

Получена: 4 сентября 2016 / Принята: 11 октября 2016 / Опубликовано online: 31 октября 2016

УДК 577.21+616.32

## **РНК СЕКВЕНИРОВАНИЕ ПРИ ПЛОСКОКЛЕТОЧНОМ РАКЕ ПИЩЕВОДА: ОТ ЗАБОРА БИОМАТЕРИАЛА ДО СЕКВЕНИРОВАНИЯ НА HISEQ2000**

**Сауле Е. Рахимова<sup>1</sup>, Улыкбек Е. Каиров<sup>2</sup>, Даурен А. Ережепов<sup>1</sup>,  
Асхат Б. Молкенов<sup>2</sup>, Мадина Р. Жалбинова<sup>1</sup>, Алия Т. Абильмажинова<sup>1</sup>,  
Юрий В. Жуков<sup>3</sup>, Марат Х. Омаров<sup>3</sup>, Айнур Р. Акильжанова<sup>1</sup>**

<sup>1</sup> Лаборатория геномной и персонализированной медицины, National Laboratory Astana, Назарбаев Университет, Астана, Казахстан;

<sup>2</sup> Лаборатория биоинформатики и вычислительной системной биологии, National Laboratory Astana, Назарбаев Университет, Астана, Казахстан;

<sup>3</sup> Онкологический центр города Астана, Астана, Казахстан

### **Резюме**

В Казахстане отмечается высокая заболеваемость плоскоклеточным раком пищевода, частота заболеваемости составляет 10,1:100 000. В структуре смертности рак пищевода среди мужчин – занимает 5-ое, среди женщин – 9-ое место в республике Казахстан. Молекулярные механизмы, способствующие инициации и прогрессированию онкопатологии еще недостаточно изучены. Тем не менее, отсутствие чувствительных и специфических биомаркеров для диагностики онкозаболеваний подчеркивает необходимость дополнительных исследований по изучению развития канцерогенеза. В этом аспекте перспективным является исследование транскриптомного профиля опухолевых клеток.

В работе представлено описание этапов пробоподготовки и РНК секвенирования на основе технологии секвенирования нового поколения. Каждый этап от момента забора биоматериала, выделения тотальной РНК, подготовка библиотеки оказывают влияние на получение качественных данных при последующем секвенировании.

**Ключевые слова:** РНК секвенирование, транскриптом, плоскоклеточный рак пищевода.

### **Summary**

## **RNA SEQUENCING OF SQUAMOUS ESOPHAGEAL CANCER: FROM THE COLLECTION OF THE BIOMATERIAL UNTIL SEQUENCING ON HISEQ2000**

**Saule E. Rakhimova<sup>1</sup>, Ulykbek E. Kairov<sup>2</sup>, Dauren A. Yerezhepov<sup>1</sup>,  
Askhat B. Molkenov<sup>2</sup>, Madina R. Zhalbinova<sup>1</sup>, Aliya T. Abilmazhinova<sup>1</sup>,  
Yuri V. Zhukov<sup>3</sup>, Marat Kh. Omarov<sup>3</sup>, Ainur R. Akilzhanova<sup>1</sup>**

<sup>1</sup> Laboratory of genomic and personalized medicine, National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan;

<sup>2</sup> Laboratory of bioinformatics and computational systems biology, National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan;

<sup>3</sup> Oncology center of Astana city, Astana, Kazakhstan.

The squamous esophageal cancer is a disease which is highly frequent in Kazakhstan, the frequency of the disease is 10,1:100 000. In the mortality structure esophageal cancer takes place among men the 5th place, among women the 9th place in the Republic of Kazakhstan. The molecular

*Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».*

mechanisms which contribute to the initiation and progression of the cancer pathology are not investigated enough. Nevertheless, the lack of the sensitive and specific biomarkers for the diagnosis of the cancer diseases, underlines about the necessity of the additional research works to study about the development of carcinogenesis. Investigation of the transcriptome profile of tumor cells is perspective in this aspect of the study.

The paper describes the stages of the sample preparation and sequencing of RNA, based on a new generation sequencing technology. Each step from the moment of the sampling of biological material, isolation of total RNA, preparation of libraries have an impact on obtaining of the high-quality data in subsequent sequencing.

**Key words:** RNA sequencing, transcriptome, squamous esophageal cancer.

Түйін

## ӨҢЕШТІҢ ЖАЛПАҚКЛЕТКАЛЫҚ ІСІГІ КЕЗІНДЕГІ РНҚ СЕКВЕНИРЛЕУ: БИОМАТЕРИАЛДЫ ЖИНАУДАН HISEQ2000 ЖҮЙЕСІНДЕ СЕКВЕНИРЛЕУГЕ ДЕЙІН

**Сауле Е. Рахимова<sup>1</sup>, Улыкбек Е. Каиров<sup>2</sup>, Даурен А. Ережепов<sup>1</sup>,  
Асхат Б. Молкенов<sup>2</sup>, Мадина Р. Жалбинова<sup>1</sup>, Алия Т. Абильмажинова<sup>1</sup>,  
Юрий В. Жуков<sup>3</sup>, Марат Х. Омаров<sup>3</sup>, Айнур Р. Акильжанова<sup>1</sup>**

<sup>1</sup> Геномдық және дербес медицина зертханасы, National Laboratory Astana, Назарбаев Университеті, Астана, Қазақстан;

<sup>2</sup> Биоинформатика және есептеуші жүйелік биология зертханасы, National Laboratory Astana, Назарбаев Университеті, Астана, Қазақстан;

<sup>3</sup> Астана қаласының Онкология орталығы, Астана, Қазақстан

Қазақстанда өңештің жалпақклеткалық ісігіне шалдығудың жоғары деңгейі тіркелуде, ауруға шалдығу жиілігі 100000 адамға шаққанда шамамен 10,1 көрсеткішті құрайды. Осы ауру Қазақстан Республикасында ерлер арасындағы өлім бойынша 5-ші, ал әйелдер арасында 9-шы орын алады. Онкопатологияның инициациясы мен прогрессиясын қоздыратын молекулалық механизмдер әлі толықтай зерттелмеген. Осыған қоса, онкоауруларды диагностикалаудың сезімтал әрі спецификалық биомаркерлердің тапшылығы қанцерогенез дамуына бағытталған қосымша зерттеулерді қажет етеді. Осы аспектіде ісік клеткаларының транскриптомдық профилдерін зерттеу үмітті тәсіл болып отыр.

Жұмыста үлгілерді дайындау мен жаңа буынды секвенирлеу технологиясы негізіндегі РНҚ секвенирлеу қадамдары сипатталған. Биоматериалды жинаудан бастап, тотальдық РНҚ-ны бөліп алу мен кітапханаларды дайындау сапасы келешектегі секвенирлеу кезінде алынған мәліметтер сапасына әсер етеді.

**Негізгі сөздер:** РНҚ секвенирлеу, транскриптом, өңештің жалпақклеткалық ісігі.

### Библиографическая ссылка:

Рахимова С.Е., Каиров У.Е., Ережепов Д.А., Молкенов А.Б., Жалбинова М.Р., Абильмажинова А.Т., Жуков Ю.В., Омаров М.Х., Акильжанова А.Р. РНҚ секвенирование при плоскоклеточном раке пищевода: от забора биоматериала до секвенирования на HiSeq2000 // Наука и Здравоохранение. 2016. №5. С. 53-63.

Rakhimova S.E., Kairov U.E., Yerezhepov D.A., Molkenov A.B., Zhalbinova M.R., Abilmazhinova A.T., Zhukov Y.V., Omarov M.Kh., Akilzhanova A.R. RNA sequencing of squamous esophageal cancer: from the collection of the biomaterial until sequencing on HiSeq2000. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016, 5, pp. 53-63.

Рахимова С.Е., Каиров У.Е., Ережепов Д.А., Молкенов А.Б., Жалбинова М.Р., Абильмажинова А.Т., Жуков Ю.В., Омаров М.Х., Акильжанова А.Р. Өңештің жалпақклеткалық ісігі кезіндегі РНҚ секвенирлеу: биоматериалды жинаудан HiSeq2000 жүйесінде секвенирлеуге дейін // Ғылым және Денсаулық сақтау. 2016. №5. Б. 53-63.

Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».

## Введение

Ежегодно диагностируется около 14 миллионов новых случаев рака пищевода в мире, по распространенности среди злокачественных новообразований он занимает 9-ое место. В структуре смертности на шестом месте. Рак пищевода составляют две основные разновидности: аденокарцинома (АК) и плоскоклеточный рак пищевода (ПКРП), каждый из них имеет определенные этиологические и патологические характеристики. Преобладающим типом является плоскоклеточный рак пищевода и составляет около 95% всех случаев [10]. Преимущественно распространен плоскоклеточный рак пищевода в развивающихся странах, в то время как заболеваемость аденокарциномой пищевода характерна для развитых стран, и связана как правило, с пищеводом Барретта [10,11]. Несмотря на высокую распространенность ПКРП, он менее изучен по сравнению с аденокарциномой пищевода. Отмечается тенденция к формированию групп высокого риска в определенных географических районах, что может свидетельствовать о генетической составляющей в данных популяциях. Это относится к странам, образующим так называемый «азиатский раковый пояс», включающий южное побережье Каспийского моря на западе до Северного Китая на востоке и страны: Иран, Афганистан, Монголию, страны Средней Азии и часть Сибирскую часть России.

В Казахстане отмечается высокая заболеваемость плоскоклеточным раком пищевода, частота заболеваемости составляет 10,1:100 000. В структуре смертности рак пищевода среди мужчин – занимает 5-ое, среди женщин – 9-ое место в РК.

Рак пищевода является мультифакториальным заболеванием. Факторы внешней среды и генетические факторы играют важную роль в патогенезе рака пищевода. В Европе и Северной Америке курение, потребление алкоголя, и повышение индекса массы тела (ИМТ) являются основными факторами риска. В отличие от них, в некоторых районах Китая,

Центральной Азии, и юга Африки заболевание связано с дефицитом питательных веществ, высоким потреблением нитрозаминбогатых или маринованных продуктов и низким социально-экономическим статусом.

К предрасполагающим факторам плоскоклеточного рака пищевода относят кератодермию [11], ахалазии [5] синдром Пламмера-Винсона [6] и т.д. Заболеваемость и прогноз рака пищевода хуже для мужчин, по сравнению с женщинами, возможно, это связано с различной экспозицией экзогенных факторов, таких как курение, прием алкоголя [9]. Плоскоклеточный рак пищевода характеризуется постепенным развитием от дисплазии, далее карциномы приводя в конечном итоге к инвазивным формам рака, с последовательностью развития процесса дисплазия-рак ассоциированы генетические изменения [1].

Характеристика генетических повреждений неразрывно связана с развитием ПКРП и пониманием молекулярных механизмов, лежащих в основе канцерогенеза и, следовательно, могут служить источником для раннего выявления опухоли, прогнозирования течения заболевания и разработки новых таргетных препаратов. Несмотря на успехи в диагностических подходах, методах лечения, большинство опухолей диагностируется на поздних стадиях, и пятилетняя выживаемость остается менее 30% [3].

Наиболее часто используемые в клинике сывороточные маркеры ПКРП включают Цитокератин 19-фрагмента [8], плоскоклеточный антиген клеточной карциномы [7], и карциноэмбриональный антиген [13], которые показывают недостаточную чувствительность 43,9%, 26,8% и 17% соответственно [12]. Плохая чувствительность и противоречивость этих маркеров обсуждался ранее, таким образом, подчеркивая необходимость поиска новых маркеров высокочувствительных и высокоспецифичных для использования в целях ранней диагностики ПКРП [14]. Дополнительные маркеры, которые, как известно, избыточно экспрессируются на ранних стадиях ПКРП включают Alcam, BPA-2, MMP13, MMP7 и MMP9 [4]. Козэкспрессия

MMP7, MMP9, MMP13, также была ассоциирована с неблагоприятным исходом в ПКРП [4]. Молекулярные механизмы, способствующие инициации и прогрессированию ПКРП еще недостаточно изучены. Тем не менее, отсутствие наличия чувствительных и специфичных биомаркеров для диагностики ПКРП подчеркивает необходимость дополнительных исследований по изучению туморогенеза плоскоклеточного рака пищевода.

Транскриптом – это совокупность всех транскриптов, синтезируемых клеткой или группой клеток, включающая мРНК, рРНК, тРНК и другие транскрибируемые некодирующие РНК. Термин «полный транскриптом» применяется к транскрипту организма в целом, либо конкретного типа клеток ткани. Так как транскриптом включает в себя все транскрипты мРНК в клетке, он отражает гены, которые активно экспрессируются в данное время.

Существует 2 основных подхода для изучения транскриптома: на основе ДНК микрочиповых технологий и технологии секвенирования следующего поколения, известный как RNA-Seq. Первый подход (DNA microarray) основан на гибридизации олигонуклеотидных зондов, с известной последовательностью на поверхности микрочипа с кДНК или мРНК (кРНК) исследуемого образца. Гибридизация олигонуклеотидного зонда при помощи флуоресценции или хемилюминисценции, данные выражаются в количественных показателях, что позволяет оценить относительное количество кДНК или мРНК с заданной последовательностью в исследуемом образце. Второй подход – секвенирование РНК, позволяет определить первичную структуру мРНК и некодирующей РНК. На современном этапе секвенируется кДНК. В отличие от микрочипового подхода, РНК секвенирование дает абсолютную информацию о наличии всех транскриптов в

исследуемом образце. В последнее время данный подход становится основным методом определения транскрипционного профиля клетки. Наибольших успехов достигла в этом направлении платформа, предлагаемая фирмой Illumina (Solexa) по сравнению с 454Life Science и SOLiD. Нами в данной работе использовался второй подход.

**Целью** данного исследования является изучение методологических подходов к проведению секвенирования полного транскриптома образцов плоскоклеточного рака пищевода.

### **Материалы и методы**

#### **Забор биологического материала**

Рекрутинг участников исследования проводился на базе ГККП «Онкологический центр» акимата г.Астана. Были рекрутированы пациенты с диагнозом плоскоклеточного рака пищевода, которым было назначено хирургическое лечение - операция тотальной/субтотальной резекции пищевода по Льюису. От каждого пациента, после ознакомления и подписания информированного согласия проведен сбор образцов ткани пищевода немедленно после резекции в РНК стабилизирующий раствор. Пациентам присваивался индивидуальный номер (ИД) с целью сохранения конфиденциальности.

#### **Выделение и очистка РНК из тканей.**

Выделение и очистка тотальной РНК из образцов опухолевой и нормальной тканей пищевода проводилась наборами RNAeasy mini kit (Qiagen), RNase free DNaseI set (Qiagen) согласно прилагаемым инструкции фирмы-производителя.

#### **Подготовка библиотек для секвенирования полного транскриптома.**

Подготовка библиотек для секвенирования полного транскриптома проводилась по протоколу TruSeq RNA Sample Preparation, Low Throughput (LT) Protocol, Illumina. Схематично протокол представлен на рисунке 1.

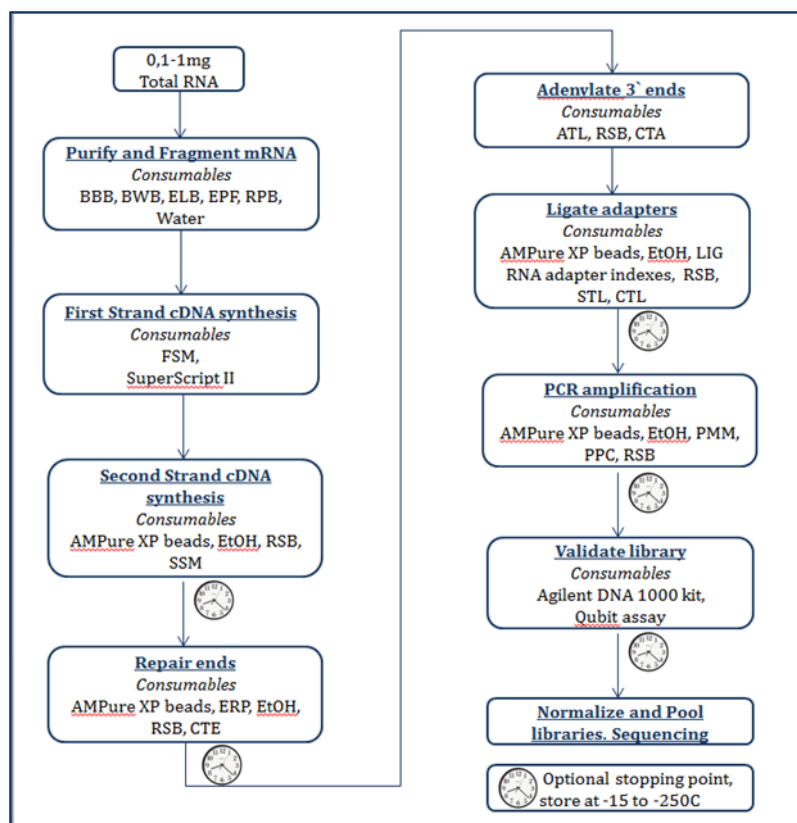


Рисунок 1 – Схема подготовки библиотеки TruSeq RNA.

Таким образом, основные этапы подготовки библиотеки включали:

- Purify and Fragment mRNA, во время данного этапа происходит двухкратное очищение mRNA, содержащего полиА концы на основе специальных магнитных шариков. При проведении второго элюирования происходит фрагментирование РНК для следующего этапа;

- Synthesize First Strand cDNA, это процесс обратной транскрипции фрагментированной РНК с участием обратной транскриптазы и рандомных праймеров-гексамеров, в результате которого происходит синтез первой цепи cDNA;

- Synthesize Second Strand cDNA, на этом этапе происходит удаление РНК матрицы и генерирование второй цепи второй цепи cDNA, с образованием в конечном итоге dsDNA;

- Perform End Repair, благодаря использованию End Repair Mix происходит затупление концов фрагментов dsDNA. Обладая 3'-5' экзонуклеазной активностью эта смесь срезает 3'-концы, одновременно благодаря полимеразной активности идет достраивание 5'-концов;

- Adenylate 3' Ends A single, благодаря использованию A-Tailing Mix происходит достраивание тупых 3'-концов библиотеки единичными нуклеотидами аденина (A), для подготовки к дальнейшему этапу лигирования с адапторами, так как адапторы на 3'-концах имеют комплементарный нуклеотид - тимин (Т);

- Ligate Adapters, целью данного процесса является лигирование мультииндекс-адапторов к концам фрагментов библиотеки ds cDNA, подготавливающий ее для этапа гибридизации на проточной ячейке;

- Enrich DNA Fragments, данный этап направлен на селективное обогащение DNA фрагментов, имеющих на своих концах адапторы на основе использования ПРЦ с коктейлем праймеров, отжигающихся на адапторах, подготавливающий ее для этапа гибридизации на проточной ячейке.

**Валидация библиотек**

Далее подготовленные библиотеки валидировали на флуориметре Qubit, биоанализаторе Agilent 2100 с целью нормализации по концентрации и подготовка для пулирования и гибридизации на

*Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».*

проточную ячейку. Qubit® dsDNA HS Assay Kit, Qubit 2.0 Fluorometer High Sensitivity DNA Analysis Kits, Bioanalyzer 2100, Agilent

#### Биоинформатический анализ.

Платформа секвенирования Illumina HiSeq2000 в результате процесса секвенирования путем синтеза, в конечном виде генерирует \*.bcl файлы после каждого цикла. Данные \*.bcl файлы одновременно конвертировались и демультимплексировались. Картирование и выравнивание сиквенсовых ридов на референсный геном человека сборки Ensembl GRC.h37.75 проводился с помощью STAR. Биоинформатический анализ выравненных сиквенсовых ридов и детекция уровня экспрессии генов и визуализация проводились с помощью DeSeq, R Bioconductor и пакета ggplot. Для анализа

транскриптомных профилей и уменьшения размерности данных применялся Метод Главных Компонент.

#### Результаты

В исследование включены биологические образцы тканей пищевода, собранные от 44 рекрутированных пациентов с диагнозом плоскоклеточного рака пищевода, проходивших хирургическое лечение на базе торакального отделения «Онкологический центр» акимата г.Астана с 2013 по 2015гг. На каждого пациента заполнялась анкета с полным диагнозом, результатами лабораторно-инструментальных исследований, гистологической характеристикой образцов. Краткая характеристика образцов приведена в таблице.

Таблица 1.

#### Характеристика образцов биоматериала, полученных от участников исследования.

	Кол-во		Возрастные группы			Средний возраст				
	N=44		>65	65-70	<70					
Мужчин	27		15	8	4	61,7				
Женщин	17		4	5	8	66,7				
TNM классификация (количество случаев) N=44										
T <sub>1</sub>		T <sub>2</sub>			T <sub>3</sub>			T <sub>4</sub>		
N <sub>1</sub> M <sub>0</sub>	N <sub>x</sub> M <sub>0</sub>	N <sub>0</sub> M <sub>0</sub>	N <sub>1</sub> M <sub>0</sub>	N <sub>2</sub> M <sub>0</sub>	N <sub>0</sub> M <sub>0</sub>	N <sub>1</sub> M <sub>0</sub>	N <sub>3</sub> M <sub>0</sub>	N <sub>x</sub> M <sub>0</sub>	N <sub>x</sub> M <sub>1</sub>	N <sub>x</sub> M <sub>0</sub>
1	3	5	2	1	7	2	4	16	1	2
Локализация опухоли, (количество случаев) N=44										
Шейный отдел пищевода		Средняя треть грудного отдела пищевода		Средняя и нижняя треть грудного отдела пищевода		Нижняя треть грудного отдела пищевода				
1		22		4		17				

В исследуемой выборке средний возраст выявления опухоли у мужчин составил 61,7 года, у женщин 66,7 лет, причем преобладали мужчины. Таким образом, чаще выявлялась T3 стадия заболевания, был случай диагностирования уже запущенной формы рака пищевода T4. По локализации преобладала средняя треть грудного отдела пищевода, в одном случае была диагностирована опухоль шейного отдела пищевода.

По гистологическому типу преобладали образцы с умеренно дифференцированным, неороговевающим типом плоскоклеточного

рака пищевода, чаще встречались образцы без некротических изъязвлений.

Для выделения ДНК из 10 образцов тканей проведено выделение и очистка от примесей ДНК тотальной РНК. Концентрации полученных РНК варьировали от 308 до 2176нг/мкл. полученные образцы тотальной РНК в количестве 1 мкг были взяты для подготовки библиотек.

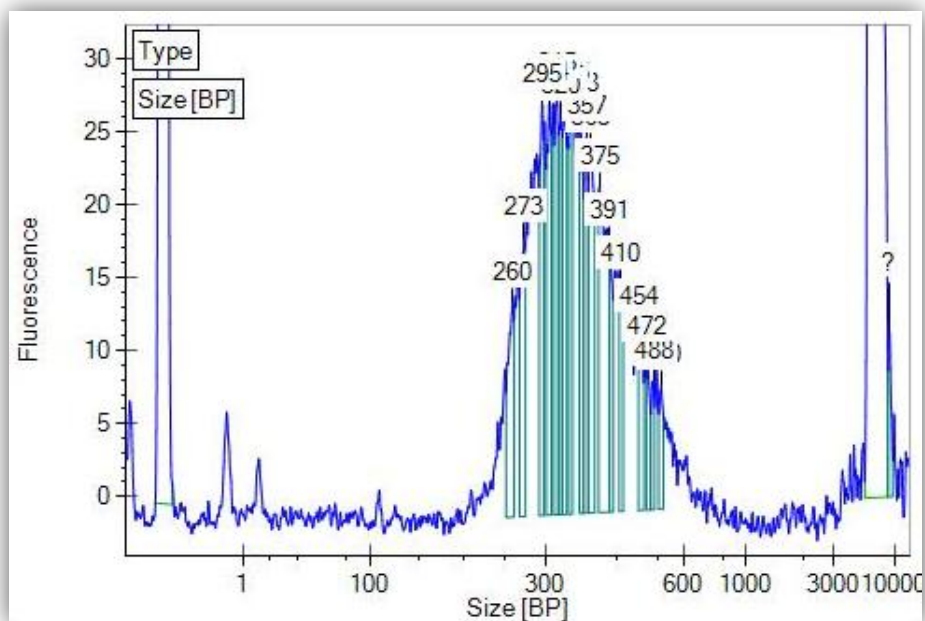
Согласно протокола производителя TruSeq RNA Sample Preparation, Low Throughput (LT) Protocol, Illumina были подготовлены 10 библиотек РНК. Основные этапы протокола включали очистку и фрагментирование РНК,

синтез первой цепи cDNA, синтез второй цепи cDNA, стадию «perform and repair», аденилирование 3' концов, лигирование адаптеров и PCR амплификацию.

Схема проведения эксперимента представлена выше. Завершающим этапом было проведение ПЦР для увеличения количества библиотеки. Критичным на данном этапе является количество циклов.

Желательно придерживаться рекомендуемых в протоколе, так как увеличение циклов ведет к обеднению разнообразия библиотеки.

Валидация библиотеки на анализаторе (LabChip GX II Touch 24, Perkin Elmer) проводилась согласно протокола фирмы производителя на чипе High Sensitivity Assay. Результаты приведены на рисунке 2.



**Рисунок 2 - Распределение фрагментов библиотеки, подготовленной на основе РНК. (LabChip GX II Touch 24, Perkin Elmer)**

Полученные библиотеки, как показано на рисунке 2 характеризовались фрагментами длиной от 250 до 470 пар оснований, целевые фрагменты были в диапазоне 300-330 пар нуклеотидов. Далее проводилась оценка концентрации библиотек на основе флуориметра Qubit 2.0

с использованием набора Qubit® dsDNA HS Assay Kit.

Расчет концентрации полученных библиотек на основе РНК был произведен по формуле (рисунок 3), согласно инструкции фирмы изготовителя. Диапазон концентраций был в пределах -40-61 наномоль.

$$\frac{\text{концентрация в нг/мкл}}{660\text{гг/моль} \cdot \text{средняя длина фрагментов библиотеки}} \cdot 10^6 = \text{концентрация в наномолях}$$

**Рисунок 3 – Формула расчета концентраций библиотек для нормализации и пулирования.**

Например, при показании флуориметра 15 нг/мкл и средней длине библиотеки 330 получаем значение 68,87 наномоль. Далее все библиотеки разводились до одной концентрации, согласно протоколов Illumina

мы разводили все библиотеки до 10 наномоль, т.е. была проведена нормализация по концентрации. Для этапа гибридизации на ячейке библиотеки объединялись в пулы по 4-5 библиотек. Далее по 2 нмоль пулированной

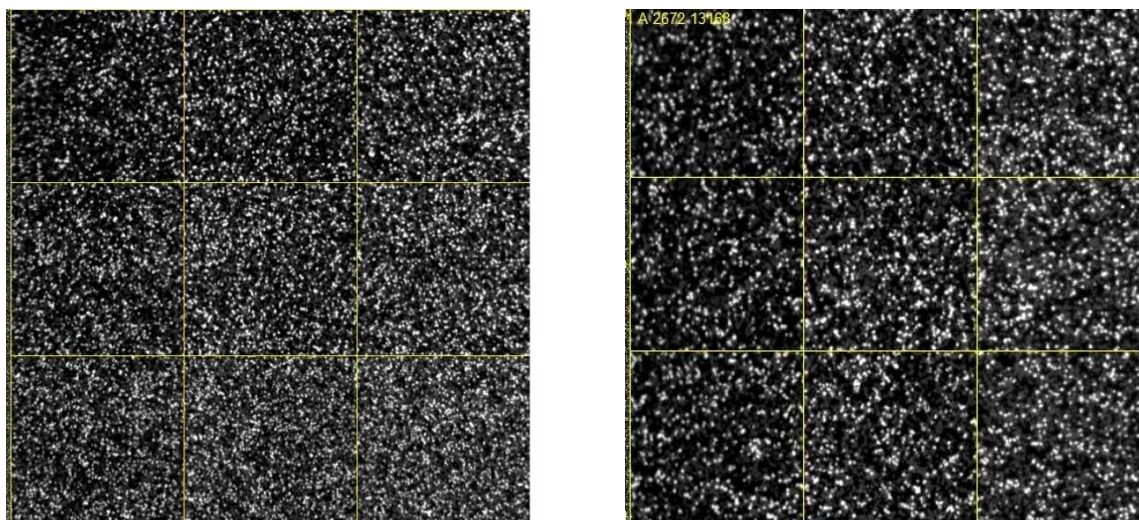
*Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».*

библиотеки отбиралось для этапа гибридизации на проточной ячейке. Гибридизация осуществлялась по протоколу PE\_Amp\_Lin\_Block\_TubeStripHyb\_v.8, входящему в стандартные протоколы прибора cBot/Illumina на ячейку, рассчитанную на 200 циклов секвенирования (HiSeq PE flow cell v3). На каждую дорожку вносился положительный контроль (библиотека вируса PhiX).

ДНК секвенирование. Секвенирование на платформе Illumina HiSeq2000 с циклической обратимой терминацией заключается в

обратимой терминации секвенирования после каждого цикла. Каждый цикл секвенирования начинается с присоединения к праймерной последовательности одного нуклеотида, комплементарного матрице.

Чтение происходит с обеих поверхностей ячейки верхней и нижней, пример картинка, формируемой прибором при проведении секвенирования представлен на рисунке 4. Каждый кластер представлен отдельной точкой на ячейке, в целом образуя картину «звездного неба».



а)

б)

**Рисунок 4 – Чтение кластеров на верхней (а) и нижней (б) поверхностях ячейки на секвенаторе HiSeq2000, Illumina**

Секвенатор нового поколения Illumina HiSeq2000 в результате процесса секвенирования путем синтеза, в конечном виде генерирует \*.bcl файлы после каждого цикла.

Высокое качество подготовленных библиотек затем было подтверждено результатами секвенирования и такими параметрами как плотность кластеров, среднее покрытие, общее количество ридов (таблица 2)

Таблица 2.

**Результаты сводного отчета по РНК секвенированию первого и второго чтения ячейки.**

Lane	read 1		read 2	
	cluster density	Q30	cluster density	Q30
#1	1092±73	89,8	1092±73	85,5
#2	1014±134	84,4	1014±134	82,6
#3	1104±58	87,6	1104±58	79,4
#4	1096±49	79,7	1096±49	55,2
#5	728±239	82,2	728±239	91,8
#6	1097±19	77,5	1097±19	76,6
#7	760±181	84,3	760±181	92,1
#8	619±138	90,7	619±138	93,2

Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».



Данные \*.bcl файлы одновременно конвертировались и демультимплексовались. Определение уровня экспрессии генов проводили с использованием HTSeq, а анализ дифференциальной экспрессии генов с помощью DESeq. Функциональный анализ дифференциально экспрессированных генов проводили с помощью Database for annotation, visualization and integrated discovery и ReViGo. Поиск сигнальных путей и биологических функций проводили на основе баз данных KEGG Pathway, Reactome Pathway. В результате картирования сиквенсовых ридов и

фильтрации данных, из более 40 тысяч генов, а также анализа дифференциально экспрессированных генов были выявлены гены с увеличенной и пониженной экспрессией в исследованных образцах рака пищевода. Проведенный анализ дифференциально экспрессированных генов специфичных для казахских пациентов в нормальной и опухолевой тканях выявил 188 высокоэкспрессированных и 287 низкоэкспрессированных генов. Наиболее значимые из них представлены в таблице 3.

Таблица 3.

**Результаты анализа высокоэкспрессированных и низкоэкспрессированных генов по базам данных сигнальных путей.**

База	Идентификатор	Кол-во	%	PValue	Гены
<b>высокоэкспрессированные гены</b>					
kegg_pathway	hsa03320: PPAR signaling pathway	4	2.7	0.01	LPL, OLR1, FABP4, MMP1
kegg_pathway	hsa04060: Cytokine-cytokine receptor interaction	6	4.05	0.05	AMH, INHBA, CCL3, IL8, CXCL3, CD70
reactome pathway	REACT_602: Metabolism of lipids and lipoproteins	4	2.7	0.03	LPL, SLCO1A2, APOE, FABP4
<b>Низкоэкспрессированные гены</b>					
kegg_pathway	hsa00980: Metabolism of xenobiotics by cytochrome P450	6	2.6	1.31E-4	GSTA1, UGT1A7, CYP3A5, ADH1B, ADH7, ALDH3A1
kegg_pathway	hsa00982: Drug metabolism	6	2.6	1.54E-4	GSTA1, UGT1A7, CYP3A5, ADH1B, ADH7, ALDH3A1
kegg_pathway	hsa00830: Retinol metabolism	4	1.73	0.01	UGT1A7, CYP3A5, ADH1B, ADH7
kegg_pathway	hsa00512: O-Glycan biosynthesis	3	1.3	0.02	GALNT5, B3GNT6, ST6GALNAC1
kegg_pathway	hsa00983: Drug metabolism	3	1.3	0.05	UGT1A7, CYP3A5, CES1

Гены с низкой и подавленной экспрессией связаны с сигналами клеточной дифференциации, процессами мышечной активности и секреции. Т.е. в исследованных нами опухолевых образцах с раком пищевода подавлена генная активность клеточной дифференциации, что является отличительной чертой опухолевых клеток и связано с потерей характерных для нормальных клеток способностей образовывать специфические тканевые

структуры или продуцировать специфические вещества. При анализе списка низкоэкспрессированных генов были обнаружены функционально значимые сигнальные пути, связанные с метаболизмом лекарств.

**Обсуждение результатов**

Целью настоящей работы являлось изучение и апробация методологического подхода к проведению РНК секвенирования образцов плоскоклеточного рака пищевода. В

*Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».*

ходе выполнения данной работы нами проведены этапы выделения РНК из тканей пищевода, подготовка библиотек для РНК секвенирования и секвенирование по технологии Next Generation Sequencing (NGS) на платформе Illumina HiSeq 2000. Высокое качество подготовленных библиотек было подтверждено в процессе анализа объективными параметрами. В работах многих исследователей наблюдается использование технологии NGS для различных аспектов исследования рака пищевода. Так в одной из работ с использованием NGS технологии выявлено, что мутации гена NOTCH1 ассоциированы со снижением выживаемости пациентов с плоскоклеточным раком пищевода и резистентностью к химиотерапии, в то время как мутации гена PIK3CA были связаны с хорошим ответом на химиопрепараты и лучшими показателями выживаемости пациентов [2]. В работе также использовалась платформа Illumina HiSeq 2000, однако в отличие от нашего исследования было проведено секвенирование полного генома и полного экзона образцов опухолевой ткани, полученных как и в нашем исследовании от пациентов с различными стадиями плоскоклеточного рака пищевода [2]. В работе других ученых мутации выявленные в генах P53, CDKN2A, FAT1, генах регуляции клеточного цикла на основе NGS технологии являются перспективными для разработки персонализированной терапии и ранней диагностики рака пищевода. [15]. В нашем исследовании также выявлена вовлеченность генов регуляции клеточного цикла.

Таким образом, технология РНК секвенирования на платформе Illumina HiSeq 2000 является перспективной во многих фундаментальных и прикладных исследованиях в качестве современного инструмента поиска предиктивных, прогностических и лечебно-диагностических таргетных генов.

#### **Выводы**

Основной целью данной работы была разработка методологического подхода РНК секвенирования на основе технологии нового поколения Illumina. Мы надеемся что данная

технология поможет лучше понять патогенез развития плоскоклеточного рака пищевода, в результате чего можно будет определить генетические маркеры для разработки персонализированной диагностики и терапии.

Для поиска новых генов-мишеней, перспективных для ранней диагностики плоскоклеточного рака пищевода планируется провести сравнительный анализ генетических маркеров, выявленных при биоинформатическом анализе с различными стадиями заболевания по TNM классификации.

Коллектив авторов подтверждает отсутствие конфликта интересов при выполнении данного исследования.

#### **Вклад авторов в исследование:**

Рахимова С.Е.– подготовка библиотек, валидация библиотек, проведение РНК секвенирования;

Каиров У.Е., Молкенов А.Б.– биоинформатический анализ;

Ережепов Д.А. – подготовка библиотек, валидация библиотек;

Жалбинова М.Р. – выделение РНК, оценка качества нуклеиновых кислот, участие в подготовке библиотек;

Абильмажинова А.Т. – выделение РНК, оценка качества нуклеиновых кислот, участие в подготовке библиотек;

Жуков Ю.В., Омаров М.Х. – хирург-онкологи, забор биоматериала, сбор клинических данных;

Акильжанова А.Р. – руководитель исследования.

Проект, в рамках которого выполнялось представленное исследование, был обсужден и принят на заседании Этической Комиссии ЧУ «Центр наук о жизни» АОО «Назарбаев Университет». Выписка из протокола №13 от 12.03.2014г.

Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».

**Литература:**

1. Asha GVD, Veerendrakumar KV. Relationship between Risk Factors and Dietary Intake of Oesophageal Cancer Patients. *J Hum Ecol.* 2005, 17: 37-45.
2. Caixia Cheng, Heyang Cui, Ling Zhang, Zhiwu Jia, Bin Song, Fang Wang, Yaoping Li, Jing Liu, Pengzhou Kong, Ruyi Shi, Yanghui Bi, Bin Yang, Juan Wang, Zhenxiang Zhao, Yanyan Zhang, Xiaoling Hu, Jie Yang, Chanting He, Zhiping Zhao, Jinfen Wang, Yanfeng Xi, Enwei Xu, Guodong Li, Shiping Guo, Yunqing Chen, Xiaofeng Yang, Xing Chen, Jianfang Liang, Jiansheng Guo, Xiaolong Cheng, Chuangui Wang, Qimin Zhan and Yongping Cui. Genomic analyses reveal FAM84B and the NOTCH pathway are associated with the progression of esophageal squamous cell carcinoma. *GigaScience.* 2016, 5:1-13.
3. Freedman N.D., Abnet C.C., Leitzmann M.F., Mouw T., Subar A.F., Hollenbeck A.R., Schatzkin A.A. Prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am J Epidemiol.* 2007, 165:1424-33.
4. Kawaguchi H., Ohno S., Miyazaki M., Hashimoto K., Egashira A., Saeki H., Watanabe M., Sugimachi K. CYFRA 21-1 determination in patients with esophageal squamous cell carcinoma: clinical utility for detection of recurrences. *Cancer.* 2000, 89:1413-7.
5. Mandard A.M., Hainaut P., Hollstein M. Genetic steps in the development of squamous cell carcinoma of the esophagus, *Mutat Res.* 2000, 462:335-42.
6. Manoel-Caetano Fda S., Borim A.A., Caetano A., Cury P.M., Silva A.E. Cytogenetic alterations in chagasic achalasia compared to esophageal carcinoma, *Cancer Genet Cytogenet.* 2004, 149:17-22.
7. Mc. Cann. J. Esophageal cancers: changing character, increasing incidence. *J Natl Cancer Inst.*, 1999, 91:497-8.
8. Metzger R., Schneider P.M., Warnecke-Eberz U., Brabender J., Holscher A.H. Molecular biology of esophageal cancer. *Onkologie,* 2004, 2:200-206.
9. Novacek G. Plummer-Vinson syndrome. *Orphanet J Rare Dis.* 2006, 1:36.
10. Parkin D.M., Pisani P., Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. *Int J Cancer,* 1999; 80: 827-41.
11. Parkin D.M., Bray F., Ferlay J., Pisani P. Global cancer statistics, 2002, *CA Cancer J Clin.* 2005, 55:74-108.
12. Shimada H., Takeda A., Arima M., Okazumi S., Matsubara H., Nabeya Y., Funami Y., Hayashi H., Gunji Y., Suzuki T., Kobayashi S., Ochiai T. Serum p53 antibody is a useful tumor marker in superficial esophageal squamous cell carcinoma. *Cancer.* 2000, 89:1677-83.
13. Wakatsuki M., Suzuki Y., Nakamoto S., Ohno T., Ishikawa H., Kiyohara H., Kiyozuka M., Shirai K., Nakayama Y., Nakano T. Clinical usefulness of CYFRA 21-1 for esophageal squamous cell carcinoma in radiation therapy. *J Gastroenterol Hepatol.* 2007, 22:715-9.
14. Yamabuki T., Daigo Y., Kato T., Hayama S., Tsunoda T., Miyamoto M., Ito T., Fujita M., Hosokawa M., Kondo S., Nakamura Y. Genome-wide gene expression profile analysis of esophageal squamous cell carcinomas. *Int J Oncol.* 2006, 28:1375-84.
15. Yasushi Sasaki, Miyuki Tamura, Ryota Koyama, Takafumi Nakagaki, Yasushi Adachi, Takashi Tokino. Genomic characterization of esophageal squamous cell carcinoma: insights from next-generation sequencing. *World J Gastroenterol.* 2016, 22(7):2284-2293.

**Контактная информация:**

**Рахимова Сауле Есламовна** - к.б.н., ведущий научный сотрудник Лаборатории геномной и персонализированной медицины, ЧУ «National Laboratory Astana», Назарбаев университет.

**Почтовый адрес:** 010000, г. Астана, проспект Кабанбай батыра, 53, кабинет 1173.

**E-mail:** saule.rakhimova@nu.edu.kz

**Телефон:** +7 7172 70 93 04

*Работа выполнена в рамках проекта «Геномный и транскриптомный профиль рака пищевода» по бюджетной программе МОН РК 0666/ПЦФ «Развитие трансляционной и персонализированной медицины для создания основ биомедицинской индустрии в Республике Казахстан на 2014-2016 гг».*