

На правах рукописи

КОРОСТИН ДМИТРИЙ ОЛЕГОВИЧ

**ПРЕДСТАВЛЕННОСТЬ ГЕНЕТИЧЕСКИХ МАРКЕРОВ
АНЕУПЛОИДИЙ ПЛОДА ВО ВНЕКЛЕТОЧНОЙ ФРАКЦИИ
КРОВИ БЕРЕМЕННЫХ ЖЕНЩИН**

1.5.7. Генетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2021

Работа выполнена в Центре высокоточного редактирования и генетических технологий для биомедицины Научно-исследовательского института трансляционной медицины Федерального государственного автономного образовательного учреждения высшего образования «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации.

Научный руководитель:

Доктор биологических наук,
профессор РАН

Ребриков Денис Владимирович

Официальные оппоненты:

Доктор биологических наук
Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт акушерства, гинекологии и репродуктологии имени Д.О. Отта» (г. Санкт-Петербург), отдел геномной медицины, руководитель

Глотов Андрей Сергеевич

Доктор биологических наук
ООО «Хайтек Генетикс» (г. Москва), генеральный директор

Глинкина Жанна Ивановна

Ведущая организация:

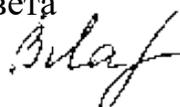
Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук Министерства науки и высшего образования Российской Федерации

Защита диссертации состоится 25 января 2022 г. в _____ часов на заседании Диссертационного совета 21.2.058.09 на базе ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России по адресу: 117997, г. Москва, ул. Островитянова, д.1

С диссертацией можно ознакомиться в Научной библиотеке ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России по адресу: 117997, г. Москва, ул. Островитянова, д.1 и на сайте www.rsmu.ru

Автореферат разослан « _____ » _____ 2021 г.

Ученый секретарь диссертационного совета
доктор медицинских наук, профессор

 **Ларина Вера Николаевна**

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования

Неинвазивный пренатальный скрининг (НИПС) анеуплоидий плода на основе анализа внеклеточной ДНК беременной женщины – группа подходов, в основном использующих технологии высокопроизводительного секвенирования ДНК, которая появилась в 2012 году. НИПС существенно превосходит классические методы скрининга анеуплоидий по чувствительности и специфичности, используемые в клинической практике. Поэтому количество проводимых неинвазивных тестов с каждым годом растет по всему миру, включая Россию. Применение НИПС снижает долю ложноотрицательных результатов, из-за которых происходит рождение детей с серьезными наследственными нарушениями, которые приводят к тяжелым социальным и экономическим проблемам как в рамках отдельной семьи, так и целого государства. Ложноположительные результаты комбинированного скрининга также имеют негативный эффект: женщин направляют на проведение инвазивной пренатальной диагностики (ИД). Рынок НИПС в значительной степени коммерциализован, и услуги, в частности, в России проводятся либо отправкой образцов за рубеж, либо путем трансфера технологий иностранных компаний в отечественных лабораториях.

С другой стороны, развитие технологий высокопроизводительного секвенирования имеет тенденцию к увеличению длины единичных прочтений (ридов), особенно среди приборов мономолекулярного секвенирования. Однако, использование этого преимущества в случае НИПС ограничено, так как внеклеточная ДНК (внДНК) представляет собой фрагментированные молекулы длиной 140–160 п.о. (пар оснований).

Для однозначного картирования рида на референсном геноме достаточно 30–40 п.о., поэтому в первом приближении секвенирование внДНК от начала до конца не имеет практического смысла. Значит, если случайно разрезать внДНК до соответствующего диапазона длин, а затем соединить фрагменты случайным образом в длинные химерные молекулы, можно будет использовать последние достижения платформ высокопроизводительного секвенирования наиболее полно в случае НИПС.

Таким образом, разработка нового варианта НИПС с использованием соответствующей пробоподготовки и биоинформатического алгоритма позволит эффективно осуществлять скрининг с учетом актуальных технологических достижений в области секвенирования ДНК.

Цель исследования

Разработать методику высокоэффективного скрининга анеуплоидий плода беременных по анализу циркулирующей в кровотоке матери внеклеточной ДНК с помощью высокопроизводительного секвенирования длинных химерных ридов.

Задачи исследования

1. Собрать коллекцию образцов венозной крови беременных, относящихся к группе высокого риска анеуплоидии по результатам скрининга I триместра и направленных на инвазивную диагностику (контрольная выборка).
2. Разработать методику конструирования библиотек ДНК, состоящих из длинных химерных ридов фрагментов внеклеточной ДНК (smash).
3. Разработать панель однонуклеотидных полиморфизмов (ОНП) для определения доли внеклеточной фетальной ДНК во фракции внеклеточной ДНК беременной.
4. Разработать методику конструирования библиотек ДНК для определения доли фетальной ДНК во фракции внеклеточной ДНК беременной (amplifet).
5. Разработать методику биоинформатической и статистической обработки результатов высокопроизводительного секвенирования для определения кариотипа плода по 13, 18 и 21 хромосомам, пола плода и доли фетальной ДНК во фракции внеклеточной ДНК беременной.
6. Провести высокопроизводительное секвенирование smash и amplifet библиотек ДНК, сконструированных из образцов контрольной выборки.
7. Провести определение значений чувствительности, специфичности и прогностических значимостей методики.

Научная новизна

Впервые разработана методика скрининга анеуплоидий плода беременных по анализу циркулирующей в кровотоке матери внеклеточной ДНК с помощью высокопроизводительного секвенирования длинных химерных ридов, включающая в себя лабораторный протокол пробоподготовки образцов биоматериала, в рамках которого также проводится подготовка библиотеки для оценки доли внеклеточной фетальной ДНК, а также биоинформатический и статистический протоколы для анализа данных.

Достоверность и обоснованность научных результатов

Степень достоверности и обоснованности научных результатов определяется достаточным количеством контрольных образцов выборки (145 образцов плазмы крови беременных), из них 82 с нормальным кариотипом, 4 с анеуплоидией по 13 хромосоме, 14 с анеуплоидией по 18 хромосоме, 45 с анеуплоидией по 21

хромосоме, современными методами исследования и верификации полученных результатов, корректными методами статистической обработки данных.

Практическая значимость работы

Разработана и внедрена в лабораторную практику методика скрининга анеуплоидий плода беременных по анализу циркулирующей в кровотоке матери внеклеточной ДНК с помощью высокопроизводительного секвенирования длинных химерных ридов, включающая в себя лабораторный протокол пробоподготовки образцов биоматериала, в рамках которого также проводится подготовка библиотеки для оценки доли внеклеточной фетальной ДНК, а также биоинформатический и статистический протоколы для анализа данных. Благодаря этому подходу значительно повышается экономическая целесообразность использования увеличивающейся в последнее время длины ридов высокопроизводительных секвенаторов последних поколений. Методика запатентована и, при получении регистрационного удостоверения на медицинское изделие Росздравнадзора, может быть использована в рутинной практике клинко-диагностических лабораторий.

Методика может стать основой для разработки новых модификаций подходов молекулярного кариотипирования высокого разрешения, осуществляемого с помощью высокопроизводительных секвенаторов, замещающих ставший уже классическим хромосомный микроматричный анализ: преимплантационное генетическое тестирование хромосомных аномалий, преимплантационное генетическое тестирование моногенных заболеваний, пренатальный скрининг хромосомных аномалий.

Получен патент «Способ определения кариотипа плода беременной женщины на основании секвенирования гибридных прочтений, состоящих из коротких фрагментов внеклеточной ДНК» (RU2717023C1).

Разработанные подходы могут быть использованы в учебном процессе при подготовке специалистов медицинского и биологического профиля.

Экономическая целесообразность использования разработанной методики превышает показатели уже существующих на рынке технологий НИПС. Благодаря этому внедрение методики НИПС в качестве более эффективной альтернативы комбинированному скринингу значительно облегчается.

Положения, выносимые на защиту

Методика конструирования длинных химерных ридов из фрагментов внеклеточной ДНК позволяет выявлять анеуплоидии плода по 13, 18 и 21 хромосомам у беременных с высоким риском анеуплоидии по результатам комбинированного скрининга на сроках гестации 11–25 недель.

Методика подготовки библиотек ДНК с использованием мультиплексной ПЦР на ОНП, отобранные по ряду критериев, позволяет проводить оценку доли внеклеточной фетальной ДНК в образце внеклеточной ДНК, выделенной из плазмы крови беременной.

Разработанные биоинформатический и статистический алгоритмы позволяют проводить совместный анализ результатов высокопроизводительного секвенирования библиотек ДНК для получения надежной информации о наличии анеуплоидий у плода по хромосомам интереса.

Внедрение результатов исследования

Результаты исследования внедрены в практическую работу Института репродуктивной генетики ФГБУ «НМИЦ АГП им. В.И. Кулакова» Минздрава России, используются в рамках создания наборов реагентов для генетического анализа фетальной ДНК человека, серийно производящихся ООО «ДНК-Технология ТС» на территории РФ, включены в программу лекций «Молекулярная биология» для студентов II курса кафедры биохимии и молекулярной биологии лечебного факультета ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, а также в практическую работу Центра высокоточного редактирования и генетических технологий для биомедицины, используются в учебном процессе в рамках выполнения курсовых и дипломных работ студентами медико-биологического факультета РНИМУ им. Н.И. Пирогова. В настоящее время ведется процесс получения регистрационного удостоверения Росздравнадзора для применения разработанной методики в клинической практике.

Получен патент № 2717023 «Способ определения кариотипа плода беременной женщины на основании секвенирования гибридных прочтений, состоящих из коротких фрагментов внеклеточной ДНК» (RU 2 717 023 C1).

Апробация работы

Результаты работы были представлены и обсуждены на 2 российских конференциях: VII Всероссийской научно-практической конференции по геномному секвенированию и редактированию (Москва, 20-21 мая 2020); Всероссийской онлайн конференции с международным участием «Геномная медицина в пренатальной диагностике, генетическом паспорте и в генной терапии» (Санкт-Петербург, 12-13 ноября 2020).

Диссертация апробирована на заседании апробационной комиссии НИИ Трансляционной медицины ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России 6 октября 2020 года, протокол № 10 и рекомендована к защите на соискание ученой степени кандидата биологических наук по специальности «03.02.07 – генетика».

Публикации

Материалы диссертации изложены в 8 работах, из них 5 статей в рецензируемых научных журналах, рекомендованных ВАК для опубликования научных результатов диссертаций на соискание ученой степени кандидата и доктора наук, а также 2 тезиса докладов в сборниках трудов конференций, получен 1 патент.

Личный вклад автора

Автором проведена систематизация литературных данных по теме диссертации, самостоятельно разработаны дизайн и программа исследования.

Автору принадлежит решающая роль в сборе контрольной выборки образцов, разработке стратегии экспериментов, отработке методик и обобщении полученных результатов. Анализ, статистическая обработка полученных данных проведены автором самостоятельно в соответствии с правилами и обеспечивают достоверность результатов и сформулированных выводов. Описание и публикация результатов исследований выполнены автором лично.

Объем и структура диссертации

Работа изложена на **126** страницах печатного текста и состоит из: введения, обзора литературы, описания материалов и методов экспериментальных исследований, результатов, обсуждения, выводов, списка использованной литературы, а также **4** приложений с развернутыми таблицами по полученным результатам. Работа включает **17** рисунков и **11** таблиц. Указатель литературы содержит **164** библиографических источника, в том числе 6 отечественных и 158 иностранных публикаций.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Для разработки методики скрининга анеуплоидий плода по 13, 18 и 21 хромосомам на ранних сроках беременности необходимо было сформировать контрольную выборку образцов, составленную из образцов с анеуплоидиями и с нормальным кариотипом. Так как анеуплоидии по 13, 18 и 21 хромосомам встречаются редко, выборка была ограничена пациентками, прошедшими скрининг I триместра и получившими значения риска анеуплоидии выше пороговых. Они были направлены в ГБУЗ «Центр планирования семьи и репродукции Департамента здравоохранения города Москвы» (ЦПСИР) на проведение инвазивного забора плодного материала (биопсия ворсин хориона или амниоцентез) для диагностики методом кариотипирования. Образцы периферической венозной крови беременных собирались в пробирки Cell-Free DNA BCT (Streck) – по 2 шт на образец, перед проведением пациенткам

инвазивного забора плодного материала. Все исследования выполнялись добровольно, по желанию женщины, после подписания соответствующих форм информированного согласия. Всего было собрано 145 образцов: 82 с нормальным кариотипом; 4 с трисомией по 13 хромосоме; 14 с трисомией по 18 хромосоме; 43 с трисомией по 21 хромосоме.

Общая схема процессирования образцов вДНК представлена на Рисунке 2.

Отделение плазмы крови проводилось по стандартной методике путем двух последовательных центрифугирований (250 g, затем 9000 g) образцов с переносом супернатанта в новые пробирки.

Выделение вДНК проводилось с помощью набора QIAamp Blood kit (Qiagen) с модификациями стандартного протокола выделения: увеличенный объем лизата (до 2–4 мл) и уменьшенный объем экстракта (до 40 мкл).

Выделенная вДНК каждого образца шла на приготовление двух типов библиотек ДНК: smash и amplifet (Рисунок 1).

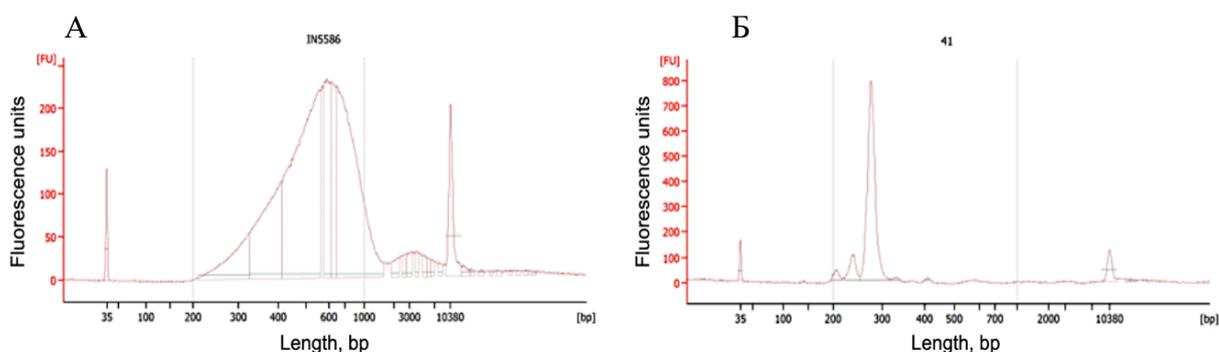


Рисунок 1. Хроматограммы контроля качества библиотек ДНК. А – smash. Б – amplifet. По оси X отложен размер в п.о., по оси Y – интенсивность флуоресценции.

В результате подготовки библиотек первого типа образуется библиотека химерных ДНК, состоящих из случайно сшитых коротких (40–50 п.о.) фрагментов вДНК. Вторая библиотека представляет собой модифицированную для высокопроизводительного секвенирования смесь ампликонов фрагментов генома, содержащих в себе однонуклеотидные полиморфизмы (ОНП). Благодаря анализу частот встречаемости аллелей каждого ОНП проводится определение доли внеклеточной фетальной ДНК (внфДНК) в образце.

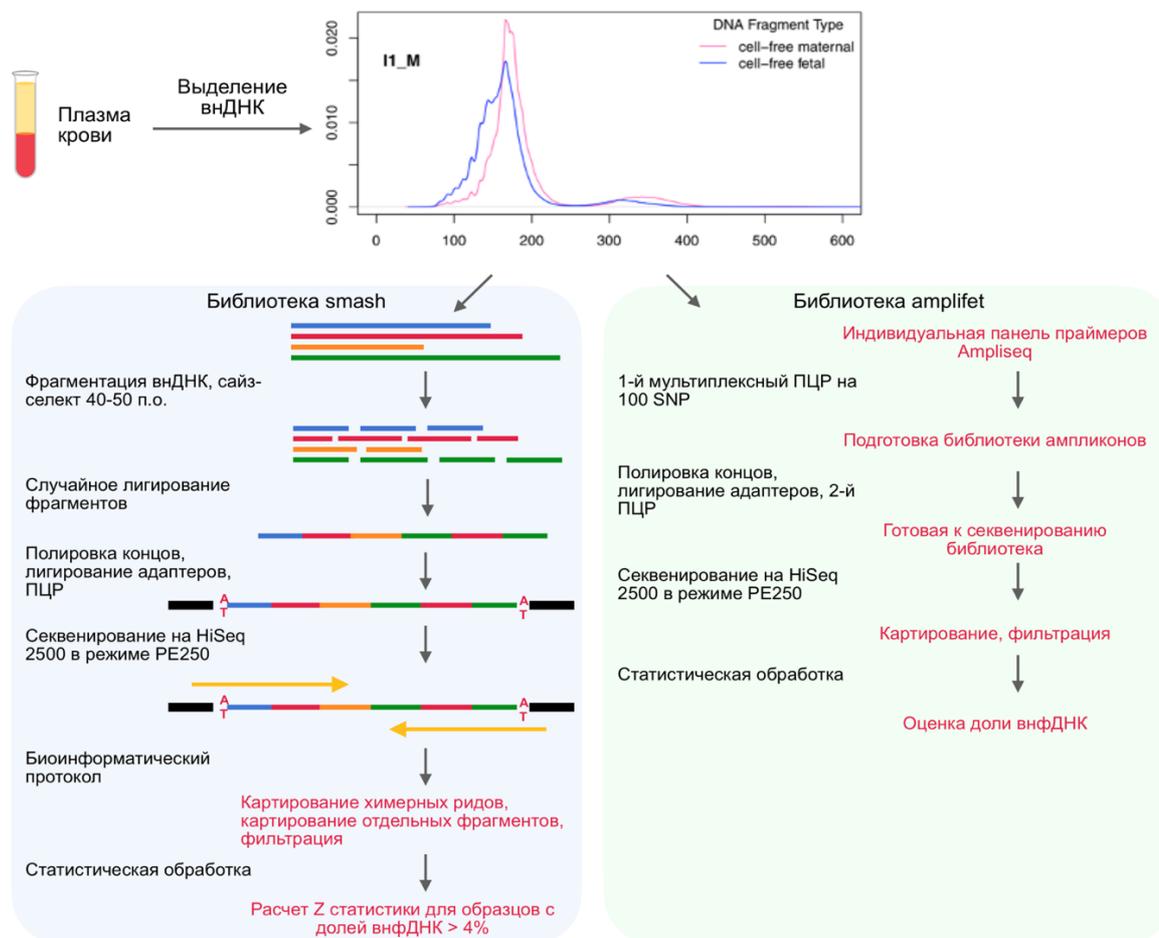


Рисунок 2. **Общая схема процессирования образца вДНК.** Выделенная из плазмы крови вДНК, прошедшая контроль качества, разделяется на две части: первая проходит пробоподготовку, секвенирование, биоинформатический и статистический анализы по протоколу smash; вторая – по протоколу amplifet.

Секвенирование проводилось на приборе Illumina HiSeq 2500 с реагентами Rapid Run v2 на 500 циклов в режиме PE250 dual-indexing.

Биоинформатическая обработка включала в себя выравнивание ридов на референсный геном в две итерации: сначала с помощью BWA версии 0.7.17 со стандартными настройками фрагменты smash ридов распределялись по геному на основании координат их картирования, а amplifet риды картировали целиком. На втором этапе сабриды, полученные в результате первого картирования smash ридов, отбирали и картировали еще раз, но уже как отдельные риды. Результаты дополнительных выравниваний, а также картирований на минус цепь, элиминировали из результатов, так как после первого картирования все экспортированные сабриды были инвертированы на плюс цепь референсного генома. Затем данные проходили систему фильтрации:

1. Если для образца имелось более одного bam файла, их объединение проводилось с помощью samtools v1.9.
2. Риды, имеющие низкое качество картирования (MAPQ < 60), отбрасывались.

3. Риды, картировавшиеся на неуникальные участки генома (определялись с помощью трека RepeatMasker в Genome Browser).

4. Для библиотек amplifet отбрасывались все риды, картировавшиеся вне регионов расположения ампликонов.

Статистическая обработка библиотек smash включала в себя расчет Z-статистики по стандартной методике. Пакет NIPTer для языка R использовался для нормировки образцов (коррекции по GC-составу, χ^2 -коррекция), сравнению контролей качества и расчета Z-статистики, проведения поправки на выбросы (были использованы как метод, основанный на отбрасывании пиков с помощью квантилей, так и процедура уменьшения вариации, основанная на критерии χ^2). При каждом запуске программы подсчёта статистики контрольная выборка формировалась случайным образом из подвыборки не имеющих трисомии образцов. Алгоритм обучался на такой выборке и определял трисомию у оставшихся образцов. Классификация представленности отдельных хромосом производилась на основании Z-статистики, вычисленной в регрессионной модели, и порогового значения Z_0 . Пороговое значение выбирается для каждой хромосомы отдельно в соответствии с критерием максимизации специфичности. Для определения доли фетальной ДНК в данных секвенирования библиотек ДНК amplifet использовался FetalQuant.

Фильтрация образцов при статистической обработке проводилась по следующим критериям:

1. доля внеклеточной фетальной ДНК (внфДНК) по расчетам составила менее 4%.
2. общее количество прошедших фильтрации фрагментов составило менее 2,5 млн на образец.

В качестве порогового значения Z-статистики был взят 3. Чтобы сделать расчет значений чувствительности, специфичности, ROC-кривой более стабильными, проводился процедура кросс-валидации результатов: тестовый датасет состоял из всех образцов с соответствующей трисомией и таким же количеством образцов с нормальным кариотипом, которые выбирались случайным образом. Оставшиеся образцы формировали обучающую выборку. Всего таких расчетов делалось 200 шт на каждую трисомию. Для каждого расчета проводилось определение чувствительности, специфичности, ROC-кривой, а затем их усреднение.

Для определения пола плода на основании фрагментов, картировавшихся на Y хромосому, использовался разработанный ранее алгоритм.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Подготовка и секвенирование библиотек *smash*

Идея использования химерных ридов основана на тенденции к появлению наборов реагентов и платформ с длинным и сверхдлинным (более 300 и более 3000 п.о., соответственно) чтением. Средняя длина вДНК составляет 140–160 п.о., а значит развитие стандартных подходов к анализу вДНК отстает от развития технологий высокопроизводительного секвенирования.

Разработанный протокол подготовки библиотек отличается от используемого в классических НИПС следующими элементами:

1. Фрагментация вДНК до коротких фрагментов;
2. Сайз-селект (отбор длин) фрагментированной вДНК с двух сторон, в результате чего в образце остаются фрагменты, длина которых распределяется в диапазоне значений 40–50 п.о.;
3. Случайное лигирование фрагментов вДНК и формирование длинных (более 300 п.о.) химерных молекул ДНК, к которым далее лигируются адаптеры для секвенирования.

В результате подобранных нами условий пробоподготовки и наличия наборов для секвенирования длинных ридов (Illumina Rapid Run PE250), удалось получить большее количество полезной для диагностики информации с каждого рида, чем в классическом НИПС. Размер фрагмента вДНК в тестовой выборке составил в среднем 43,8 п.о., таким образом, информативность одного прямого или обратного химерного рида длиной 250 п.о. в 5,7 раза больше, чем для классического НИПС рида. Так как определить размер вставки пары ридов путем ее картирования на референсный геном в нашем случае оказалось невозможным, мы попробовали оценить исходные размеры отсекуемых длинных химерных молекул, слив их инструментом FLASH: в результате получилось, что гораздо большая часть ридов имеет длину около 250 п.о. Значит объединить пары ридов не удалось, и длина химерной молекулы для них превышает 460 п.о., что коррелирует с данными на гистограммах распределения длин библиотек, полученных на приборе Bioanalyzer 2100 в процессе пробоподготовки (Рисунок 1 А). На Рисунке 3 приведено распределение длин нефильтрованных фрагментов в парах ридов.

В процессе разработки было введено определение фильтрованных фрагментов – это сабриды из химерных длинных ридов, прошедшие через этапы фильтрации из соответствующего раздела Материалов и методов экспериментальных исследований.

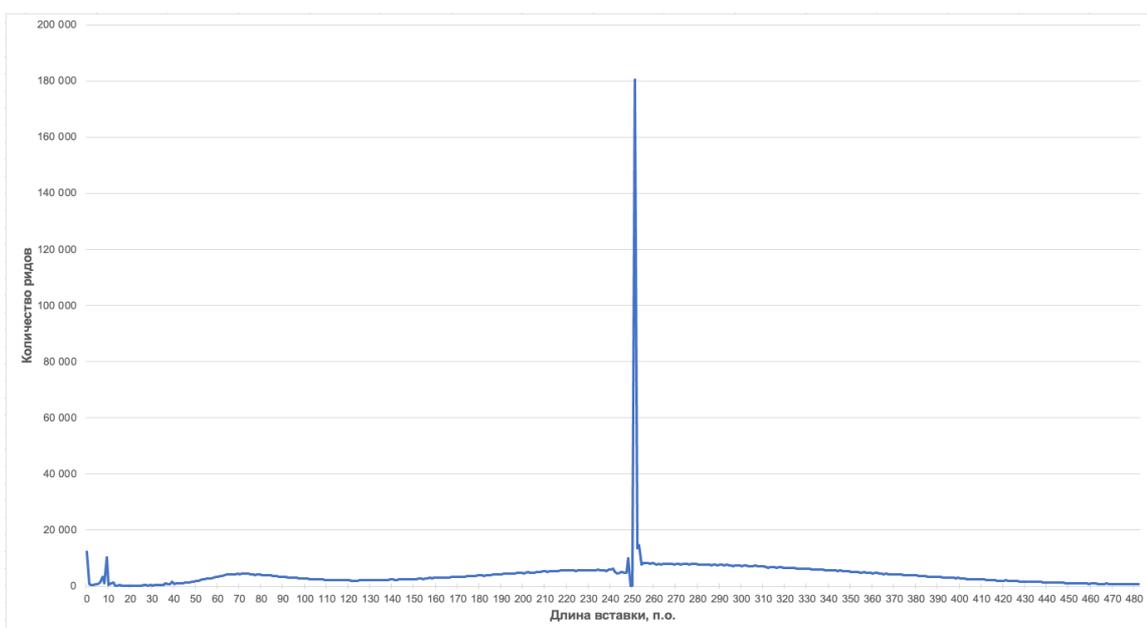


Рисунок 3. Распределение длин вставки для парноконцевых ридов smash-библиотеки. По оси X длина вставки в п.о., по оси Y отложено количество ридов соответствующей длины.

Так как суммарная длина регионов в треке RepeatMasker 1 586 326 530 нуклеотидов, что занимает примерно половину генома человека, можно допустить, что общее количество фрагментов, приходящееся на химерный рид, до фильтраций составляет удвоенное значение фильтрованных. Химерные риды прочитаны парноконцевым способом по 250 п.о. в обе стороны. Таким образом, парноконцевое чтение одной химерной молекулы PE250 дает информацию примерно о 9 фрагментах генома с учетом всех фильтраций (Рисунок 4).

Подбор ОНП, секвенирование и анализ библиотек amplifet

На основании анализа баз данных 1000 Genomes и gnomAD были отобраны ОНП. Фильтрация ОНП была выполнена с помощью пакета программ VCFtools. Маркеры в геноме человека отбирали по следующим критериям:

1. Не учитывали инсерции, делеции, хромосомы X, Y, митохондриальную, р-плечо 6 хромосомы;
2. Учитывали только диаллельные полиморфизмы;
3. Учитывали ОНП только с идентификаторами «rs»;
4. Минимальная частота аллеля (MAF) не менее 0,4;
5. Минимальная вероятность действия на ОНП отбора по Харди-Вайнбергу: $p < 0,00001$ (т.е. отбор почти не действует);

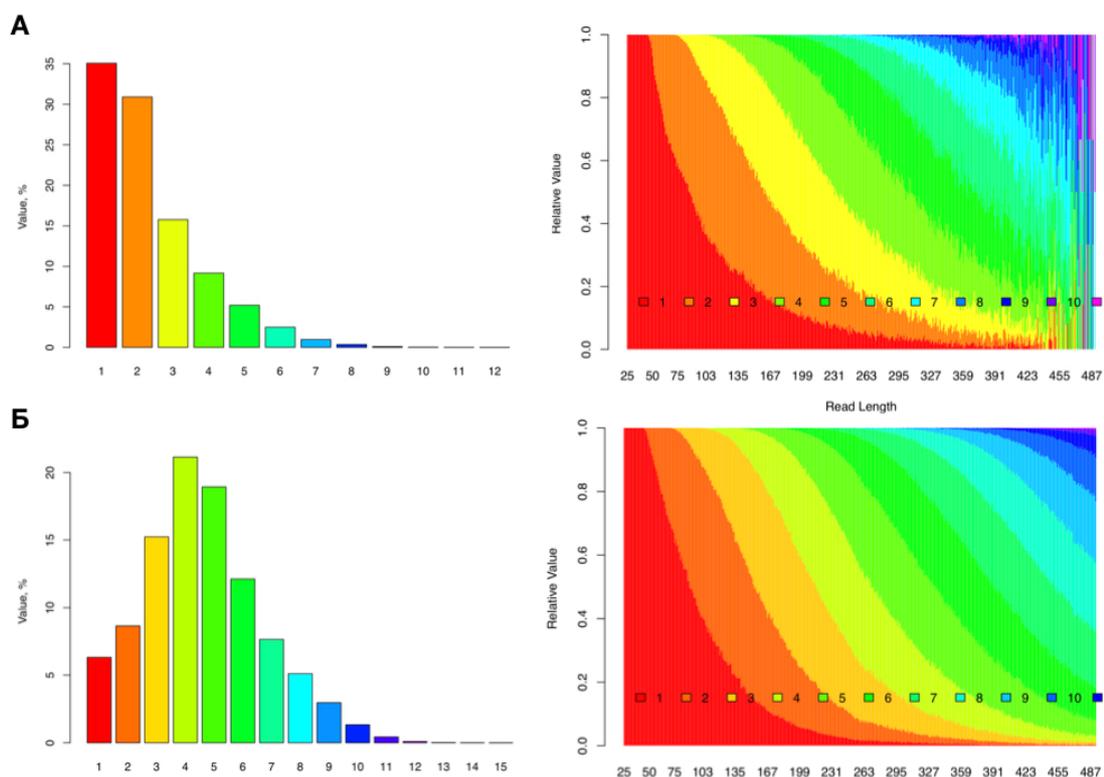


Рисунок 4. Гистограммы, демонстрирующие метрики входящих в химерные риды фрагментов для библиотек *smash* в процессе разработки (А) и финальной версии протокола пробоподготовки (Б). На левой гистограмме приведены доли ридов, содержащих 1, 2 и т. д. фрагментов в своем составе (доля отложена по Y, количество фрагментов – по X). На правой гистограмме изображена зависимость доли ридов, содержащих соответствующее количество фрагментов, от длины рида. Риды, различающиеся количеством фрагментов, обозначены на гистограмме своим цветом, соответствие цвета и количества фрагментов приведено на правых гистограммах в квадратах.

6. Учитывали только несцепленные ОНП. Их определяли с использованием метода скользящего окна при следующих параметрах: коэффициент корреляции $r^2 < 0,5$; размер скользящего окна равен 50 ОНП; шаг равен 5 ОНП.

Далее оставили для дальнейшего анализа только те ОНП, которые были общими для образцов из европейских популяций.

К 610 отобранным ОНП подобрали праймеры с помощью программы AmpliSeq Designer (www.ampliseq.com) с размером ампликона не более 140 п.о. Проанализировали нуклеотидный профиль вокруг отобранных праймеров и удалили те, которые потенциально могут иметь проблемы при амплификации (шпильки, гомополимеры и т.д.). Критерии последовательной фильтрации:

1. отсутствие гомополимеров в праймере (4 и больше одинаковых нуклеотида подряд);
2. удаление повторов (например, «АТАТАТА...»);
3. удаление праймеров с большим количеством GC на 3' конце;
4. GC-состав праймера должен быть в диапазоне от 0,4 до 0,6.

В итоге была сформирована панель из праймеров на 102 ОНП, синтез которой заказали в ThermoFisher как custom Ampliseq panel.

В результате секвенирования библиотек amplifet 96 маркеров оказались рабочими. В среднем доля внфДНК в образцах составила 12,5% (Рисунок 5), что согласуется с результатами, полученными другими авторами для этого диапазона сроков беременности.

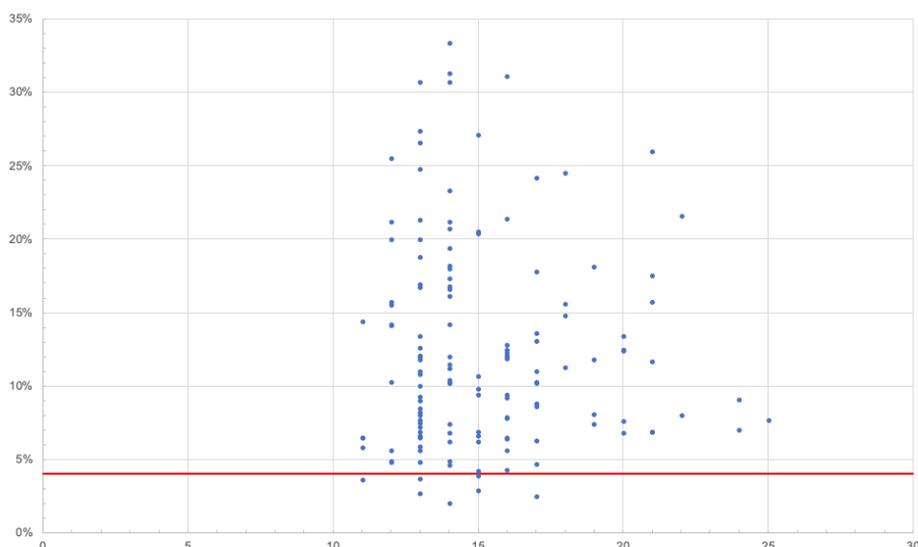


Рисунок 5. Значение доли фетальной ДНК (ось Y) в зависимости от срока гестации для 139 образцов контрольной выборки (ось X), рассчитанное на результатах секвенирования библиотек amplifet. Красная сплошная линия – доля фетальной ДНК, равная 4%.

На сегодня определение доли внфДНК в образце внДНК беременной является стандартом, так как при значениях доли ниже пороговых частота ложноотрицательных результатов НИПС существенно возрастает. Мы использовали подход с генотипированием образца внДНК по несцепленным, неассоциированным с известными фенотипами диаллельным ОНП, частота встречаемости минорного аллеля в которых не ниже 0,4. В этом случае на выборке из 96 ОНП гарантированно будут встречаться такие точки, в которых генотип матери будет гомозиготным, а плода – гетерозиготным.

Недостаток подхода с оценкой доли внфДНК по анализу панели ОНП – усложнение процесса пробоподготовки, так как в лаборатории из одного образца внДНК нужно готовить две библиотеки фрагментов ДНК (smash и amplifet). Широко используемый способ на основе анализа распределения длин фрагментов внДНК в первом приближении кажется нереализуемым в разработанной методике, потому что мы фрагментируем тотальную внДНК до длины около 40 п.о. Однако, это не так: в химерных молекулах ДНК, которые получаются в результате пробоподготовки, представленность концов внДНК скорее всего сохраняется. Значит задача имплементации подхода анализа распределения длин внДНК сводится к обучению алгоритма на основании выборки образцов с измеренной

другим способом долей внфДНК (например, с помощью секвенирования панели ОНП). В дальнейшем планируется провести этот анализ. Аналогично планируется изучить возможность оценки доли внфДНК на основании анализа трека намотки на нуклеосомы: если используемая в smash ферментативная фрагментация ДНК происходит случайно в любом из участков молекулы внДНК, обучение алгоритма сводится к анализу достаточно объемной выборки образцов. Известно, что как ферментативная, так и физическая – с помощью небулизации или соникации/кавитации – фрагментация ДНК происходит неслучайно и имеет паттерны последовательностей, по которым разрезание происходит чаще. Однако, на сегодня не удалось найти литературных данных, которые бы изучали вопрос расположения разреза относительно концов ДНК определенной длины, поэтому необходимо проводить собственное исследование.

Определение минимального количества фрагментов в образце

Для определения минимального количества фильтрованных фрагментов, необходимых и достаточных для определения анеуплоидий по 13, 18 и 21 хромосомам, были проведены расчеты чувствительности и специфичности для образцов, имеющих от 1 до 3 млн фильтрованных фрагментов с шагом 0,5 млн фрагментов. Таким образом, оптимальным для всех исследуемых анеуплоидий является не менее 2,5 млн фильтрованных фрагментов на образец (Таблица 1); поэтому из дальнейшего анализа были исключены образцы, имеющие менее 2,5 млн фрагментов.

Таблица 1. Расчеты чувствительности и специфичности определения анеуплоидий по 13, 18 и 21 хромосомам для образцов, имеющих разное количество фильтрованных фрагментов.

Анеуплоидная хромосома	Минимальное количество фильтрованных фрагментов	Чувствительность	Чувствительность σ	Специфичность	Специфичность σ
13	1 000 000	1	0	0,99536	0,011605
13	1 500 000	1	0	0,98883	0,030454
13	2 000 000	1	0	0,9952	0,020135
13	2 500 000	1	0	0,99169	0,032054
13	3 000 000	1	0	0,99556	0,012202
18	1 000 000	1	0	0,98154	0,023683
18	1 500 000	1	0	0,97256	0,04133
18	2 000 000	1	0	0,98368	0,020579
18	2 500 000	1	0	0,98342	0,031326
18	3 000 000	1	0	0,9925	0,01991
21	1 000 000	0,96537	0,025367	0,98735	0,014686
21	1 500 000	0,97143	0,023793	0,99017	0,026724
21	2 000 000	0,97517	0,016273	0,99388	0,015848
21	2 500 000	0,99935	0,0068607	0,9914	0,022761
21	3 000 000	1	0	0,9925	0,01991

Фильтрация образцов контрольной выборки

В соответствии с определенным выше критерием минимального количества фильтрованных фрагментов, а также отсечкой не менее 4% по минимальной доле фетальной ДНК в образце, образцы были отфильтрованы. Таким образом, в конечную выборку попали 83 образца (нормальный кариотип – 48 шт, 58%; с трисомией по 21 хромосоме – 25 шт, 30 %; с трисомией по 18 хромосоме – 6 шт, 7%; с трисомией по 13 хромосоме – 4 шт, 5%).

Расчет чувствительности и специфичности

Наложив все указанные выше ограничения на контрольную выборку образцов, проводится дальнейший расчет Z -статистики и других статистических показателей методики: чувствительности, специфичности, PPV, NPV, а также были построены ROC-кривые.

Для расчета чувствительности и специфичности теста использовался порог 3 для Z -статистики. Для увеличения количества образцов в выборке в целях получения робастных оценок использовалась процедура кросс-валидации. Результаты расчета Z -статистики для соответствующих анеуплоидий приведены на Рисунке 6.

В результате расчета Z -статистики для всех трех типов анеуплоидий группы нормальных и трисомных по соответствующей хромосоме образцов кластеризовались относительно значения $Z=3$, а значит методика надежно отличает здоровый и анеуплоидный кариотип плода.

Чувствительность – это способность диагностического метода давать правильный результат, который определяется как доля истинно положительных результатов среди всех проведенных тестов. Специфичность – это способность диагностического метода не давать при отсутствии заболевания ложноположительных результатов, который определяется как доля истинно отрицательных результатов среди здоровых лиц в группе исследуемых. Вместе чувствительность и специфичность являются важнейшими операционными характеристиками информативности диагностического метода.

Расчет чувствительности и специфичности для анеуплоидий приведен в Таблице 2.

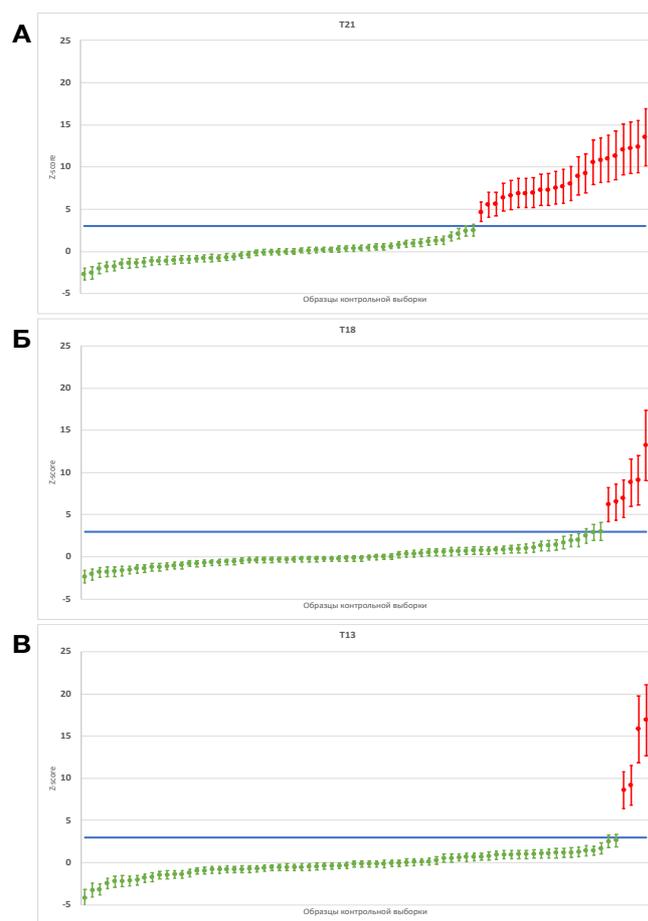


Рисунок 6. Гистограммы расчета Z -статистики в рамках кросс-валидации образцов контрольной выборки (А – трисомия по 21, Б – трисомия по 18, В – трисомия по 13 хромосоме). Зеленым обозначены образцы, не имеющие соответствующей анеуплоидии, красным образцы с анеуплоидией. Усами отложено стандартное отклонение значения Z для образца в симуляциях. Синяя горизонтальная линия на гистограммах – $Z=3$.

Таблица 2. Чувствительность и специфичность методики на контрольной выборке образцов. Доверительные интервалы рассчитаны на основании метода Вильсона.

	Трисомия по 21 хромосоме	Трисомия по 18 хромосоме	Трисомия по 13 хромосоме
Чувствительность	99,93%	100%	100%
двухсторонний 95% доверительный интервал (CI)	(85,58–99,99%)	(60,96–100%)	(51,01–100%)
Специфичность	99,14%	98,34%	99,17%
двухсторонний 95% доверительный интервал (CI)	(90,29–99,93%)	(91,16–99,7%)	(92,7–99,91%)

Полученные результаты чувствительности и специфичности согласуются со значениями, полученными другими исследователями. Для исследуемых анеуплоидий на контрольной выборке образцов были построены ROC-кривые (Рисунок 7). Они наглядно демонстрируют качество работы классификатора, «есть/нет анеуплоидия».

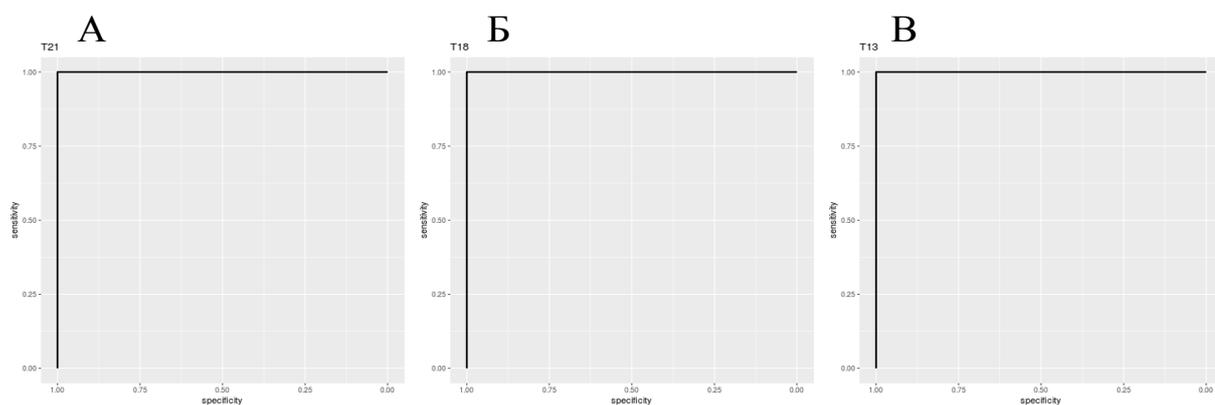


Рисунок 7. ROC-кривые для определения анеуплоидии по 21 (А), 18 (Б) и 13 (В) хромосомам. По оси Y отложена чувствительность, по оси X специфичность.

Расчет прогностических значимостей

Положительная прогностическая значимость (PPV) и отрицательная прогностическая значимость (NPV) относятся к операционным характеристикам теста и дополняют метрики чувствительности и специфичности методики. Результаты их расчета приведены в Таблице 3.

Прогностическая ценность по результатам теста в клинической практике сильно зависит от распространенности заболевания в исследуемой группе пациентов, которая может существенно отличаться от средней по популяции распространенности заболевания, что снижает ценность проведенного исследования. Чтобы нивелировать этот эффект, был проведен расчет PPV и NPV для соответствующих частот анеуплоидий, комбинируя образцы в контрольной выборке таким образом, чтобы получить указанные ниже частоты анеуплоидий.

Таблица 3. Расчет PPV и NPV для разработанной методики в зависимости от распространенности соответствующей анеуплоидии.

Анеуплоидия	Распространенность	PPV	NPV
Трисомия по 21 хромосоме	0,05%	5,491%	100,000%
	0,10%	10,415%	100,000%
	0,20%	18,880%	100,000%
	0,50%	36,853%	100,000%
	1,00%	53,983%	99,999%
	1,50%	63,881%	99,999%
Трисомия по 18 хромосоме	2,00%	70,328%	99,999%
	0,03%	1,777%	100,000%
	0,05%	2,928%	100,000%
	0,10%	5,693%	100,000%
	0,20%	10,782%	100,000%
	0,30%	15,358%	100,000%
	0,40%	19,496%	100,000%
Трисомия по 13 хромосоме	0,50%	23,255%	100,000%
	0,01%	1,190%	100,000%
	0,02%	2,352%	100,000%
	0,05%	5,680%	100,000%
	0,10%	10,755%	100,000%
	0,20%	19,437%	100,000%

Важно отметить, что корректная интерпретация результатов расчета прогностических характеристик должна учитывать распространенность анеуплоидий в популяции: например, 22,0 (95% CI 21,7–22,4), 5,0 (95% CI 4,8–5,1) и 2,0 (95% CI 1,9–2,2) для анеуплоидий по 21, 18 и 13 хромосомам на 10 000 новорожденных.

Поэтому в случае применения разработанной методики НИПС на всех беременных без ввода ограничений по результатам риска комбинированного скрининга доля ложноположительных результатов будет значительна. Чтобы убедиться в корректности приведенных в работе расчетов, необходимо проводить широкомасштабную валидацию методики на очень крупной выборке беременных, включающей в себя женщин со всем спектром значений риска анеуплоидии по результатам комбинированного теста.

Определение пола плода

Определение пола плода проводилось по расчетам нормированного покрытия Y-хромосомы на результатах секвенирования библиотек smash. Результаты представлены на Рисунке 8.

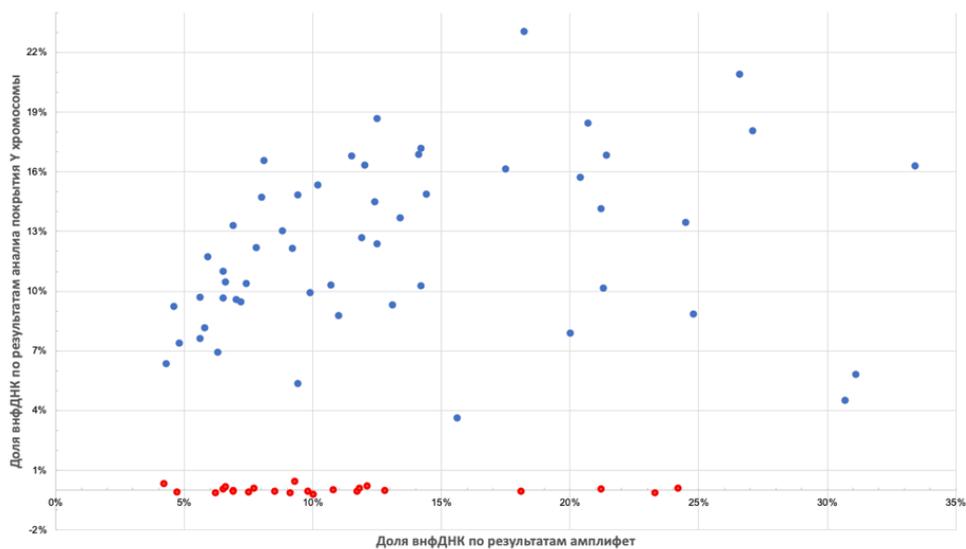


Рисунок 8. Доля внфДНК, рассчитанная по амплифет (ось X) и по Y хромосоме (ось Y). Образцы мужского пола обозначены синим цветом, женского – красным цветом.

Если доля внфДНК превышает 4%, а доля по Y хромосоме не более 1%, плод относится к женскому полу. Также заметно, что результаты доли внфДНК для образцов мужского пола по amplifet и Y хромосоме имеют тенденцию коррелировать. Мы предполагаем, что дальнейшие настройки системы расчетов по Y хромосоме могут хорошо согласовываться с другими способами определения доли фетальной ДНК. Также возможно, что аналогичный подход по сравнению покрытия X хромосомы и аутосом может быть использован для определения доли внфДНК у беременных девочками и заменить более трудоемкий подход с

использованием amplifet библиотек. Недостаток такого способа – ошибочные результаты, которые будут получаться в случае анеуплоидий по половым хромосомам.

Оценка себестоимости решения

Важнейшим ограничением на пути широкого внедрения НИПС в клиническую практику, помимо высоких требований к квалификации проводящего исследование персонала, является себестоимость. Проводить пусть и эффективное, но дорогое исследование в масштабах целой страны оказывается экономически неэффективно до порога, который составляет около 100 \$ (долларов США) по оценке авторов.

Были проведены расчеты себестоимости этапов процессирования образца биоматериала по разработанной нами методике от его забора до получения файлов fastq в зависимости от используемой платформы для секвенирования.

В случае использования наиболее современного и производительного секвенатора NovaSeq 6000 себестоимость анализа оказывается близка к пороговой (105,6 \$).

Было проведено сравнение себестоимости от забора биоматериала до получения fastq файлов разработанной методики НИПС с существующими на отечественном рынке аналогами (Таблица 4). Решение на базе smash/amplifet оказывается в 1,5–2,5 раза дешевле аналогов.

Таблица 4. Сравнение себестоимости методик НИПС на рынке России.

Название	Платформа для секвенирования	Режим секвенирования	Количество образцов в запуске	Стоимость анализа одного образца, \$	Стоимость анализа одного образца, Р
НИПС в ведущем роддоме РФ	Thermo Ion S5	SE200	9	265	19 875
Veriseq	Illumina NextSeq 550	PE36	24	360	27 000
Veriseq	Illumina NextSeq 550	PE36	48	240	18 000
Veriseq	Illumina NextSeq 550	PE36	96	241	18 090
Smash+amplifet	Illumina HiSeq 2500	PE250	136	170	12 721
Smash+amplifet	Illumina NovaSeq 6000	PE250	318	105	7 902

Предполагается, что эта характеристика методики позволит существенно расширить объем НИПС в России и значительно улучшить качество программ скрининга I триместра для всех беременных.

ВЫВОДЫ

1. Методика конструирования библиотек из длинных химерных ридов ДНК smash, состоящих из фрагментов внеклеточной ДНК, подходит для проведения анализа анеуплоидий плода по 13, 18 и 21 хромосомам.
2. Панель ОНП и методика конструирования библиотек ДНК amplifet для определения доли фетальной ДНК во фракции внеклеточной ДНК беременной позволяют вычислять долю фетальной ДНК.
3. Методики биоинформатической и статистической обработки результатов высокопроизводительного секвенирования для определения кариотипа плода по 13, 18 и 21 хромосомам, пола плода и доли фетальной ДНК во фракции внеклеточной ДНК беременной показали высокую эффективность при анализе образцов контрольной выборки.
4. Значения чувствительности, специфичности, положительной и отрицательной прогностических значимостей разработанной технологии определения кариотипа плода по 13, 18 и 21 хромосомам, пола плода и доли фетальной ДНК во фракции внеклеточной ДНК беременной соответствуют мировым аналогам.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. С учетом последних тенденций в мнении компетентных сообществ проведение НИПС может быть рекомендовано всем беременным вне зависимости от результатов комбинированного скрининга.
2. Необходимым и достаточным для надежного определения анеуплоидий по 13, 18, 21 хромосомам с помощью разработанной методики является получение не менее 2,5 млн фильтрованных фрагментов. В зависимости от используемой платформы и типа реагентов для секвенирования количество «сырых» ридов может варьировать.
3. Определение доли внфДНК в образце вДНК является обязательным, так как значительно сокращает количество ложноотрицательных результатов скрининга, возникающих в образцах пациентов, содержащих внфДНК ниже порогового значения.
4. Себестоимость проведения НИПС по разработанной методике существенно ниже представленных на рынке РФ аналогов, что позволяет ускорить внедрение высокоэффективных молекулярно-генетических технологий в клиническую практику.
5. НИПС не является диагностическим методом, так как источником внфДНК для анализа служит плацента, геном которой может быть генетически гетерогенен и не соответствовать полностью геному плода.

ПЕРСПЕКТИВЫ ДАЛЬНЕЙШЕЙ РАЗРАБОТКИ ТЕМЫ

1. Более технологичным подходом для процессирования образцов в лаборатории является оценка доли внфДНК на основании прямого анализа ридов библиотек ДНК smash без подготовки второй библиотеки типа amplifet. Возможным решением, во-первых, является сравнение нуклеотидного профиля концов фрагментов по аналогии с уже используемыми подходами на основе анализа треков намотки ДНК на нуклеосомы, который отличается у эмбриона и матери. Во-вторых, анализ данных с помощью нейросетей также с высокой вероятностью позволит построить алгоритм анализа регионов интереса в геноме. Для обоих перечисленных решений необходимо увеличение контрольной выборки образцов с определенной на основании анализа ОНП долей внфДНК до 1500–2000 образцов.

2. В качестве альтернативы избавлению от библиотек типа amplifet из п.1 возможно модифицировать панель олигонуклеотидов для первой амплификации внДНК таким образом, что помимо оценки доли внфДНК будет возможно проводить определение у плода носительства часто встречающихся патогенных вариантов, связанных с развитием ряда наследственных заболеваний. Таким образом спектр скринируемых наследственных патологий можно значительно расширить.

3. Важным является внедрение уникальных молекулярных индексов (UMI – unique molecular index) как решения, позволяющего проводить корректную дедупликацию результатов секвенирования и отделять природные дубликаты от ПЦР-дубликатов, возникающих в процессе подготовки библиотек ДНК. Благодаря этому улучшится как точность определения доли внфДНК, так и разрешающая способность скрининга на носительство патогенных вариантов из п.2.

4. Представляется необходимым ввести в анализ определение анеуплоидий по другим хромосомам, в частности, половым, так как они имеют большое клиническое значение. Как и в п.1, для этого необходима соответствующая контрольная выборка, содержащая в себе десятки образцов с анеуплоидиями по половым хромосомам, подтвержденные инвазивными методами.

5. Разработанная нами методика теоретически позволяет детектировать частичные анеуплоидии как хромосом интереса, так и других хромосом, а также дубликации и делеции. Необходимо исследовать вопрос их клинической значимости в зависимости от происхождения – материнского или плодового/плацентарного.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Каретникова Н.А., Шубина Е.С., Баранова Е.Е., **Коростин Д.О.**, Екимов А.Н., Парсаданян Н.Г., Гус А.И., Бахарев В.А., Трофимов Д.Ю., Тетруашвили Н.К.,

Сухих Г.Т. Неинвазивная пренатальная диагностика анеуплоидий методом высокопроизводительного секвенирования (NGS) в группе женщин высокого риска //Акушерство и гинекология. – 2015. – №. 4. – С. 5-10.

2. Сухих Г.Т., Трофимов Д.Ю., Барков И.Ю., Донников А.Е., Шубина Е.С., **Коростин Д.О.**, Екимов А.Н., Бахарев В.А., Каретникова Н.А., Баранова Е.Е., Тетруашвили Н.К., Ким Л.В., Павлович С.В., Скрыбин К.Г., Прохорчук Е.Б., Мазур А.М., Сангаев С.С, Пантюх К.С., Чеканов Н.Н. Неинвазивный пренатальный ДНК-скрининг анеуплоидий плода по крови матери методом высокопроизводительного секвенирования (Клинические рекомендации) //Акушерство и гинекология. – 2016. – №. 6. – С. 3-22.

3. Г.Т. Сухих, Д.Ю. Трофимов, И.Ю. Барков, А.Е. Донников, Е.С. Шубина, **Д.О. Коростин**, А.Н. Екимов, А.Ю. Гольцов, В.А. Бахарев, Н.А. Каретникова, П.И. Боровиков, Н.К. Тетруашвили, Л.В. Ким, А.С. Гата, С.В. Павлович, К.Г. Скрыбин, Е.Б. Прохорчук, А.М. Мазур, К.С. Пантюх. Новые подходы к проведению пренатального скрининга хромосомной патологии: ДНК-скрининг по крови матери //Акушерство и гинекология. – 2016. – №. 8. – С. 72-78.

4. **Коростин Д. О.**, Плахина Д. А., Белова В. А. Неинвазивный пренатальный молекулярный скрининг: особенности внедрения в клиническую практику //Вестник Российского государственного медицинского университета. – 2019. – №. 3.

5. Belova V., Plakhina D., Evfratov S., Tsukanov K., Khvorykh G., Rakitko A., Konoplyannikov A., Ilinsky V., Rebrikov D., **Korostin D.** NIPT technique based on the use of long chimeric DNA reads //Genes. – 2020. – Т. 11. – №. 6. – С. 590.

6. Плахина Д.А., Белова Д.А., Евфратов С.А., Ракитько А.С., Цуканов К.В., Ильинский В.В., Ребриков Д.В., **Коростин Д.О.** Применение длинных химерных ридов ДНК для неинвазивного пренатального скрининга //Геномная медицина в пренатальной диагностике, генетическом паспорте и в геномной терапии: Сб. науч. тр. / Под ред. чл.-корр. РАН В.С. Баранова, д.б.н. А.С. Глотова.- Новосибирск: Академиздат, 2020, с. 37-41. ISBN 978-5-6045338-4-0

7. В.А.Белова, Д.А.Плахина, А.С.Ракитько, К.В.Цуканов, **Д.О.Коростин**. Длинные химерные риды ДНК как способ снижения себестоимости неинвазивного пренатального скрининга //Сборник тезисов 8-й Всероссийской научно-практической конференции по геномному секвенированию и редактированию (NGS-2020). 20-21 мая 2020 года / Москва, с.12. ISBN 978-5-88458-503-4

8. Патент № 2717023, Российская Федерация. Способ определения кариотипа плода беременной женщины на основании секвенирования гибридных прочтений, состоящих из коротких фрагментов внеклеточной ДНК / **Коростин Д.О.**, Плахина Д.С., Евфратов С.В., Ракитько А.С., Ильинский В.В.

СПИСОК СОКРАЩЕНИЙ

внДНК – внеклеточная ДНК

внфДНК – внеклеточная фетальная ДНК

ИД – инвазивная пренатальная диагностика

НИПС – неинвазивный пренатальный скрининг

ОНП –однонуклеотидный полиморфизм

п.о. – пар оснований

рид – единичное прочтение в высокопроизводительном секвенаторе

AUC – area under ROC curve, площадь под ROC-кривой, количественная интерпретация ROC кривой

CI – confidence interval, доверительный интервал

NPV – negative predictive value, отрицательная прогностическая значимость

PPV – positive predictive value, положительная прогностическая значимость

ROC кривая – receiver operating characteristic, график качества бинарной классификации

SD – standard deviation, стандартное отклонение, σ

UMI – unique molecular index, уникальный молекулярный индекс