

Наименование проекта: ИРН АР19678041 «Разработка программного обеспечения для идентификации tandemных повторов при полногеномном секвенировании»

Актуальность:

За последние десятилетия кардинально изменилось представление роли повторяющихся последовательности в геноме, и из разряда «мусорной ДНК», повторяющие элементы оказывают большое влияние на функционирование и эволюцию геномов их хозяев, способствуя генетическому разнообразию и появлению новых регуляторных элементов. Дальнейшее развитие технологии секвенирования и в особенности секвенирование третьего поколения значительно способствует в изучении tandemных повторов, что привело к появлению новых данных для детального изучения. Установлено, что на долю коротких tandemных повторов приходится около 7% генома человека. Широкая представленность в геномах эукариот и прокариот, и их высокая скорость изменчивости, как один из ключевых факторов эволюции генома и генетической диверсификации, повторы будут систематически оцениваться на предмет их роли. Поскольку известно, что многие из этих элементов активируются при заболеваниях, возникает потенциал для персонализированной медицины и диагностики болезней относительно генетических изменений и ожидаемых последствий при анализе tandemных повторов и выявление ассоциаций биомаркеров и регулированием биологических процессов в организме. В связи с этим разработка продвинутых и простых в использовании биоинформатических инструментов для идентификации разнообразных форм tandemных повторов, является **актуальной задачей.**

Цель предлагаемого проекта заключается в разработке биоинформационного приложения с открытым доступом для идентификации и анализа вариабельности tandemных повторов, в том числе, в исходных данных при полногеномном секвенировании третьего поколения.

Ожидаемые и достигнутые результаты:

В рамках проекта будут применяться биоинформатические алгоритмы для идентификации родственных последовательностей с разным уровнем дивергенции, а также языки и средства программирования при разработке структуры и интерфейса приложения. Подтверждение достоверности результатов, получаемых в разрабатываемом программном обеспечении, будет проводится стандартными лабораторными молекулярно-генетическими методами, включая полногеномное секвенирование геномов прокариот и эукариот, и методы для дифференциации tandemных повторов с помощью капиллярного электрофореза. **Основным результатом** реализуемого проекта будет программное обеспечение с открытым доступом и пользовательским интерфейсом для идентификации tandemных повторов при полногеномном секвенировании. Программное обеспечение позволит идентифицировать

разнообразии целевых локусов с тандемными повторами, в том числе, в исходных данных полногеномного секвенирования и проводить статистический анализ идентифицированных вариантов.

В ходе реализации проекта, будут опубликованы не менее 2 (двух) статей и (или) обзоров в рецензируемых научных изданиях, входящих в 1 (первый) и (или) 2 (второй) квартиль по импакт-фактору в базе Web of Science и (или) имеющих процентиль по CiteScore в базе Scopus не менее 65 (шестьдесят пять); либо не менее 1 (одной) статьи или обзора в рецензируемом научном издании, входящем в 1 (первый) квартиль в базе Web of Science или процентиль по CiteScore в базе Scopus не менее 95 (девяносто пять). Все биоинформационные коды, скрипты будут размещены в постоянных, открытых репозиториях, а также размещены на Github со свободным доступом.

Члены исследовательской группы:

руководитель проекта – Исмаилова Айсулу Абжаппаровна, ГНС, PhD, ассоциированный профессор

ORCID: [0000-0002-8958-1846](https://orcid.org/0000-0002-8958-1846)

Scopus/WoS (Hirsch Index = 3): Scopus Author ID: [56145830200](https://orcid.org/56145830200)

исследовательская группа:

1) **Календарь Руслан Николаевич**, ГНС, к.б.н., биолог-генетик, Профессор (Биология), Доцент генетики (университет Хельсинки)

ORCID: [0000-0003-3986-2460](https://orcid.org/0000-0003-3986-2460)

Scopus/WoS (Hirsch Index = 34): ResearcherID: [D-9751-2012](https://orcid.org/D-9751-2012)

Scopus ID: [6602789279](https://orcid.org/6602789279)

2) **Бельдеубаева Жанар Толеубаевна**, ВНС, PhD

ORCID: [0000-0003-4056-6220](https://orcid.org/0000-0003-4056-6220)

Scopus/WoS (Hirsch Index =2): Scopus Author ID: [56951278600](https://orcid.org/56951278600)

3) **Сатыбалдиева (Сатекбаева) Айжан Жанабековна**, ВНС, PhD

ORCID: [0000-0001-5740-7934](https://orcid.org/0000-0001-5740-7934)

Scopus/WoS (Hirsch Index =2): Scopus Author ID: [56145597900](https://orcid.org/56145597900)

4) **Шевцов Владислав Александрович**, СНС, магистр технических наук докторант кафедры «Информационные системы» НАО «Казахский агротехнический университет им. С. Сейфуллина»

ORCID: [0000-0001-6202-2123](https://orcid.org/0000-0001-6202-2123)

Scopus/WoS (Hirsch Index =3): Scopus Author ID: [57216896596](https://orcid.org/57216896596)

5) **Голенко Екатерина Сергеевна**, СНС, магистр технических наук

ORCID: [0000-0002-4643-4571](https://orcid.org/0000-0002-4643-4571)

6) **Вакансия**, ВНС, IT архитектор, программист

7) **Вакансия**, НС, докторант

Информация для потенциальных пользователей:

Область применения разработанного программного обеспечения: биоинформатика, медицинская и сельскохозяйственная генетика, генетика

микроорганизмов. **Результаты этого проекта** имеют большое значение, в том числе, для фундаментальных наук. Программное обеспечение позволит эффективно идентифицировать тандемных повторов и установлению ассоциаций между разнообразием тандемных повторов с генетическими заболеваниями человека и с генетическим разнообразием микроорганизмов и их патогенностью. Реализация проекта позволит усилить направление биоинформатики в ведущем вузе страны и создаст платформу для специализации и профориентации обучающихся.