

Представленность альтернативных изоформ РНК гена SLC4A11 в эндотелии роговицы здоровых людей и пациентов с эндотелиальной дистрофией роговицы Фукса

Ковалева П.А.^{1*}, Котова Е.С.¹, Антонова О.П.², Малюгин Б.Е.², Ткаченко И.С.², Шарова Е.И.¹, Скородумова Л.О.¹

¹ ФНКЦ ФХМ имени ак. Ю.М. Лопухина ФМБА России, Москва, Россия

² НМИЦ «МНТК "Микрохирургия глаза" им. ак. С.Н. Федорова» МЗ РФ, Москва, Россия

* polina_a_kovaleva@mail.ru

Ключевые слова: дистрофия Фукса; SLC4A11; роговица; 5'RACE

Мотивация и цель: Первичная дистрофия эндотелия роговицы Фукса (ДФ) является наследственным заболеванием глаза, характеризующимся снижением числа клеток эндотелия роговицы, выполняющих функцию поддержания водного баланса стромы роговицы. Известно, что ДФ является мультилокусным заболеванием. Наиболее распространенным каузальным вариантом (около 70 %) среди популяций европейского происхождения является экспансия тринуклеотидных повторов CTG18.1 в гене TCF4. По литературным данным часть случаев ДФ обусловлена вариантами в гене SLC4A11, белковый продукт которого является ионным транспортером и может вносить вклад в поддержание водного баланса стромы роговицы [1]. На данный момент большое количество литературных данных, связанных с функционированием белкового продукта SLC4A11 в эндотелии роговицы, было основано на данных о существовании трех основных РНК изоформ, отличающихся 5'-концевыми участками и кодируемых ими белковых продуктов. В большинстве статей на основе различных данных и предположений в качестве основной изоформы РНК называлась изоформа 2, а в качестве основной изоформы белка – ее продукт. Также в некоторых работах исследования проводились на изоформах РНК 3 и 1 и их белковых продуктах [2, 3]. Кроме того были получены свидетельства, что основной формой белка SLC4A11 в эндотелии роговицы является укороченный с N-конца полипептид (далее-v2short), который считают результатом старта трансляции на изоформе РНК 2 не с первого, а с третьего старт-кодона, находящегося в той же рамке считывания, что и первый [4]. Таким образом, целью нашей работы стал анализ представленности альтернативных изоформ РНК гена SLC4A11 в эндотелии роговицы здоровых людей и пациентов с эндотелиальной дистрофией роговицы Фукса.

Методы и алгоритмы: Аннотационный gtf-файл с полным набором транскриптов SLC4A11 формировался на основе представленных транскриптов в базе данных Refseq и проекта Gencode с дополнением транскриптов, полученных de novo StringTie (2.2.1). Также для транскриптов гена SLC4A11, полученных StringTie, проводился сравнительный анализ аннотации пакетом GffCompare (0.12.6) и для предсказания открытых рамок считывания (ORF) и нахождения потенциальных кодирующих последовательностей в транскриптомных данных использовали пакет TransDecoder (5.5.0). Полученный аннотационный gtf-файл использовали

для оценки уровня экспрессии генов и изоформ с помощью пакета Salmon (1.3.0) для двух сетов транскриптомных данных, представленных образцами эндотелия роговицы доноров и пациентов с ДФ [5, 6]. После проводился анализ дифференциальной экспрессии гена SLC4A11 с помощью пакета EdgeR (2.0.7). Для подтверждения данных об уровне экспрессии изоформ, полученных биоинформатически, было проведено выявление 5'-концевых участков транскриптов, образующихся в эндотелии роговицы человека методом быстрой амплификации 5'-концевых фрагментов 5'RACE (Rapid Amplification of cDNA Ends). При этом из образцов эндотелия роговицы 5 здоровых доноров и 5 пациентов с ДФ была выделена РНК, которая далее использовалась для получения наборов ампликонов, соответствующих 5'-концевым участкам транскриптов SLC4A11, содержащихся в этих образцах. Обратная транскрипция с переключением матрицы проводилась набором SMARTer RACE 5'/3' Kit (Takara). На матрице полученных образцов кДНК проводились 2 раунда нестед ПЦР с ПЦР-супрессией. После этого проводилось секвенирование полученных ампликонов по Сэнгеру и на Illumina MiSeq для определения и аннотации транскриптов, которые экспрессируются в эндотелии роговицы.

Результаты: Мы провели предварительный анализ покрытия прочтениями локуса гена SLC4A11, используя опубликованные транскриптомы (тотальные и поли-А) эндотелия здоровых доноров [5, 7] и наибольшая плотность прочтений соответствовала 5'-области РНК-изоформы, представленной в базе данных RefSeq, как NM_001400277, или изоформа 6, которая не соответствовала ни одной из трех основных изоформ SLC4A11, фигурирующих в литературных данных. Также высокая плотность прочтений приходилась на 5'-область третьей изоформы. При этом покрытие прочтениями 5'-области, соответствующей 2 и 1 изоформам, было довольно низким.

Затем была проведена оценка уровня экспрессии изоформ SLC4A11 для двух сетов транскриптомных данных, представленных образцами эндотелия роговицы доноров и пациентов с ДФ [5, 6].

De novo сборка транскриптомов гена SLC4A11 выявила две дополнительные изоформы MSTRG.58852.10 и MSTRG.58852.21. Пакетом TransDecoder для этих MSTRG.58852.10 и MSTRG.58852.21 были выявлены белок-кодирующие последовательности. Однако их кодирующая роль оценивается нами как мало значимая, так как MSTRG.58852.21 кодирует довольно короткий полипептид, а у MSTRG.58852.10 после стоп-кодона идут экзон-экзонные сочленения, это указывает на то, что для данной изоформы в клетке скорее всего реализуется механизм nonsense-mediated decay.

При оценке представленности изоформ в двух наборах данных эндотелия роговицы наиболее представленной изоформой оказалась изоформа 3, второй по представленности – изоформа 6 (NM_001400277), которая не соответствовала ни одной из трех основных изоформ SLC4A11. Третьей по представленности оказалась не кодирующая изоформа NR_174471. Высокая представленность изоформы 6 интересна тем, что она кодирует v2short, но при этом его старт трансляции является первым ATG-кодоном данной мРНК. Представленность изоформ 1 и 2 была оценена как довольно низкая.

Секвенирование по Сэнгеру продуктов, полученных методом быстрой амплификации 5'-концевых фрагментов целевых транскриптов гена SLC4A11, показало, что образцы соответствовали 5'-концевому участку изоформы 6 и

некодирующей изоформы NR_174471. Предварительный анализ NGS-секвенирования одного из образцов, показал, что в нем наиболее представленной является 5'-конец изоформы 6, гораздо менее представлен конец изоформы 3, практически нет прочтений, соответствующих изоформам 1 и 2 (1 прочтение) и есть небольшое количество прочтений, соответствующих сплайсированным транскриптам, начинающимся внутри второго интрона 1-3 и 6 изоформ, о промоторной активности этого участка генома пока не обнаружено никаких данных.

Выводы: Мы получили свидетельства преимущественной экспрессии РНК изоформ SLC4A11 3 и 6 в эндотелии роговицы и довольно низкой представленности в нем изоформы 2. Эти данные не только помогут уточнить информацию о преимущественно образующихся в эндотелии роговицы белковых изоформах SLC4A11, но также охарактеризовать функционирование промоторного региона SLC4A11 в эндотелии роговицы.

Финансирование: Исследование поддержано грантом Российского научного фонда (№ 22-75-10157).

Study of the SLC4A11 alternative RNA isoforms expression in the corneal endothelium of healthy individuals and patients with Fuchs corneal endothelial dystrophy

Kovaleva P.A.^{1*}, Kotova E.S.¹, Antonova O.P.², Malyugin B.E.², Tkachenko I.S.², Sharova E.I.¹, Skorodumova L.O.¹

¹ Lopukhin FRCC PCM, Moscow, Russia

² The S. Fyodorov Eye Microsurgery Federal State Institution, Moscow, Russia

* polina_a_kovaleva@mail.ru

Key words: FECD; SLC4A11; cornea; 5'RACE

Motivation and Aim: Primary Fuchs' endothelial corneal dystrophy (FECD) is an inherited eye disease characterized by a reduction in the number of corneal endothelial cells, which are responsible for maintaining the water balance in the corneal stroma. FECD is known to be a multilocus disease. The most common causative variant (approximately 70 %) in populations of European descent is the expansion of the trinucleotide repeat CTG18.1 in the TCF4 gene. According to the literature, a proportion of FECD cases are caused by variants in the SLC4A11 gene, the protein product of which is an ion transporter and may contribute to maintaining water balance in the corneal stroma [1]. To date, much of the literature on the function of the SLC4A11 protein product in the corneal endothelium has been based on the existence of three major RNA isoforms that differ in their 5'-ends and the protein products encoded by them. Most papers, based on different data and assumptions, identified isoform 2 as the major RNA isoform and its product as the major protein isoform. Some papers have also looked at RNA isoforms 3 and 1 and their protein products [2, 3]. In addition, it has been shown that the major form of SLC4A11 protein in the corneal endothelium is an N-short polypeptide (hereafter referred to as v2short), which is considered to be the result of translation start of RNA isoform 2 not from the first but from the third start codon, which is in the same reading frame as the first [4]. Therefore, the aim of our work was to analyze

the expression of alternative RNA isoforms of the SLC4A11 gene in the corneal endothelium of healthy subjects and patients with FECD.

Methods and Algorithms: An annotation gtf-file with the complete set of SLC4A11 transcripts was generated based on the submitted transcripts in the Refseq database and the Gencode project with the addition of transcripts obtained de novo by StringTie (2.2.1). Similarly, for the SLC4A11 gene transcripts obtained by StringTie, the annotation was compared using the GffCompare package (0.12.6) and the TransDecoder package (5.5.0) was used to predict open reading frames (ORFs) and find potential coding sequences in the transcriptome data. The resulting annotated gtf-file were used to assess gene expression levels and isoforms using the Salmon package (1.3.0) for two sets of transcriptomic data represented by corneal endothelial samples from donors and patients with FECD [5, 6]. Differential expression of the SLC4A11 gene was analyzed using the EdgeR package (2.0.7). To confirm the bioinformatic data on isoform expression levels, the 5'-end regions of transcripts produced in human corneal endothelium were identified by rapid amplification of 5'-end fragments using 5'RACE (Rapid Amplification of cDNA Ends). RNA was isolated from corneal endothelial samples of 5 healthy donors and 5 patients with FECD and used to obtain sets of amplicons corresponding to 5'-end fragments of SLC4A11 transcripts present in these samples. Reverse transcription with matrix switching was performed using the SMARTer RACE 5'/3' Kit (Takara). 2 rounds of nested PCR with PCR suppression were performed on the matrix of the obtained cDNA samples. The resulting amplicons were then sequenced by Sanger approach and on Illumina MiSeq to identify and annotate the transcripts expressed in the corneal endothelium.

Results: We performed a preliminary analysis of read coverage of the SLC4A11 gene locus using published transcriptomes (total and poly-A) of healthy donor endothelium [5, 7] and the highest density of reads corresponded to the 5' region of the RNA isoform represented in the RefSeq database as NM_001400277 or isoform 6, which does not correspond to any of the three major SLC4A11 isoforms in the literature. There was also a high density of reads in the 5' region of the third isoform. However, the coverage of reads in the 5' region corresponding to isoforms 2 and 1 was rather low.

We then evaluated the expression level of SLC4A11 isoforms for two sets of transcriptomic data represented by corneal endothelial samples from donors and patients with FECD [5, 6].

The expression level of SLC4A11 isoforms was then evaluated for two sets of transcriptomic data represented by corneal endothelial samples from donors and FECD patients [5, 6].

De novo assembly of the SLC4A11 gene transcriptome revealed two additional isoforms, MSTRG.58852.10 and MSTRG.58852.21. The protein coding sequences for these MSTRG.58852.10 and MSTRG.58852.21 were identified using the TransDecoder package. However, we consider their coding role to be of little importance since MSTRG.58852.21 encodes a rather short polypeptide and MSTRG.58852.10 has exon-exon junctions after the stop codon, indicating that the nonsense-mediated decay mechanism is most likely realized in the cell for this isoform.

When evaluating the representation of isoforms in two sets of corneal endothelial data, isoform 3 was the most represented isoform, the second most represented isoform was isoform 6 (NM_001400277), which did not correspond to any of the three major isoforms of SLC4A11. The third most abundant isoform was the non-coding isoform NR_174471. The high representation of isoform 6 is interesting because it encodes

v2short, but its translation start is the first ATG codon of this mRNA. The representation of isoforms 1 and 2 was estimated to be rather low.

Sanger sequencing of the products obtained by rapid amplification of 5'-end fragments of the target transcripts of the SLC4A11 gene showed that the samples corresponded to the 5'-end of isoform 6 and the non-coding isoform NR_174471. Preliminary NGS sequencing analysis of one of the samples showed that the 5' end of isoform 6 is most abundant, the end of isoform 3 is much less abundant, there are virtually no reads corresponding to isoforms 1 and 2 (1 read), and there are a small number of reads corresponding to spliced transcripts starting within the second intron of isoforms 1-3 and 6. No data on the promoter activity of this genomic region have been found.

Conclusion: We obtained evidence for the predominant RNA expression of SLC4A11 isoforms 3 and 6 in the corneal endothelium and a rather low representation of isoform 2. These data will not only help to clarify the information about the protein isoforms of SLC4A11 predominantly produced in the corneal endothelium, but also to characterize the function of the promoter region of SLC4A11 in the corneal endothelium.

Funding: The study is supported by Russian Science Foundation, grant No. 22-75-10157.

Список литературы/References

1. Tsedilina T.R., Sharova E., Iakovets V., Skorodumova L.O. Systematic review of SLC4A11, ZEB1, LOXHD1, and AGBL1 variants in the development of Fuchs' endothelial corneal dystrophy. *Front Med.* 2023;10:1153122. doi 10.3389/fmed.2023.1153122
2. Hara S., Tsujikawa M., Kawasaki S., Nishida K. Homeostasis of SLC4A11 protein is mediated by endoplasmic reticulum-associated degradation. *Experimental Eye Res.* 2019;188:107782. doi 10.1016/j.exer.2019.107782
3. Kao L., Azimov R., Shao X.M. et al. Multifunctional ion transport properties of human SLC4A11: comparison of the SLC4A11-B and SLC4A11-C variants. *Am J Physiol Cell Physiol.* 2016;311(5):C820-C830. doi 10.1152/ajpcell.00233.2016
4. Malhotra D., Loganathan S.K., Chiu A.M., Lukowski C.M., Casey J.R. Human corneal expression of SLC4A11, a gene mutated in endothelial corneal dystrophies. *Sci Rep.* 2019;9(1):9681. doi 10.1038/s41598-019-46094-y
5. Nikitina A.S., Belodedova A.V., Malyugin B.E. et al. Dataset on transcriptome profiling of corneal endothelium from patients with Fuchs endothelial corneal dystrophy. *Data Brief.* 2019;25:104047. doi 10.1016/j.dib.2019.104047
6. Chu Y., Hu J., Liang H. et al. Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease. *Nucleic Acids Res.* 2020;48(12):6740-6758. doi 10.1093/nar/gkaa422
7. Chen Y., Huang K., Nakatsu M.N., Xue Z., Deng S.X., Fan G. Identification of novel molecular markers through transcriptomic analysis in human fetal and adult corneal endothelial cells. *Hum Mol Genet.* 2013;22(7):1271-1279. doi 10.1093/hmg/dd527