Университет ИТМО

Сергушичев Алексей Александрович

Методы вычислительного анализа метаболических моделей для интерпретации транскриптомных и метаболомных данных

Диссертация на соискание ученой степени кандидата технических наук

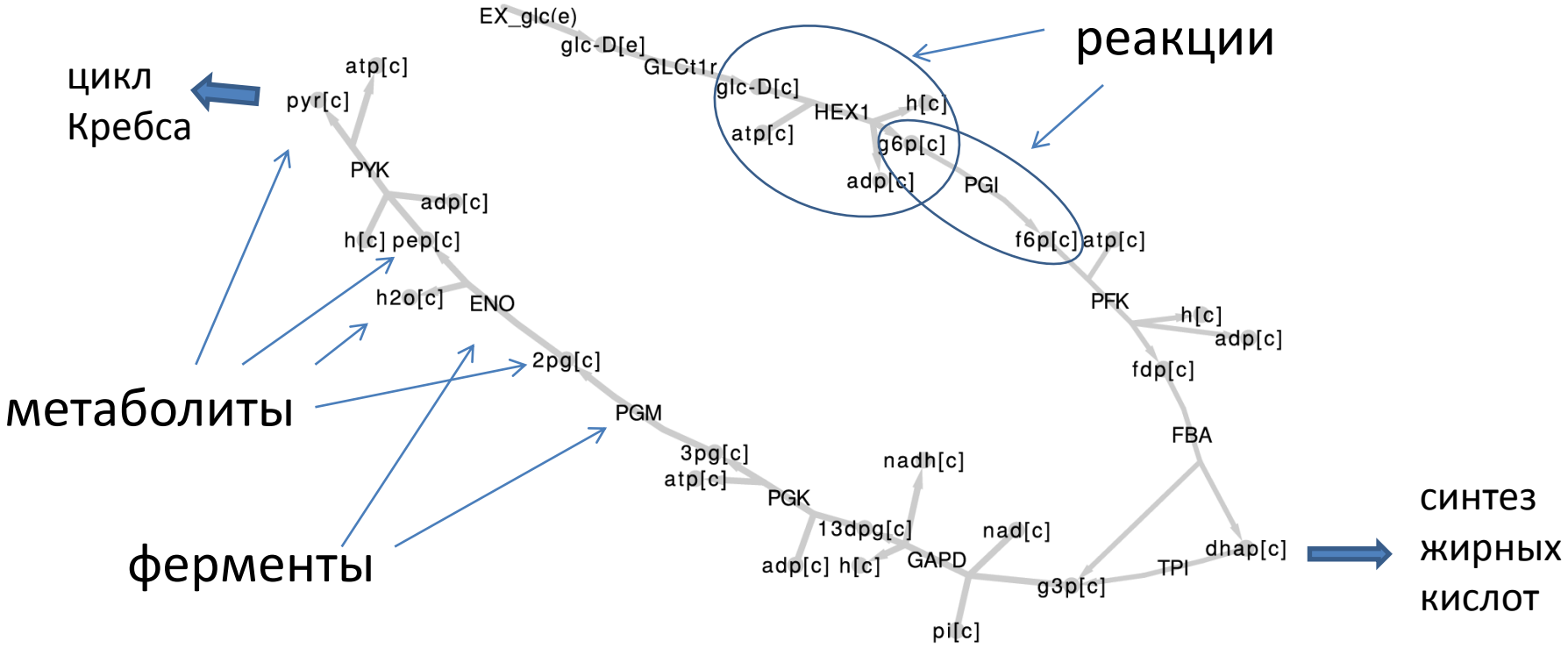
Специальность 05.13.18 — Математическое моделирование, численные методы и комплексы программ

Научный руководитель – Артемов Максим, PhD

22 декабря 2016, Санкт-Петербург

Метаболизм и метаболические модели

метаболические модели описывают биохимические реакции, возможные к клетке



метаболический путь расщепления глюкозы

Регуляция метаболизма важна в работе иммунной системы и развитии опухолей

nature REVIEWS IMMUNOLOGY

Full text access provided to V by Th

Search

Journal home > Archive > Foreword > Full Text

JOURNAL CONTENT

Journal home
Advance online publication
Current issue
Archive
Web Focuses
Article Series
Posters
Calendars

Journal information

Guide to Nature Reviews Immunology
Online submission
Guidelines for referees
About the journal
Subscribe
Feedback for editors

NPG services

Help
Authors and Referees
Librarian gateway
Advertising information
work@npg
Reprints and permissions

Foreword

Nature Reviews Immunology **11**, 81-83 (February 2011) | doi:10.1038/nri2922

FOCUS ON: Metabolism and immunology

Immunometabolism: an emerging frontier

Diane Mathis[1] & Steven E. Shoelson[2]

top

Immunometabolism is an emerging field of investigation at the interface between the historically distinct disciplines of immunology and metabolism. Accelerating interest in this area is being fuelled by the obesity epidemic and the relatively recent realization that obesity affects the immune system and promotes inflammation, and that obesity-induced inflammation potentially promotes a variety of chronic conditions and diseases. The multilevel interactions between the metabolic and immune systems suggest pathogenic mechanisms that may underlie many of the downstream complications of obesity and offer substantial therapeutic promise.

"To lengthen thy life, lessen thy meals." Benjamin Franklin, Poor Richards Almanac (1737).

It has long been recognized that effector cells of the immune system are required to ward off tumours and infectious agents. Likewise, it is well known that regulatory cells of the immune system rein in such responses, as well as guarding against immune dysregulation, such as that which occurs in allergy and autoimmunity. Even greater respect for this powerful homeostatic system has emerged over the past few years with the increasing appreciation that immune cells also affect important non-immune functions, including neurodegeneration, cardiovascular function and metabolism. This

nature REVIEWS CANCER

Search

Journal home > Archive > Review > Full Text

JOURNAL CONTENT

Journal home
Advance online publication
Current issue
Archive
Web Focuses
Article Series
Podcasts
Posters
Calendars

Journal information

Guide to Nature Reviews Cancer
Online submission
Guidelines for referees
About the journal
Subscribe
Feedback for editors

NPG services

Help
Authors and Referees
Librarian gateway
Advertising information
work@npg

Review

Nature Reviews Cancer **11**, 85-95 (February 2011) | doi:10.1038/nrc2981

Regulation of cancer cell metabolism

Rob A. Cairns[1,2], Isaac S. Harris[1,2] & Tak W. Mak[1] About the authors

top

Interest in the topic of tumour metabolism has waxed and waned over the past century of cancer research. The early observations of Warburg and his contemporaries established that there are fundamental differences in the central metabolic pathways operating in malignant tissue. However, the initial hypotheses that were based on these observations proved inadequate to explain tumorigenesis, and the oncogene revolution pushed tumour metabolism to the margins of cancer research. In recent years, interest has been renewed as it has become clear that many of the signalling pathways that are affected by genetic mutations and the tumour microenvironment have a profound effect on core metabolism, making this topic once again one of the most intense areas of research in cancer biology.

View At a Glance

Over the past 25 years, the oncogene revolution has stimulated research, revealing that the crucial phenotypes that are characteristic of tumour cells result from a host of mutational events that combine to alter multiple signalling pathways. Moreover, high-throughput sequencing data suggest that the mutations leading to tumorigenesis are even more numerous and heterogeneous than previously thought[1,2]. It is now clear that there are thousands of point mutations, translocations, amplifications and deletions that may contribute to cancer development, and that the mutational range can differ even

http://www.nature.com/nrc/journal/v11/n2/full/nrc2981.html
http://www.nature.com/nri/journal/v11/n2/full/nri2922.html

Становятся доступнее технологии получения больших объемов данных

- Транскриптомные данные дают информацию о ферментах
- Метаболомные – о метаболитах (веществах)
- Для клеток человека/мыши описано ~2000 реакций
- **Необходимы вычислительные методы для интерпретации этих данных**

Цель диссертационной работы

- Целью работы: разработка и программная реализация набора эффективных **вычислительных методов анализа метаболических моделей** для идентификации регулируемых метаболических путей и их взаимосвязей по транскриптомным и метаболомным экспериментальным данным.

Задачи работы: разработка методов для разных уровней описания моделей

1. Разработка и реализация эффективного метода идентификации регулируемых путей в метаболических моделях на основе анализа представленности, **без использования информации о связях между реакциями**
2. Разработка и реализация эффективного метода идентификации регулируемых путей и их взаимосвязей в метаболических моделях **на основе подхода поиска активного модуля**
3. Разработка и реализация эффективного метода идентификации регулируемых путей и их взаимосвязей в метаболических моделях на основе подхода поиска активного **модуля с использованием информации об атомной структуре метаболитов**

Новые научные результаты

1. Метод FGSEA для проведения эффективного взвешенного анализа представленности функциональных наборов генов
2. Метод GAM для выделения активных метаболических модулей с помощью анализа сети метаболических реакций
3. Метод GATOM для выделения активных метаболических модулей с помощью анализа графа атомных переходов

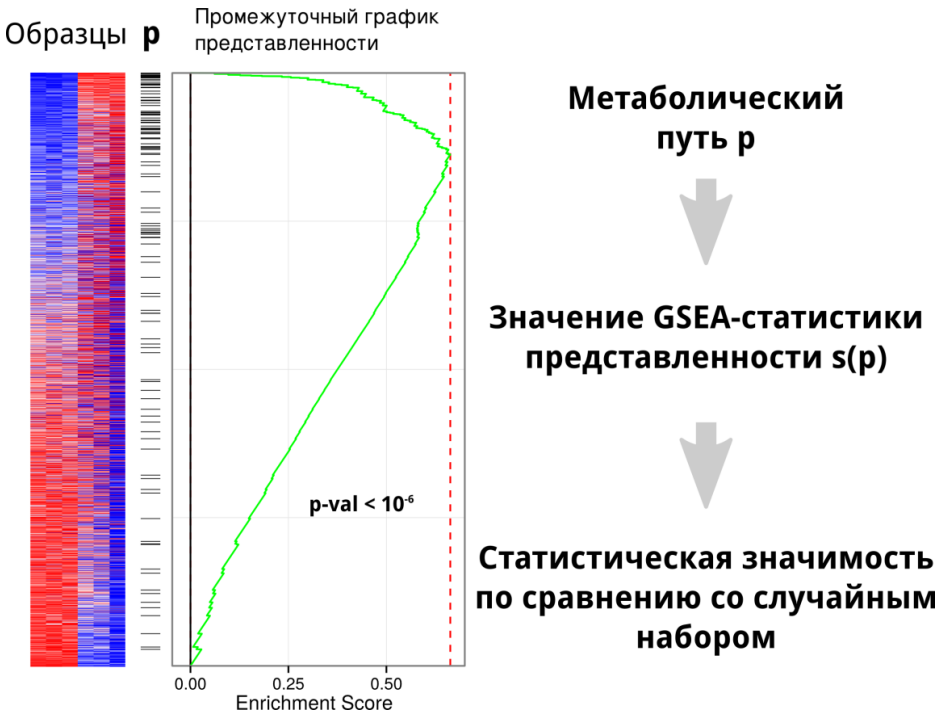
Задача 1: поиск регулируемых путей без информации о связях между реакциями

- Рассматриваются два состояния клеток: например, контрольное и после воздействия
- Даны транскриптомные данные об индивидуальной регуляции ферментов
- Дан список метаболических путей: какие ферменты используются в каждом
- **Какие из путей** имеют признаки совместной регуляции?

Взвешенный анализ представленности GSEA позволяет оценить степень совместной регуляции

Проблема – метод медленный: время работы – O(nmKlogK),

n – число случайных наборов, *m* – число входных наборов, *K* – максимальный размер



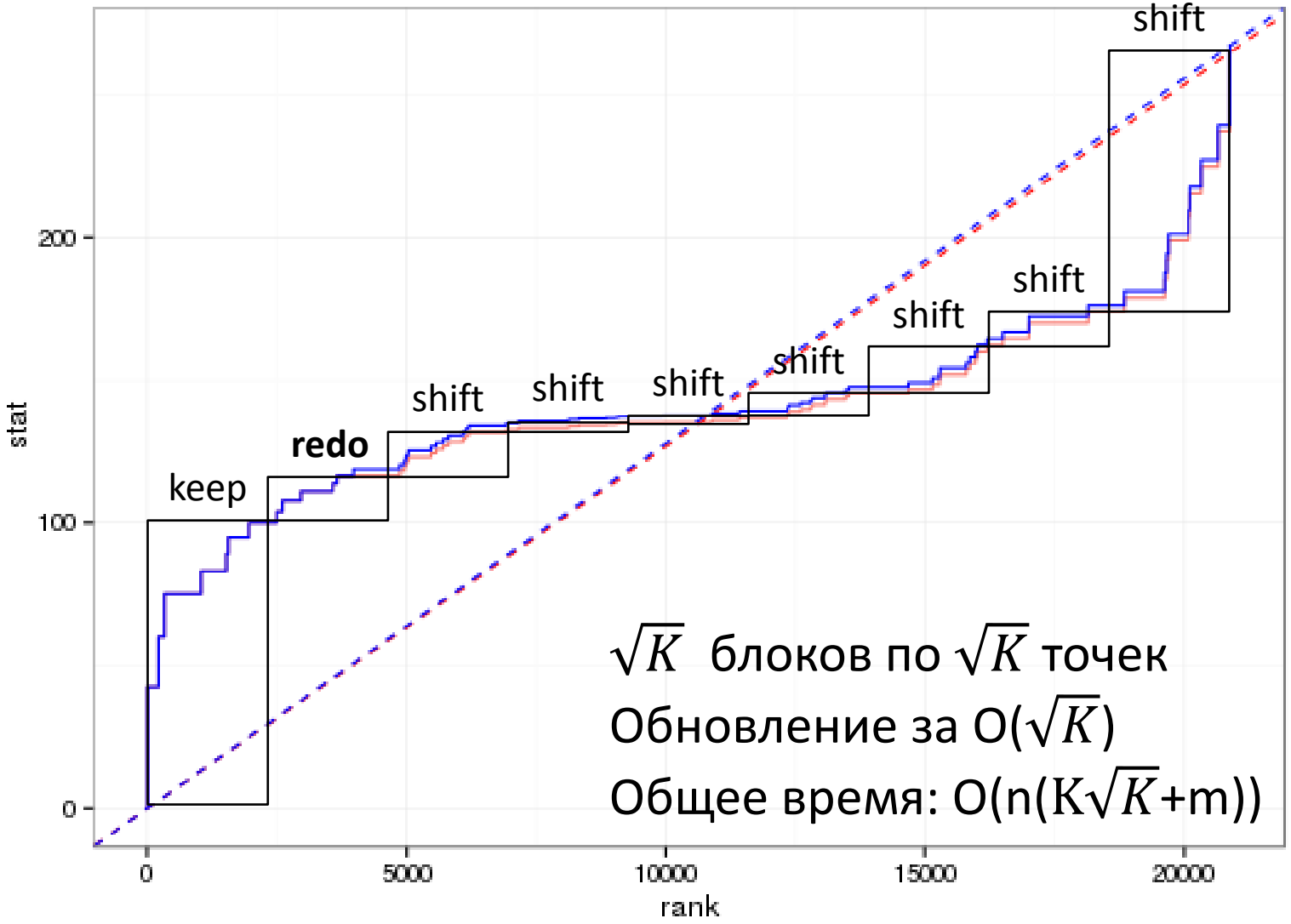
* GSEA = Gene Set Enrichment Analysis

Метод описан в Subramanian et al. 2005, больше 10000 цитирований

Предложен метод быстрого взвешенного анализа представленности

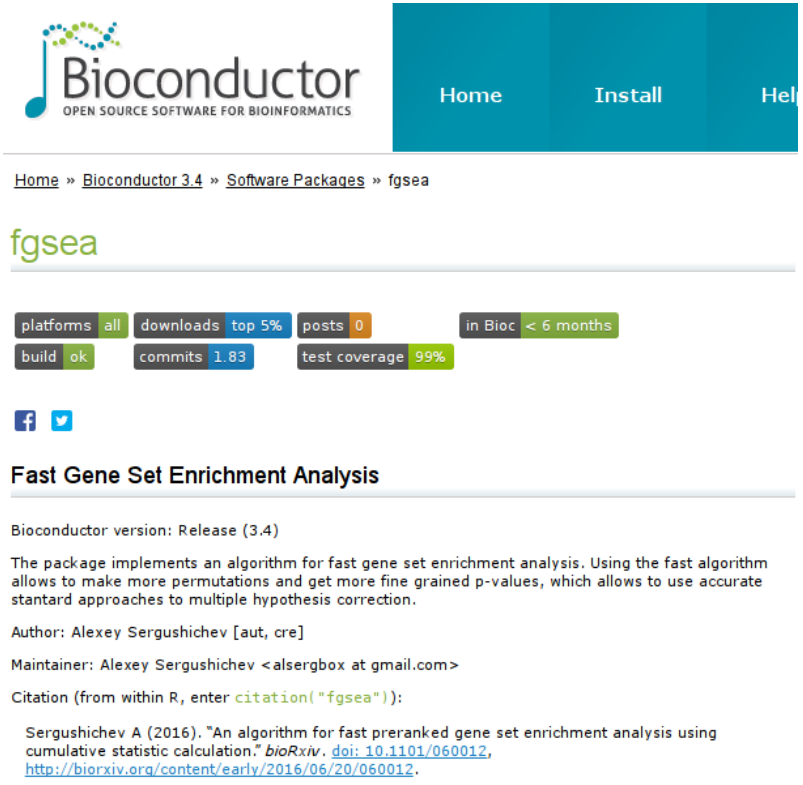
- Разработан метод FGSEA (Fast Gene Set Enrichment), позволяющий вычислять фоновые распределения GSEA-статистики одновременно для всех входных наборов
- Метод основан на разработанном алгоритме быстрого обновления GSEA-статистики при добавлении гена в набор (кумулятивного вычисления)

Статистику можно быстро обновлять с помощью корневой эвристики



Реализация метода FGSEA

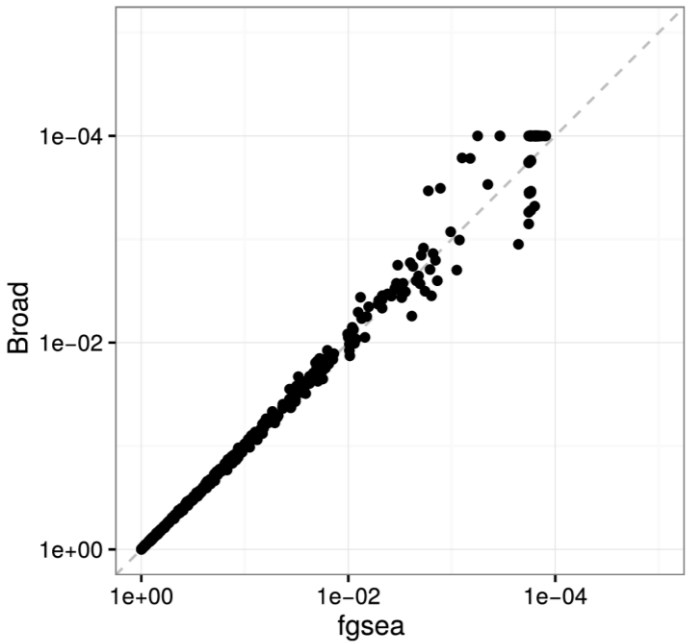
- Метод реализован в виде программного пакета на языке R
- Пакет принят в открытую библиотеку R/Bioconductor
- Пакет входит в 5% наиболее загружаемых



https://bioconductor.org/packages/fgsea

Экспериментальное исследование: сравнение с референсной реализацией

	n = 1000 наборов	n = 10000 наборов
Broad	96 с	1048 с
fgsea	0,8 с	5,5 с



При выполнении в один поток достигается ускорение на два порядка

P-значения, полученные обоими методами согласуются друг с другом

Процессор: AMD Phenom II X6 1090T 3,2 ГГц

Внедрение метода FGSEA

- Внедрен в рабочий процесс компании Immuneering (Кэмбридж, США, http://immuneering.com/)
- Использовался в статье *Lampropoulou V.,* ***Sergushichev A.*** et al. Itaconate Links Inhibition of Succinate Dehydrogenase with Macrophage Metabolic Remodeling and Regulation of Inflammation // Cell Metabolism (IF=17.56). — 2016

Пример результата анализа активации Т-клеток

Метаболический путь	P -значение	Поправленное P -значение	Значение GSEA-статистики
Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	1.77E-05	0.000411	0.812715
Gluconeogenesis, oxaloacetate => fructose-6P	1.83E-05	0.000411	0.818604
Glycolysis, core module involving three-carbon compounds	3.75E-05	0.000563	0.840485
Adenine ribonucleotide biosynthesis, IMP => ADP,ATP	0.000293	0.003039	0.752229
C5 isoprenoid biosynthesis, mevalonate pathway	0.000338	0.003039	0.812125

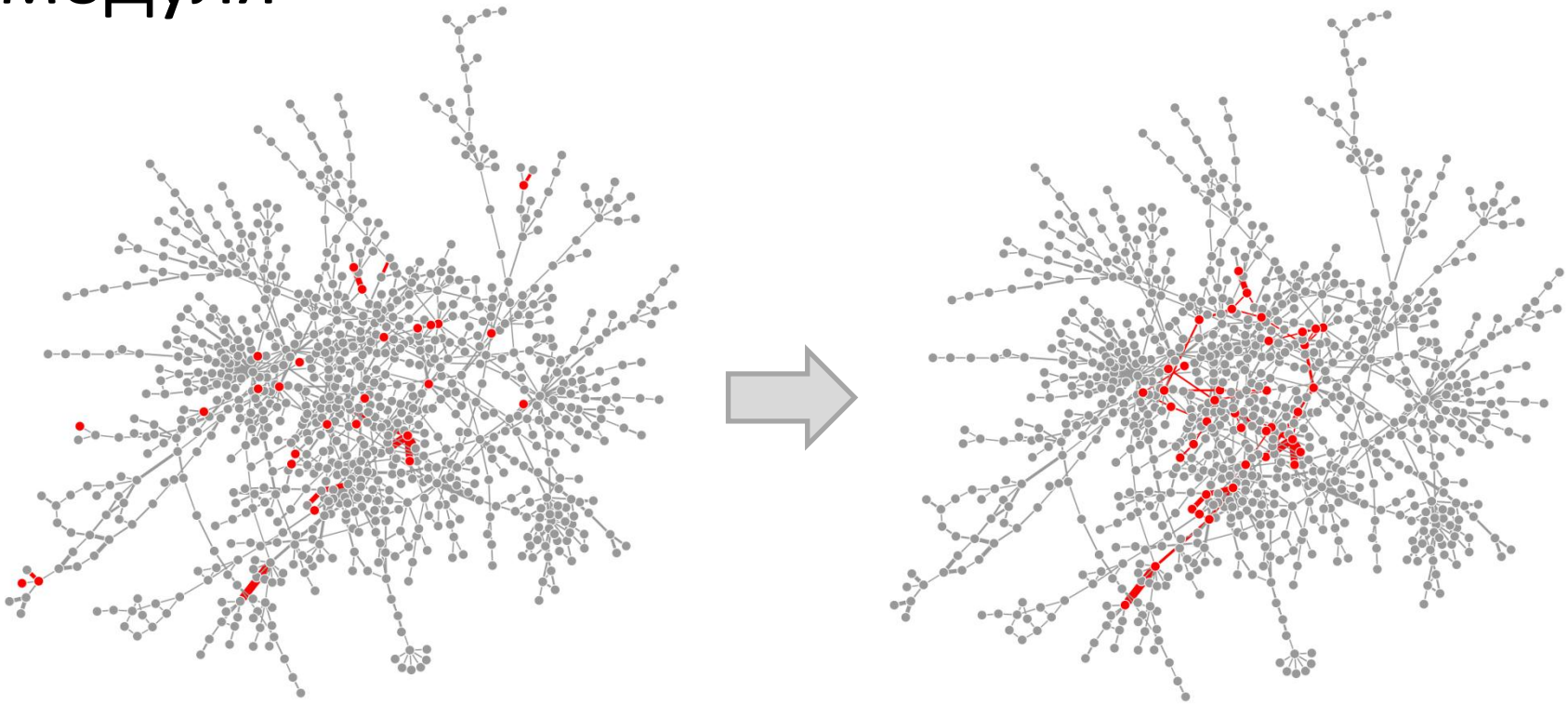
Выводы (1)

1. Предложен метод *FGSEA* для беспорогового анализа представленности
2. Разработан эффективный алгоритм кумулятивного вычисления *GSEA*-статистики
3. Проведены экспериментальные исследования, подтверждающие высокую скорость работы метода
4. Разработанный метод реализован в виде программного пакета для языка *R* и доступен в библиотеке *R/Bioconductor* под свободной лицензией

Задача 2: поиск регулируемых путей и их связей с помощью анализа сети реакций

- Даны транскриптомные и метаболомные данные о индивидуальной регуляции ферментов и метаболитов, соответственно
- Дана сеть метаболических реакций (возможно, с аннотацией путями)
- Какой **фрагмент** сети реакций имеет признаки совместной регуляции?
- Можно обнаружить регуляцию известных путей, их **взаимосвязи**, а также **новые пути**

Используется идея поиска активного модуля

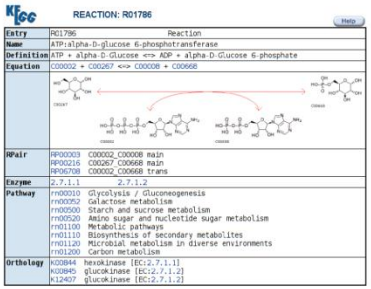


- Формально: задача поиска связного подграфа максимального веса (MWCS)

* MWCS – Maximum Weight Connected Subgraph

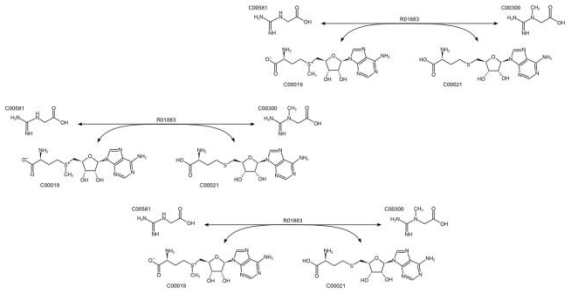
Предложен метод GAM

KEGG database



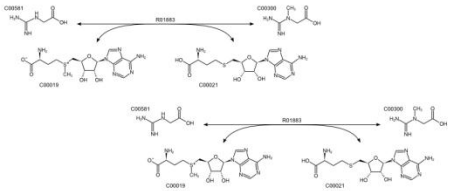
A

Species-specific reactions

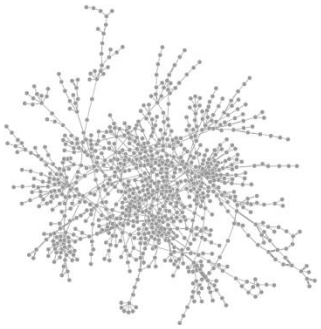


B

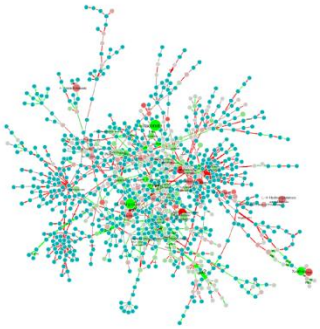
Data-specific reactions



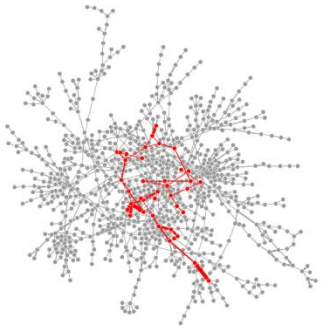
C



D



E



F

Final module

Most regulated subgraph

Scored reaction graph

Reaction graph

Предложено использовать сведение к задаче обобщенному варианту MWCS

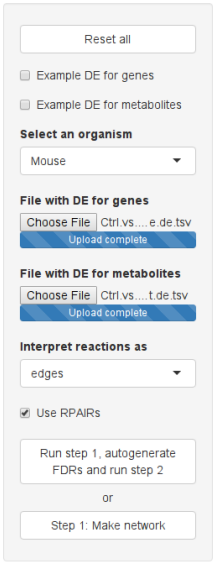
- GMWCS: Generalized Maximum-Weight Connected Subgraph
- Дан связный граф G = (V, E) и весовая функция ω : (V ∪ E) → R
- Веса назначаются по входным данным
- Найти связный подграф $\tilde{G} = (\tilde{V}, \tilde{E})$ с максимальным весом:

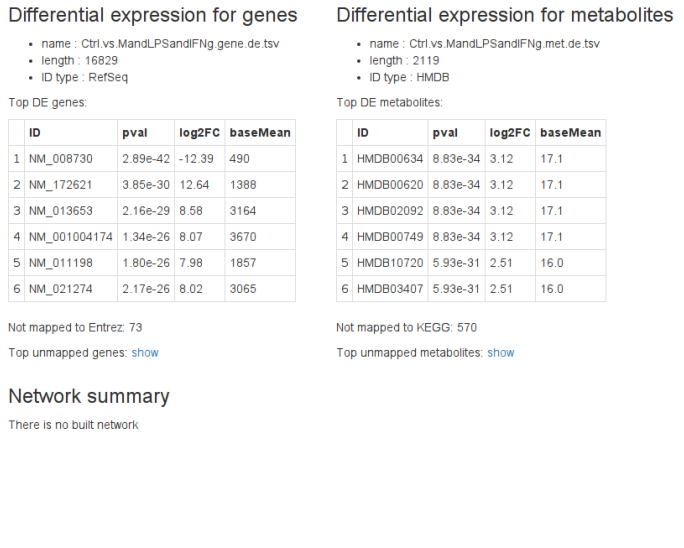
$$\Omega(\tilde{G}) = \sum_{v \in \tilde{V}} \omega(v) + \sum_{e \in \tilde{E}} \omega(e) \to max$$

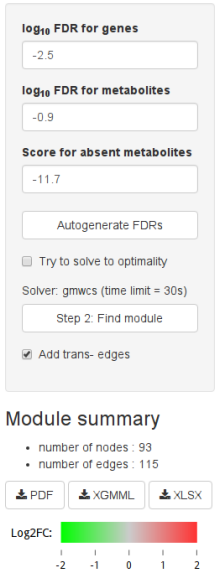
Для задачи GMWCS разработан точный решатель

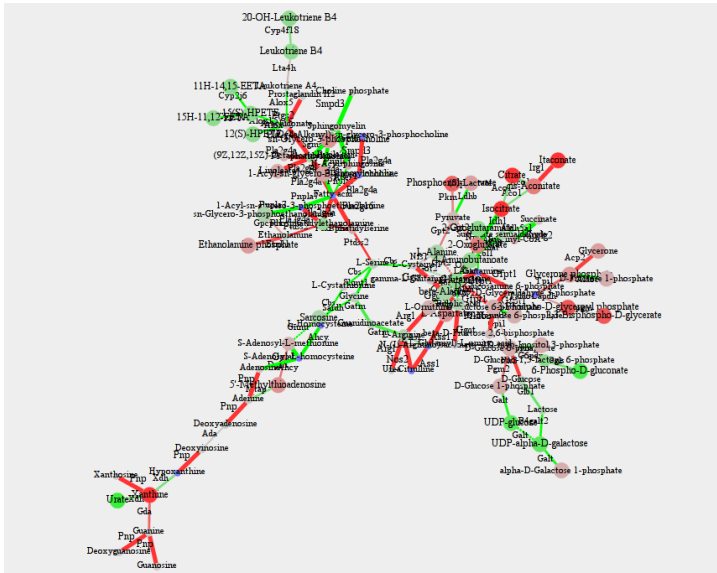
- Задача NP-трудна
- Разработан **точный** решатель, использующий сведение к задаче целочисленного линейного программирования и библиотеку CPLEX
- Многие экземпляры, возникающие в методе GAM, решаются до оптимальности за десять секунд на обычном компьютере

Разработан веб-сервис Shiny GAM http://genome.ifmo.ru/shiny/gam





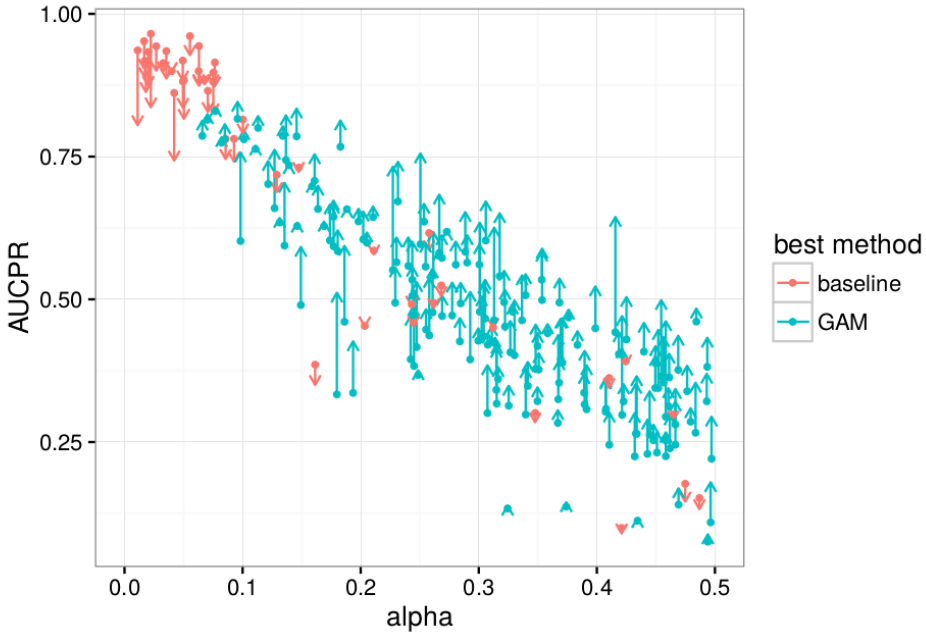


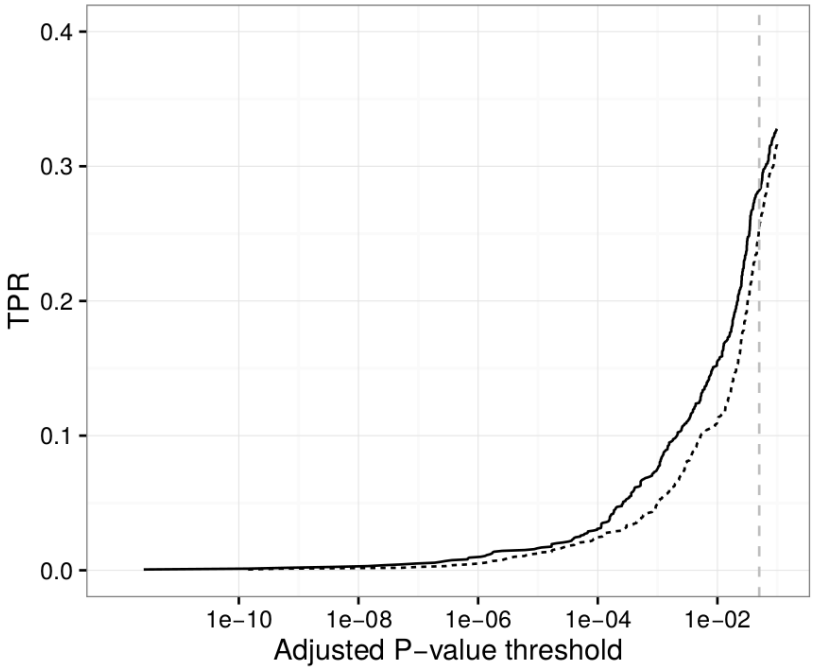


- Собраны наборы реальных данных
- Собраны экземпляры задач GMWCS (выложены на сайт соревнования DIMACS11)

http://dimacs11.zib.de/downloads.html

Экспериментальное исследование

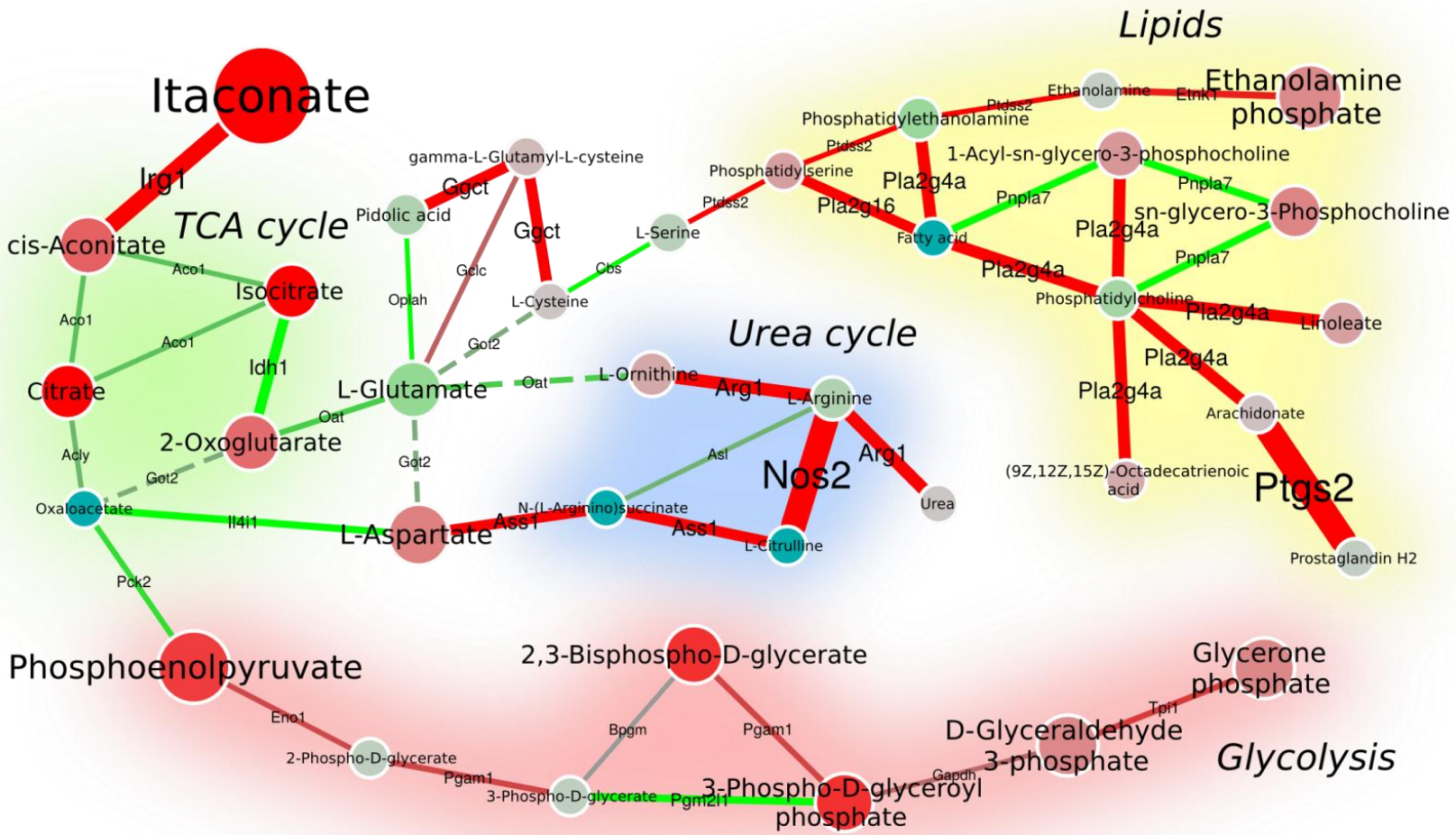




В большинстве случаев метод GAM лучше находит индивидуальную регуляцию генов

Метод GAM находит не меньше регулируемых путей по сравнению с базовым методом

Пример: сравнение неактивированных и активированных макрофагов



Внедрение метода GAM

- Внедрен в рабочий процесс компании Elucidata (Кэмбридж, США, http://www.elucidata.io/)
- Использовался в статье *Jha A. K., Huang S. C.,* ***Sergushichev A.,*** et al. Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization // Immunity (IF=24.08). — 2015
- **Независимое** использование в статье *Liu X.* et al. Metformin Targets Central Carbon Metabolism and Reveals Mitochondrial Requirements in Human Cancers // Cell Metabolism (IF=17.56). — 2016

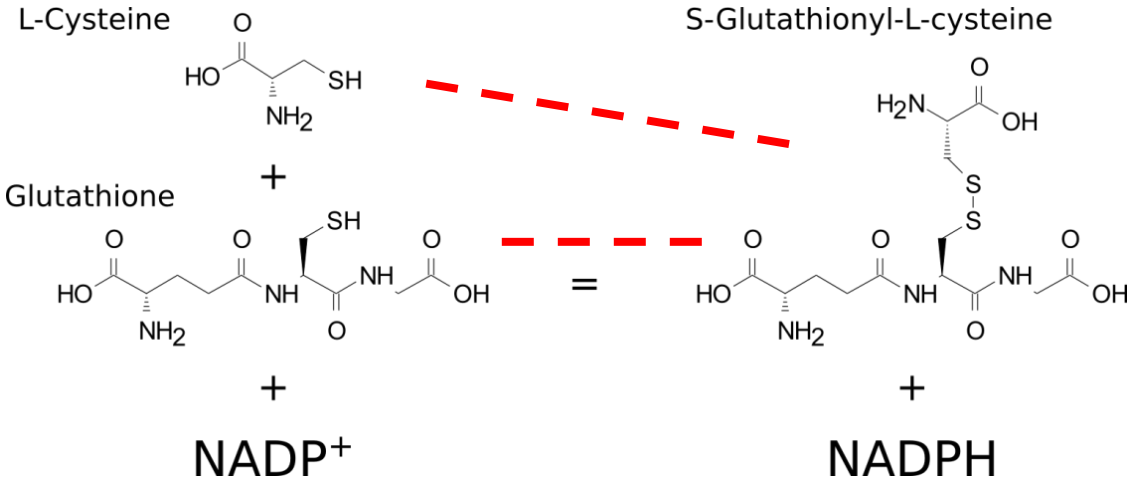
Выводы (2)

1. Разработан метод GAM для выделения активного модуля в сети метаболических реакций с помощью сведения к задаче GMWCS
2. Для задачи GMWCS разработан точный решатель
3. Разработан веб-сервис Shiny GAM , реализующий предложенный метод
4. Проведено экспериментальное исследование подтверждающее эффективность метода
5. Приведен пример, демонстрирующий применение метода к реальным экспериментальным данным

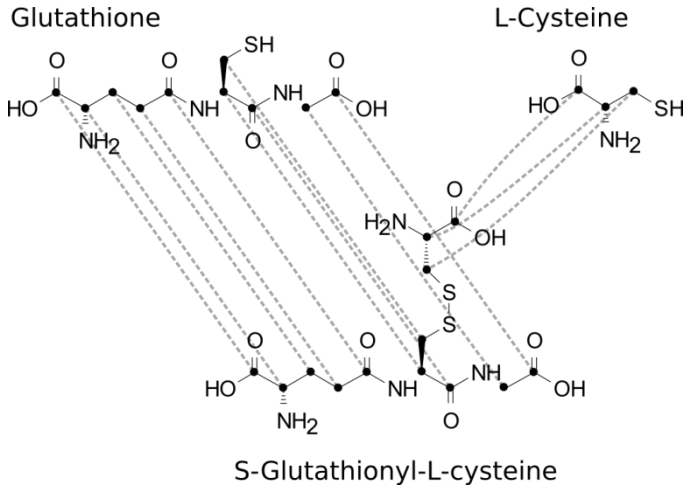
Задача 3: поиск регулируемых путей и их связей с учетом структуры метаболитов

- Даны транскриптомные и метаболомные данные о индивидуальной регуляции ферментов и метаболитов, соответственно
- Дан граф переходов атомов углерода в метаболических реакциях
- Какой **фрагмент графа** имеет признаки совместной регуляции?

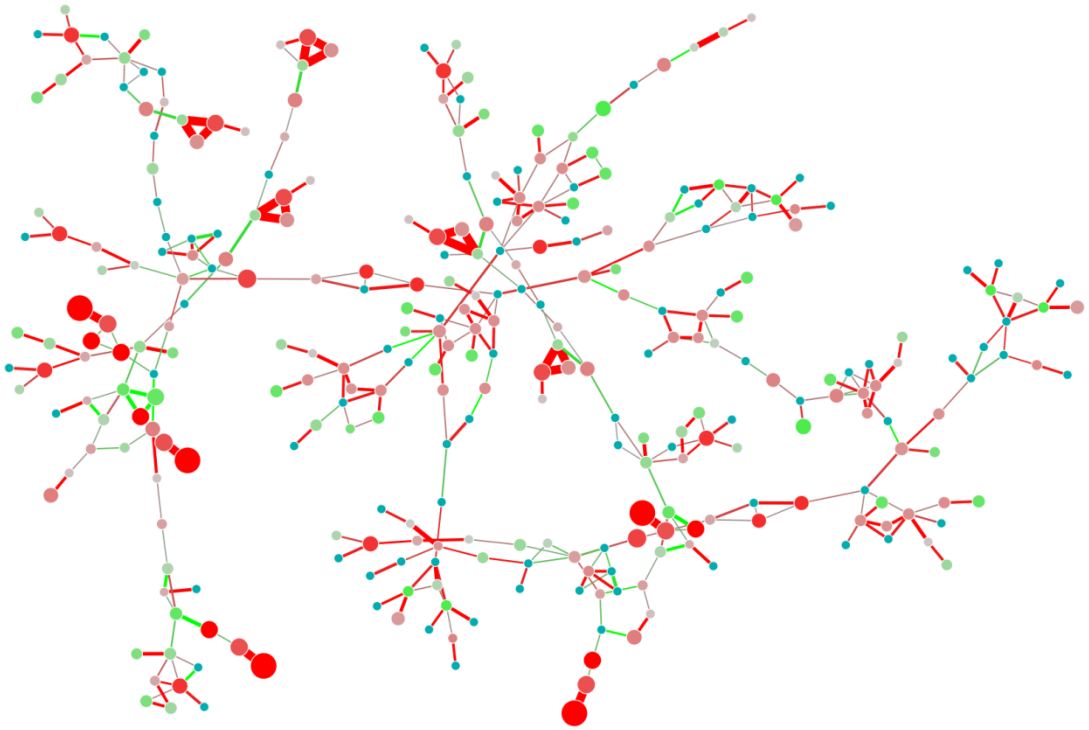
В графе атомных переходов меньше «плохих» соединений



Любой путь в графе атомных переходов соответствует возможной последовательности реакций => соответствие метаболическим путям



Граф «расслаивается»: каждому веществу соответствует несколько вершин



- Повторение одинаковых весов несколько раз с разной «плотностью» делает решения GMWCS не качественными

Предложен сигнальный вариант задачи GMWCS – SGMWCS

- Дано:
  - граф G = (V, E),
  - множество сигналов S,
  - весовая функция $\omega: S \rightarrow R$,
  - разметка вершин и ребер $\sigma: (V \cup E) \rightarrow S$,
  - $\omega(s) < 0 \Rightarrow |\sigma^{-1}(s)| = 1$
- Найти связный подграф $\tilde{G} = (\tilde{V}, \tilde{E})$ с максимальным весом:

$$\Omega(\tilde{G}) = \sum_{s \in \sigma(\tilde{V} \cup \tilde{E})} \omega(s) \rightarrow max$$

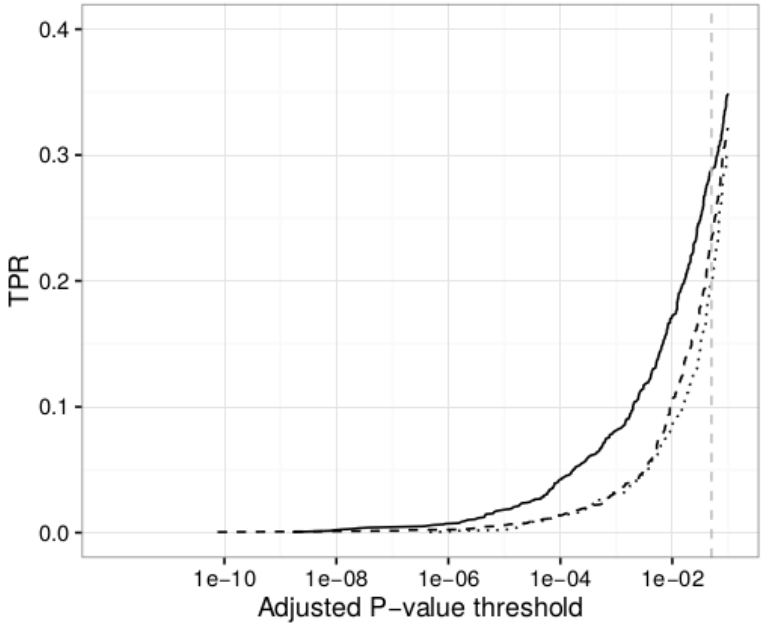
* SGMWCS = Signal GMWCS

Для задачи SGMWCS разработан точный решатель

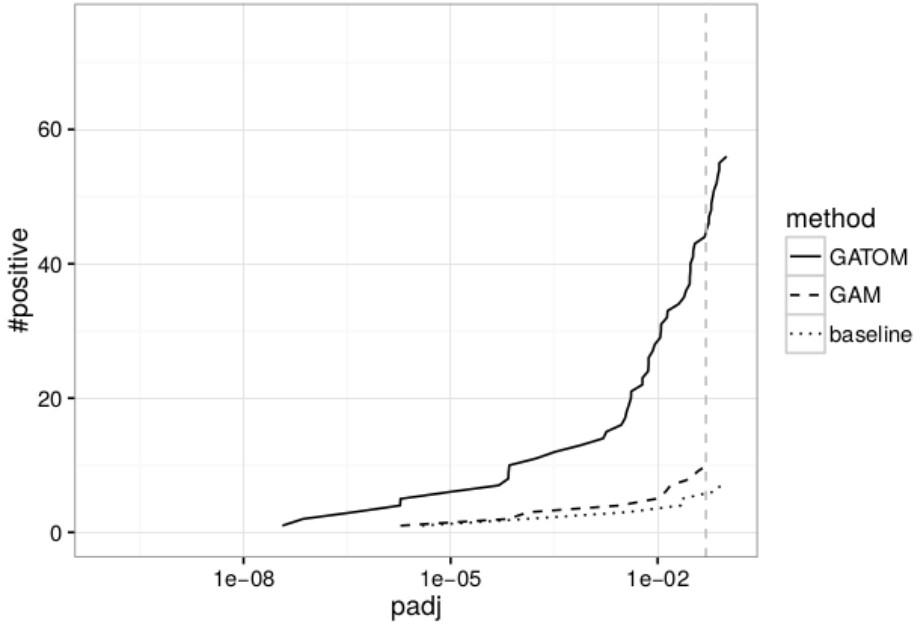
- Задача NP-трудна как и GMWCS
- Разработан **точный** решатель, на основе решателя для GMWCS
- Многие реальные экземпляры решаются до оптимальности за минуту на обычном компьютере

Экспериментальное исследование

- Метод GATOM находит больше путей на сгенерированных и реальных данных по сравнению с базовым методом и GAM

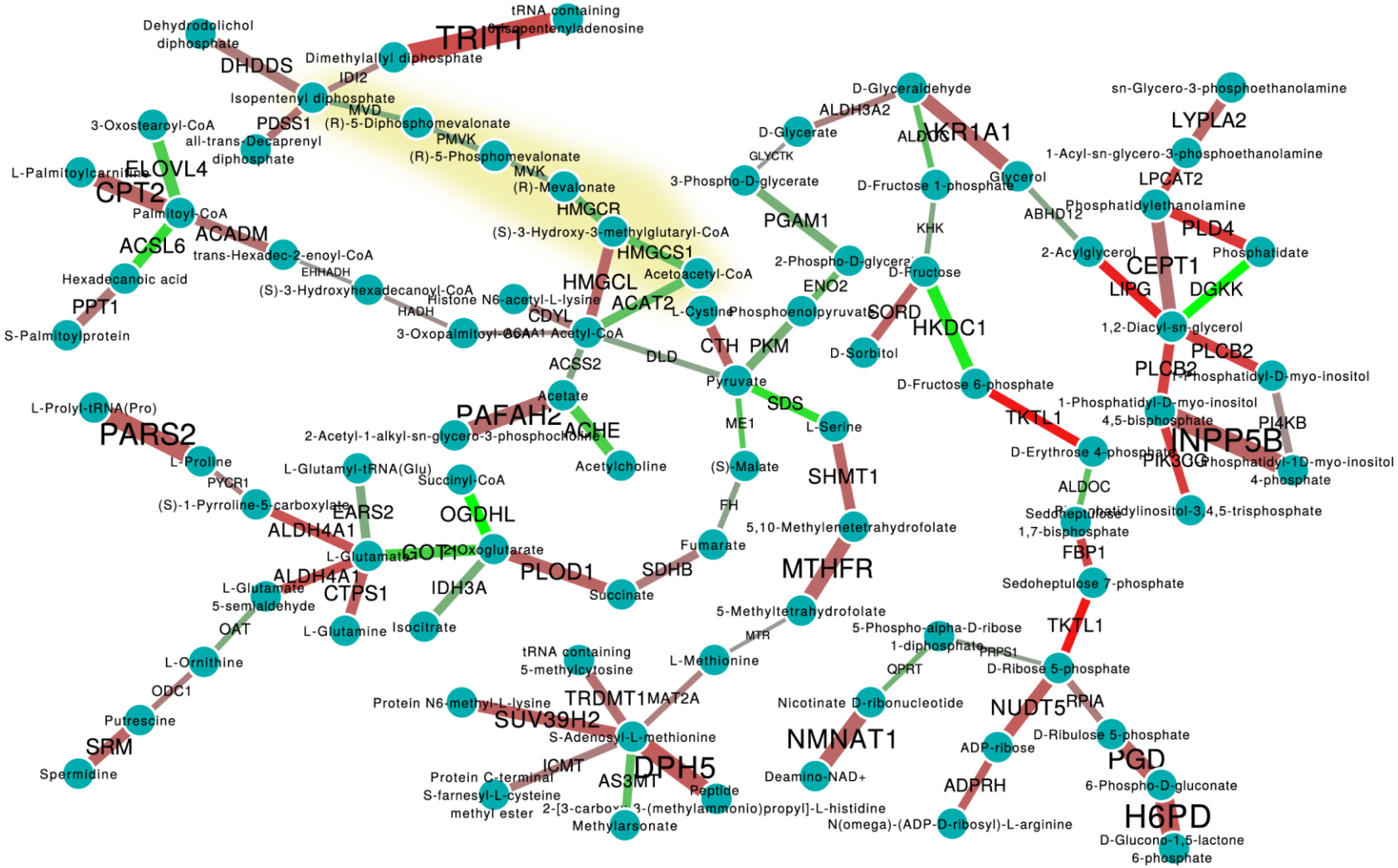


сгенерированные данные



реальные данные

Пример: сравнение образцов глиомы с мутацией в гене TP53 и без мутации



Внедрение метода GATOM

- Использовался в статье *Vincent E. E.,* ***Sergushichev A. A.*** et al. Mitochondrial Phosphoenolpyruvate Carboxykinase Regulates Metabolic Adaptation and Enables Glucose-Independent Tumor Growth // Molecular Cell (IF=13.96). — 2015
- Использовался в статье *Izreig S., Samborska B., Johnson R. M.,* ***Sergushichev A.*** et al. The miR-17-92 microRNA Cluster Is a Global Regulator of Tumor Metabolism // Cell Reports (IF=7.87). — 2016

Выводы (3)

1. Разработан метод GATOM для выделения активного метаболического модуля с помощью анализа графа атомных переходов
2. Сформулирована задача SGMWCS , позволяющая учитывать структуру графа атомных переходов
3. Для задачи SGMWCS разработан точный решатель
4. Разработанный метод включен в веб-сервис Shiny GAM
5. Проведено экспериментальное исследование подтверждающее эффективность метода
6. Приведен пример, демонстрирующий применение метода к реальным экспериментальным данным

Результаты

1. Разработан метод FGSEA для проведения эффективного взвешенного анализа представленности функциональных наборов генов
2. Разработан метод GAM для выделения активных метаболических модулей с помощью анализа сети метаболических реакций
3. Разработан метод GATOM для выделения активных метаболических модулей с помощью анализа графа атомных переходов

Публикации

1. ***Сергушичев А. А.*** Алгоритм кумулятивного вычисления статистики представленности набора генов // Научно-технический вестник информационных технологий, механики и оптики. — 2016

2. *Loboda A. A., Artyomov M. N.,* ***Sergushichev A. A.*** Solving generalized maximum-weight connected subgraph problem for network enrichment analysis // Workshop on Algorithms in Bioinformatics. — 2016

3. *Izreig S., Samborska B., Johnson R. M.,* ***Sergushichev A.*** et al. The miR-17-92 microRNA Cluster Is a Global Regulator of Tumor Metabolism // Cell Reports (IF=7.87). — 2016

4. ***Sergushichev A. A.*** et al. GAM: a web-service for integrated transcriptional and metabolic network analysis // Nucleic Acids Research (IF=9.11). — 2016

5. *Lampropoulou V.,* ***Sergushichev A.*** et al. Itaconate Links Inhibition of Succinate Dehydrogenase with Macrophage Metabolic Remodeling and Regulation of Inflammation // Cell Metabolism (IF=17.56). — 2016

6. *Vincent E. E.,* ***Sergushichev A. A.*** et al. Mitochondrial Phosphoenolpyruvate Carboxykinase Regulates Metabolic Adaptation and Enables Glucose-Independent Tumor Growth // Molecular Cell (IF=13.96). — 2015

7. *Jha A. K., Huang S. C.,* ***Sergushichev A. A.,*** et al. Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization // Immunity (IF=24.08). — 2015

8. ***Сергушичев А. А.*** Программа для быстрого анализа представленности метаболических путей по упорядоченному списку генов с весами // Свидетельство №2016 660664 от 20.09.2016

Конференции

1. 16th Workshop on Algorithms in Bioinformatics (WABI 2016), Оорхус, Дания
2. Всероссийская научная конференция по проблемам информатики «СПИСОК 2016», СПбГУ, Матмех
3. Moscow Conference on Computational Molecular Biology (MCCMB'15), 2015, Москва
4. Cold Spring Harbor Laboratory meeting on Systems Biology: Networks, 2015, Колд-Спринг-Харбор, США
5. IV международная научно-практическая конференция «Постгеномные методы анализа в биологии, лабораторной и клинической медицине», 2014, Казань
6. Metabolism and Immunity: A Rediscovered Frontier, 2014, Дублин, Ирландия

Университет ИТМО

Сергушичев Алексей Александрович

Методы вычислительного анализа метаболических моделей для интерпретации транскриптомных и метаболомных данных

Диссертация на соискание ученой степени кандидата технических наук

Специальность 05.13.18 — Математическое моделирование, численные методы и комплексы программ

Научный руководитель – Максим Артемов, PhD

22 декабря 2016, Санкт-Петербург