

Анализ механизмов большого депрессивного расстройства на основе интеллектуального анализа текстов

Тарасова О.^{1*}, Бизюкова Н.¹, Сухачёв В.¹, Иванов С.^{1, 2}, Филимонов Д.¹,
Поройков В.¹

¹ Институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия

² Российский национальный исследовательский медицинский университет им. Н.И. Пирогова, Москва, Россия

* olga.a.tarasova@gmail.com

Ключевые слова: большое депрессивное расстройство; патогенез; интеллектуальный анализ текстов; эпигенетическая регуляция; антидепрессанты

Мотивация и цель: Депрессивные расстройства приводят к снижению качества жизни миллионов людей. Совокупность факторов, ведущих к возникновению и развитию этого психического заболевания, включает в себя генетическую предрасположенность, факторы среды (в том числе стресс и травмы), а также наличие соматических заболеваний [1]. В недавних исследованиях было показано, что ранняя психическая травматизация ребенка приводит к изменению эпигенетической регуляции генов [2], что может оказывать влияние на развитие ряда психических заболеваний, включая большое депрессивное расстройство (БДР). Целью нашей работы является исследование совокупности причин большого депрессивного расстройства с применением методов интеллектуального анализа текстов.

Методы и алгоритмы: С применением ранее разработанных алгоритмов отбора аннотаций релевантных научных публикаций, извлечения наименований биологических и химических объектов из их текстов и анализа взаимосвязи между извлеченными объектами [3, 4] нами была выявлена совокупность наименований генов, белков, микроРНК, которые потенциально ассоциированы с БДР, и определен ряд молекулярных механизмов этого заболевания, а также причин, не относящихся к изменениям на молекулярном уровне. Отбор релевантных публикаций производился на основе ключевых слов по запросу в БД библиографической информации PubMed. Распознавание наименований проводилось с применением метода на основе условных случайных полей (CRF) [4], библиотеки HunFlair для языка Python [5], регулярных выражений и словарей. Исследование связей между терминами проводилось на основе классификации предложений, содержащих и не содержащих указание на наличие взаимосвязи между терминами, с использованием наивного байесовского классификатора, применение которого к задаче распознавания наименований было исследовано нами ранее [3].

Результаты: По результатам запросов в БД PubMed было извлечено свыше 3000 записей, соответствующих аннотациям текстов публикаций. По результатам извлечения ассоциаций между объектами наименований белков, генов и микроРНК и их влияния на развитие БДР были идентифицированы гены дофаминергической и серотонинергической систем, глюкокортикоидного рецептора, изменение уровня экспрессии которых может быть обусловлено

воздействием неблагоприятных обстоятельств (например, сенсорная депривация) в раннем детском возрасте и, таким образом, может непосредственно влиять на развитие депрессии при воздействии неблагоприятных факторов во взрослом возрасте (формирование порочного круга). Дополнительно, были выявлены взаимосвязи между эпигенетическими механизмами регуляции рецепторов эстрогена, которые могут вносить вклад в развитие психических заболеваний. При распознавании в текстах наименований факторов, которые вносят вклад в развитие БДР, но не могут быть отнесены к молекулярным механизмам развития депрессии, нами были найдены также такие факторы, как тип привязанности ребенка, наличие в его близком окружении благоприятного объекта (близкого взрослого), а также наличие сопутствующих расстройств, таких как пограничное расстройство личности или тревожно-депрессивные расстройства, которые могут как сопровождаться более тяжелой симптоматикой, так и обосновывать необходимость дифференциальной диагностики у пациентов с БДР.

Выводы: С применением интеллектуального анализа текстов исследованы потенциальные механизмы развития большого депрессивного расстройства. Выявленные причины включают в себя совокупность биологических и психологических факторов. Некоторые из найденных молекулярных механизмов, такие как регуляция генов дофаминергической системы, подтверждены в более ранних исследованиях, в то время как идентифицированные взаимосвязи между регуляцией эстрогенового и глюкокортикоидного рецепторов и развитием депрессии у взрослых, а также оценка соотношения психологических и наследственных факторов в динамике у пациентов с большим депрессивным расстройством представляют интерес.

Финансирование: Работа выполнена при поддержке гранта Российского научного фонда № 24-25-00453.

An analysis of major depressive disorders using automated text analysis

Tarasova O.^{1*}, Biziukova N.¹, Sukhachev V.¹, Ivanov S.^{1, 2}, Filimonov D.¹, Poroikov V.¹

¹ *Institute of Biomedical Chemistry, Moscow, Russia*

² *Pirogov Russian National Research Medical University, Moscow, Russia*

* olga.a.tarasova@gmail.com

Key words: major depressive disorders; text mining; pathogenesis; epigenetic regulation; antidepressants

Motivation and Aim: Major depressive disorder (MDD) can have a significant impact on the quality of life for millions of people. The development of MDD may be influenced by genetic and environmental factors, such as stress and trauma, as well as the comorbidities of a patient [1]. In recent studies, an influence of early childhood psychological trauma on epigenetic gene regulation was shown [2]. The purpose of our study is to examine the impact of several biological and psychological factors on mental disorders, such as MDD, using text-mining approach.

Methods and Algorithms: Based on earlier developed algorithms of extracting biological and chemical named entities from the texts of scientific publications [3, 4] we carried

out named entity recognition of objects from their texts and analyzing the relationship between the extracted objects. We identified a set of genes, proteins and microRNAs that are associated with MDD and may be involved in the molecular mechanisms of the disorder. We selected relevant texts of publications based on keywords such as “major depressive disorder and epigenetics regulation”, “major depressive disorder and comorbidities”, “major depressive disorder and genetic polymorphisms”. Named entity recognition was carried out based on the CFR [4] and HunFlair [5] packages. The relationships between the terms were extracted based on the classification of sentences into those that contain and do not contain an indication of a relationship between terms, using a naïve Bayesian classifier. We demonstrated the applicability of naïve Bayesian classifier for the task of name recognition earlier [3].

Results: Based on queries to NCBI PubMed we extracted 3,000 of abstracts. Extraction of associations between names of protein, gene and microRNA allowed us to reveal their possible impact into the development of MDD. In particular, we identified genes of the dopaminergic and serotonergic systems and the glucocorticoid receptor, that are associated with altered expression levels that may occur as a result of adverse circumstances (e. g. deprivation) in early childhood and thus may play a role in the development of depression in adults in stressful situation or in difficult life circumstances. Such changes of gene expression levels can represent the so-called vicious circle of pathology.

Additionally, we identified relationships between epigenetic mechanisms of estrogen receptor regulation that may be responsible for the development of mental illness. The factors that contribute to the development of major depressive also include a particular attachment type of the child, the ability to form object relationship that depend on a early interaction of a child with an adult. The factors that may impact on the development of MDD may also include comorbid disorders such as borderline personality disorder or anxiety-depressive disorders, which both may be accompanied by more severe symptomatology and should be differentiated from other psychiatric disorder.

Conclusion: We evaluated the potential mechanisms for the development of major depressive disorder using text mining approach. The revealed reasons include a combination of biological and psychological factors. In particular, we found some of the molecular mechanisms responsible for the development of MDD, including gene regulation of the dopaminergic system. The role of these genes in the development of MDD is discussed in earlier studies [3]. At the same time the identified relationships between estrogen and glucocorticoid receptor regulation and the development of depression in adults is interesting for further investigation. The relationships between psychological and hereditary factors in the dynamics in patients with major depressive disorder are also of interest.

Funding: The study is supported by Russian Science Foundation grant No. 24-25-00453.

Список литературы/References

1. Uher R. et al. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. *Depress Anxiety*. 2014;31(6):459-471. doi 10.1002/da.22217
2. Thumfart K.M. et al. Epigenetics of childhood trauma: Long term sequelae and potential for treatment. *Neurosci Biobehav Rev*. 2022;132:1049-1066. doi 10.1016/j.neubiorev.2021
3. Tarasova O.A. et al. Chemical named entity recognition in the texts of scientific publications using the naïve Bayes classifier approach. *J Cheminform*. 2022;14(1):55. doi 10.1186/s13321-022-00633-4
4. Tarasova O.A. et al. Extraction of data on parent compounds and their metabolites from texts of scientific abstracts. *J Chem Inf Model*. 2021;61(4):1683-1690. doi 10.1021/acs.jcim.0c01054
5. Weber L. et al. HunFlair: an easy-to-use tool for state-of-the-art biomedical named entity recognition. *Bioinformatics*. 2021;37(17):2792-2794. doi 10.1093/bioinformatics/btab042