективность препарата можно предсказать, оценивая его способность компенсировать патологические изменения (SD) во всей совокупности сигнальных путей – сигналоме

$$SD = \frac{\prod_{i=1}^N [AGEL]_i}{\prod_{j=1}^M [RGEL]_j}$$



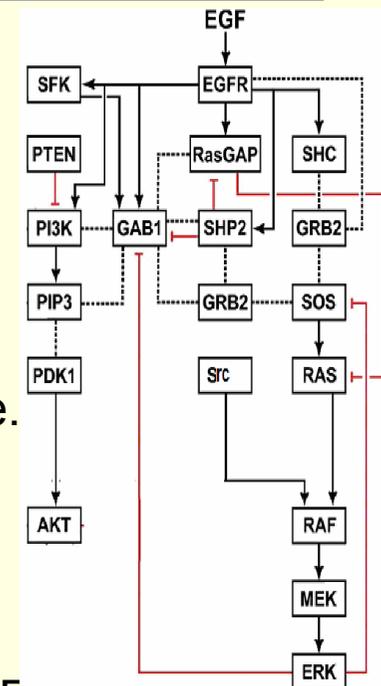
SD – signaling disturbance

$[AGEL]_i$ – activator gene # i expression level

$[RGEL]_j$ – repressor gene # j expression level

Два приближения, стоящие в основе системы OncoFinder (Buzdin et al, 2014)

1. Учет только общие функции каждого генного продукта. В этом случае граф белок-белкового взаимодействия в каждом сигнальном пути полагают в виде двух параллельных цепочек событий: одна приводит к активации, а другая – к ингибированию сигнального пути.
2. Уровень экспрессии каждого из белков – переносчиков сигнала в пути представляется в состоянии покоя полагают много меньшим, чем с состоянии активации (т.е. в состоянии покоя имеет место глубокононасыщенное состояние каждого из белков-переносчиков сигнала) – (Kiyatkin et al, 2006, Borisov et el, 2009, Kuzmina and Borisov, 2011)



Проверка допущения (1) для кинетической модели пути EGFR (Kuzmina and Borisov, 2011) для девяти транскриптомов глиосаркомы (Griesinger, 2013)

$$w_j^{(1)} = \frac{1}{T} \int_0^T \frac{\partial \ln[EFF(t)]}{\partial \ln C_j^{tot}} dt \quad SD = \frac{\sum_{i=1}^N [AGEL]_i}{\sum_{j=1}^M [RGEL]_j} \quad w_j^{(1)} = \frac{1}{T} \int_0^T \frac{\partial \ln[EFF(t)]}{\partial \ln C_j^{tot}} dt \quad H_{ij} = \frac{\partial^2}{\partial C_i^{tot} \partial C_j^{tot}} \sum_k \frac{([EFF(C^{tot}, t_k)] - [EFF]_k^{exp})^2}{\sigma_k^2}$$

$$\frac{PAS_{EGFR}^{(1)}}{PAS_{EGFR}} = 0,7 \pm 0,3 \quad \frac{PAS_{EGFR}^{(2)}}{PAS_{EGFR}} = 1,0 \pm 0,6$$

Оценка патологических изменений в сигнальных путях у индивидуального больного

Pathway activation score

$$PAS_p = \sum_n \zeta_{ARR_{np} \square BTIF_n \square g} (CNR_n)$$

n – номер белка

Cancer(case)-normal ratio CNR_n есть отношение уровней экспрессии гена, кодирующего белок n , к норме (среднему значению у контрольной группы),

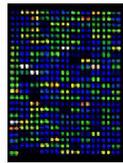
BTIF: beyond tolerance interval flag

Activator/repressor role:

$$H \xi_m = \lambda_m \square \xi_m$$

Концепция работы системы

Анализ транскриптома



1 Microarray Per Well
24 Microarrays Per Slide
96 Microarrays Total



Повышенно и
пониженно
экспрессиро-
ванные гены

БД
путей

БД методов
терапии

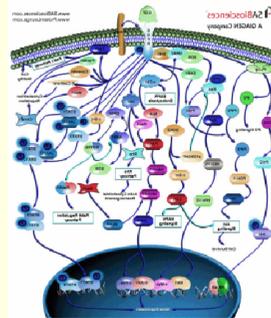
Концепция работы системы

Определение
патологически
экспрессиро-
ванных генов

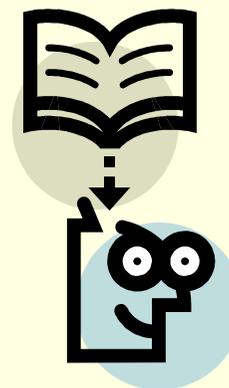
Повышенно и
пониженно
экспрессиро-
ванные гены

Степень
патологических
изменений в
путях
(PAS)

БД методов
терапии



Концепция работы системы



Список рекомендуемых методов терапии



Концепция работы системы

Анализ
транскриптома

Определение
патологически
экспрессиро-
ванных генов

Повышенно и
пониженно
экспрессиро-
ванные гены

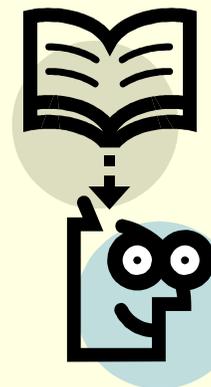
БД
путей

Степень
патологических
изменений в
путях (PAS)

БД методов
терапии

Список
рекомендуемых
методов терапии

-
-
-
-



Программа OncoFinder

От крупномасштабных данных по генной экспрессии –
к подбору лекарства

The screenshot shows the OncoFinder web application interface. The browser address bar displays the URL: `ec2-54-91-51-249.compute-1.amazonaws.com/user/nicolasborissoff/`. The application header includes the OncoFinder logo and navigation links for 'Biochem database', 'Drugs database', and 'Medical DB'. The user profile for 'nicolasborissoff' is visible, featuring a profile picture and an 'Edit profile' button. Below the profile, a 'Users' list includes: admin, Mikhail, aliper, nadterekhanova, ksenia, and baranovamarila. The 'Projects' section, titled 'Projects' with a '+ Create new project' button, lists three projects:

- canada_renal**: Type: Scientific, Created at: Nov. 11, 2014, 9:56 p.m. (20 items)
- gladysheva**: Type: Medical, Created at: Nov. 10, 2014, 7:29 p.m. (4 items)
- moscow_canada**: Type: Scientific, Created at: May 30, 2014, 4:45 p.m. (8 items)

The browser's taskbar at the bottom shows the Windows Start button, several application icons, and the system tray with the date '07.12.2014' and time '22:06'.

Программа OncoFinder: БД путей

The screenshot shows a web browser window displaying the OncoFinder Biochem DataBase interface. The browser's address bar shows the URL `ec2-54-91-51-249.compute-1.amazonaws.com/db/pathways/`. The page features a navigation bar with icons for 'Biochem database', 'Drugs database', and 'Medical DB'. Below this, the 'Biochem DataBase' section is active, with tabs for 'Human Pathways', 'Mouse Pathways', and 'Metabolism Pathways'. A search bar is present with the text 'Pathway name' and a 'Go!' button. The main content area displays a list of pathways, each with a small circular diagram icon and associated statistics:

- 0_IL_6_Pathway_IKBs_Degradation**
AMCF: 0.0
Genes: 13
Nodes: 4
- AHR_Pathway**
AMCF: 1.0
Genes: 27
Nodes: 19
- AHR_Pathway_AHR_Degradation**

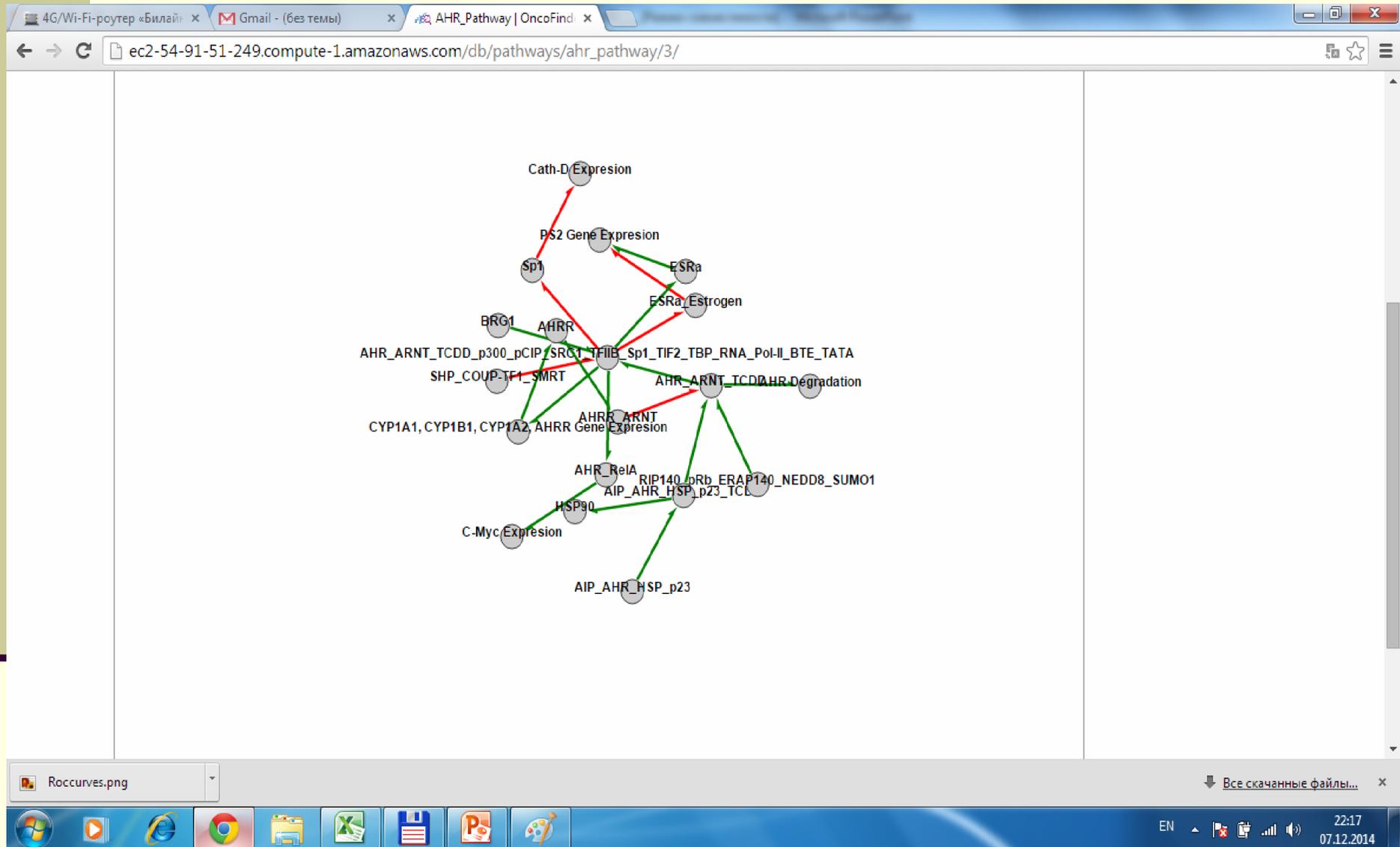
The Windows taskbar at the bottom shows the system tray with the date '07.12.2014' and time '22:08'. A file explorer window titled 'Roccurves.png' is also visible.

Программа OncoFinder: БД генов

The screenshot shows a web browser window with the OncoFinder application. The address bar shows the URL: `ec2-54-91-51-249.compute-1.amazonaws.com/db/pathways/0_il_6_pathway_ikbs_degradation/302/`. The page title is "0_IL_6_Pathway_IKBs_Degradation" and the database is identified as "Human Pathways | AMCF = 0.0". There are navigation tabs for "Genes", "Nodes and Components", and "Graphic Scheme". A table lists 13 genes with their symbols and associated ARR values.

#	SYMBOL	ARR
1	AKT1	1.0
2	CHUK	1.0
3	IKBKB	1.0
4	IKBKE	1.0
5	IKBKG	1.0
6	NFKB1	1.0
7	NFKB2	1.0
8	NFKBIA	1.0
		1.0

Программа OncoFinder: графы путей



Программа OncoFinder: БД лекарств

4G/Wi-Fi-роутер «Билайн» x Gmail - (без темы) x Drugs database | OncoFinder x

ec2-54-91-51-249.compute-1.amazonaws.com/db/drugs/

ONCO FINDER

Biochem database Drugs database Medical DB

чувси

Drugs DataBase

Total amount = 195 drugs

Drug name Go!

Abemaciclib (LY2835219)

- o DataBase: genego
- o Type: inhibitor
- o Targets: CDK4, CDK6

Afatinib

- o DataBase: genego
- o Type: inhibitor
- o Targets: EGFR, HER2

Aflibercept

Microsoft Excel - Статья 2

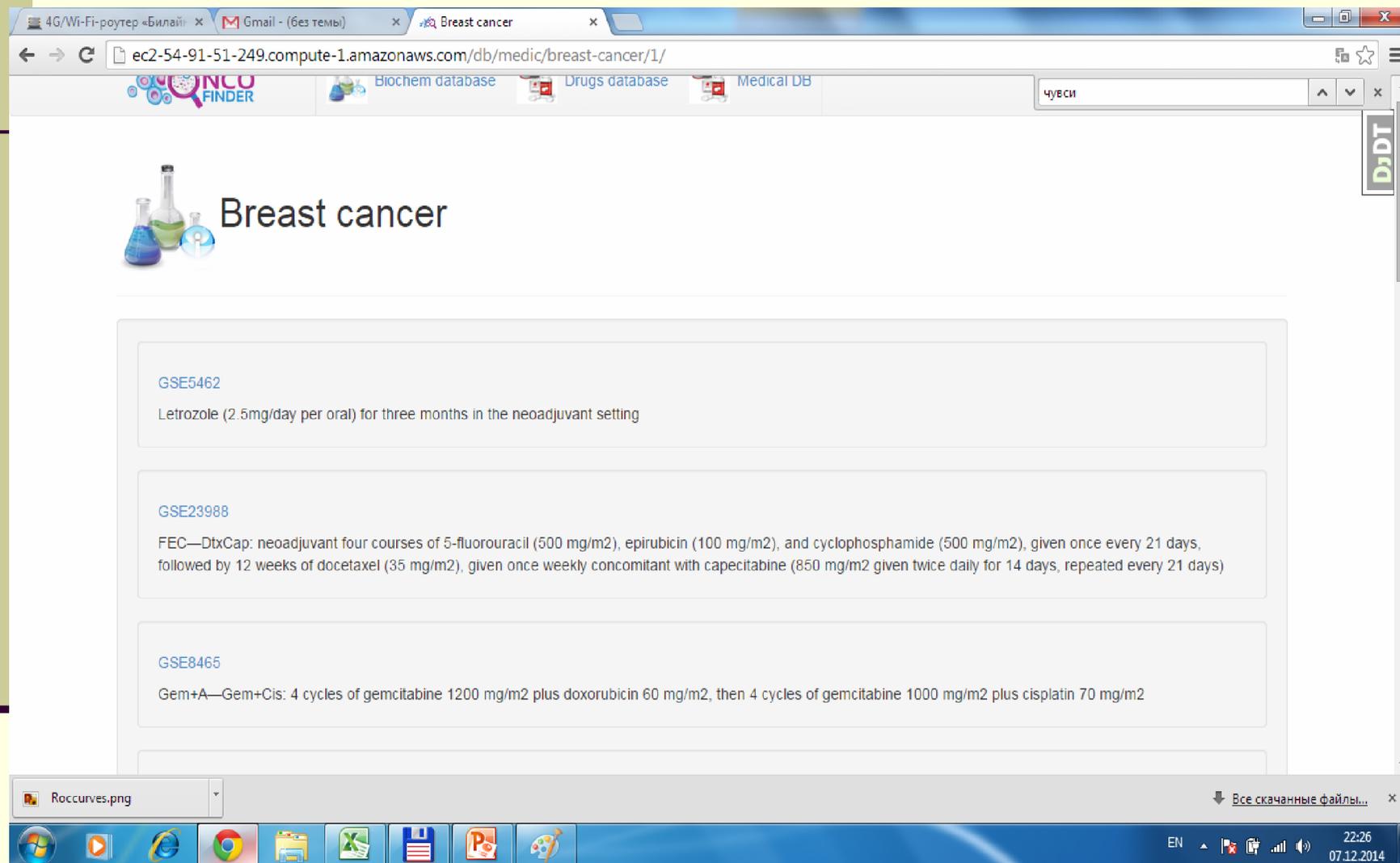
Drug Name	Target	Type	GenBank	UniProt	PubMed	Chemical Structure
Abemaciclib	CDK4, CDK6	Inhibitor	U01103	P05512, P05513	25481111	CC1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3C4=CC=CC=C4C5=CC=CC=C5C6=CC=CC=C6C7=CC=CC=C7C8=CC=CC=C8C9=CC=CC=C9C10=CC=CC=C10C11=CC=CC=C11C12=CC=CC=C12C13=CC=CC=C13C14=CC=CC=C14C15=CC=CC=C15C16=CC=CC=C16C17=CC=CC=C17C18=CC=CC=C18C19=CC=CC=C19C20=CC=CC=C20C21=CC=CC=C21C22=CC=CC=C22C23=CC=CC=C23C24=CC=CC=C24C25=CC=CC=C25C26=CC=CC=C26C27=CC=CC=C27C28=CC=CC=C28C29=CC=CC=C29C30=CC=CC=C30C31=CC=CC=C31C32=CC=CC=C32C33=CC=CC=C33C34=CC=CC=C34C35=CC=CC=C35C36=CC=CC=C36C37=CC=CC=C37C38=CC=CC=C38C39=CC=CC=C39C40=CC=CC=C40C41=CC=CC=C41C42=CC=CC=C42C43=CC=CC=C43C44=CC=CC=C44C45=CC=CC=C45C46=CC=CC=C46C47=CC=CC=C47C48=CC=CC=C48C49=CC=CC=C49C50=CC=CC=C50C51=CC=CC=C51C52=CC=CC=C52C53=CC=CC=C53C54=CC=CC=C54C55=CC=CC=C55C56=CC=CC=C56C57=CC=CC=C57C58=CC=CC=C58C59=CC=CC=C59C60=CC=CC=C60C61=CC=CC=C61C62=CC=CC=C62C63=CC=CC=C63C64=CC=CC=C64C65=CC=CC=C65C66=CC=CC=C66C67=CC=CC=C67C68=CC=CC=C68C69=CC=CC=C69C70=CC=CC=C70C71=CC=CC=C71C72=CC=CC=C72C73=CC=CC=C73C74=CC=CC=C74C75=CC=CC=C75C76=CC=CC=C76C77=CC=CC=C77C78=CC=CC=C78C79=CC=CC=C79C80=CC=CC=C80C81=CC=CC=C81C82=CC=CC=C82C83=CC=CC=C83C84=CC=CC=C84C85=CC=CC=C85C86=CC=CC=C86C87=CC=CC=C87C88=CC=CC=C88C89=CC=CC=C89C90=CC=CC=C90C91=CC=CC=C91C92=CC=CC=C92C93=CC=CC=C93C94=CC=CC=C94C95=CC=CC=C95C96=CC=CC=C96C97=CC=CC=C97C98=CC=CC=C98C99=CC=CC=C99C100=CC=CC=C100

Рoccurves.png

Все скачанные файлы...

EN 22:20 07.12.2014

Программа OncoFinder: БД схем лечения

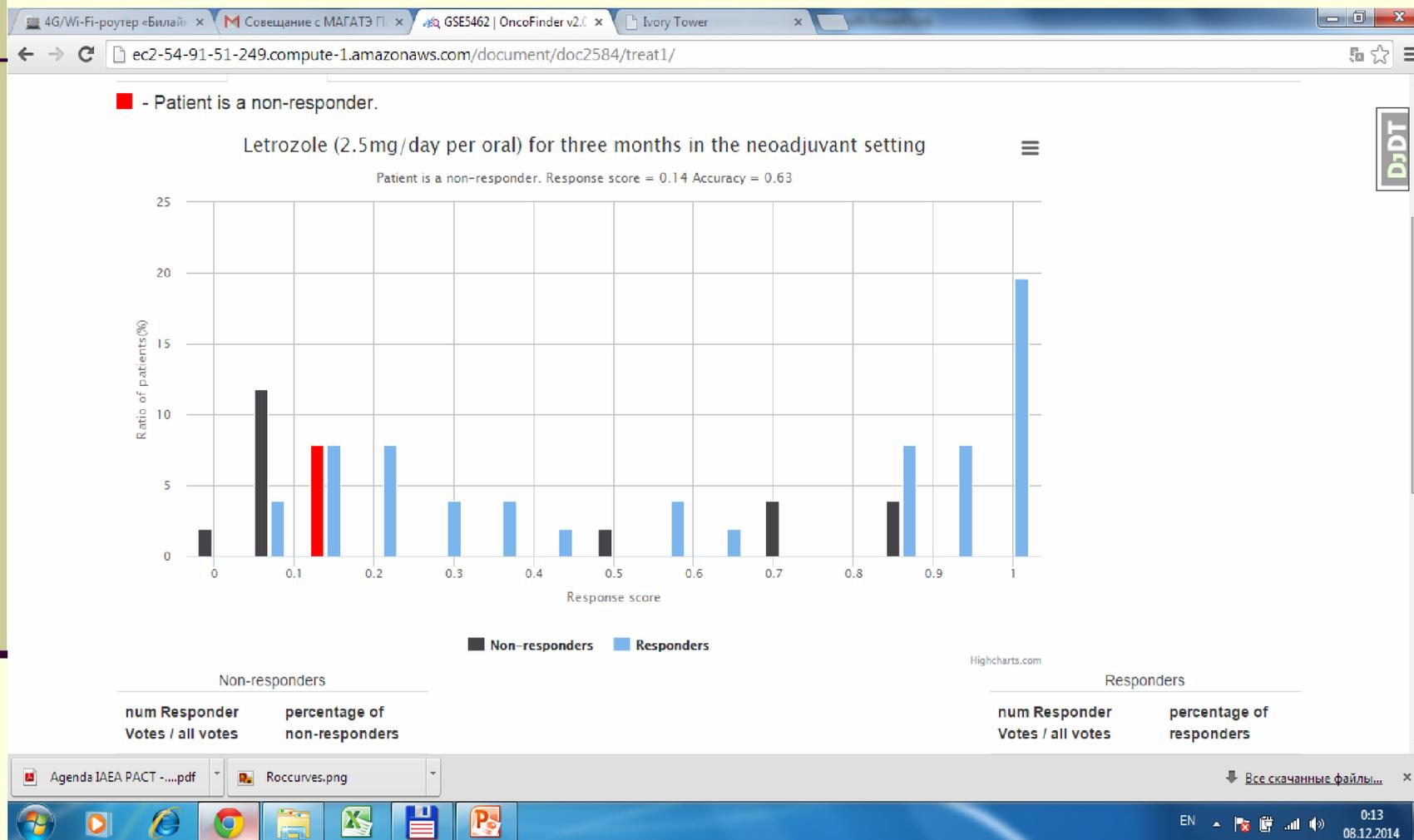


The screenshot displays a web browser window with the OncoFinder application. The browser's address bar shows the URL `ec2-54-91-51-249.compute-1.amazonaws.com/db/medic/breast-cancer/1/`. The application header includes the OncoFinder logo and navigation links for 'Biochem database', 'Drugs database', and 'Medical DB'. The main content area is titled 'Breast cancer' and features a list of treatment regimens:

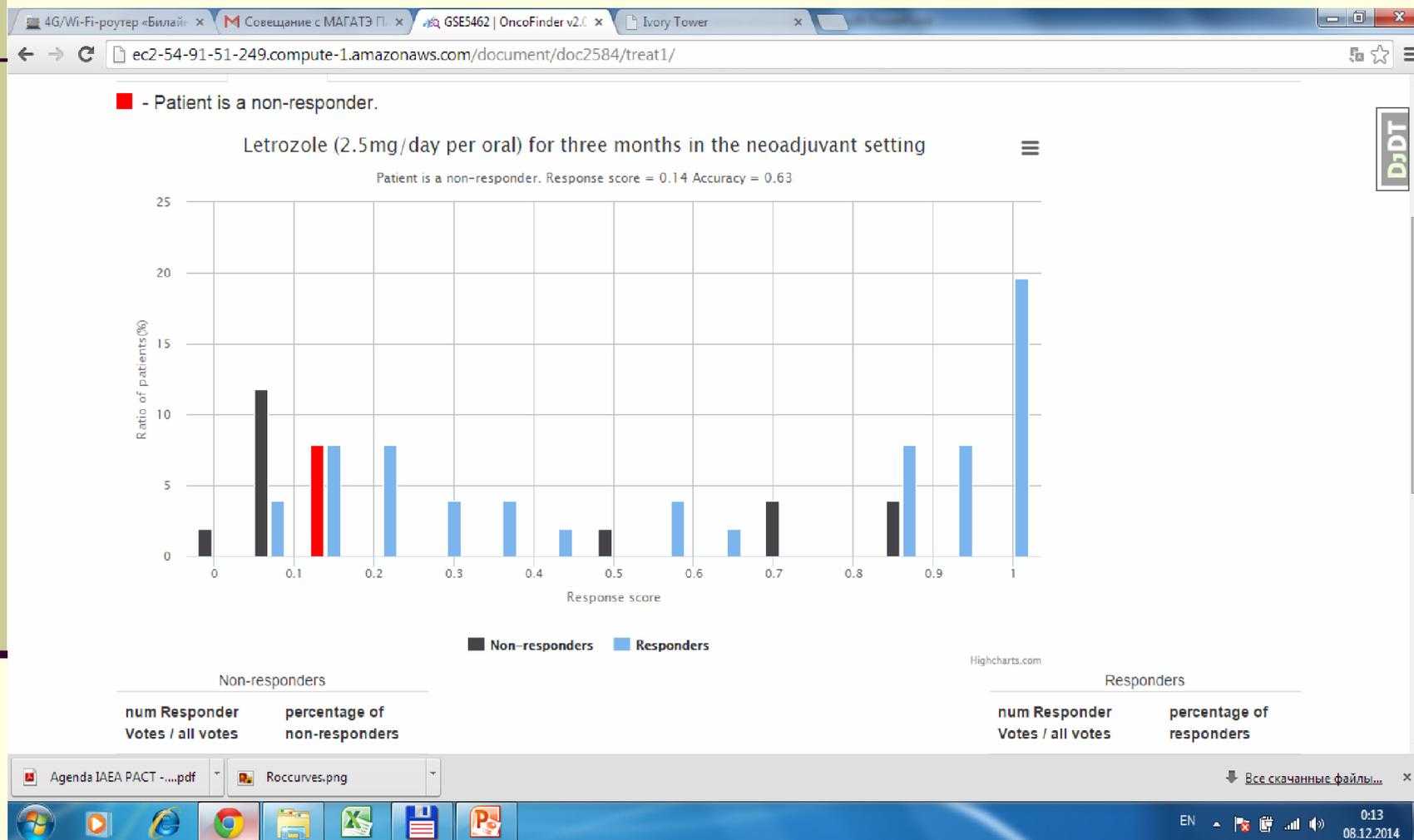
- GSE5462**
Letrozole (2.5mg/day per oral) for three months in the neoadjuvant setting
- GSE23988**
FEC—DtxCap: neoadjuvant four courses of 5-fluorouracil (500 mg/m²), epirubicin (100 mg/m²), and cyclophosphamide (500 mg/m²), given once every 21 days, followed by 12 weeks of docetaxel (35 mg/m²), given once weekly concomitant with capecitabine (850 mg/m² given twice daily for 14 days, repeated every 21 days)
- GSE8465**
Gem+A—Gem+Cis: 4 cycles of gemcitabine 1200 mg/m² plus doxorubicin 60 mg/m², then 4 cycles of gemcitabine 1000 mg/m² plus cisplatin 70 mg/m²

The Windows taskbar at the bottom shows the system tray with the date 07.12.2014 and time 22:26, along with various application icons.

Программа OncoFinder: представление результатов



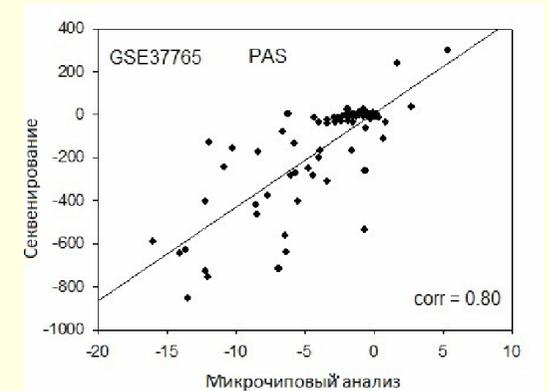
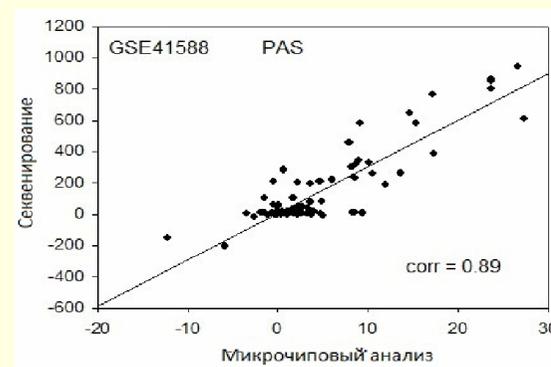
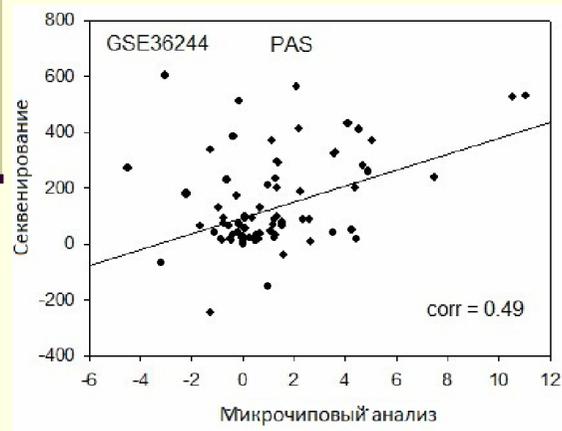
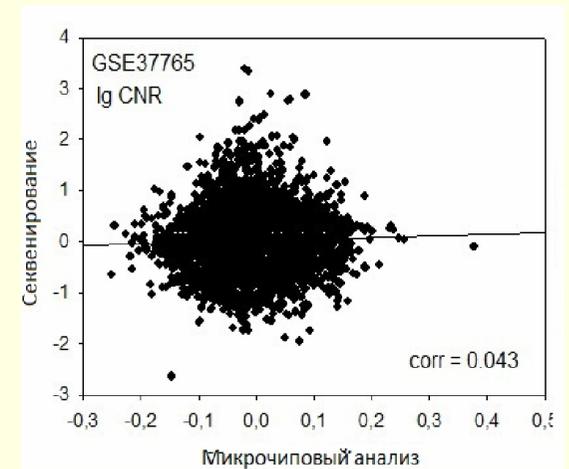
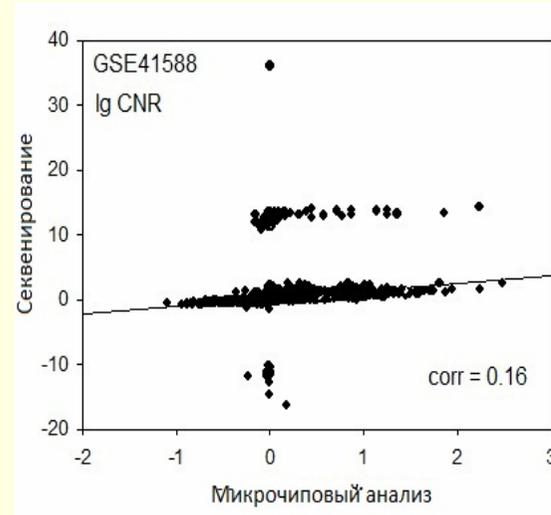
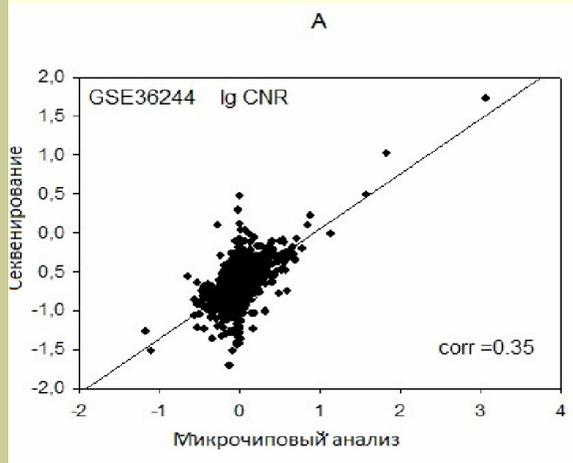
Программа OncoFinder: представление результатов



OncoFinder позволяет выделять сигнал на фоне шума при сравнении результатов исследования транскриптома различными методами: коэффициенты корреляции между значениями, полученными с помощью микрочиповой гибридизации РНК и РНК-секвенирования (Buzdin et al, 2014)

Образцы		Транскриптомный уровень (lg CNR для различных генов)	Сигналомный уровень (значения PAS для различных путей)
GSE36244, 24 ч после воздействия VaP	Образец 1	0,35	0,49
	Образец 2	0,10	0,47
	Среднее	0,22	0,49
GSE41588, 5 мкМ 5-аза-дезоксцитидина	Образец 1	0,16	0,89
	Образец 2	0,049	0,88
	Образец 3	0,047	0,80
	Среднее	0,082	0,87
	Образец 1	0,18	0,79
	Образец 3	0,098	0,75
	Образец 4	0,12	0,80
	Образец 5	-0,029	0,21
	Образец 8	0,043	0,80
	Среднее	0,068	0,77

Облака значений транскриптомного (верхний ряд) и сигнального (нижний ряд) профилей, полученных методом гибридизации на микрочипах (по оси абсцисс) и NGS (по оси ординат) (Buzdin et al, 2014).



Могут ли сигнальные пути использоваться как маркеры рака?

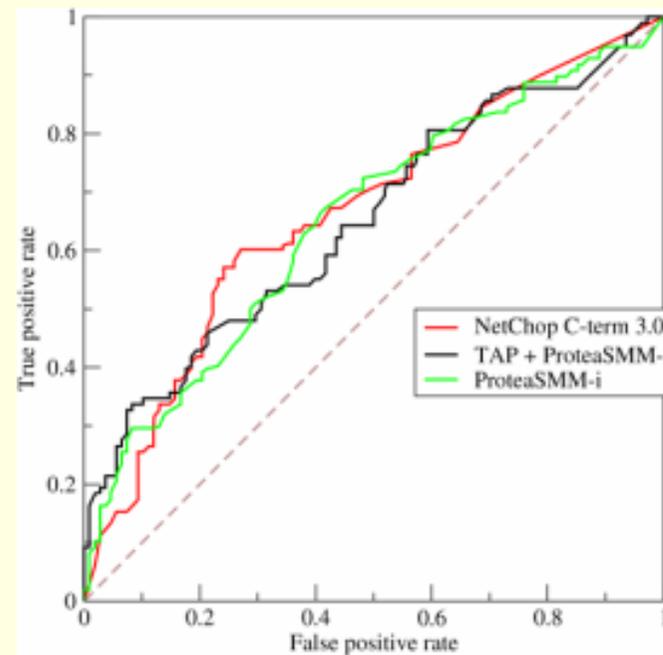
$$AUC = \frac{Sn + Sp}{2} = \frac{TPR + TNR}{2}$$

Area Under Curve:
площадь под кривой
ошибок

AUC > 0.7: хороший маркер

AUC > 0.95: прекрасный маркер

AUC = 1: идеальный маркер



На уровне сигнальных путей (AUC1) уровне хороших маркеров видов рака существенно больше, чем на уровне отдельных генов (AUC2): Borisov et al., 2014

Нозологическая форма	AUC1 >0.7 ^а	AUC2 >0.7 ^б	AUC1/2 >0.7; AUC1>AU C2 ^в	AUC1/2 >0.7; AUC2>AU C1 ^г	AUC1 >0.75 ^д	AUC2 >0.75 ^е	AUC1/2 >0.75; AUC1>AU C2 ^ж	AUC1/2 >0.75; AUC2>AU C1 ^з
Базально-клеточный рак			40	1			23	0
Рак мочевого пузыря	40	5			23	0		
Глиобlastома	20	23			10	9		
Рак печени	66	68	66	12	59	5	59	0
Аденокарцинома легких	17	0	17	0	7	0	7	0
Чешуйчатоклеточный рак ротовой полости			32	1			21	0
Первичная меланома	32	2			21	0		
Рак простаты (хорошо дифференцированный)			5	0			2	0
Рак почки	5	0	25	0	2	0	13	0
	25	0			13	0		
			28	5			16	0
	28	8			16	0		
			19	6			10	0

OncoFinder находит сигналомные маркеры рака мочевого пузыря относительно нормы (Lezhnina et al., 2014)

Нами был предложен новый метод проверки устойчивости сигналомных маркеров, служащий для нивелирования влияния случайно выбираемого нормального референтного уровня. Согласно этому методу, значения PAS, вычисленные для раковых образцов относительно нормального уровня N1, сравнивают со значениями PAS для какого-либо другого нормального уровня N2, вычисленными относительно того же уровня N1. Пути, оказавшиеся при данном сравнении дифференциальными, считают маркерными для данной нозологической формы относительно нормального состояния.

В двух лабораториях (ИБХ РАН и Университет г. Летбридж, Канада) на одинаковых платформах (Illumina HT12-4) исследовано 17 образцов рака мочевого пузыря. Найдено 44 маркерных пути, дифференциально активированных в результатах обеих лабораторий. Из них 10 (23%) повышено активированы, а 34 (77%) ингибированы.

Сравнение двух предиктивных показателей, DS1 и DS2, для клинической эффективности препарата трастузумаба

Номер набора образцов в хранилище GEO	42822	37946
Число больных со значимыми клиническими улучшениями	12	27
Число больных без значимых клинических улучшений	11	21
Линейное ядро SVM		
DS1 для больных со значимыми клиническими улучшениями	210 ± 140	360 ± 150
DS1 для больных без значимых клинических улучшений	110 ± 80	320 ± 110
DS2 для больных со значимыми клиническими улучшениями	15 ± 3	13 ± 4
DS2 для больных без значимых клинических улучшений	10 ± 6	12 ± 5

FloWPS – FLOating Window Projective Separator, модификация метода опорных векторов (SVM)

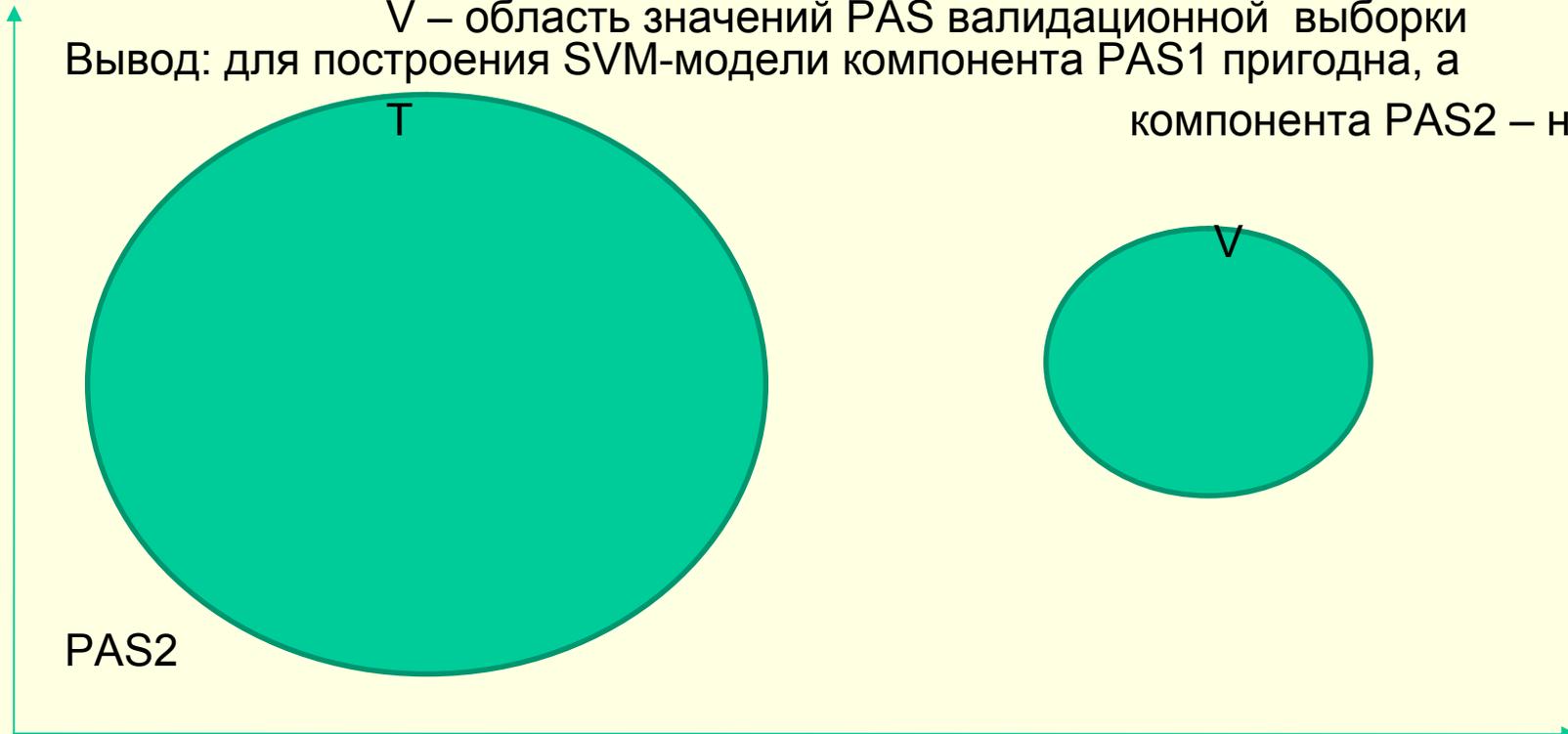
PAS1

T – область значений PAS тренировочной выборки

V – область значений PAS валидационной выборки

Вывод: для построения SVM-модели компонента PAS1 пригодна, а

компонента PAS2 – нет



Валидация FloWPS на раке почки

Тренировочная выборка (Т)

227 клеточных линий

4 нормировки (гладкая мускулатура стенок аорты, печень, нейроглия)

Пять квинтилей по IC₅₀

Квинтиль

Score

1

0

2

25

3

50

4

75

5

100

Валидационная выборка (V)

Число образцов

Из них ответчики

Не-ответчики

Пазопаниб

7

4

3

Сорафениб

28

13

15

Сунитиниб

15

5

10

AUC для FloWPS

Линейное ядро SVM

Нормировка Т-выборки	Аортная	Глиальная	Печеночная	Усредненная	
Пазопаниб		1	1	0,833333	1
Сорафениб		0,82	0,89	0,78	0,8
Сунитиниб		0,94	0,94	0,84	0,86

Полиномиальное (кубический сплайн) ядро SVM

Нормировка Т-выборки	Аортная	Глиальная	Печеночная	Усредненная	
Пазопаниб		1	1	1	1
Сорафениб		0,78	0,78	0,86	0,84
Сунитиниб		1	1	0,96	0,94

С точки зрения теории алгоритмов машинного обучения:

Переход от транскриптомного уровня представления информации к сигнальному означает добавления (нулевого) слоя данных в архитектуру многослойного неклассического перцептрона

Новые методы машинного обучения, разрабатываемые для системы OncoFinder:

Разделения ответчиков и неответчиков на терапию с помощью машины опорных векторов с «плавающим окном»

Настройка k -тов ARR для того, чтобы величины PAS были маркерами нозологических форм и/или ответов на методы лечения.

Применение генетических алгоритмов для гармонизации результатов обследования на разных экспериментальных платформах.