

РЕКОНСТРУКЦИЯ ПУТЕЙ СИГНАЛЬНОЙ ТРАНСДУКЦИИ, УЧАСТВУЮЩИХ В РАЗВИТИИ РАССЕЯННОГО СКЛЕРОЗА У ЧЕЛОВЕКА

Р.Н. Шарипов*

*ООО «Институт системной биологии», Новосибирск
Институт цитологии и генетики СОРАН, Новосибирск*

*e-mail: shrus79@gmail.com

И.С. Евшин

*ООО «Институт системной биологии», Новосибирск
Новосибирский государственный университет, Новосибирск*

М.Л. Филипенко

Институт химической биологии и фундаментальной медицины СОРАН, Новосибирск

Ю.В. Кондрахин

*Конструкторско-технологический институт вычислительной техники СОРАН,
Новосибирск*

ООО «Институт системной биологии», Новосибирск

У.А. Боярских

Институт химической биологии и фундаментальной медицины СОРАН, Новосибирск

Г.М. Дымшиц

*Институт цитологии и генетики СОРАН, Новосибирск
Новосибирский государственный университет, Новосибирск*

Ф.А. Колпаков

*Конструкторско-технологический институт вычислительной техники СОРАН
ООО «Институт системной биологии», Новосибирск*

Аннотация

Рассеянный склероз – заболевание человека, характеризующееся поражением миелиновой оболочки нейронов центральной нервной системы (ЦНС) и ведущее к инвалидности или гибели в молодом возрасте. Картина поражения и симптомы позволяет отнести его к аутоиммунным заболеваниям, однако этиология его не ясна. С целью выявления ключевых компонентов сигнальных путей, принимающих участие в развитии данной патологии, был проведен мета-анализ восьми наборов микрочиповых данных. В результате анализа выявлен паттерн генной экспрессии, характеризующий рассеянный склероз, регулирующие ее ключевые транскрипционные факторы, а также их белковое окружение как часть комплексной сигнальной сети, реконструированной методами математической статистики.

ВВЕДЕНИЕ

Рассеянный склероз (РС) – заболевание человека, проявляющееся, в основном, в молодом возрасте, характеризующееся повреждением или утратой миелиновой оболочки нейронов центральной нервной системы (ЦНС) и ведущее к инвалидности. Диагностика РС затруднена из-за разной степени и частоты проявления симптомов у больных, их периодичности с последующей ремиссией или постепенном прогрессировании, а также из-за того, что симптоматика может быть схожа с другими заболеваниями. Среди возможных

причин развития РС рассматриваются аутоиммунная реакция, инфекция или генетическая предрасположенность [1].

Исследование механизмов развития РС является актуальным особенно на фоне постоянно растущего числа пациентов с этим диагнозом, которое связывают, в том числе, и с повышением качества диагностики, а также ухудшением экологии в развитых и развивающихся странах. Это накладывает на традиционно повышенный риск заболевания РС среди населения стран Северного полушария [2,3]. Современные методы исследования предоставляют более широкие возможности для изучения механизмов развития РС по сравнению с более ранними методами, что дает больше материала для работы исследователей. Международные базы данных (например, GEO (Gene Expression Omnibus) [4], ArrayExpress [5]) содержат достаточно большой объем данных для РС, полученных в ходе массовых экспериментов с использованием современных высокопроизводительных технологий, таких как платформы для исследования генной экспрессии.

Согласно данным литературы, в манифестации РС играют важную роль сигнальные пути, регулирующие иммунный ответ. С его нарушением связывают разрушение миелиновой оболочки нейронов, наблюдаемое у больных РС. Геномный анализ ассоциаций показал значимость полиморфизма локусов ряда генов главного комплекса гистосовместимости I и II, например, 6p21.3 и 13q31.3, и генов иммунной системы для прогнозирования риска развития РС и тяжести его протекания (обзор в [6]). Поскольку миелиновая оболочка нейронов формируется глиальными клетками, то предполагается, что сигнальные пути, регулирующие апоптоз, также играют важную роль в развитии данной патологии [7].

АНАЛИЗ МИКРОЧИПОВЫХ ДАННЫХ

В рамках работы по выявлению молекулярных факторов и путей сигнальной трансдукции, играющих ключевую роль в клетках больных РС по сравнению с контролем, нами проведен анализ генной экспрессии с использованием восьми выборок микрочиповых данных из базы данных GEO: GSE14895, GSE15245, GSE16461, GSE17449, GSE19285, GSE23832, GSE24427, GSE26104. Все данные были получены на мононуклеарных клетках периферической крови больных РС и контрольных доноров. Согласно недавним исследованиям, состояние клеток крови в достаточной степени отражает состояние ЦНС при РС [8], что упрощает исследовательскую работу в силу большей доступности данного биоматериала и низкой инвазивности для пациента при его заборе в отличие от проб, взятых из ЦНС. Используемые в работе первичные данные были получены с применением семейства чипов Affymetrix платформы HG-U133 (версии A, A 2.0, A и B и Plus 2.0).

Обработка первичных данных была проведена с применением платформы BioUML (Biological Universal Modeling Language, [9]), разрабатываемой нами для поддержки широкого спектра исследований в области биоинформатики и системной биологии. BioUML позволяет загружать данные пользователя в базу, проводить статистическую обработку, сохранять результаты и работать с ними по сети Интернет, а также создавать формальное описание путей сигнальной трансдукции и создавать модели биологических систем с использованием данных международных баз или введенных пользователем. В настоящее время большое распространение в мире получила бесплатная платформа для поддержки научных исследований R/Bioconductor [10], поэтому в BioUML была реализована

возможность работы с его библиотеками и скриптами, что еще больше расширило его возможности.

Для анализа были отобраны пробы, полученные от пациентов с четко выраженной формой РС, которые не получали медикаментозное лечение. На начальном этапе анализа было проверено качество первичных данных. После этого, данные были загружены в базу для анализа с помощью BioUML (Рис. 1). Так как микрочипы даже внутри одной платформы отличаются по набору проб, был применен разработанный нами ранее метод мета-анализа IDURO (Identification of Down- and Up-Regulated Objects) [11], основанный на гипергеометрическом анализе данных микрочиповых экспериментов, который позволил свести отдельные результаты для каждого набора данных воедино.

РЕЗУЛЬТАТЫ

В результате мета-анализа были выявлены гены со значимо повышенной или пониженной по сравнению с контролем экспрессией. На рисунке 2 представлен фрагмент таблицы с генами, экспрессия которых значимо понижена у больных РС по сравнению с нормой.

The screenshot displays the BioUML web interface. The top navigation bar includes the URL 'platform.genexplain.com/bioutilweb/#de=data/Projects/sclerosis/Data/Results/ISB_renamed/GSE13732_HGU133Plus20_normc' and a search filter 'Research: sclerosis'. The left sidebar shows a tree view of data files under 'Databases', 'Data', and 'Analyses'. The main content area is divided into two panes. The top-right pane shows a table of gene expression data for GSE13732, with columns for 'ID', 'cis001_13732', 'cis002_13732', 'cis003_13732', and 'cis004_13732'. The bottom-right pane shows a 'Filters' and 'Columns' section with a 'Template to construct the filtering expression' dropdown and a 'Columns (double-click to paste)' list containing the same cis IDs.

ID	cis001_13732	cis002_13732	cis003_13732	cis004_13732
1007 s at	6.44646	7.0089	7.03019	6.79913
1053 at	5.09943	4.91172	5.39961	5.53742
117 at	5.28006	5.25065	5.0298	5.10993
121 at	7.31417	7.27408	7.30946	7.40608
1255 q at	2.72919	2.85313	2.50942	2.75342
1294 at	9.90606	9.85731	9.86869	9.02175
1316 at	4.88102	4.86556	4.74708	5.11167
1320 at	3.3227	3.0712	3.17679	3.29477
1405 i at	8.98519	9.20388	9.43552	9.15454
1431 at	4.96542	4.26043	4.75397	4.45061
1438 at	4.04964	4.36828	4.06597	4.68133
1487 at	7.07908	6.82379	6.74471	6.52187
1494 f at	5.497	5.30112	5.5429	5.58912

Рис. 1. Пользовательский интерфейс программы BioUML (веб). Окно вверху слева отражает список загруженных в систему данных по генной экспрессии. Окно вверху справа – данные набора GSE13732.

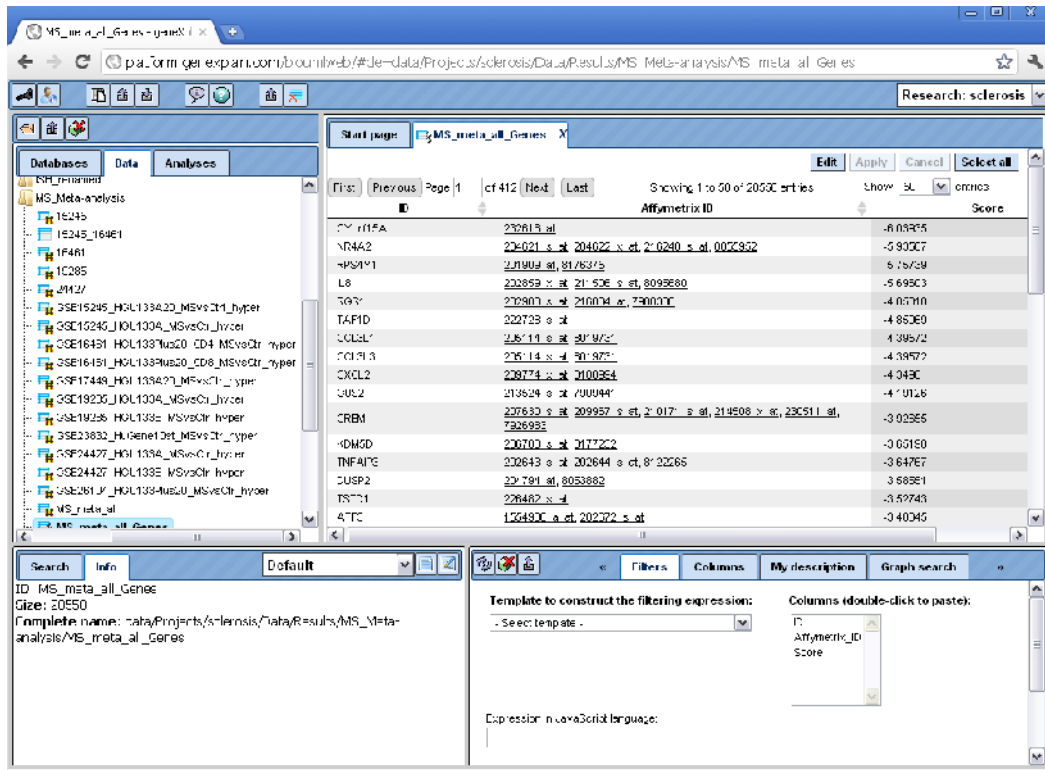


Рис. 2. Результаты мета-анализа микрочиповых данных. Пользовательский интерфейс программы BioUMLweb. Окно сверху слева отражает список собранных данных по генной экспрессии. Окно сверху справа – данные одной из таблиц списка.

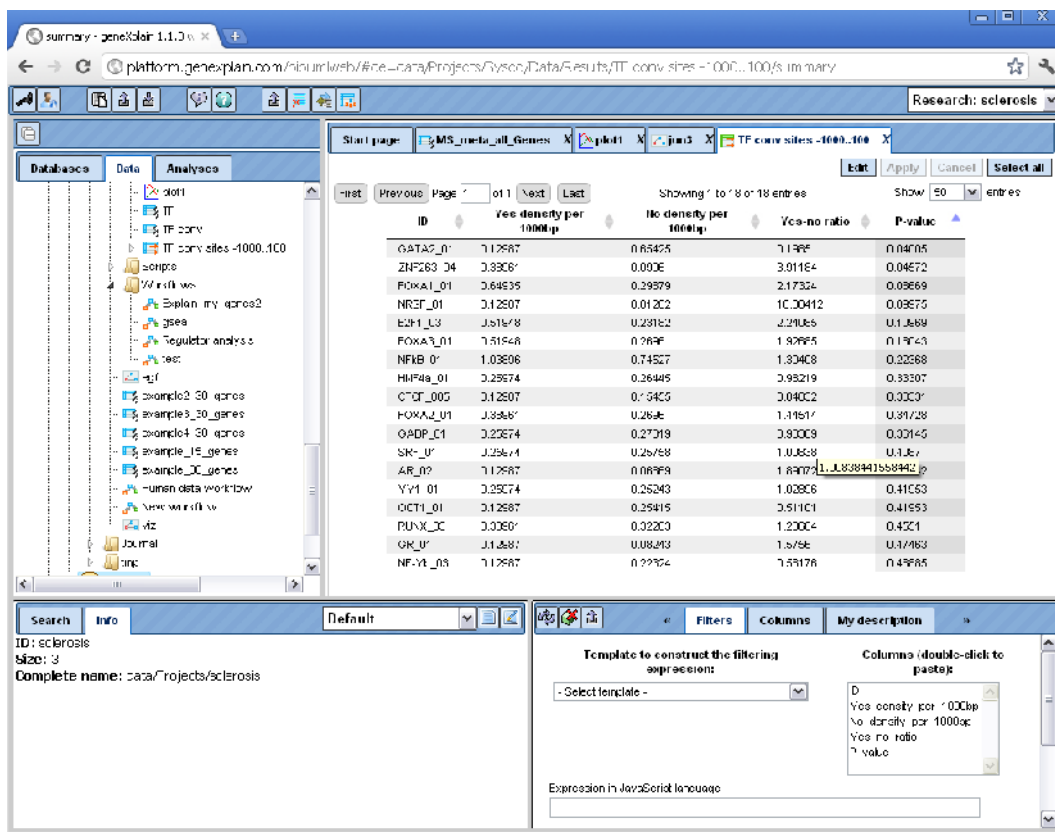


Рис. 3. Результаты протоморного анализа генов, экспрессия которых значительно отличается больных рассеянным склерозом от здоровых людей.

Из выявленных генов были отобраны по 1000 генов, которые были классифицированы согласно GeneOntology (биологические процессы). Эти гены соответствовали, в том числе, категориям «воспаление», «апоптоз» и «иммунные реакции» (p-value <0.01). Классификация по путям сигнальной трансдукции показала активацию, в том числе, HIF1-, Fas-, TNFR1- и p53- путей, регулирующих воспаление и апоптоз (p-value <0.01). В ходе промоторного анализа выявленных генов был определен набор регулирующих их транскрипционных факторов (ТФ).

Промоторный анализ отобранных 2000 генов с применением весовых матриц TRANSFAC (Biobase, GmbH), а также созданных нами ранее, выявил ТФ, потенциально играющие ключевую роль в формировании рассеянного склероза (Рис. 3). В промоторном анализе были задействованы весовые матрицы для более чем тысячи ТФ человека. Поиск производился на интервале [-1000; +1000] относительно старта транскрипции.

ТФ играют важную роль в регуляции генной экспрессии, однако они, в свою очередь, регулируются различными белками, в том числе, белками-ферментами (различными трансферазами, пептидазами, гидролазами и др.). Сигнальные пути передают сигналы от рецепторов внутрь клетки, в ядро, активируя или блокируя определенные ТФ. В ходе дальнейшего анализа, опираясь на информацию международных баз данных о взаимодействиях между белковыми молекулами, была проведена реконструкция сетей регуляторов ТФ (Табл. 1), что позволит в дальнейшем более детально исследовать наиболее и наименее активные цепочки сигнальных путей, уровень активности которых, возможно, был изменен как результат формирования РС.

Таблица 1. Список белков-ферментов, регулирующих ТФ, выявленных в промоторном анализе.

Master molecule name	FDR	Z-Score
MKK6(h){p}	0.0	77.528
HEPTP-xbb2(h)	0.002	11.276
MEKK2(h)	0.002	11.229
Cot-58kDa(h)	0.002	11.146
carm1(h)	0.006	8.464
JKAP(h)	0.004	8.304
mTOR:riCTOR	0.002	8.094
MEK1(h){p}	0.006	8.057
Cot-52kDa(h)	0.006	7.846
MKK3-isoform2(h)	0.006	7.163
MKK3-isoform3(h)	0.006	6.659
ТАК1c(h)	0.008	6.303
ТАК1d(h)	0.006	6.041
MKK6-isoform2(h)	0.006	6.010
MEK(h){p}	0.006	5.837
MEK2(h){p}	0.006	5.812
senp1-isoform2(h)	0.006	5.765
ТАК1b(h)	0.008	5.739
hipk2-isoform3(h)	0.008	5.473
PP2Cbeta1(h)	0	-21.219

ЗАКЛЮЧЕНИЕ

В результате проведенной работы нами были выявлены паттерны генной экспрессии у больных РС, выявлены ТФ, регулирующие дифференциально экспрессированные гены, а также проведена реконструкция путей сигнальной трансдукции, эффекторами которых являются выявленные ТФ. В дальнейшем будет проведен более детальный анализ реконструированных путей с целью идентификации ключевых факторов – мишеней для корректирующей терапии.

Работа поддержана в рамках междисциплинарного проекта №17 СО РАН и проекта №33 программы «Фундаментальные науки – медицине» Президиума РАН.

ЛИТЕРАТУРА:

1. Rosati G. The prevalence of multiple sclerosis in the world: an update. *Neur. Sci.* 2001. V. 22. № 2. P. 117–139.
2. Thompson A.J. Multiple sclerosis--a global disorder and still poorly managed. *Lancet Neurol.* 2008. V. 7. № 12. P. 1078-1079.
3. Kurtzke J.F. Epidemiology and etiology of multiple sclerosis. *Phys. Med. Rehabil. Clin. N. Am.* 2005. V. 16. № 2. P. 327-349.
4. Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>
5. ArrayExpress, <http://www.ebi.ac.uk/arrayexpress/>
6. Oksenberg J.R., Baranzini S.E. Multiple sclerosis genetics--is the glass half full, or half empty? *Nat. Rev. Neurol.* 2010. V. 6. № 8. P. 429-437.
7. Jana A, Pahan K. Sphingolipids in multiple sclerosis. *Neuromolecular Med.* 2010. V. 12. № 4. P. 351-361.
8. Achiron A, Gurevich M. Peripheral blood gene expression signature mirrors central nervous system disease: the model of multiple sclerosis. *Autoimmun Rev.* 2006. V. 5. № 8. P. 517-522.
9. BioUML, <http://www.biouml.org>
10. R/Bioconductor, <http://www.bioconductor.org>
11. Kondrakhin Y.V., Sharipov R.N., Kel A.E., Kolpakov F.A. Identification of differentially expressed genes by meta-analysis of microarray data on breast cancer. *In Silico Biol.* 2008 V. 8. № 5-6. P. 383-411.