

Выявление дифференциально метилированных при бронхиальной астме и туберкулезе легких генов (реанализ данных)

Бабушкина Н.*, Огородников С., Зарубин А., Брагина Е.

Научно-исследовательский институт медицинской генетики, Томский НИИЦ РАН, Томск, Россия

* nad.babushkina@medgenetics.ru

Ключевые слова: астма; туберкулез; метилирование ДНК

Мотивация и цель: Феномен дистропии, характеризующийся редким сочетанием на фенотипическом уровне двух патологий у одного индивида, сложен для изучения. Ряд исследований свидетельствует о возможной дистропии аллергических болезней и туберкулеза (см., например, [1]). Так, показано, что диагноз и астмы, и туберкулеза у одного индивида встречается в разных популяциях с частотой не более чем 1–5 % [1, 2]. Иммунологические характеристики астмы и туберкулеза показывают, что в основе дистропии «астма-туберкулез» может лежать разнонаправленная поляризация иммунного ответа с преимущественной активностью Th2 при астме и Th1 при туберкулезе [3, 4]. Изучение изменчивости регуляции генов при этих патологиях дает дополнительную информацию, важную для понимания феномена дистропии «астма-туберкулез».

Методы и алгоритмы: Проведен реанализ двух наборов данных метилирования ДНК, представленных в общедоступном репозитории функциональных геномных данных Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). GSE118469 получен при изучении туберкулеза у жителей Тайваня: пациенты с туберкулезом ($n = 12$) без сопутствующих инфекционных патологий (в том числе ВИЧ) и злокачественных новообразований, контрольная группа – здоровые индивиды ($n = 6$) [5]. Из набора данных GSE104471 была использована информация о профилях метилирования у 12 пациентов с аллергической бронхиальной астмой и 12 здоровых индивидов без астмы. Все участники этого исследования являются взрослыми некурящими европеоидами неиспаноязычного происхождения [6]. Анализ метилирования в обоих исследованиях проведен на чипах Infinium HumanMethylation450K BeadChip v1.2 [5, 6]. В реанализ из обоих наборов данных взята информация о метилировании ДНК из мононуклеаров периферической крови.

Реанализ проводили с использованием языка программирования R-4.3.3 (R Foundation, Австрия) в среде разработки RStudio-2021.09.1-372 (Rstudio PBC, США). Из наборов данных GSE118469 и GSE104471 был создан общий массив данных с информацией о метилировании ДНК как у больных с астмой и с туберкулезом, так и в контрольных образцах. Далее с применением программного пакета minfi [7] полученный массив подвергали квантильной нормализации, удаляли слабо сработавшие зонды и пробы полиморфных вариантов с частотой встречаемости минорного аллеля больше 5 %, исключали данные с половых хромосом. Поиск дифференциально метилированных у больных астмой и

туберкулезом CpG-сайтов осуществляли при помощи limma [8]. Для исключения межэтнических различий этнос использовался в качестве дополнительной группирующей переменной, после получения взвешенной оценки по каждому из признаков объясняемая ими дисперсия вычиталась. Сходный подход использовали и при деконволюции по клеточному составу, проведенной с использованием программного пакета EpiDISH [9].

Результаты: В результате был получен список CpG-сайтов, дифференциально метилированных (ДМС) у больных с двумя патологиями – астмой и туберкулезом. Поправку на множественные сравнения ($\text{adj } P < 0.05$) прошли 6653 сайта; из них $|\Delta\beta| > 20\%$ выявляется для 62 сайтов. Прошедшие оба фильтра сайты относятся к 53 генам (ДМГ).

Анализ сверхпредставленности (over-representation analysis, ORA)) с помощью Web Gestalt (<https://www.webgestalt.org/>) показал вовлеченность этих генов в биологические процессы (в категории GO Biological Processes), связанные с иммунным ответом (сигнальный путь рецепторов клеточной поверхности; сигнальный путь, опосредованный антигенным рецептором; позитивная регуляция CD8+; ответ на цитокины; продукция цитокинов; адаптивный иммунный ответ; дифференцировка альфа-бета-Т-клеток, иммунный ответ, активация альфа-бета-Т-клеток), три из них проходят поправку FDR: активация альфа-бета-Т-клеток (GO:0046631, $p = 0.010031$), иммунный ответ (GO:0006955, $p = 0.010298$), дифференцировка альфа-бета-Т-клеток (GO:0046632, $p = 0.010298$). В эти три пути вовлечены в общей сложности 17 генов (*CX3CR1*, *CD247*, *CYFIP1*, *ETS1*, *FASLG*, *IGSF6*, *IL19*, *LY9*, *HLA-E*, *PRDM1*, *PVRIG*, *RFTN1*, *RORA*, *RUNX1*, *SYK*, *SURF4*, *ZC3HAV1*), причем 5 из них (*LY9*, *PRDM1*, *RORA*, *RUNX1*, *SYK*) участвуют во всех трех биологических процессах.

Согласно результатам обогащения, полученным с помощью ресурса Metascape (metascape.org), наиболее представленными процессами являются также активация альфа-бета-Т-клеток (GO:0046631) и регуляция активации альфа-бета-Т-клеток (GO:0046634), цитокиновый сигналинг в иммунной системе (Reactome Gene Sets: R-HSA-1280215), цитоксичность, опосредованная NK-клетками (KEGG Pathway: hsa04650), путь дифференцировки клеток Th17 (WikiPathways: WP5130). Анализ ассоциаций с патологиями и физиологическими состояниями (по Metascape) показал связь выявленных ДМГ с количеством эозинофилов и лимфоцитов, средним уровнем гемоглобина, заболеваниями иммунной системы (в том числе – синдром Бехчета, васкулит, первичный синдром Шегрена, иммунная тромбоцитопеническая пурпура, обыкновенная волчанка, красная волчанка, дискоидная красная волчанка, болезнь Либмана–Сакса, астма у взрослых, реакция «трансплантат против хозяина») и связанными с ними заболеваниями (лимфаденопатия, моноклональные гаммапатии), а также с карциномой яичников, лимфоидным лейкозом, венозной тромбоэмболией, отеком суставов. Помимо этого, выявленные ДМГ ассоциированы в первую очередь с CD56+ NK-клетками, селезенкой и тимусом (по результатам анализа в Metascape).

Ресурс FUMA GWAS (<https://fuma.ctglab.nl/>) указывает на наличие ассоциаций с различными патологиями/физиологическими показателями для 19 из выявленных ДМГ (*AFF3*, *ANKH*, *APBA2*, *CX3CR1*, *ETS1*, *FASLG*, *GIMAP4*, *HLA-E*, *IKZF1*, *IL19*, *MPHOSPH9*, *RORA*, *SLA*, *SPATA13*, *SYK*, *TG*, *UBAC2*, *ZC3HAV1*, *CD247*). Так, выявлены ассоциации с эозинофилами (количество/процентное содержание среди белых клеток), тромбокрит, соотношение альбумин-глобулин; температур;

рассеянный склероз, эрозия костей при ревматоидном артрите, аутоиммунные заболевания щитовидной железы, применение медикаментов (препараты для щитовидной железы, адренергические средства). Кроме того, показаны ассоциации с астмой: как в целом (*CD247, FASLG, AFF3, ANKH, IKZF1, TG, ETS1, UBAC2, RORA*), так и с аллергической формой (*FASLG, IKZF1, ETS1, RORA*).

Следует отметить, что распределение ДМС не симметрично – все 62 сайта с $|\Delta\beta| > 20\%$ имеют отрицательное значение. При расчете $\Delta\beta$ проводилось вычитание среднего уровня метилирования у больных с туберкулезом из среднего уровня метилирования у больных с астмой. Поэтому можно утверждать, что у пациентов с астмой происходит относительное снижение уровня метилирования ДМС, что может свидетельствовать об усилении активности ДМГ. Поскольку связь между уровнем экспрессии и уровнем метилирования нелинейна и к тому же различается в зависимости от расположения CpG-сайтов в пределах гена, напрямую говорить об увеличении экспрессии ДМГ будет опрометчивым.

Выводы: Полученные нами данные свидетельствуют о том, что при развитии таких дистропных патологий, как астма и туберкулез, происходит изменение активности прежде всего именно тех генов, которые вовлечены как в развитие иммунного ответа в целом, так и в активацию и дифференциацию альфа-бета-Т-клеток. Полученные данные полностью укладываются в гипотезу поляризации Т-клеточного иммунного ответа при дистропии «астма-туберкулез».

Финансирование: Работа выполнена в рамках государственного задания Министерства науки и высшего образования (№ 122020300041-7).

Identification of differentially methylated genes in bronchial asthma and pulmonary tuberculosis (re-analysis of data)

Babushkina N.*, Ogorodnikov S., Zarubin A., Bragina E.

Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

* nad.babushkina@medgenetics.ru

Key words: asthma; tuberculosis; DNA methylation

Motivation and Aim: The phenomenon of dystropy is characterized by a rare combination at the phenotypic level of two pathologies in one individual and is difficult to study. A number of studies indicate a possible dystrophy of allergic diseases and tuberculosis (see, for example, [1]). Thus, it has been shown that a combination of both asthma and tuberculosis in one individual occurs in different populations with a frequency no more than 1–5 % [1, 2]. The immunological characteristics of both diseases show that the basis of the dystrophy “asthma-tuberculosis” may be multidirectional polarization of the immune response with predominant activity of Th2 in asthma and Th1 in tuberculosis [3, 4]. Studying the variability of gene regulation in these pathologies provides important information for understanding the phenomenon of “asthma-tuberculosis” dystrophy.

Methods and Algorithms: Two DNA methylation datasets were re-analyzed from the Gene Expression Omnibus public functional genomic data repository (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). Microarray dataset GSE118469 was obtained in a study of tuberculosis in residents of Taiwan: patients with tuberculosis ($n = 12$) without concomitant infectious (including HIV) and malignant neoplasms, the control group – healthy individuals ($n = 6$) [5]. We analyzed information on methylation profiles of

12 patients with allergic bronchial asthma and 12 healthy individuals without asthma from GSE104471 dataset. All participants are non-Hispanic white non-smoking adults [6]. Infinium HumanMethylation450K BeadChip v1.2 was used for methylation analysis in both studies [5, 6]. In the re-analysis, DNA methylation data from PBMCs are taken from both datasets.

Re-analysis was carried out using the R-4.3.3 programming language (R Foundation, Austria) in the RStudio-2021.09.1-372 development environment (Rstudio PBC, USA). Data from datasets GSE118469 and GSE104471 were integrated in common dataset with information on DNA methylation in both patients with asthma and tuberculosis, and in control samples. Next, using the minfi software package [7], the obtained dataset was subjected to quantile normalization. Weakly working probes and samples of polymorphic variants with a minor allele frequency of more than 5 % were removed and data from sex chromosomes were excluded. The differentially methylated CpG sites in patients with asthma and tuberculosis were searched using limma [8]. To exclude interethnic differences, ethnicity was used as an additional grouping variable; after obtaining a weighted estimate for each of the attributes, the variance explained by them was subtracted. A similar approach was used for deconvolution by cellular composition, carried out using the EpiDISH software package [9].

Results: As a result, a list of differentially methylated CpG sites (DMS) was obtained in patients with two pathologies – asthma and tuberculosis. Adjustment for multiple comparisons ($\text{adj_P} < 0.05$) passed 6653 sites; of these, $|\Delta\beta| > 20\%$ are identified for 62 sites. The sites that passed both filters belong to 53 genes (DMG).

Over-Representation analysis (ORA) using Web Gestalt (<https://www.webgestalt.org/>) showed the involvement of these genes in biological processes (in the GO Biological Processes category) associated with the immune response; activating cell surface receptor signaling pathway; antigen receptor-mediated signaling pathway; positive regulation of CD8-positive; alpha-beta T cell activation; response to cytokine; cytokine production; adaptive immune response; alpha-beta T cell differentiation; immune response), three of them undergo FDR correction: alpha-beta T cell activation (GO:0046631, $p = 0.010031$), immune response (GO:0006955, $p = 0.010298$), alpha-beta T cell differentiation (GO:0046632, $p = 0.010298$). A total of 17 genes are involved in these three pathways (*CX3CR1*, *CD247*, *CYFIP1*, *ETS1*, *FASLG*, *IGSF6*, *IL19*, *LY9*, *HLA-E*, *PRDM1*, *PVRIG*, *RFTN1*, *RORA*, *RUNX1*, *SYK*, *SURF4*, *ZC3HAV1*), and 5 of them (*LY9*, *PRDM1*, *RORA*, *RUNX1*, *SYK*) are involved in all three biological processes. According to the enrichment results obtained using the Metascape resource (metascape.org), the most represented processes are also alpha-beta T cell activation (GO:0046631), as well as regulation of alpha-beta T cell activation (GO:0046634), Cytokine Signaling in Immune system (Reactome Gene Sets: R-HSA-1280215), Natural killer cell mediated cytotoxicity (KEGG Pathway: hsa04650), Th17 cell differentiation pathway (WikiPathways: WP5130). Analysis of associations with pathologies and physiological conditions (according to Metascape) showed the connection of identified DMGs with Eosinophil and Lymphocyte count, Mean Hemoglobin, Immune System Diseases (including Behcet Syndrome, Vasculitis, Primary Sjögren's syndrome, Immune thrombocytopenic purpura, Lupus Vulgaris, Lupus Erythematosus, Discoid, Lupus Erythematosus, Libman-Sacks Disease, Adult onset asthma, Graft-vs-Host Disease) and related diseases (Lymphadenopathy, Monoclonal Gammopathies), as well as Ovarian Carcinoma, Lymphoid leukemia, Venous Thromboembolism, Joint swelling. In

addition, the identified DMGs are associated primarily with Cell-specific: CD56+ NKCells, spleen and thymus (based on Metascape analysis).

The FUMA GWAS resource (<https://fuma.ctglab.nl/>) indicates associations with various pathologies/physiological indicators for 19 of the revealed DMGs (*AFF3*, *ANKH*, *APBA2*, *CX3CR1*, *ETS1*, *FASLG*, *GIMAP4*, *HLA-E*, *IKZF1*, *IL19*, *MPHOSPH9*, *RORA*, *SLA*, *SPATA13*, *SYK*, *TG*, *UBAC2*, *ZC3HAV1*, *CD247*). Thus, associations were identified with Eosinophil (counts/percentage of white cells), plateletcrit, Albumin-globulin ratio; Temperament; Multiple sclerosis, Bone erosion in rheumatoid arthritis, Autoimmune thyroid disease, Medication use (thyroid preparations, adrenergics). In addition, associations with asthma are shown: both in general (*CD247*, *FASLG*, *AFF3*, *ANKH*, *IKZF1*, *TG*, *ETS1*, *UBAC2*, *RORA*) and with the allergic form (*FASLG*, *IKZF1*, *ETS1*, *RORA*).

Of note that the DMS distribution is not symmetric – all 62 sites with $|\Delta\beta| > 20\%$ have a minus value. When calculating $\Delta\beta$, the average methylation level in patients with tuberculosis was subtracted from the average methylation level in patients with asthma. Therefore, it can be argued that in patients with asthma there is a relative decrease in the level of DMS methylation, which may indicate increased DMG activity. Since the relationship between the level of expression and the level of methylation is not linear, moreover, it differs depending on the location of CpG sites within the gene, it would be rash to directly talk about an increase in DMG expression.

Conclusion: Our data directly indicate that in development of dystropic pathologies such as asthma and tuberculosis, there is a change in the activity primarily of those genes that are involved both in the development of the immune response in general and in alpha-beta T cell activation and differentiation. The data obtained are fully fit into the hypothesis of polarization of the T-cell immune response in “asthma-tuberculosis” dystropy.

Funding: This work was supported by the Ministry of Science and Higher Education (project No. 122020300041-7).

Список литературы/References

1. Fekih L. et al. Asthme et tuberculose: association bénéfique ou maléfique? [Tuberculosis in patients with asthma]. *Rev Mal Respir.* 2010;27(7):79-684. doi 10.1016/j.rmr.2010.06.010
2. Cernat T. et al. Simultaneous occurrence of other diseases among prison inmates with tuberculosis. *Curr Health Sci J.* 2010;36(3):143-147
3. Brighenti S., Ordway D.J. Regulation of immunity to tuberculosis. *Microbiol Spectr.* 2016;4(6). doi 10.1128/microbiolspec.TB2-0006-2016
4. Hammad H., Lambrecht B.N. The basic immunology of asthma. *Cell.* 2021;184(9):2521-2522. doi 10.1016/j.cell.2021.04.019. Erratum for: *Cell.* 2021;184(6):1469-1485
5. Chen Y.C. et al. Whole genome dna methylation analysis of active pulmonary tuberculosis disease identifies novel epigenotypes: PARP9/miR-505/RASGRP4/GNG12 gene methylation and clinical phenotypes. *Int J Mol Sci.* 2020;21(9):3180. doi 10.3390/ijms21093180
6. Yang I.V. et al. The nasal methylome: a key to understanding allergic asthma. *Am J Respir Crit Care Med.* 2017;195(6):829-831. doi 10.1164/rccm.201608-1558LE
7. Aryee M.J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30(10):1363-1369. doi 10.1093/bioinformatics/btu049
8. Smyth G.K. Limma: linear models for microarray data. In: Gentleman R. (Ed.). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* 2005;397-420. doi 10.1007/0-387-29362-0_23
9. Teschendorff A.E. et al. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics.* 2017;18(1):105. doi 10.1186/s12859-017-1511-5