

УДК 575.11:616.89-008.454-053.2

ПРИОРИТИЗАЦИЯ ПРОЦЕССОВ-КАНДИДАТОВ ПРИ УМСТВЕННОЙ ОТСТАЛОСТИ И АУТИЗМЕ НА ОСНОВЕ ДАННЫХ МОЛЕКУЛЯРНОГО КАРИОТИПИРОВАНИЯ О ВАРИАЦИЯХ ЧИСЛА КОПИЙ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК

^{1,2}Зеленова М.А., ^{1,2}Ворсанова С.Г., ^{1,2}Юров Ю.Б., ^{1,2}Васин К.С.,
¹Шмитова Н.С., ^{1,2,3}Юров И.Ю.

¹ФГБНУ «Научный центр психического здоровья», Москва;

²Обособленное структурное подразделение «Научно-исследовательский
клинический институт педиатрии имени академика Ю.Е. Вельтищева»

ГБОУ ВПО РНИМУ им. Н.И. Пирогова Минздрава России, Москва;

³Российская медицинская академия непрерывного профессионального
образования Минздрава России, Москва, e-mail: ivan.iourov@gmail.com

Определение патогенности вариаций генома, выявляемых при помощи современных молекулярно-цитогенетических методов, не представляется возможным без биоинформатического анализа. Данные о вариациях числа копий последовательностей ДНК (CNV), исследованные при помощи алгоритмов приоритизации и фильтрации, могут быть использованы для выявления нарушенных биологических процессов, в том числе вносящих вклад в патогенез заболеваний мозга. Определение таких процессов представляет высокую значимость, так как позволяет предложить возможное терапевтическое вмешательство на основании идентифицированных нарушений. В настоящей работе описывается алгоритм приоритизации данных молекулярного кариотипирования, предназначенный для анализа вариаций числа копий последовательностей ДНК (CNV) и их возможных функциональных последствий в контексте изменений в геномных сетях. Алгоритм основан на последовательном применении этапов обработки данных о CNV, полученных с помощью молекулярного кариотипирования. Вышеупомянутый алгоритм был использован в группе из 191 ребенка с умственной отсталостью, аутизмом и врожденными пороками развития. Было идентифицировано 13 кластеров-кандидатов, состоящих из 39 геномных сетей и 475 генов. Показано, что данный алгоритм может применяться как для исследования индивидуальных геномных вариаций, способствуя выбору генетически обоснованной терапевтической тактики, так и для определения молекулярных механизмов болезней мозга.

Ключевые слова: молекулярное кариотипирование, CNV, приоритизация, биоинформатический анализ, головной мозг, умственная отсталость

PRIORITIZATION OF CANDIDATE PROCESSES IN INTELLECTUAL DISABILITY AND AUTISM USING MOLECULAR KARYOTYPING DATA ON COPY NUMBER VARIATIONS

^{1,2}Zelenova M.A., ^{1,2}Vorsanova S.G., ^{1,2}Yurov Yu.B., ^{1,2}Vasin K.S.,
¹Shmitova N.S., ^{1,2,3}Iourov I.Yu.

¹FSBSI «Mental Health Research Center», Moscow;

²Academician Yu.E. Veltishchev Research Clinical Institute of Pediatrics, N.I. Pirogov Russian
National Research Medical University, Ministry of Health of the Russian Federation, Moscow;

³FSBEI FPE «Russian Medical Academy of Continuous Professional Education»
of the Ministry of Healthcare of the Russian Federation, Moscow, e-mail: ivan.iourov@gmail.com

Determining the pathogenicity of genomic variations uncovered by molecular cytogenetic techniques is currently not feasible without bioinformatic analysis. Data on copy number variations (CNV), analyzed using algorithms for prioritization and filtration, may be used to identify disturbed biological processes, specifically those contributing to brain diseases. Identification of such processes is of great importance, since it allows the appointment of therapeutic correction based on the identified data. In this paper, we describe an algorithm for the prioritization of molecular karyotyping data, designed to analyze CNVs and their potential functional consequences in the context of genomic networks disruption. The presented algorithm is based on the consistent application of the processing steps to the initial CNV data obtained after molecular karyotyping. The abovementioned algorithm was tested in a group of 191 children with intellectual disability, developmental delay, autistic spectrum disorder and congenital malformations. This allowed us to identify 13 candidate clusters consisting of 39 genomic networks and 475 genes. The algorithm is to be used both for an individual patient, facilitating the selection of genetically based therapeutic strategies, and in studies aimed at determining the molecular mechanisms of brain disorders.

Keywords: molecular karyotyping, CNV, prioritization, bioinformatic analysis, brain, intellectual disability

Применение современных биоинформатических и молекулярно-генетических технологий показало, что гены, ассоциированные с нарушением работы головного мозга,

часто кодируют белки, взаимодействующие в рамках одной функции или биологического процесса [1]. Количество генов, описанных при нарушениях функционирования мозга

у детей, составляет более 1000 [2–4]. Мутации во многих из них ассоциируют с нарушением синаптических функций, нейрогенеза, нарушении транскрипции и репарации ДНК [5–10]. Отмечая большое количество ассоциированных генов, низкую воспроизводимость результатов относительно рекуррентных мутаций в различных исследованиях, а также феномен «бремени мутаций» (изменения в отдельных генах не приводят к значимому эффекту, однако большое количество мутаций в этих же генах, например, одной геномной сети, может привести к нарушению ее функционирования), различные авторы обращают внимание на значимость выявления процессов-кандидатов для разработки потенциальной терапии [8, 11]. В качестве примера можно привести сигнальную сеть Notch, которая играет роль в развитии органов и тканей в целом, но также вовлечена в поддержание клеток-предшественников эмбриональной нервной системы, способствует дифференциации астроцитов, а также участвует в регуляции функционирования нервной системы во взрослом возрасте. Нарушения в данной геномной сети ассоциированы, в частности, с лобно-височной лобарной дегенерацией, нарушениями обучения и памяти (у мышей). При нарушениях этого геномного пути предлагаются два основных направления терапии: блокировка сигнального пути (на уровне лигандов и/или рецепторов; на уровне транскрипционного комплекса) или его активация [12, 13]. Определение процессов-кандидатов является значимым направлением в современной медицинской генетике, предлагающим новые возможности для терапии генетически обусловленных заболеваний [14]. Таким образом, в настоящее время актуальность исследования вариаций генома, приводящих к изменению биологических процессов и вызывающих нарушения функционирования головного мозга, не вызывает сомнений.

Цель исследования

Целью настоящей работы являлись описание и апробация алгоритма приоритизации процессов-кандидатов заболеваний мозга на основе анализа данных о CNV, полученных с помощью молекулярного кариотипирования.

Материалы и методы исследования

Были исследованы образцы ДНК, выделенной из лимфоцитов периферической крови 374 детей с аутистическими расстройствами и/или задержкой психомоторного/психоречевого развития или умственной отсталостью, врожденными пороками и/или микроаномалиями развития. Молекулярное кариотипирование проводилось при помощи методов SNP array (352 пациента) и array CGH (22 пациента),

с применением микроматриц (чипов) Affymetrix CytoScan HD и Nimblegen 12×135K соответственно. Биоинформатический анализ проводился по описанному ранее протоколу [15, 16].

Результаты исследования и их обсуждение

В настоящей работе описывается алгоритм интерпретации данных молекулярного кариотипирования, который позволяет на основе пошагового применения стратегий приоритизации, ранжирования, составления интерактома и выявления геномных сетей получить кластеры процессов-кандидатов, ассоциированных с определенными фенотипическими проявлениями заболевания. В группе из 374 пациентов с нарушением функционирования головного мозга (ГМ), выражающимся как расстройство аутистического спектра и/или задержка психомоторного/психоречевого развития или умственная отсталость с врожденными пороками и/или микроаномалиями развития, патогенных и вероятно патогенных вариаций генома не было обнаружено в 26 (7%) случаях. Патология генома в виде анеуплоидии, структурных аномалий хромосом, сложных хромосомных перестроек, потери гетерозиготности, характерной для детей от кровнородственных браков [17], а также потери гетерозиготности в импринтированных генах, были выявлены в 157 случаях (42%). Данные случаи были исключены из дальнейшего исследования. В 51% случаев (191 пациент) были выявлены CNV, являющиеся патогенными или вероятно патогенными и затрагивающие от 1 до 500 тыс. пн. Для изучения молекулярных механизмов нарушения психики у детей из данной группы был предложен алгоритм интерпретации данных молекулярного кариотипирования [8, 15]. Алгоритм основан на пошаговом применении стратегий приоритизации, ранжирования, составления интерактома и выявления геномных сетей, и позволяет в результате получить набор процессов-кандидатов и кластеров процессов, предположительно вносящих вклад в определенное заболевание (рисунок). Выявление CNV, имеющих значение для фенотипических проявлений, проводилось с помощью анализа следующих параметров: рекуррентность; упоминание в базах данных патогенных и непатогенных геномных вариаций; определение генов, локализованных в участках CNV; функций данных генов и заболеваний, ассоциированных с CNV/генами. Далее анализировалась экспрессия генов, локализованных в участках патогенных и вероятно патогенных CNV, в нервных клетках. Следующим шагом являлось выявление взаимодействий генов, имеющих повышен-

ную экспрессию в клетках головного мозга. На основании выявленных взаимодействий составлялся так называемый «объединенный интерактом», включающий в себя все взаимодействующие элементы из исследуемого на данном этапе массива. Для всех генов объединенного интерактома определяли геномные сети при помощи нескольких баз данных. Далее полученные геномные сети ранжировали, основываясь на соотношении общего количества генов, входящих в геномную сеть, и генов, задействованных в этой геномной сети в исследуемой выборке. Исходя из результатов предыдущего шага, выявляли значимые геномные сети, которые объединяли в кластеры по функции и по вовлеченности в подобные процессы. Вышеобозначенные кластеры рассматривали в качестве элементов патогенетического каскада молекулярных и клеточных процессов, являющегося причиной нарушений психики при болезнях мозга (таблица).

Применение данного алгоритма к исследуемой группе из 191 пациента позволило показать, что недифференцированное расстройство аутистического спектра и умственная отсталость могут быть ассоциированы с нарушением определенных кластеров процессов, геномных сетей и генов. Было выявлено 475 генов, входящих в 39 геномных сетей и составляющих 13 кластеров процессов и «молекулярных путей» (pathways): «функциональное везикулярное», «репарация ДНК», «транскрипция», «нейродегенератив-

ные заболевания», «сигнальный путь ErbB», «протеасома», «В-лимфоциты», «макромолекулярные взаимодействия», «функциональное актина», «митоз», «старение», «сигнальная сеть Notch», «сигнальная сеть TP53».

Наибольшее количество генов (35 из 475) было вовлечено в кластер «Протеасома». Протеасома выполняет значимую функцию, заключающуюся в разложении отработанных или поврежденных белков путем протеолиза. Нарушение функции протеасомных комплексов вызывает снижение протеолитической активности, в результате чего накапливаются поврежденные или неправильно сформированные белки. Это может приводить к развитию характерных изменений при нейродегенеративных, сердечно-сосудистых и аутоиммунных заболеваниях, а также при воспалительных реакциях и системных ответах на повреждение ДНК [18, 19]. В кластер «Нейродегенеративные заболевания» входят гены, ассоциированные с различными тяжелыми заболеваниями и выполняющие значимые биологические функции. Ген *DCTN1* вовлечен в формирование веретена деления и аксонов; ген *FUS* участвует в регуляции экспрессии генов и поддержании целостности генома; ген *CDK5* вносит вклад в различные процессы, такие как синаптическая пластичность и миграция нейронов; ген *GRN* регулирует рост клетки, а *OPTN* участвует в транспорте веществ через мембрану. Геномная сеть p53 состоит из генов, реагирующих на широкий спектр сигналов стресса.

Использованные электронные ресурсы

Номер этапа	Использованные ресурсы	Ссылка на ресурс
1	UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly	http://genome.ucsc.edu/
	OMIM (Online Mendelian Inheritance in Man)	https://www.omim.org/
	NCBI Map Viewer	https://www.ncbi.nlm.nih.gov/mapview/
	Database of Genomic Variants – catalogue of structural variation in the human genome	http://dgv.tcag.ca/dgv/app/home
2	BioGPS (Genomics Institute of the Novartis Research Foundation)	http://biogps.org/#goto=welcome
	Genatlas (Genatlas Universite Paris Descartes)	http://genatlas.medecine.univ-paris5.fr/
3.	STRING.db (Search Tool for the Retrieval of Interacting proteins database)	http://string.db.org/
	BioGRID (Biological General Repository for Interaction Datasets)	https://thebiogrid.org/
	NCBI gene	https://www.ncbi.nlm.nih.gov/gene
5.	KEGG (Kyoto Encyclopedia of Genes and Genomes)	http://www.genome.jp/kegg/
	Gene Ontology (Gene Ontology Consortium)	http://www.geneontology.org/
	REACTOME	http://www.reactome.org/
	NCBI biosystems	https://www.ncbi.nlm.nih.gov/Structure/biosystems/docs/biosystems_about.html



Алгоритм приоритизации данных молекулярного кариотипирования

Изучение молекулярных сетей, в которые вовлечены белки, ассоциированные с умственной отсталостью, позволило продемонстрировать, что многие из них принадлежат к подробно описанным сигнальным путям [20].

Рассматривая потенциальные подходы к лечению, можно обратить внимание на применение методов, модулирующих их активность. Например, большая часть мутаций, влияющих на малые ГТФазы, приводит к гиперактивации пути транспорта данных ферментов. Уменьшение трансдукции сигнала возможно при использовании либо общих ингибиторов малых ГТФаз, либо более специфических ингибиторов,

направленных на нижележащие киназы [21, 22]. Для уменьшения уровня активации Ras у модельных животных с нейрофиброматозом первого типа было успешно использовано ингибирование фарнезил-трансферазы. Данное заболевание является комплексным нервно-психическим нарушением, обусловленным потерей функции RasGAP, NF1 [23]. Мутации в других генах, ассоциированных с умственной отсталостью, могут инактивировать геномные сети типа APIX/Rac/PAK3/LIMK или IL1RAPL1/JNK. В этих случаях терапевтические стратегии могут быть связаны с активацией нижележащих эффекторов. Многие геномные пути, ассоциированные с нарушением функционирования ГМ,

также вовлечены в развитие онкологических заболеваний, и разработка терапии при их нарушении проводится прежде всего в этом направлении. Например, активирующие мутации в генах, участвующих в геномной сети EgrB (члены семейства EGFR), приводят к гиперактивации сигнального пути PI3K-AKT-mTOR, контролирующего клеточный метаболизм, пролиферацию и подвижность. Ингибирование данного сигнального пути восстанавливает рецепторную активность и компенсирует отсутствие нисходящих сигналов от рецептора [24].

Необходимо отметить, что, имея представление о механизмах нарушений психики, определенных с помощью биоинформатического анализа, можно рассматривать две основные платформы для дальнейшей разработки терапевтических стратегий: (1) изучение структурных изменений белков, вовлеченных в нарушенные «молекулярные пути» (pathways); (2) оценка эффекта мутаций на определенные биологические процессы или их каскадов. Таким образом, изучение геномных сетей и взаимодействия генов является необходимым для полноценного анализа результатов молекулярного кариотипирования, что делает биоинформатический анализ неотъемлемой частью интерпретации его результатов [25].

Выводы

Показано, что недифференцированная умственная отсталость и расстройство аутистического спектра могут быть ассоциированы с нарушением 13 кластеров процессов и геномных сетей (нейродегенерация, функционирование актина, функционирование протеасомы, сигнальный путь ERBB, регуляция транскрипции, сигнальная сеть TP53, сигнальная сеть Notch, старение, митотическое деление клетки, репарация ДНК, функционирование везикул, макромолекулярные взаимодействия, В-лимфоциты). Применение методов молекулярного кариотипирования совместно с биоинформатическим анализом позволяет не только выявлять хромосомные и эпигенетические аномалии, но и определять процессы, лежащие в основе недифференцированных форм заболеваний, связанных с патологией головного мозга, что способствует разработке терапевтических стратегий индивидуального лечения. Применение данного метода обладает высокой эффективностью для выявления процессов, лежащих в основе недифференцированных форм заболеваний, связанных с патологией головного мозга.

Список литературы

1. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities // *Annual review of genetics*. – 2011. – Т. 45. – P. 81–104.
2. Gonzalez-Mantilla A.J., Moreno-De-Luca A., Ledbetter D.H., Martin C.L. A cross-disorder method to identify novel candidate genes for developmental brain disorders // *JAMA psychiatry*. – 2016. – № 73(3). – P. 275–283.
3. Feuk L., Carson A.R., Scherer S.W. Structural variation in the human genome // *Nat. Rev. Genet.* – 2006. – № 7. – P. 85–97.
4. Grayton H.M., Fernandes C., Rujescu et al. Copy number variations in neurodevelopmental disorders // *Progress in neurobiology*. – 2012. – № 99(1). – P. 81–91.
5. Chelly J., Khelifaoui M., Francis F., Cherif B., Bienvu T. Genetics and pathophysiology of mental retardation // *European Journal of Human Genetics*. – 2006. – № 14(6). – P. 701–713.
6. Kaindl A.M., Passemard S., Kumar P. et al. Many roads lead to primary autosomal recessive microcephaly // *Progress in Neurobiology*. – 2010. – № 90. – P. 363–383.
7. Iourov I.Y., Vorsanova S.G., Zelenova M.A. et al. Genomic copy number variation affecting genes involved in the cell cycle pathway: implications for somatic mosaicism // *International Journal of Genomics*. – 2015. – V. 2015. – Article ID 757680.
8. Yurov Y., Vorsanova S., Iourov I. Network-Based Classification of Molecular Cytogenetic Data // *Current Bioinformatics*. – 2017. – Т. 12, № 1. – С. 27–33.
9. Gissen P., Maher E. R. Cargos and genes: insights into vesicular transport from inherited human disease // *Journal of Medical Genetics*. – 2007. – Т. 44. – №9. – P.545–555.
10. van Reeuwijk J., Brunner H.G., van Bokhoven H. Glyc-O-genetics of Walker-Warburg syndrome // *Clinical Genetics*. – 2005. – № 67. – P. 281–289.
11. Oti M., Huynen M.A., Brunner H.G. Phenome connections // *Trends in Genetics*. – 2008. – № 24. – P. 103–106.
12. Siebel C., Lendahl U. Notch signaling in development, tissue homeostasis, and disease // *Physiological reviews*. – 2017. – Т. 97, № 4. – P.1235–1294.
13. Aste-Amézaga M., Zhang N., Lineberger J.E. et al. Characterization of Notch1 antibodies that inhibit signaling of both normal and mu-tated Notch1 receptors // *PLoS One*. – 2010. – 5: e9094.
14. Heng H., Regan S. A systems biology perspective on molecular cytogenetics // *Current Bioinformatics*. – 2017. – Т. 1, № 12(1). – P. 4–10.
15. Iourov I.Y., Vorsanova S.G., Yurov Y.B. In silico molecular cytogenetics: a bioinformatic approach to prioritization of candidate genes and copy number variations for basic and clinical genome research // *Molecular Cytogenetics*. – 2014. – № 7(1). – P. 98.
16. Юров И.Ю., Ворсанова С.Г., Зеленова М.А. и др. Биоинформатическая технология оценки функциональных последствий геномных вариаций // *Фундаментальные исследования*. – 2015. – № 2–19. – С. 4209–4214.
17. Iourov I. Y., Vorsanova S.G., Korostelev S.A et al. Long contiguous stretches of homozygosity spanning shortly the imprinted loci are associated with intellectual disability, autism and/or epilepsy // *Molecular cytogenetics*. – 2015. – Т. 8, № 1. – С. 77.
18. Lodish H., Berk A., Matsudaira P. et al. *Molecular cell biology* (5th ed.). – N.Y.: W.H. Freeman and CO. – 2004. – № 3. – P. 66–72.
19. Schmidt M., Finley D. Regulation of proteasome activity in health and disease // *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. – 2014. – Т. 1843, № 1. – P. 13–25.
20. Chen Y., Wang W., Zhou Y. et al. In silico gene prioritization by integrating multiple data sources // *PLoS One*. – 2011. – № 6(6). – P.21137.
21. Pavlowsky A., Chelly J., Billuart P. Emerging major synaptic signaling pathways involved in intellectual disability // *Molecular psychiatry*. – 2012. – Т. 17, № 7. – С. 682–693.
22. Giannandrea M., Bianchi V., Mignogna M.L. et al. Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly // *American Journal of Human Genetics*. – 2010. – № 86. – P. 185–195.
23. Shilyansky C., Lee Y.S., Silva A.J. Molecular and cellular mechanisms of learning disabilities: a focus on NF1 // *Annual reviews of neuroscience*. – 2010. – № 33. – P. 221–243.
24. Wee P., Wang Z. Epidermal growth factor receptor cell proliferation signaling pathways // *Cancers*. – 2017. – Т. 9. – № 5. – С. 52.
25. Iourov I.Y., Vorsanova S.G., Voinova V.Y., Yurov Y.B. 3p22.1p21.31 microdeletion identifies CCK as Asperger syndrome candidate gene and shows the way for therapeutic strategies in chromosome imbalances // *Molecular Cytogenetics*. – 2015. – № 8. – P. 82.