

Кечин Андрей Андреевич, к.б.н., a.a.kechin@gmail.com, руководитель работы, НГУ, старший преподаватель ФЕН и ИМПЗ (по совместительству); ИХБФМ СО РАН, научный сотрудник Лаборатории фармакогеномики

Боробова Виктория Сергеевна, студент ФЕН НГУ, 2-й курс магистратуры, кафедра молекулярной биологии и биотехнологии, работа не курсовая и не дипломная, научный руководитель – Кечин А.А., обучение заканчивается в 2022 году

Работать будут Кечин А.А. и Боробова В.С. Кечин А.А. уже работал на кластере.

- Название (тема) работы.

Механизмы канцерогенеза BRCA-зависимого рака яичников

- Постановка задачи (что именно должно быть сделано, какие результаты должны быть получены).

Цель: выявить новые механизмы канцерогенеза BRCA-зависимого рака яичников. Задачи: собрать выборку образцов свежемороженой опухолевой ткани, выделить высокомолекулярную ДНК, определить крупные перестройки, выявить новые потенциальные мишени для терапии. Непосредственно на кластере будет проводиться преобразование сигнала изменения тока ионов через нанопору в последовательность нуклеотидов. Для того, чтобы этот процесс шел достаточно быстро, требуется CUDA.

- Современное состояние проблемы, с ссылками на источники, предпочтительно доступные в сети Интернет (существующие работы, ваши или других коллективов, на которые вы будете опираться).

Рак яичников является одним из наиболее смертоносных типов опухолей, достигающий более, чем 20 % смертности, в первый год после постановки диагноза [1]. Столь высокая смертность, несмотря на хороший первоначальный ответ большинства пациенток (более 75 %), связана в первую очередь с быстрым развитием резистентности к используемым химиопрепаратам [2]. В связи с этим в литературе появляется все больше работ, направленных на поиск новых мишеней для терапии. Одно из наиболее перспективных направлений – это применение ингибиторов тирозинкиназ, некоторые из которых (например, пазопаниб) уже показали высокую эффективность для лечения пациенток с резистентностью к препаратам платины [3], для других (гефитиниб, эрлотиниб, лапатиниб, топотекан) – показано лишь усиление побочных эффектов [4–7]. Тем не менее, за последние годы предложено несколько новых перспективных мишеней для ингибиторов тирозинкиназ [8–11]. В связи с чем поиск новых мишеней с использованием технологии секвенирования последнего поколения (с протяженными прочтениями) может позволить определить новые ранее пропускаемые события канцерогенеза. Это подтверждает и недавняя работа по полногеномному секвенированию с использованием коротких прочтений клеточных линий рака яичников, для которых была показана корреляция наличия впервые выявленных крупных перестроек и чувствительностью к различным химиопрепаратам [12]. Применение секвенирования с протяженными прочтениями позволит значительно расширить список таких корреляций.

При исследовании соматических мутаций, приводящих к развитию рака яичников как при BRCA-зависимом (с герминальной или соматической мутацией в гене *BRCA1* или *BRCA2*), так и BRCA-независимом типах рака, точечные мутации в основных генах контроля деления, клеточного цикла и репарации ДНК (*KRAS*, *NRAS*, *EGFR*, *PIK3CA*, *PTEN*, *BRAF*) встречаются достаточно редко, особенно при высокодифференцированной серозной аденокарциноме [13]. Исключением является ген *TP53*, в котором мутации наблюдаются у более, чем 70% всех пациенток [13]. Наличие последней в большинстве исследуемых опухолевых клеток говорит о её достаточно

раннем появлении при малигнизации клеток, однако сама по себе такая мутация не приводит к потере контроля над частотой делений. В то же время она ведет к нарушению контроля за целостностью генома и, как следствие, накоплению большого числа крупных перестроек, что и было показано для рака яичников при анализе данных TCGA (база данных секвенирования образцов от пациентов с различными онкологическими заболеваниями) [14]. Примерами уже известных клинически значимых крупных перестроек при раке яичников являются амплификация генов *PIK3CA* (38%), *CCNE1* (циклин E1, 22%), *AURKA* (аврора А киназа, 20%) и *KRAS* (15%) [14]. Результатом крупных перестроек может быть активация тирозинкиназ через формирование химерных генов. В недавней работе было показано, что «сшивание» различных генов может быть представлено по-отдельности не менее, чем у 50% пациенток с диагнозом рак яичников, что суммарно может достигать 77% пациенток [15]. Кроме того, механизмами развития резистентности могут быть амплификация *CCNE1*, а также нарушение функций (чаще крупными перестройками) генов *RB1*, *NF1*, *RAD51B* и *PTEN* [16].

По большинству данных, полученных на небольших выборках, *PTEN* может теряться примерно в половине всех случаев высокодифференцированной серозной аденокарциномы [17], что расходится с результатами анализа данных TCGA, согласно которым такое событие наблюдалось только в 7% случаев [18]. Такие различия в результатах выявления крупных перестроек могут быть связаны с особенностями хранящихся в TCGA данных. Во-первых, большинство образцов представляли собой парафиновые гистологические блоки, ДНК которых значительно модифицируется и фрагментируется при фиксации материала, что усложняет выявление крупных перестроек. Во-вторых, при извлечении информации из больших массивов данных без возможности подтверждения результата поиск ведется «с максимальной специфичностью», то есть таким образом, чтобы найти максимум новой информации при минимальной контаминации ложной информацией. В-третьих, все образцы секвенировались с помощью коротких прочтений (не более, чем по 250 нуклеотидов с каждой стороны). Представленные на сегодняшний день в литературе результаты исследования крупных перестроек при раке яичников в большинстве случаев представляют собой анализ данных TCGA, что приводит к смещению в имеющихся представлениях о встречаемости крупных перестроек в геномах опухолевых клеток. Использование же свежезамороженной опухолевой ткани одновременно с длинными прочтениями Oxford Nanopore позволит получить **новые данные** о встречаемости и клиническом значении крупных перестроек при BRCA-зависимом раке яичников.

На сегодняшний день, применимость данной технологии секвенирования для исследования образцов опухоли уже показана во многих работах: для исследования крупных перестроек в гене *BRCA1* [19], поиска крупных перестроек в генах *SMAD4*, *CDKN2A* [20]. Несомненным преимуществом новой технологии является и одновременное определение метилированных цитозинон [21], что позволяет оценивать активность промоторов клинически значимых генов. О необходимости использования технологий секвенирования с протяженными прочтениями для расширения наших знаний о механизмах канцерогенеза говорится во многих обзорах [22, 23]. Некоторые недавно проведенные исследования образцов ДНК из опухоли уже доступны для чтения в базе данных bioRxiv или medRxiv [24, 25], что говорит о **высокой актуальности** исследуемой проблемы, однако достоверность и правильность выводов этих источников пока не была оценена внешними рецензентами.

Таким образом, наши знания о крупных перестройках и их связи с механизмами канцерогенеза при раке яичников являются ограниченными, поскольку получены в большинстве случаев с помощью подходов, не позволяющих их надежно идентифицировать. В связи с этим, исследование крупных перестроек при BRCA-зависимом раке яичников является **актуальной и перспективной задачей**.

Для исследования механизмов канцерогенеза при BRCA-зависимом раке яичников будет использоваться уже собранная выборка образцов свежемороженой опухолевой ткани высокодифференцированной серозной аденокарциномы и соответствующие образцы цельной крови от тех же пациенток. Для всех из них уже было проведено секвенирование генов BRCA1, BRCA2 и TP53. Будут отобраны 4 образца опухолевой ткани и соответствующих образцов цельной крови, для которых была найдена мутация в генах *BRCA1* и *TP53* и которые имеют наибольшую представленность опухолевых клеток. Для этого совместно с коллегами из лаборатории будут приготовлены гистологические срезы с окраской гематоксилин-эозином для оценки содержания на срезах опухолевых клеток. Далее для отобранных образцов будет проведено выделение высокомолекулярной геномной ДНК с использованием легкоплавкой агарозы, как это было сделано, например, Gabrieli и соавт. [19]. Контроль длин получаемых фрагментов будет осуществляться с помощью пульс-гель-электрофореза, доступного для использования у коллег из Института молекулярной и клеточной биологии СО РАН.

Для каждой пары отобранных образцов будут приготовлены библиотеки с использованием реагентов Oxford Nanopore Technologies, New England Biolabs и Beckman Coulter (полный список реагентов см. в п. 4.11). Далее для каждого из образцов опухоли будет проведено секвенирование на одной проточной ячейке SpotON Flow Cell (R9.4) Single (Oxford Nanopore Technologies, по одной ячейке на образец) на приборе MiniON с программой MinKNOW. Образцы ДНК из цельной крови будут секвенироваться на одной проточной ячейке путем пулирования отдельно приготовленных библиотек и с добавлением индексов. Предполагается, что на каждый образец опухоли будет прочитано до 50 миллиардов нуклеотидов, что даст возможность прочитать каждый геном с покрытием до 15 прочтений на позицию. Этого будет достаточно для выявления как крупных перестроек, так и точечных мутаций с высокой представленностью в опухоли (не менее 7–10%). Тестирование ДНК из мононуклеарных клеток крови для фильтрации герминальных вариантов будет проведено с более низким покрытием (около 5х на каждый образец), чего будет достаточно для выявления герминальных крупных перестроек.

Первичная обработка данных с прибора MiniON будет осуществляться с использованием стандартной программы Oxford Nanopore Technologies – Albacore. Преобразование прочтений в FASTQ-формат и их картирование будут проводиться программами poretools и BWA, соответственно. Выявление изменений числа копий и других крупных перестроек, а также определение статуса метилирования будет выполнено с помощью программ QDNaseq, DNAsour, LAST и собственными Python-скриптами, используя общедоступные алгоритмы (такие, как скрытая марковская модель) [26]. Точечные мутации будут выявлены с помощью программ Pises и Freebayes и, сравнивая результаты секвенирования с данными, полученными ранее на приборе MiniSeq Illumina, будут определены параметры для фильтрации ложноположительных мутаций. Часть (около 10-20) выявленных крупных перестроек будут подтверждены секвенированием по Сэнгеру, для чего будут разработаны праймеры, фланкирующие место перестройки. Результаты сравнения двух методов будут использованы для определения ложноположительных мутаций и параметров их фильтрации с составлением ROC-кривых.

Для всех выявленных крупных перестроек будут предсказаны эффекты на работу затрагиваемых ими генов. Среди выявленных крупных перестроек предполагается найти новые химерные гены, приводящие к активации одних киназ, регулирующих работу клеток, а также крупные перестройки, нарушающие работу других. Высока вероятность, что среди новых нарушений сигнальных путей клетки найдутся такие, которые могут стать мишенью для разработки новых таргетных противоопухолевых препаратов либо маркером при выборе тактики лечения пациента. Суммируя результаты исследования (точечные мутации, метилирование промоторов и крупные перестройки) для каждого из образцов будет предположен механизм канцерогенеза.

Список литературы

1. Каприн А.Д., Старинский В.В., Петрова Г.В. Состояние онкологической помощи населению России в 2018 году // 2019.
2. Katopodis P., Chudasama D., Wander G., et al. Kinase Inhibitors and Ovarian Cancer. // *Cancers (Basel)*. – 2019. – V. 11. – No. 9. – .
3. Dinkic C., Eichbaum M., Schmidt M., et al. Pazopanib (GW786034) and cyclophosphamide in patients with platinum-resistant, recurrent, pre-treated ovarian cancer - Results of the PACOVAR-trial // *Gynecol. Oncol.* – 2017. – V. 146. – No. 2. – P. 279–284.
4. Posadas E.M., Liel M.S., Kwitkowski V., et al. A phase II and pharmacodynamic study of gefitinib in patients with refractory or recurrent epithelial ovarian cancer // *Cancer*. – 2007. – V. 109. – No. 7. – P. 1323–1330.
5. Hirte H., Oza A., Swenerton K., et al. A phase II study of erlotinib (OSI-774) given in combination with carboplatin in patients with recurrent epithelial ovarian cancer (NCIC CTG IND.149) // *Gynecol. Oncol.* – 2010. – V. 118. – No. 3. – P. 308–312.
6. Garcia A.A., Sill M.W., Lankes H.A., et al. A phase II evaluation of lapatinib in the treatment of persistent or recurrent epithelial ovarian or primary peritoneal carcinoma: A gynecologic oncology group study // *Gynecol. Oncol.* – 2012. – V. 124. – No. 3. – P. 569–574.
7. Lheureux S., Krieger S., Weber B., et al. Expected Benefits of Topotecan Combined With Lapatinib in Recurrent Ovarian Cancer According to Biological Profile // *Int. J. Gynecol. Cancer*. – 2012. – V. 22. – No. 9. – P. 1.
8. Yu Y., Suryo Rahmanto Y., Shen Y.-A., et al. Spleen tyrosine kinase activity regulates epidermal growth factor receptor signaling pathway in ovarian cancer // *EBioMedicine*. – 2019. – V. 47. – P. 184–194.
9. Yu Y., Suryo Rahmanto Y., Lee M.-H., et al. Inhibition of ovarian tumor cell invasiveness by targeting SYK in the tyrosine kinase signaling pathway // *Oncogene*. – 2018. – V. 37. – No. 28. – P. 3778–3789.
10. Lane D., Matte I., Laplante C., et al. CCL18 from ascites promotes ovarian cancer cell migration through proline-rich tyrosine kinase 2 signaling // *Mol. Cancer*. – 2016. – V. 15. – No. 1. – P. 58.
11. Song G., Chen L., Zhang B., et al. Proteome-wide Tyrosine Phosphorylation Analysis Reveals Dysregulated Signaling Pathways in Ovarian Tumors. // *Mol. Cell. Proteomics*. – 2019. – V. 18. – No. 3. – P. 448–460.
12. Papp E., Hallberg D., Konecny G.E., et al. Integrated Genomic, Epigenomic, and Expression Analyses of Ovarian Cancer Cell Lines. // *Cell Rep*. – 2018. – V. 25. – No. 9. – P. 2617–2633.
13. Garziera M., Roncato R., Montico M., et al. New Challenges in Tumor Mutation Heterogeneity in Advanced Ovarian Cancer by a Targeted Next-Generation Sequencing (NGS) Approach. // *Cells*. – 2019. – V. 8. – No. 6. – .
14. Ciriello G., Miller M.L., Aksoy B.A., Senbabaoglu Y., Schultz N., Sander C. Emerging landscape of oncogenic signatures across human cancers. // *Nat. Genet*. – 2013. – V. 45. – No. 10. – P. 1127–33.
15. Yu Y.-P., Liu P., Nelson J., et al. Identification of recurrent fusion genes across multiple cancer types // *Sci. Rep*. – 2019. – V. 9. – No. 1. – P. 1074.
16. Patch A.-M., Christie E.L., Etemadmoghadam D., et al. Whole-genome characterization of chemoresistant ovarian cancer // *Nature*. – 2015. – V. 521. – No. 7553. – P. 489–494.
17. Martins F.C., Santiago I. de., Trinh A., et al. Combined image and genomic analysis of high-grade serous ovarian cancer reveals PTEN loss as a common driver event and prognostic classifier // *Genome*

Biol. – 2014. – V. 15. – No. 12. – P. 526.

18. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma // Nature. – 2011. – V. 474. – No. 7353. – P. 609–615.

19. Gabrieli T., Sharim H., Fridman D., Arbib N., Michaeli Y., Ebenstein Y. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). // Nucleic Acids Res. – 2018. – V. 46. – No. 14. – P. e87.

20. Norris A.L., Workman R.E., Fan Y., Eshleman J.R., Timp W. Nanopore sequencing detects structural variants in cancer. // Cancer Biol. Ther. – 2016. – V. 17. – No. 3. – P. 246–53.

21. Euskirchen P., Bielle F., Labreche K., et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing // Acta Neuropathol. – 2017. – V. 134. – No. 5. – P. 691–703.

22. Sakamoto Y., Sereewattanawoot S., Suzuki A. A new era of long-read sequencing for cancer genomics // J. Hum. Genet. – 2020. – V. 65. – No. 1. – P. 3–10.

23. Mantere T., Kersten S., Hoischen A. Long-Read Sequencing Emerging in Medical Genetics // Front. Genet. – 2019. – V. 10. – P. 426.

24. Aganezov S., Goodwin S., Sherman R., et al. Comprehensive analysis of structural variants in breast cancer genomes using single molecule sequencing // bioRxiv. – 2019P. 847855.

25. Valle-Inclan J.E., Stangl C., Jong A.C. de., et al. Rapid identification of genomic structural variations with nanopore sequencing enables blood-based cancer monitoring // medRxiv. – 2019P. 19011932.

26. Simpson J.T., Workman R.E., Zuzarte P.C., David M., Dursi L.J., Timp W. Detecting DNA cytosine methylation using nanopore sequencing // Nat. Methods. – 2017. – V. 14. – No. 4. – P. 407–410.

- Научная новизна, практическая значимость работы (для чего она выполняется, в чём смысл и практическая / потенциальная польза).

На сегодняшний день основным материалом для исследования опухолевой ткани являются парафиновые гистологические блоки, фиксированные в формалине (так называемые formalin fixed paraffin embedded – FFPE). Такой материал обладает многими преимуществами, в первую очередь, связанными с удобством их транспортировки и хранения. Однако фиксация нарушает целостность молекул ДНК, а также вводит различные модификации в азотистые основания. В данном проекте предполагается использование свежемороженых образцов опухолевой ткани – редко используемый тип образцов во всем мире, однако именно это позволит выделить геномную ДНК в виде протяженных фрагментов (до сотен тысяч пар оснований).

Поэтому уникальность материала (высокодифференцированная серозная аденокарцинома), тип образцов (свежемороженая опухолевая ткань), платформа, на которой будет проводиться полногеномное секвенирование образцов опухоли (MinION) с протяженными прочтениями, совместно обеспечивают высокую **научную новизну** проекта. Согласно данным литературы до момента написания проекта, таких работ не проводилось нигде в мире. Все это связано с относительно недавним появлением данной технологии секвенирования и редким использованием образцов свежемороженой опухолевой ткани.

Работы выполняются по Гранту Президента МК-4082.2021.1.4, номер соглашения 075-15-2021-200

Еще одна небольшая работа будет сделана по этому же типу секвенирования, но секвенироваться будут геномы бактерий. Эта работа делается без поддержки гранта.

Результаты

Задачей второго года работы по проекту было выделение высокомолекулярной ДНК из образцов опухоли рака яичников, проведение секвенирования на платформе MinION (Oxford Nanopore) и анализ данных с определением точечных мутаций, крупных перестроек и гиперметилированных промоторов генов с целью идентификации нарушений в сигнальных путях регуляции клеточного цикла (химерные гены, смена или выключение промоторов генов, делеции или амплификации генов), а также статистическая обработка полученных данных и анализ результатов исследования механизмов канцерогенеза BRCA-зависимого рака яичников.

За второй год были выполнены следующие работы:

- выделена высокомолекулярная ДНК из образца опухоли рака яичников;
- проведен контроль качества образца ДНК;
- приготовлена библиотека для секвенирования на приборе MinION (Oxford Nanopore);
- проведено секвенирование и проанализированы данные;
- выявлены крупные геномные перестройки и гены с гиперметилированными промоторами;
- сконструированы праймеры для подтверждения нескольких перестроек в геноме и проведена проверка их наличия;
- проведена статистическая обработка полученных данных, предложены мишени для дальнейшего их исследования на расширенной выборке образцов.

Выделение высокомолекулярной ДНК и контроль качества

ДНК была выделена из образца опухоли рака яичников стандартным фенол-хлороформным методом, который использовался и в первый год выполнения проекта. На всех этапах были исключены резкие взбалтывания раствора, а также быстрые пипетирования для исключения фрагментации ДНК. Количество и качество выделенных образцов было оценено на приборе NanoDrop OneC (Thermo Scientific) согласно инструкции производителя. Соотношения A260/A280 и A260/A230 составили 1,87 и 2,03, соответственно. Дополнительно концентрация была измерена флуорометрически на приборе Qubit 4 (Thermo Scientific) – 476 нг/мкл. Распределение длин фрагментов ДНК в

образце было определено путем электрофоретического разделения в 1% агарозном геле. Все фрагменты в образце имели длину значительно превышающую 3000 п.о. Таким образом, образец ДНК подходил для анализа по всем необходимым требованиям к качеству: соотношения A260/A280, A260/A230, концентрации и распределению длин фрагментов.

Приготовление NGS-библиотек, секвенирование и качественная оценка полученных данных

NGS-библиотеки для секвенирования были приготовлены с использованием реагентов NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (New England BioLabs), Flow Cell Priming Kit (Oxford Nanopore) и Ligation Sequencing Kit (Oxford Nanopore). Первоначально в ячейке было оценено число активных пор – 1193. Секвенирование было проведено на приборе MinION Mk1B (Oxford Nanopore) в течение 40 часов, после чего ячейка была промыта, библиотека была дозагружена библиотекой и проведено еще 24 часа секвенирования. Всего было получено 5,42 миллиона прочтений с суммарной длиной прочтений, прошедших фильтр, 9,52 миллиардов пар оснований (Гб). N50 всех прочтений составил около 3 тысяч пар оснований. Преобразование сырого сигнала в нуклеотиды проводилось на вычислительном кластере НГУ с использованием технологии CUDA, что позволило сократить время обработки с недели до нескольких часов. При этом применялся алгоритм высокоточного преобразования, который был выбран оптимальным в первый год работы над проектом.

Анализ полученных данных: точечные мутации

Во второй год работы над проектом было определено, что результаты картирования новой версии программы `minimap2` практически не отличаются от результатов картирования `bwa mem`, но при этом скорость работы была в 26 раз выше (**рисунок 1**).

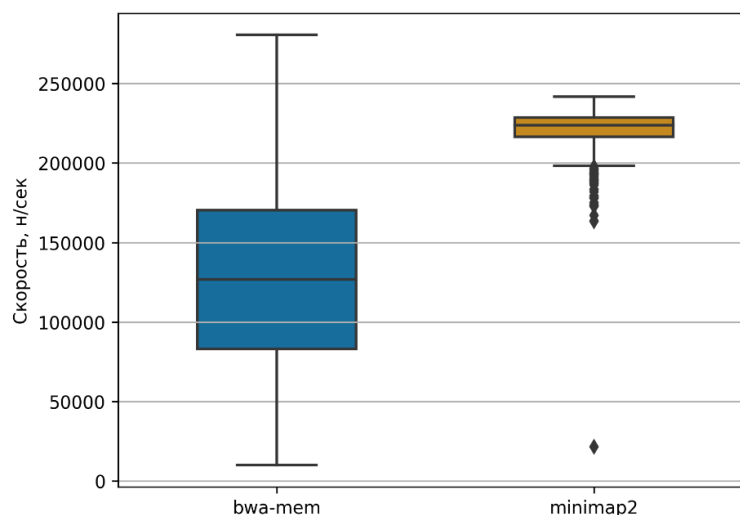


Рисунок 1. Скорость картирования протяженных прочтений Oxford Nanopore на референсный геном hg38 программами bwa mem и minimap2. Показан вариант сравнения, при котором прочтения разделялись по 4000 на файл. Без деления относительная разница составляла 26 раз.

Поскольку большинство ложно положительных мутаций были делециями и инсерциями, для поиска соматических мутаций был проведен анализ только замен нуклеотидов. Из всего набора выявленных вариантов были отобраны только те, которые имели процент прочтений с альтернативным аллелем (VAF) не менее 40%. Такой процент был выбран по той причине, что нам необходимы только драйверные события, а следовательно, представленные в большинстве клонов опухоли. Далее были отобраны несинонимичные замены и варианты, приводящие к образованию стоп-кодона, в генах тирозинкиназ. Следующим этапом необходима валидация выявленных вариантов и оценка их влияния на функции кодируемых ими белков.

Анализ данных: метилирование промоторов

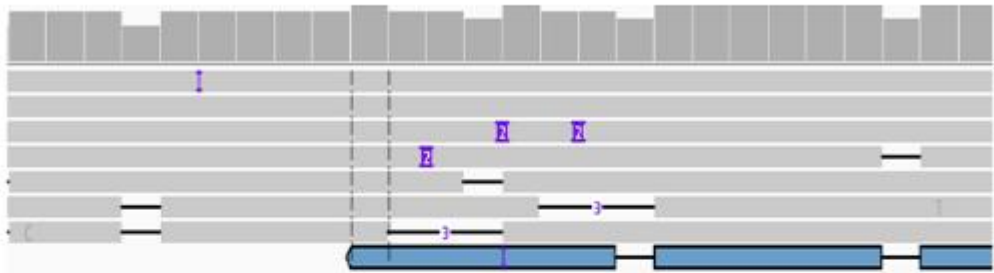
Для анализа статуса метилирования по CpG-островкам была использована программа megalodon, которая использует файлы с сырым сигналом (fast5) и определяет по каждому CpG-островку в геноме процент метилированных цитозинов. Полученный файл был обработан собственным скриптом Python: используя GFF-файл из базы данных NCBI, для каждого гена подсчитывался медианный процент метилирования CpG-островков, а также число и процент CpG-островков с метилированием не менее 50%. Статус метилирования промоторов генов был определен для образца ДНК из лейкоцитов крови и для образца ДНК из опухолевой ткани. После этого были отобраны гены, медиана процента метилирования CpG-островков которых изменилась не менее, чем на 75%. Были выбраны две группы генов: с увеличением метилирования в опухоли (842 гена) и уменьшением (1344 гена). После этого был проведен анализ обогащения генами метаболических и сигнальных путей на

платформе WebGestalt (www.webgestalt.org) (**таблица 2**). Стоит отметить, что у большого числа генов изменился статус метилирования промоторов как в большую так и в меньшую сторону, однако обе группы генов были связаны с передачей сигнала. При этом в образце опухоли значительно снизился уровень метилирования промоторов генов, связанных с передачей сигнала от G-белков, что, по-видимому, помогает усилить уровень транскрипции таких генов. В то же время, учитывая, что сравнение проводилось против образца ДНК из лейкоцитов крови, полученные результаты требуют валидации.

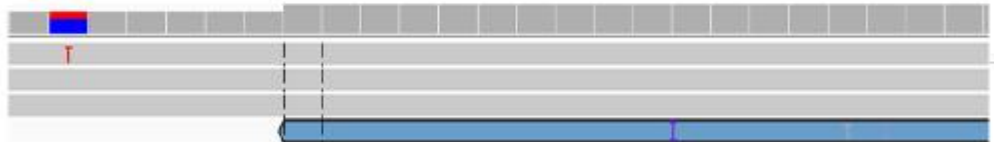
Анализ данных: крупные геномные перестройки

Поиск крупных геномных перестроек оказался осложнен тем, что существующие программы либо неправильно определяют позицию перестройки, либо неправильно определяют ее показатели качества (число прочтений с перестройкой или их процент от общего числа) (**рисунки 2-3**). За первый год выполнения проекта данная проблема не была обнаружена, поскольку существующие программы сравнивались только по числу и типам выявляемых перестроек.

```
chr1 81688614 Sniffles2.BND.375ES0 N N[chr3:174397054[
60 GT PRECISE;SVTYPE=BND;SUPPORT=1;COVERAGE=5,6,6,5,5;STRAND=-;AF=0.16
7;CHR2=chr3;STDEV_POS=0 GT:GQ:DR:DV 0/0:3:5:1
```



```
Hap name: null
Dist: 0
Read name = 01a4559a-a4ba-461f-a4ea-510c5e8dc751
Read length = 921bp
-----
Mapping = Supplementary @ MAPQ 60
Reference span = chr1:81,688,614-81,689,569 (-) = 956bp
Cigar = 603H4M1I3M1D6M1D37M1I6M1I55M3D...6M1I2M1I38M1D18M1D5M1D75M2499H
Clipping = Left 603 hard; Right 2,499 hard
```



```
Hap name: null
Dist: 0
Read name = 01a4559a-a4ba-461f-a4ea-510c5e8dc751
Read length = 567bp
-----
Mapping = Supplementary @ MAPQ 60
Reference span = chr3:174,397,055-174,397,636 (-) = 582bp
Cigar = 12H10M1I19M3D12M1I8M2D55M2I1M...2D3M2I44M1D23M2D10M1I10M3444H
Clipping = Left 12 hard; Right 3,444 hard
```

Рисунок 2. Пример выравнивания прочтений, указывающих на некорректное обозначение перестройки на хромосоме 3. Сверху приведена строка из файла с результатами sniffles, снизу – выравнивание.

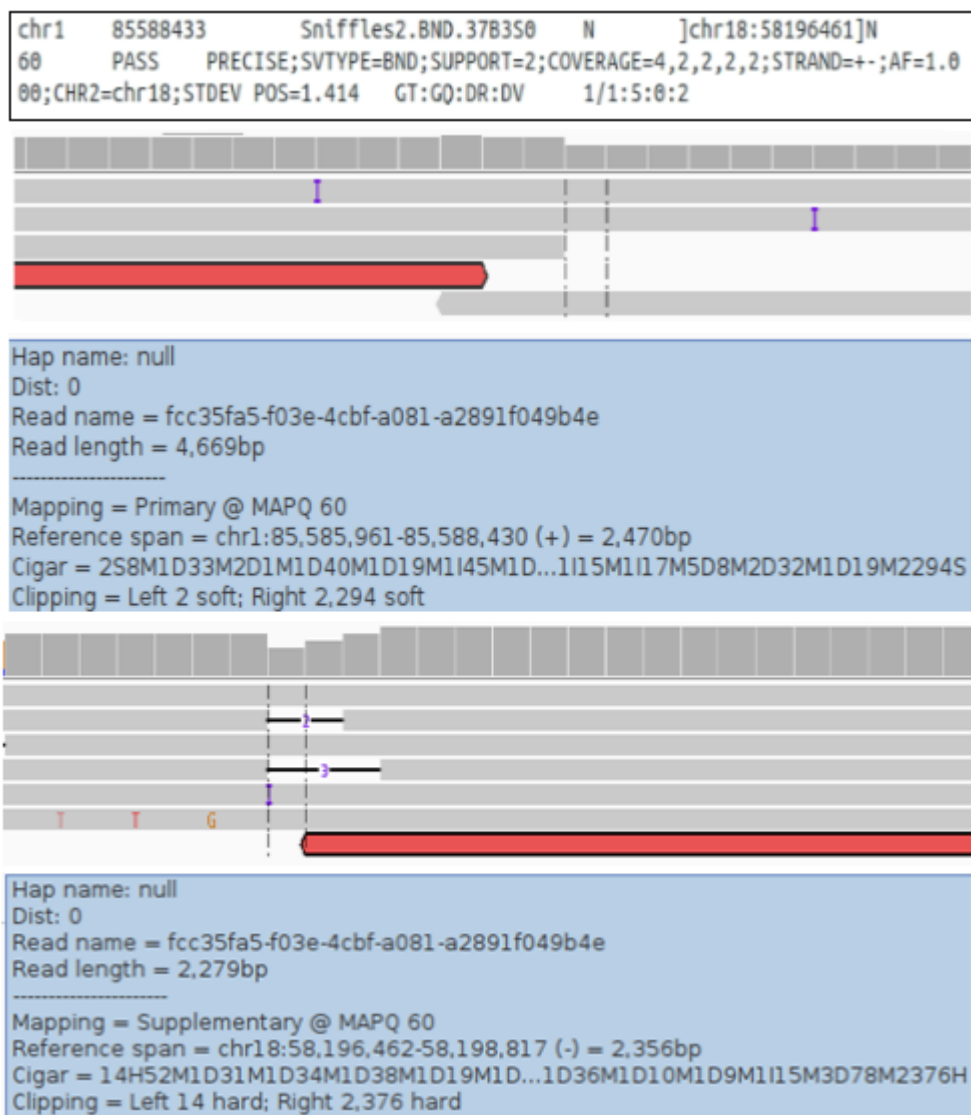


Рисунок 3. Пример выравнивания прочтений с некорректно указанными значениями частоты аллеля (AF), покрытия (COVERAGE) и количества прочтений, подтверждающих перестройку (SUPPORT). Сверху приведена строка из файла с результатами sniffles, снизу – выравнивание.

Для трёх перестроек были сконструированы праймеры для их валидации в образце (таблицы 3 и 4).

Таблица 1. Последовательности олигонуклеотидов, использовавшихся для валидации наличия перестройки.

Олигонуклеотид	Последовательность 5'-3'
chr3_18380420_norm_F	ctctctatgggcagtcggtgattCTTTGGAAAGTTGGTGTTTCGAT
chr3_18380420_norm_R	ctgcgtgtctccgactcagTGTTAAAGTTTCACCCTCACATTC
chr7_107050351_norm_F	ctctctatgggcagtcggtgattTATTACTTTGGGCTTTGGAAGAA GG
chr7_107050351_norm_R	ctgcgtgtctccgactcagCAGACACAGCACATCAATTTATG
chr3_7726666_norm_F	ctctctatgggcagtcggtgattTGTA AAAAACATTTAATTCAAC TACCCT

chr3_7726666_norm_R	ctgcgtgtctccgactcagAGTCATGATGTGTGCAACTTG
chr19_23845152_norm_F	ctctctatgggcagtcggtgattTCATTGCTAGTGATTTTTTGTAT TGCT
chr19_23845152_norm_R	ctgcgtgtctccgactcagGCTTACATGTTAAGGACCTGACAT
chr3_18380420_chr2_46039365_F	ctctctatgggcagtcggtgattCTTAAATGAAAGAGAAAAGAAA AGCCA
chr3_18380420_chr2_46039365_R	ctgcgtgtctccgactcagGGCATTCCAGGGAGAGGA
chr7_107050351_chr5_15675402_F	ctctctatgggcagtcggtgattAGGTAGACTCAGAAATGACTCC T
chr7_107050351_chr5_15675402_R	ctgcgtgtctccgactcagGCTTGCTTGAGTAACTTACACTT
chr3_7726666_chr11_76718830_F	ctctctatgggcagtcggtgattTGTA AAAAACATTTAATTC AAC TACCCT
chr3_7726666_chr11_76718830_R	ctgcgtgtctccgactcagCGTACACGTACACATATACATATA CA
chr19_23845152_chr1_28188968_F	ctctctatgggcagtcggtgattCTTTTTTTTATTACTGACTCAAT CATGC
chr19_23845152_chr1_28188968_R	ctgcgtgtctccgactcagCCAGATGTTGTGCCAATTTAAAAG

Таблица 2. Ожидаемые длины ампликонов при подтверждении наличия перестройки.

Ампликон	Длина
chr3_18380420_norm	300
chr7_107050351_norm	300
chr3_7726666_norm	271
chr19_23845152_norm	300
chr3_18380420_chr2_46039365	300
chr7_107050351_chr5_156754002	300
chr3_7726666_chr11_76718830	294
chr19_23845152_chr1_28188968	288

Подтвердилась только перестройка с переносом фрагмента хромосомы 19 на хромосому 1 (**рисунок 4**). Однако для использования технологии секвенирования с протяженными прочтениями Oxford Nanopore с целью детекции крупных геномных перестроек требуется разработка нового программного обеспечения с валидацией альтернативными методами, поскольку текущие не позволяют отобрать наиболее вероятные.

Нами была начата работа по разработке собственной программы для выявления крупных геномных перестроек, однако это не входило в планы проекта и на текущий момент ее результат позволяет только определить геномные границы перестройки, которые в дальнейшем должны быть аннотированы по влиянию на гены.

Эффект от использования кластера в достижении целей работы

Благодаря использованию кластера стало возможным конвертировать сырой сигнал секвенирования в нуклеотиды с использованием технологии CUDA, что сократило время на данную процедуру с недели до нескольких часов.

Перечень публикаций, содержащих результаты работы (если есть). Если имеется, указать DOI, импакт-фактор журнала (Thomson Reuters, РИНЦ,...)

1. Кечин А.А., Корюков М.А., Боробова В.С., Филипенко М.Л. Особенности выявления мутаций в данных полногеномного секвенирования с протяженными прочтениями // Сборник трудов XVIII Российской конференции с международным участием (Новосибирск, 5-8 декабря 2022 г.) / Под ред. С.А. Рылова, Ю.И. Молородова, А.А. Жирнова, Ю.Н. Синявского. – Новосибирск: ФИЦ ИВТ, 2022. – С. 109–114. <https://doi.org/10.25743/DIR.2022.16.63.019>.
2. Кечин А.А., Боробова В.С., Корюков М.А., Субботина К.В., Боярских У.А., Тархов А.В., Филипенко М.Л. Технология секвенирования Oxford Nanopore: перспективы и возможности в молекулярной онкологии // VII Всероссийская конференция по молекулярной онкологии (г. Москва, 21–23 декабря 2022).

Аннотация

Массовое параллельное секвенирование (NGS) позволяет проводить поиск патогенных мутаций во всем геноме пациентов, повышая шанс определить именно те мутации, которые и связаны с развитием или прогрессированием заболевания. С появлением секвенирования третьего поколения, с протяженными прочтениями, стало возможным выявление не только точечных и коротких мутаций, но и протяженных перестроек генома. Целью работы стала отработка подходов к выявлению точечных и протяженных мутаций в данных полногеномного секвенирования с протяженными прочтениями технологией Oxford Nanopore. Для этого был секвенирован один образец ДНК из лейкоцитов крови человека. Всего при секвенировании в течение 48 часов было получено 4,12 миллиона прочтений с суммарной длиной прочтений 15,8 миллиардов п.о. Технология секвенирования Oxford Nanopore подразумевает регистрацию изменения силы тока при прохождении цепи ДНК через пору, из-за чего при секвенировании накапливается огромное количество данных, которые должны быть конвертированы в нуклеотидные последовательности. Для этого применяются алгоритмы, основанные на машинном обучении. В то же время, опыт применения данной технологии показал, что для быстрой конвертации необходимым является применение программно-аппаратной архитектуры параллельных вычислений CUDA. Для последующего выявления точечных и протяженных мутаций нами были использованы несколько свободно доступных программ: BWA и minimap2 – для картирования полученных прочтений на геном; PISCES и LongShot – для поиска точечных мутаций; Sniffles и NanoVar – для выявления крупных перестроек. С помощью ROC-анализа нами были определены оптимальные параметры выявления точечных вариаций программой PISCES: минимально допустимое качество прочитанного основания – 6, минимальное качество варианта – 2; минимальное покрытие – 4. Для части выявленных крупных перестроек (более 100 тысяч различных перестроек до фильтрации по их качеству) было проведено сравнение числа и типа мутаций были сконструированы праймеры для их валидации в образце ДНК.