

Анализ событий альтернативного сплайсинга в децидуальных клетках плаценты человека

Гавриленко М.М.*, Трифонова Е.А., Бабовская А.А., Зарубин А.А., Степанов В.А.

Научно-исследовательский институт медицинской генетики Томского национального

исследовательского медицинского центра РАН, Томск, Россия

* mmgavrilenko@gmail.com

Ключевые слова: альтернативный сплайсинг; полнотранскриптомное секвенирование; транскрипт; плацента; децидуальные клетки; MAJIQ; rMATS; SGSeq

Мотивация и цель: Изучение альтернативного сплайсинга (АС) представляет значительный интерес, поскольку данный процесс является фундаментальным в молекулярной биологии, в ходе которого из одной пре-мРНК может быть получено множество транскриптов мРНК, что приводит к экспрессии различных изоформ белка из одного гена. Известно, что АС регулирует экспрессию генов, участвующих в развитии и функционировании плаценты. Изменение экспрессии плацентарных генов может нарушить такие важные процессы, как дифференцировка трофобласта, ангиогенез, транспорт питательных веществ и иммунный гомеостаз, что может привести к осложнениям беременности. Децидуальные клетки (ДК) являются основной клеточной популяцией материнской части плаценты и играют ключевую роль в поддержании физиологической беременности [1]. Немаловажным является выбор биоинформатического инструмента для анализа событий АС, которых достаточно много на сегодняшний день. Методы анализа, основанные на событийном подходе, обладают более высокой чувствительностью и специфичностью даже при низких уровнях экспрессии. В данном случае обычно используются такие инструменты, как MAJIQ [2], rMATS [3], SGSeq [4].

Методы и алгоритмы: Глубокое секвенирование проведено для 8 образцов плацентарной ткани, полученной от женщин с физиологическим течением беременности. Расчетное число прочтений на образец составляло не менее 60 млн. Анализ качества полученных сиквенсов выполнен с помощью программы FastQC. Последовательности адаптеров удалены с помощью программы Trimmomatic. Выравнивание полученных транскриптомных данных на референсный геном человека (GRCh38) выполнено с помощью программы STAR. Далее для анализа событий АС выбраны программы MAJIQ, rMATS, SGSeq, так как они существенно обновлены в 2023–2024 гг. и имеют возможность визуализации.

Результаты: С помощью программы MAJIQ идентифицировано 3501 событие АС для 2731 гена; с помощью программы rMATS идентифицировано 66 687 событий для 14 784 генов; с помощью программы SGSeq – 15 782 события для 5616 генов. При анализе результатов полученных данных важно отметить, что количество генов, подверженных АС, значительно варьировало. Идентифицировано 15 887 уникальных плацентарных генов, подверженных альтернативному сплайсингу. При поиске общих генов показана низкая репликация результатов (33.9 %), так как всего 5387 генов выявлены хотя бы в двух программах одновременно. Низкая репликация и высокая вариативность в количестве идентифицированных событий

и генов могут быть связаны с различными статистическими и алгоритмическими подходами, заложенными разработчиками в программы для анализа альтернативного сплайсинга.

Следующим шагом все события были классифицированы согласно 7 типам событий АС для программ MAJIQ и SGSeq (пропуск экзона, альтернативные 5' или 3' сайты сплайсинга, удержание интрона, альтернативные первый и последний экзоны и взаимоисключающие экзоны) и 5 типам событий для программы rMATS (без альтернативных первого и последнего экзонов). Наиболее распространенным типом АС является пропуск экзона (46.4 %), что показано для всех трех программ и согласуется с давно известным положением о преобладании пропуска экзона у млекопитающих. Далее наблюдается вариативность. В программе MAJIQ альтернативный первый экзон является вторым по популярности событием АС (24.8 %), в rMATS – альтернативный 3'-сайт сплайсинга (14.1 %), тогда как в SGSeq – удержание интрона (28.6 %), что подтверждается новыми данными о росте событий удержания интрона [5]. Наименее редким событием являются взаимоисключающие экзоны (1.3 %), что также показано для всех трех программ. В целях более детальной функциональной аннотации генов, подверженных АС, был проведен анализ 1857 общих для всех трех программ генов. Анализ реконструированной с использованием программы STRING белок-белковой сети включал в себя 386 белковых продуктов (high confidence = 0.999) и 98 кластеров, из которых 8 кластеров включали в себя 8 генов и более, 57 кластеров по 2 гена, количество генов в остальных кластерах варьировало от 3 до 5 включительно. Далее проведена подробная аннотация первых по количеству генов 3 кластеров в базах данных GeneOntology, KEGG, Reactome с помощью онлайн-ресурса WebGestalt (<https://www.webgestalt.org/>).

Кластер 1 включает 26 генов, центральное место с максимальным числом и силой взаимодействий ($\text{node_degree} \geq 10$, $\text{score} \geq 0.99$) занимает ген *ACTL6A*. Согласно результатам функциональной аннотации, сигнальные пути кластера 1 преимущественно связаны с модификацией хроматина и регуляцией транскрипции. Во время беременности изменения хроматина играют важную роль в регуляции экспрессии генов в развивающемся плоде, плаценте и тканях матери. Одним из примеров модификации хроматина во время беременности является метилирование ДНК. Известно, что изменения в метилировании материнской ДНК могут привести к гестационному диабету, анемии, преждевременным родам и преэклампсии [6].

Кластер 2 включает 17 генов, центральное место с максимальным числом и силой взаимодействий ($\text{node_degree} \geq 8$, $\text{score} \geq 0.99$) занимает ген *SRC*. Согласно результатам функциональной аннотации, сигнальные пути кластера 2 преимущественно связаны с межклеточными сигналами, миграцией клеток и межклеточной адгезией. Молекулы клеточной адгезии играют решающую роль в обеспечении адгезии и миграции клеток трофобласта в эндометрий и последующей инвазии в кровеносные сосуды матери [7]. Нарушение регуляции этих процессов адгезии может способствовать развитию таких состояний, как преэклампсия или задержка роста плода [8].

Кластер 3 включает 15 генов, центральное место с максимальным числом и силой взаимодействий ($\text{node_degree} \geq 5$, $\text{score} \geq 0.99$) занимают гены *TBK1*, *TRAF3*. Согласно результатам функциональной аннотации, сигнальные пути кластера 3 преимущественно связаны с иммунной системой, апоптозом и сигнальным путем

NF-κB (ядерный фактор каппа В), который является ключевым фактором транскрипции, участвующим в регуляции иммунных и воспалительных реакций, пролиферации, дифференцировке и выживании клеток. Адекватная регуляция активности NF-κB имеет решающее значение для поддержания иммунного гомеостаза матери и плода и обеспечения успешного исхода беременности. Апоптоз также способствует поддержанию иммунной толерантности на уровне взаимодействия матери и плода, устраняя активированные иммунные клетки и предотвращая чрезмерные воспалительные реакции, которые могут нанести вред развивающемуся плоду. Нарушение регуляции этих процессов может способствовать возникновению самопроизвольного аборта и преждевременных родов [9].

Выводы: В исследовании проведена оценка событий АС в децидуальных клетках плаценты при физиологической беременности с помощью трех биоинформатических программ MAJIQ, rMATS, SGSeq. Было идентифицировано разное количество генов и событий АС каждой из программ, а также продемонстрирована низкая репликация результатов, что может быть связано с различными математическими подходами, заложенными разработчиками в алгоритмы программ, в связи с чем целесообразно использовать несколько программ для идентификации событий АС и для последующего анализа выбирать те гены, которые являются общими для всех примененных в исследовании программ. Во всех трех программах наиболее часто встречающимся из выявленных событий является пропуск экзона, а самым редким – взаимо-исключающий пропуск экзонов. Анализ реконструированной геномной сети позволил выявить биологические пути, ассоциированные с иммунной системой, клеточной миграцией, межклеточными контактами и регуляцией экспрессии. Следует отметить, что ряд генов, подверженных АС в ДК, ассоциирован с осложнениями беременности. Полученные результаты подтверждают важность альтернативного сплайсинга, который существенно увеличивает транскрипционное разнообразие и представляет собой значимый механизм регуляции генов в децидуальных клетках.

Финансирование: Исследование выполнено за счет средств государственного задания по теме ФНИ № 122020200083-8.

Analysis of alternative splicing events in human placental decidual cells

Gavrilenko M.M.*, Trifonova E.A., Babovskaya A.A., Zarubin A.A., Stepanov V.A.

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

* mmgavrilenko@gmail.com

Key words: alternative splicing; whole transcriptome sequencing; transcript; placenta; decidual cells; MAJIQ; rMATS; SGSeq

Motivation and Aim: The alternative splicing (AS) study is of considerable interest, because this process is fundamental in molecular biology. Multiple mRNA transcripts can be obtained from a single pre-mRNA during AS. It leads to the expression of various protein isoforms from a single gene. It is known that AS regulates the expression of genes involved in the development and functioning of the placenta. Changes in the placental genes expression can disrupt important processes such as trophoblast

differentiation, angiogenesis, nutrient transport and immune homeostasis, which can lead to pregnancy complications. Decidual cells (DC) are the main cell population of the placenta maternal part and have a key role in maintaining physiological pregnancy [1]. It is also important to choose a bioinformatics tool for analyzing AS events among a large number of them. Event-based analysis methods have higher sensitivity and specificity even at low expression levels. In this case, tools such as MAJIQ [2], rMATS [3], SGSeq [4] are usually used.

Methods and Algorithms: Deep sequencing was performed for 8 placental tissue samples obtained from women with a physiological pregnancy. The calculated number of readings per sample was at least 60 million. The quality analysis of the received transcriptomic data was performed using the FastQC. The adapter sequences are deleted using the Trimmomatic. The alignment to the human reference genome (GRCh38) was performed using the STAR. The MAJIQ, rMATS and SGSeq programs were selected for the analysis of AS events, because they have the ability to visualize and were significantly updated in 2023–2024.

Results: As a result of the study, 3501 AS events for 2731 genes were identified with MAJIQ; 66687 events for 14784 genes were identified with rMATS; 15782 events for 5616 genes were identified with SGSeq. It is important to note that the number of genes affected by AS varied significantly. 15887 unique placental genes subject to alternative splicing have been identified. Low replication of the results was shown (33.9 %), because only 5387 genes were identified in at least two programs. Low replication and high variability in the number of identified events and genes may be associated with various statistical and algorithmic approaches embedded by developers in programs for analyzing AS.

All events were classified according to 7 types of AS events for the MAJIQ and SGSeq programs (exon skipping, alternative 5' or 3' splicing sites, intron retention, alternative first and last exons and mutually exclusive exons) and 5 types of events for the rMATS program (without alternative first and last exons). The most common type of AS is exon skipping (46.4 %), which is shown for all three programs. It is consistent with the long-known statement on the predominance of exon skipping in mammals. Further, there is variability. The alternative first exon is the second most popular AS event (24.8 %) for MAJIQ, the alternative 3'-splicing site (14.1 %) for rMATS and intron retention (28.6 %) for SGSeq. It is confirmed by new data on the growth of intron retention events [5]. The least rare event is mutually exclusive exons (1.3 %), which is also shown for all three programs.

In order to provide a more detailed functional annotation of genes affected by AS, 1857 genes common to all three programs were analyzed. The analysis of the protein-protein network reconstructed using the STRING program included 386 protein products (high confidence = 0.999) and 98 clusters, of which 8 clusters included 8 or more genes, 57 clusters consist of 2 genes, the genes number in the other clusters varied from 3 to 5 genes inclusive. Next, a detailed annotation of the first 3 clusters in terms of the genes number in the GeneOntology, KEGG, Reactome databases was carried out using the WebGestalt online resource (<https://www.webgestalt.org/>).

Cluster 1 includes 26 genes. The *ACTL6A* gene is the central gene with the maximum number and strength of interactions (node_degree ≥ 10 , score ≥ 0.99). According to the results of the functional annotation, the cluster 1 signaling pathways are mainly associated with chromatin modification and transcription regulation. During pregnancy,

chromatin changes play an important role in regulating gene expression in the developing fetus, placenta, and maternal tissues. One example of chromatin modification during pregnancy is DNA methylation. It is known that changes in maternal DNA methylation can lead to gestational diabetes, anemia, premature birth and preeclampsia [6].

Cluster 2 includes 17 genes. The *SRC* gene is the central gene with the maximum number and strength of interactions ($\text{node_degree} \geq 8$, $\text{score} \geq 0.99$). According to the results of the functional annotation, the cluster 2 signaling pathways are mainly associated with intercellular signals, cell migration and intercellular adhesion. Cell adhesion molecules play a crucial role in providing adhesion and trophoblast cells migration into the endometrium and subsequent invasion into the mother blood vessels [7]. A disorder of these adhesion processes regulation can contribute to the preeclampsia and fetal growth restriction development [8].

Cluster 3 includes 15 genes. The *TBK1* and *TRAF3* are the central genes with the maximum number and strength of interactions ($\text{node_degree} \geq 5$, $\text{score} \geq 0.99$). According to the results of the functional annotation, cluster 3 signaling pathways are predominantly associated with the immune system, apoptosis and the NF-kappaB signaling pathway, which is a key transcription factor involved in the regulation of immune and inflammatory reactions, cell proliferation, differentiation and survival. Adequate regulation of NF-kappaB activity is crucial for maintaining maternal and fetal immune homeostasis and provides a successful pregnancy outcome. Apoptosis also helps maintain immune tolerance at the level of maternal-fetal interaction by eliminating activated immune cells and preventing excessive inflammatory reactions that can harm the developing fetus. A violation these biological processes can contribute to the spontaneous abortion and premature birth [9].

Conclusion: In the study was performed assessment of AS events in placental decidual cells during physiological pregnancy using three bioinformatics programs MAJIQ, rMATS, SGSeq. A different number of AS genes and events were identified for each programs. Also low replication of the results was demonstrated. There are may be due to different mathematical approaches embedded by the developers in the programs algorithms. Therefore it is advisable to use several programs to identify AS events and for subsequent analysis select those genes that are common to all tools. In all three programs, the most common of the identified events is an exon skip, and the rarest is a mutually exclusive exons. The reconstructed gene network analysis revealed biological pathways associated with the immune system, cell migration, intercellular contacts and regulation of gene expression. It should be noted that a number of genes susceptible to AS in DC are associated with pregnancy complications. The results obtained confirm the importance of alternative splicing, which significantly increases transcriptional diversity and represents a significant mechanism for gene regulation in decidual cells.

Funding: The research was carried out at the expense of the state assignment (fundamental scientific research No. 122020200083-8).

Список литературы/References

1. Schatz F., Guzeloglu-Kayisli O., Arlier S. Kayisli U.A., Lockwood C.J. The role of decidual cells in uterine hemostasis, menstruation, inflammation, adverse pregnancy outcomes and abnormal uterine bleeding. *Hum Reprod Update*. 2016;4:497-515. doi 10.1093/humupd/dmw004
2. Vaquero-Garcia J., Barrera A., Gazzara M.R. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife*. 2016;5:e11752. doi 10.7554/elifesciences.11752
3. Shen S., Park J. W., Lu Z.X. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA*. 2014;111(51):E5593-E5601. doi 10.1073/pnas.1419161111

4. Goldstein L.D., Ca Y., Pau G. et al. Prediction and quantification of splice events from RNA-seq data. *PLoS One*. 2016;11(5):e0156132. doi 10.1371/journal.pone.0156132
5. Monteuis G., Wong J.J., Bailey C.G. et al. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res*. 2019;47(22):11497-11513. doi 10.1093/nar/gkz1068
6. Das J., Maitra A. Maternal DNA methylation during pregnancy: a review. *Reprod Sci*. 2021;28(10):2758-2769. doi 10.1007/s43032-020-00456-4
7. Hohn H.P., Denker H.W. Experimental modulation of cell-cell adhesion, invasiveness and differentiation in trophoblast cells. *Cells Tissues Organs*. 2002;172(3):218-236. doi 10.1159/000066965
8. Barbitoff Y.A., Tsarev A.A., Vashukova E.S. et al. A data-driven review of the genetic factors of pregnancy complications. *Int J Mol Sci*. 2020;21(9):3384. doi 10.3390/ijms21093384
9. Sharp A.N., Heazell A.E., Crocker I.P. et al. Placental apoptosis in health and disease. *Am J Reprod Immunol*. 2010;64(3):159-169. doi 10.1111/j.1600-0897.2010.00837.x