



НАЦИОНАЛЬНОЕ
РУКОВОДСТВО

РОССИЙСКОЕ ОБЩЕСТВО МЕДИЦИНСКИХ ГЕНЕТИКОВ

МЕДИЦИНСКАЯ ГЕНЕТИКА

Под редакцией
академика РАН Е.К. Гинтера,
академика РАН В.П. Пузырева,
члена-корреспондента РАН
С.И. Куцева



Москва
ИЗДАТЕЛЬСКАЯ ГРУППА
«ГЭОТАР-Медиа»
2022

ОГЛАВЛЕНИЕ

Участники издания.....	5
Предисловие.....	9
Список сокращений и условных обозначений	11
Введение. Краткий исторический обзор развития общей и медицинской генетики (<i>Гинтер Е.К., Пузырев В.П.</i>).....	15
Часть I. Научные основы медицинской генетики.....	31
Глава 1. Геном человека (<i>Скоблов М.Ю.</i>)	33
Глава 2. Гены человека: структура и функция (<i>Скоблов М.Ю.</i>)	44
Глава 3. Изменчивость генома человека (<i>Спарбер П.А., Скоблов М.Ю.</i>)	56
Глава 4. Хромосомы человека (<i>Рубцов Н.Б.</i>)	68
Глава 5. Митоз, мейоз, оогенез, сперматогенез (<i>Курило Л.Ф., Коломиец О.Л., Черных В.Б.</i>)	122
Глава 6. Генетика развития человека (<i>Баранов В.С., Кузнецова Т.В.</i>)	147
Глава 7. Эпигенетика человека: норма и патология (<i>Залетаев Д.В., Немцова М.В., Стрельников В.В.</i>)	179
Глава 8. Генетика соматических клеток и болезни человека (<i>Лебедев И.Н.</i>)	202
Глава 9. Популяционная генетика человека (<i>Балановская Е.В., Балановский О.П.</i>).....	216
Глава 10. Фармакогенетика (<i>Сычев Д.А.</i>).....	255
Глава 11. Иммуногенетика человека (<i>Гинтер Е.К., Писарев В.М.</i>).....	285
Глава 12. Система генов HLA: структура и функции (<i>Хаитов Р.М., Кофиади И.А.</i>)	308
Часть II. Клиническая генетика.....	327
Глава 13. Клиническая диагностика наследственных болезней (<i>Асанов А.Ю., Демикова Н.С.</i>)	329
Глава 14. Биохимические методы диагностики наследственных заболеваний (<i>Байдакова Г.В., Куркина М.В., Захарова Е.Ю.</i>).....	347
Глава 15. Молекулярно-генетические методы диагностики наследственных болезней (<i>Щагина О.А., Поляков А.В.</i>).....	360
Глава 16. Методы цитогенетической диагностики хромосомных болезней (<i>Лебедев И.Н.</i>)	374
Глава 17. Моногенные наследственные болезни (<i>Гинтер Е.К.</i>).....	398
Глава 18. Моногенные наследственные болезни в популяциях человека (<i>Зинченко Р.А.</i>).....	423
Глава 19. Наследственные болезни обмена веществ (<i>Захарова Е.Ю.</i>)	467
Глава 20. Митохондриальные заболевания (<i>Цыганкова П.Г., Иткис Ю.С., Крылова Т.Д., Бычков И.О., Захарова Е.Ю.</i>)	474
Глава 21. Заболевания, обусловленные экспансией tandemных повторов (<i>Иллариошкин С.Н.</i>)	492
Глава 22. Геномный импринтинг и наследственные заболевания у человека (<i>Залетаев Д.В., Немцова М.В., Стрельников В.В.</i>)	516

Глава 23. Хромосомные болезни (<i>Лебедев И.Н.</i>)	548
Глава 24. Генетика нарушений формирования пола, развития половой системы и репродуктивной функции (<i>Черных В.Б.</i>)	607
Глава 25. Многофакторные заболевания (<i>Пузырев В.П.</i>)	643
Глава 26. Врожденные пороки развития. Тератогенные синдромы (<i>Демикова Н.С., Жулева Л.Ю., Зубкова М.В.</i>)	685
Глава 27. Генетика и таргетная терапия опухолей (<i>Стрельников В.В., Немцова М.В., Кекеева Т.В., Залетаев Д.В.</i>)	710
Часть III. Лечение и профилактика наследственных болезней	731
Глава 28. Лечение наследственных болезней обмена веществ (<i>Куцев С.И., Захарова Е.Ю., Михайлова С.В.</i>)	733
Глава 29. Генная и генно-клеточная терапия (<i>Куцев С.И., Смирнихина С.А.</i>)	750
Глава 30. Медико-генетическое консультирование: организационные принципы и методические основы (<i>Козлова С.И., Прытков А.Н., Жулева Л.Ю.</i>)	770
Глава 31. Пренатальная диагностика врожденной и наследственной патологии (<i>Шилова Н.В., Мальмберг О.Л.</i>)	799
Глава 32. Преимплантационное генетическое тестирование (<i>Лебедев И.Н.</i>)	830
Глава 33. Массовый скрининг на наследственные болезни (<i>Куцев С.И., Захарова Е.Ю., Ижевская В.Л., Чумакова О.В.</i>)	852
Глава 34. Этические вопросы медицинской генетики (<i>Ижевская В.Л.</i>)	864
Предметный указатель	890

ЧАСТЬ I

НАУЧНЫЕ ОСНОВЫ МЕДИЦИНСКОЙ ГЕНЕТИКИ

Глава 1

Геном человека

ВСТУПЛЕНИЕ

Термин «геном» на сегодняшний день является одним из самых популярных научных терминов в мире. Действительно, несмотря на большой размер генома человека и на его сложность как биологического объекта, с геномом научились проводить множество сложнейших манипуляций, среди которых секвенирование, редактирование и даже синтез генома *de novo*. Все это стало возможным в последнее десятилетие благодаря появлению различных молекулярных технологий.

Сам термин «геном» был предложен Г. Винклером в 1920 г. для обозначения совокупности всего генетического материала конкретного организма. Сегодня у всех эукариотических организмов и у человека, в частности, принято выделять ядерный и митохондриальный геномы. Ядерный геном человека состоит из 22 аутосом и двух половых хромосом и имеет размер в 3,2 млрд пар нуклеотидов. Митохондриальный геном имеет только одну митохондриальную кольцевую хромосому длиной 16 569 пар нуклеотидов.

История секвенирования генома человека поучительна. Сама идея прочесть геном человека родилась в 1986 г., и возглавил проект «Геном человека» один из открывателей двуцепочечной структуры ДНК Джеймс Уотсон. Стоимость проекта была оценена в 3 млрд долларов, а сам проект был рассчитан на 15 лет при участии в проекте целого ряда стран: США, Германии, Франции, Великобритании, Китая, Японии и др., в том числе и России. В программе «Геном человека» участвовали более 400 российских исследователей из 30 научных учреждений, ими были опубликованы сотни научных работ, описывающих те или иные участки генома человека и модельных животных. В базах данных российскими учеными было зарегистрировано более миллиона нуклеотидных пар фрагментов ДНК человека и многие сотни генетических маркеров, имеющих большое значение для детального анализа генома человека.

Основными целями проекта «Геном человека» были определение последовательности нуклеотидов всей ДНК человека, а также идентификация и картирование генов. Первая («черновая») версия генома человека была опубликована в 2001 г. в журнале *Nature* (International Human Genome Sequencing Consortium, 2001). В результате секвенирования были обнаружены и картированы

последовательности 22 585 белок-кодирующих генов, которые суммарно занимали около 1,5% длины всего генома. Данные проекта были представлены в свободном доступе на сайте Национального центра биотехнологической информации США (National Center for Biotechnological Information, NCBI).

Помимо генов, были идентифицированы и картированы другие элементы генома (рис. 1.1, см. цв. вклейку): интроны, регуляторные последовательности, уникальная некодирующая ДНК, различные типы повторяющихся последовательностей.

Сейчас мировая биологическая наука находится в эре геномного секвенирования, которое стало очень доступным инструментом исследования. На сегодняшний день на сервере NCBI доступны данные о последовательностях геномов 47 186 организмов, среди которых 9199 — эукариотические. Ежегодно это число увеличивается на тысячи. Однако секвенирование дает лишь общие представления о том, как реализуется генетическая информация. Функциональное исследование генетических элементов в норме и при патологии на сегодняшний день является самым важным и до конца не решенным вопросом.

МЕТОДЫ ИССЛЕДОВАНИЯ ГЕНОМА

Исследование функционирования генома в виде отдельных генетических элементов или их совокупности эффективно реализуется при сравнительном анализе нормы и патологических состояний, что является основным направлением медицинской генетики. На сегодняшний день существует широкий спектр различных инструментов, позволяющих проводить такие исследования на молекулярном уровне для разных целей. Ниже будут рассмотрены основные подходы к исследованию генома.

Генетические карты

Одним из первых инструментов в исследовании генома были генетические карты. Генетическая карта — это схема взаимного расположения и относительных расстояний элементов генома, таких как гены или генетические маркеры. Она была придумана как инструмент, который должен облегчить картирование всех генов человека, особенно генов наследственных заболеваний.

При создании генетической карты генома человека использовались разные генетические маркеры, как полиморфные, так и уникальные. Самым распространенным, простым и удобным в использовании полиморфным маркером являются микросателлитные, или короткие, tandemные повторы (short tandem repeats — STR). Амплификация таких повторов методом ПЦР на уникальных фланкирующих праймерах давала информацию не только о наличии такого участка в исследуемом образце, но и о его аллельном состоянии. STR маркеры до сих пор активно используются в широком спектре задач медицинской генетики: при картировании генов наследственных заболеваний, идентификации личности, оценке генетического разнообразия.

STS маркеры (sequence tagged sites) представляют собой небольшие (100–200 п.н.) уникальные участки генома человека, которые амплифицируются с помощью ПЦР. Они были разработаны и активно использовались во время проекта по секвенированию генома человека для создания физической карты генома, позволяя устанавливать корректное расположение контигов (фрагментов ДНК) друг относительно друга.

Генетические карты сцепления являются наименее точными из всех имеющихся типов генетических карт, и их можно рассматривать только в качестве первого

приближения к реальным физическим картам. Тем не менее построение и использование генетических карт помогло за короткое время в геноме человека картировать тысячи генов.

Секвенирование по Сэнгеру

Метод секвенирования был предложен Фредериком Сэнгером в 1977 г. и по сей день является золотым стандартом исследования нуклеотидной последовательности генома (Sanger и др., 1977). Большая часть современной ДНК-диагностики, направленной на установление последовательности конкретного участка генома, осуществляется этим методом. С тех пор метод претерпел множество различных усовершенствований, и на сегодня он осуществляется на автоматических секвенаторах, где продукты реакции секвенирования с флуоресцентными метками разделяются с помощью капиллярного электрофореза. Единственным минусом этого метода является ограничение по длине фрагмента ДНК, который можно секвенировать за один раз. За одно «прочтение» можно определить последовательность длиной до 1000 нуклеотидов.

В 1977 г. был секвенирован первый геном бактериофага X174, длиной 5386 нуклеотидов (Sanger и др., 1977). В 1995 г. — первый полный геном клеточного организма — бактерии *Haemophilus influenzae*, вызывающей некоторые формы пневмонии и менингита, общей длиной 1 830 137 нуклеотидов (Fleischmann и др., 1995). В 1998 г. был секвенирован уже первый геном многоклеточного животного, круглого червя *Caenorhabditis elegans*, длиной 98 млн нуклеотидов (C. elegans Sequencing Consortium, 1998).

Для секвенирования геномов большого размера, включая геном человека, применялась следующая стратегия: геном разрезали на множество частей длиной примерно в 150 000 нуклеотидов с помощью двух или более рестриктаз. Эти частично перекрывающиеся фрагменты встраивали в искусственные кольцевые хромосомы (BAC — bacterial artificial chromosome), которые, в свою очередь, встраивали в бактерии. С помощью бактерий эти хромосомы размножаются, и в результате получается множество копий одного и того же фрагмента молекулы ДНК. Каждый такой фрагмент затем секвенировали отдельно, как правило, разбив его на еще меньшие по длине фрагменты. Прочитанные фрагменты наносили на карту хромосом, их последовательность устанавливали по перекрывающимся фрагментам. Данный подход был успешно реализован в ходе проекта «Геном человека».

Секвенирование методом «дробовика»

В связи с вышеописанными особенностями подхода, лежащего в основе проекта «Геном человека», сам проект реализовывался крайне медленными темпами. В то же время известный молекулярный биолог и генетик Крейг Вентер заявил, что его компания Celera Genomics, основанная в 1998 г., выполнит секвенирование генома человека раньше на 4 года, чем завершится международный проект. На проект потребуется всего 300 млн долларов, и реализован он будет с помощью новой технологии секвенирования «whole genome shotgun» — чтение случайных коротких фрагментов генома. Данное заявление вызвало бурную реакцию ученых и общественности, в частности еще и потому, что также было заявлено, что компания Celera Genomics собиралась заработать на проекте, создав базу данных последовательностей генома человека, которая была бы платной для коммерческих фармацевтических компаний.

Крейг Вентер сдержал озвученные им обещания, продемонстрировав вначале эффективность своего подхода при секвенировании генома плодовой мушки дро-

зофилы (Adams и др., 2000), а потом уже и при секвенировании генома человека. Появление такого коммерческого соперника подстегнуло и международный проект. В результате в 2001 г. было опубликовано сразу две версии предварительных вариантов генома человека (Venter и др., 2001; International Human Genome Sequencing Consortium, 2001).

NGS-секвенирование

Сейчас метод «дробовика» уже давно ушел в прошлое, и на смену ему пришли технологии секвенирования нового поколения (NextGen). За последние 10 лет они стремительно развивались в условиях активной конкуренции между ведущими коммерческими компаниями. Все это в результате привело к появлению современных NGS-секвенаторов, работающих на разных принципах, но быстро и качественно производящих огромное количество данных. С каждым годом, при существующей конкуренции производителей, стоимость секвенирования уменьшается, при этом качество и количество получаемых данных неуклонно растут. Это приводит к активному распространению технологий секвенирования для решения широкого спектра задач — от фундаментальных до прикладных, в частности в области медицинской генетики.

В основе технологий массового параллельного секвенирования лежит подход одновременного многократного прочтения ДНК матрицы. Многократность прочтения дает высокую точность получаемых данных. Так, на сегодняшний день принято получать данные при секвенировании полного генома с покрытием в 30 раз, то есть каждый нуклеотид в среднем должен быть просеквенирован и прочитан 30 раз.

Наиболее распространенными вариантами NGS-секвенирования являются более дешевые подходы по секвенированию панелей генов или экзона. Панели генов представляют собой набор генов, объединенных, как правило, какой-то одной нозологией или группой заболеваний. Например, существует совсем небольшая панель из двух генов для проведения анализа у пациентов с туберозным склерозом или же, наоборот, огромная неспецифическая онкопанель, куда входят сотни различных генов, имеющих отношение к канцерогенезу. Экзом представляет собой самую большую универсальную панель генов, содержащую экзоны большинства генов человека. Экзом секвенируют со средним покрытием 60 раз, получая качественные данные, способные выявить варианты нуклеотидной последовательности в кодирующей части генов.

Биоинформатический анализ генома

Секвенирование генома уже стало решенной технической задачей, при этом получают сотни гигабайтов информации. Для медицинской генетики основной задачей является поиск мутации в полученных данных при секвенировании ДНК пациента с наследственным заболеванием. Это весьма непростая задача, решением которой занимается относительно молодая наука «клиническая биоинформатика». Сейчас создаются и постоянно совершенствуются различные алгоритмы и программы, базы данных, серверы, кластеры, все то, без чего невозможен на сегодняшний день анализ генома после его секвенирования.

Биоинформатический анализ генома человека условно можно разделить на две следующие друг за другом задачи. Первая, техническая, когда полученные прочтения необходимо картировать на референсный геном человека и затем определить отличия в виде нуклеотидных замен или же протяженных делеций/инсерций. Это очень важная ресурсоемкая задача, правильность выполнения которой в большой

степени определяет успех следующей, второй, задачи — интерпретации полученных данных.

Сейчас с помощью NGS-секвенирования патогенные варианты нуклеотидной последовательности находят в среднем в 40% случаев. Одна из причин этого кроется в большой вариабельности генома, его полиморфности. При секвенировании генома здорового человека можно обнаружить около 3 млн однонуклеотидных позиций, по которым он будет отличаться от любого другого человека. Большая часть из них не имеет никакого биологического смысла, зато другая часть определяет индивидуальность человека. Разобраться в функциональной значимости этих полиморфных позиций — очень непростая задача, которую ученые-генетики пытаются решить до сих пор. Сложная она потому, что многие проявления фенотипа определяются не одним геном, а, как правило, несколькими (многими), их комбинацией. Так, например, за форму носа отвечают аллели 4 генов, цвет глаз — аллели трех генов, и так далее.

Для поиска патогенных вариантов клинические биоинформатики применяют различные подходы, пытаются фильтровать огромное количество полученных данных, используя различные критерии. Большая часть потенциально патогенных вариантов уникальна, поэтому поиск и интерпретация их требуют хорошего понимания как медицинских, так и молекулярно-биологических основ формирования фенотипа.

Индивидуальный геном

По завершении проекта «Геном человека» и при появлении современных NGS-секвенаторов наступила эра секвенирования индивидуальных геномов, которая активно продолжается и по сей день. Референсная последовательность генома человека была получена при секвенировании ДНК, выделенной из крови разных здоровых людей.

В 2007 г. был секвенирован первый полный индивидуальный геном Крейга Вентера (Levy et al., 2007). Анализ полученных данных показал наличие около 3 млн однонуклеотидных отличий, не считая большого количества крупных геномных вариаций. Год спустя был опубликован полный диплоидный геном Джеймса Уотсона, который содержал 3,3 млн однонуклеотидных замен по сравнению с референсным геномом человека (Wheeler et al., 2008). Более 10 000 замен были найдены в белокодирующих генах и представляли собой миссенс-варианты.

После Вентера и Уотсона были секвенированы еще несколько индивидуальных геномов, и было достоверно показано, что в среднем два индивидуума отличаются друг от друга на 3 млн однонуклеотидных замен. Также значимой отличительной частью являются участки генома, присутствующие в разном числе копий и порядке (copy number variants — CNV). В целом у 2 индивидов 99,5% последовательности генома является одинаковой.

Сегодня число индивидуальных отсекуированных геномов во всем мире уже перевалило за миллион. NGS-технология оказалась настолько мощной и чувствительной, что успешно удалось отсекуировать более 1000 геномов древних людей, умерших много тысяч лет назад. Помимо секвенирования геномов, также запущены проекты по получению метаданных, позволяющих максимально полноценно описать как норму, так и патологические состояния. В результате накапливаются данные по секвенированию транскриптомов, эпигеномов, микробиомов, протеомов. Можно уверенно сказать, что генетика человека сейчас стоит на пороге вхождения в эру персонализированной медицины, что должно привести к совсем другому уровню понимания происходящих процессов в организме.

Проекты по секвенированию геномов

Секвенирование персональных геномов здоровых людей, безусловно, является интересной фундаментальной задачей для установления разнообразия человеческого вида, эволюционных механизмов и многих других. Однако эти данные также очень востребованы и в медицинской генетике, для того чтобы при анализе генома пациента исключать из рассмотрения частые варианты нуклеотидной последовательности, встречающиеся в популяции.

Для исследования генетического полиморфизма в начале 2008 г. был запущен международный проект «1000 геномов». Целью его было провести секвенирование минимум одной тысячи здоровых людей из различных популяций. В результате к 2015 г. были опубликованы данные о геномах 2504 человек, представляющих 26 разных популяций (Adam и др., 2015). В ходе данного проекта удалось охарактеризовать широкий спектр генетических вариаций, в общей сложности было найдено более 88 млн генетических вариаций, среди них 84,7 млн однонуклеотидных полиморфизмов (single nucleotide polymorphism — SNP), 3,6 млн коротких инсерций и делеций, 60 000 структурных вариантов. По оценкам данного проекта, было охарактеризовано более 99% SNP, встречающихся с частотой >1%.

Сейчас многие страны имеют аналогичные проекты по секвенированию геномов как здоровых, так и больных различными заболеваниями. Так, например, в Великобритании в конце 2018 г. был успешно закончен проект «100 000 Genomes Project» по секвенированию геномов больных и их здоровых родственников с редкими заболеваниями. В группу больных включены люди как с редкими заболеваниями, так и с некоторыми видами опухолей и инфекционных заболеваний.

В 2015 г. в России стартовал аналогичный проект «Российские геномы», целью которого было секвенирование 3000 геномов здоровых индивидуумов из разных этносов, проживающих на территории РФ. Как заявляют участники проекта, за 5 лет удалось просеквенировать и проанализировать всего 60 геномов.

Типы полиморфизмов

Как уже было сказано, два генома отличаются друг от друга несколькими миллионами SNP. В геноме человека в среднем один SNP встречается на 300 нуклеотидов. Такая частая встречаемость их в геноме сделала SNP самыми эффективными и широко распространенными генетическими маркерами. Важно отметить, что полиморфизмами принято считать такие варианты нуклеотидной последовательности, которые встречаются в популяции чаще, чем в 1% случаев. Если частота полиморфного варианта меньше, то его называют мутацией.

Очевидно, что большая часть полиморфизмов находится во внегенных областях. Долгое время считалось, что они не несут никакого функционального значения. Однако, как будет сказано ниже, в геноме существует множество разных регуляторных последовательностей, оказывающих влияние на экспрессию генов. Для многих полиморфизмов экспериментально было показано, что в ряде случаев их расположение в регуляторном регионе может влиять на уровень экспрессии регулируемого гена.

Совсем небольшая доля SNP попадает в регионы белок-кодирующих генов. В части случаев SNP может являться миссенс-заменой. В этих случаях миссенс-замена может приводить к небольшому изменению функции белка. Именно такое большое количество событий в разных генах (а также в их регуляторных участках) с небольшими эффектами и является одной из основ существующего фенотипического разнообразия популяции.

Поиск и исследование таких функционально значимых полиморфизмов являются одними из интересных задач генетики человека. Существует целое направление исследований, называемое «полногеномный поиск ассоциаций» (genome-wide association studies — GWAS), связанное с исследованием ассоциаций между геномными вариантами и фенотипическими признаками. Наиболее известными полиморфизмами стали миссенс-варианты в генах, кодирующих ферменты биотрансформации I и II фазы (цитохром P450, глюкуронилтрансфераза и др.), а также транспортеров лекарственных средств (ЛС) (ABCB1, SLCO1B1 и др.). Было показано, что носительство конкретных генетических маркеров влияет на эффективность и безопасность фармакотерапии, как правило, путем изменения фармакокинетики ЛС либо путем модуляции фармакодинамики ЛС. Все это привело к появлению новой области, сочетающей медицинскую генетику и клиническую фармакологию, — фармакогенетики.

Помимо отдельных нуклеотидов, в ДНК полиморфными также являются небольшие инсерции и делеции, а также изменение числа микросателлитных (1–6 п.н. STR — short tandem repeats) и минисателлитных (7–190 п.н. VNTR — variable number tandem repeats) tandemных повторов. О повторах отдельно будет сказано ниже.

Другим распространенным видом генетического полиморфизма является CNV. В этом случае в популяции наблюдаются вариации числа копий протяженных участков ДНК, которые могут нести в себе функциональные гены. Происходит это за счет несбалансированных хромосомных перестроек, таких как делеции и дупликации.

ФУНКЦИОНАЛЬНОСТЬ ГЕНОМА

Проект ENCODE

Если секвенирование генома является вопросом относительно хорошо решенным, то его функционирование еще долгое время будет изучаться. В 2003 г. с этой целью был запущен международный проект ENCODE, который продолжается до сих пор. Название проекта расшифровывается, как энциклопедия ДНК элементов (Encyclopedia of DNA Elements).

Основная цель проекта ENCODE — определить роль белок-некодирующей части генома, большая часть которого традиционно рассматривалась как «мусорная ДНК». Уже первые пилотные результаты проекта ENCODE показали, что около 80% генома являются функциональными. Эти данные были восприняты научной общественностью с большим скептицизмом. Действительно, если участок генома проявляет какую-то активность, выявляемую различными экспериментами, то совсем еще не значит, что этот участок может быть действительно функционально важен для клетки или организма. Однако очевидно, что такие полномасштабные исследования функционального анализа генома необходимы не только для фундаментального понимания регуляции экспрессии генов, но и для поиска причин возникновения заболеваний.

При исследовании функционирования генома в проекте ENCODE применяются следующие методы:

- различные варианты секвенирования РНК, позволяющие не только определить структуру и уровень экспрессии транскриптов, но также их локализацию;
- методы CLIP-seq и RIP-seq, позволяющие выявить РНК-взаимодействия с белками;

- различные варианты методов определения полногеномного метилирования;
- метод ChIP-seq, позволяющий выявлять участки связывания с транскрипционными факторами (ТФ) и другими ДНК-связывающими белками;
- различные варианты методов, позволяющих определять статус компактизации хроматина;
- набор методов, выявляющих конформационные ДНК-ДНК-взаимодействия, описывающие пространственную организацию хроматина в ядре клетки.

Весь перечисленный арсенал методов применяется на большой коллекции из более чем 10 000 образцов различных тканей человека и клеточных линий. Что также важно, полученные данные находятся в открытом доступе. С помощью геномных браузеров удобно проводить анализ исследуемого региона для построения гипотез относительно его функционирования.

Регуляторные последовательности

У прокариот (архей и бактерий), как правило, существует линейная зависимость между размером генома и количеством генов. Однако у эукариотических организмов зависимости между размером генома и сложностью организма нет. Такая значительная избыточность некодирующих нуклеотидных последовательностей и изменчивость количества ДНК у близких видов названа С-парадоксом. Секвенирование геномов различных организмов привело к формулированию и другого явления — G-парадокса, описывающего несоответствие количества белок-кодирующих генов и сложности фенотипа. На основании этого возникла гипотеза, что «эволюционное качество» достигается не количеством генов, а их регуляцией.

Структура генов и их функции будут подробно рассмотрены в отдельной главе. Но стоит заметить, как уже было сказано выше, белок кодирующие гены занимают в геноме человека около 1,5%, при этом регуляторных последовательностей на порядок больше. Действительно, для «получения» и существования такого сложного многоклеточного организма, как человек, в котором насчитывается минимум 400 различных типов клеток, существенно различающихся по структуре и функциям, необходимо иметь огромный потенциал в регуляции молекулярных и клеточных процессов. Этот потенциал реализуется на многих уровнях функционирования биомолекул в клетке, однако базовым является уровень генома, который изначально определяет доступность и возможность транскрипции генов.

Известно, что активность и экспрессия белок-кодирующих генов могут изменяться за счет различных регуляторных элементов ДНК, таких как промотор, транскрипционные регуляторные последовательности, области структуры хроматина, модификации гистонов.

Базовым регуляторным элементом является промотор. Это участок ДНК, который приводит к инициации транскрипции конкретного гена. Как правило, он располагается рядом с точкой инициации транскрипции (TSS — transcription start site) и может быть до нескольких тысяч нуклеотидов длиной. По данным ENCODE, промоторные последовательности РНК-полимеразы II в геноме человека составляют 4,2%.

Также в поиск и аннотирование промоторов большой вклад внес японский проект FANTOM5. В ходе его были получены данные о картированных TSS, полученных в коллекции из почти тысячи образцов первичных клеток человека, тканей и клеточных культур. В результате было получено более 100 000 TSS, которые свидетельствовали о существовании промоторов в этих участках. Детальный анализ показал, что для некоторых генов было найдено два или более промоторов, дающих тем самым альтернативные начала генов. Но для большей доли промоторов не было найдено генов, которые бы они регулировали, их еще предстоит найти и охарактеризовать.

Часто в последовательности промотора встречаются так называемые CpG-островки. Это участки ДНК, характеризующиеся высокой частотой динуклеотидов CG. В геноме человека их насчитывается около 31 000. Известно, что метилирование цитозина в положении 5 в островках CpG приводит к изменению уровня экспрессии гена. Такая регуляция экспрессии является широко распространенной и называется эпигенетической. Подробно она будет рассмотрена в отдельной главе.

На активность промоторов могут влиять другие регуляторные элементы — энхансеры, сайленсеры и инсуляторы. Энхансеры представляют собой небольшие участки ДНК (от 50 до 1500 нуклеотидов длиной), с которыми могут связываться белки ТФ и тем самым приводить к увеличению транскрипции конкретного гена. По данным ENCODE, в геноме человека таких последовательностей насчитывается более миллиона, и суммарно они занимают 15,9% всей длины генома. Аналогичным образом выглядят сайленсеры — последовательности ДНК, с которыми связываются факторы транскрипции, приводящие к понижению или к полному подавлению экспрессии гена. В геноме человека их насчитывается около 1,5 млн.

Влияние энхансеров или сайленсеров на промотор могут регулировать инсуляторы, также являющиеся цис-регуляторными элементами. Обычно их длина варьирует от 300 до 2000 п.н. Если инсулятор расположен между промотором и каким-то другим регуляторным элементом, то связывание определенных белков с последовательностью инсулятора может приводить к блокированию взаимодействия с промотором. Одним из самых известных инсуляторов является белок CTCF. По данным ENCODE, он активно связывается с множеством последовательностей в геноме, приводя к формированию петель хроматина. Суммарно таких участков в геноме насчитывается около 0,7%.

Все эти регуляторные последовательности оказывают влияние как на уровень экспрессии гена, так и на его профиль. Важно понимать, что наблюдаемая экспрессия гена является результирующей совокупности всех регуляторных взаимодействий. Для некоторых генов таких взаимодействий насчитывается около сотни. Эти взаимодействия различны в разных тканях и на разных стадиях развития организма, что и приводит к различной экспрессии регулируемых генов.

Повторяющиеся последовательности

Большую часть генома человека составляют повторяющиеся последовательности, или просто повторы. Они представляют собой последовательности ДНК, которые встречаются во множестве копий по всему геному. Такие последовательности присутствуют в геномах всех организмов в большом количестве. В геноме человека суммарно их насчитывается около 5 млн, и они покрывают собой половину генома.

Повторяющиеся элементы делят на разные типы в зависимости от их структуры и/или способа умножения. Первый тип — тандемные повторы, в котором повторяющиеся элементы расположены в массивах тандемно повторяющихся последовательностей. Тандемные повторы делят на три класса согласно длине повторяющейся последовательности: микросателлитные повторы содержат ДНК-повторы от 1 до 9 нуклеотидов, минисателлитные — от 10 до 100 нуклеотидов и макросателлитные — более 100 нуклеотидов.

Микросателлитные повторы (STR) являются высокополиморфными участками генома, так как из-за особенности их структуры ДНК-полимераза способна при репликации как увеличивать их число, так и уменьшать. Поэтому они являются широко распространенными молекулярными маркерами в генетических и геномных исследованиях, о чем уже было сказано выше.

Но в отдельных случаях изменение числа микросателлитных повторов может приводить к возникновению наследственных заболеваний, к так называемым болезням экспансии повторов. В зависимости от их местоположения в геноме реализуется различный молекулярный механизм патогенеза заболевания. Так, увеличение числа CGG-повторов до более 200 копий в 5'-нетранслируемой области гена *FMR1* приводит к гиперметилированию данного гена и, как следствие, к отсутствию его экспрессии. Это приводит к возникновению синдрома ломкой X-хромосомы, или синдрома Мартина–Белла. Другим известным примером является болезнь Гентингтона, вызванная гетерозиготной экспансией тринуклеотидных повторов CAG, кодирующих глутамин, в гене *HTT*, который кодирует белок гентингтин. При увеличении числа повторов свыше 37 образуется токсичный белок, приводящий к дегенерации нейронов.

Другие классы tandemных повторов также могут изменять свое число, приводя к возникновению наследственных заболеваний. Одним из известных примеров является мышечная дистрофия Ландузи–Дежерина. Уменьшение числа макросателлитных повторов D4Z4 до 10 копий приводит к активации экспрессии располагающегося в них транскрипционного фактора DUX4, который запускает патогенетические процессы мышечной атрофии.

Другой тип повторов — диспергированные повторы, являющиеся по своей сути транспозонами. Их делят на 2 класса: ДНК-транспозоны и ретротранспозоны. Они являются мобильными генетическими элементами, способными перемещаться (транспозиция) в геноме и размножаться. Считается, что у человека транспозоны составляют до 45% всей последовательности ДНК.

Ретротранспозоны являются самым большим классом диспергированных повторов, в геноме человека они занимают 42% и способны самовоспроизводиться в геноме за счет реакции обратной транскрипции. Для этого вначале ретротранспозон может транскрибироваться в РНК, которая с помощью фермента обратной транскриптазы переводит РНК в комплементарную ДНК, способную встроиться в геном.

Ретротранспозоны делятся на несколько основных групп:

- ретротранспозоны с длинными концевыми повторами (Long terminal repeats, LTRs), HERV, MER4;
- ретротранспозоны с длинными диспергированными повторами (LINEs, long interspersed elements);
- ретротранспозоны с короткими диспергированными повторами (SINEs, short interspersed elements).

ДНК-транспозоны способны перемещаться по геному за счет вырезания и встраивания с использованием фермента транспозазы, которую они кодируют в своей последовательности. В геноме человека содержится около 300 000 копий ДНК-транспозонов, и они составляют около 3%. Среди семейств ДНК-транспозонов двумя основными являются MER1 и MER2, которые встречаются в геноме наиболее часто.

Существуют различные механизмы, которые подавляют активность транспозонов, препятствуя их перемещению и размножению. Однако технологии полногеномного секвенирования выявляют новые места встраивания транспозонов, которые иногда могут приводить к возникновению различных заболеваний. Так, в 2014 г. впервые были исследованы геномы 290 образцов опухолей и найдено порядка 3000 разных перестановок повторяющихся элементов из L1. Как было показано, в 2% случаев встраивание происходило в функционально значимые участки генома.

Другой масштабной работой являлся анализ данных почти 10 000 экзомов, полученных при секвенировании семей, состоящих из родителей с больными деть-

ми с нарушением развития. В результате было определено, что в среднем каждый индивидуум имеет около 33 встраиваний транспозонов или процессированных псевдогенов относительно референсного генома. 11 таких событий оказались новыми (*de novo*), а в 4 случаях они затронули гены, которые часто связывают с пороками развития (ПР).

ЗАКЛЮЧЕНИЕ

Геном человека сейчас является активно изучаемым объектом. Развивающиеся технологии массового параллельного секвенирования и биоинформатического анализа приближают нас к пониманию структуры и функционирования генома как в норме, так и при патологии. И сейчас видно, как они сложно устроены.

В начале 2000-х была очень популярной теория о существовании мусорной ДНК, которая говорила о том, что функционально важна только небольшая доля генома, представляющая белок кодирующие гены. Эта теория подтверждалась и экспериментами на мышах, которым делетировали протяженные участки генома, не содержащие гены, и мыши при этом оставались здоровыми.

Первые данные проекта ENCODE показали, что до 80% генома человека могут проявлять активность. И хотя эти данные были восприняты научным сообществом весьма критично, последующие данные этого и других проектов выявляют все больше и больше транскрибирующихся и регуляторных последовательностей. Однако их реальная функция будет исследоваться еще не один десяток лет.

Геном человека, как и любой другой, постоянно эволюционирует за счет рекомбинации и активных мутационных процессов. Несмотря на то что скорость накопления мутаций у разных эукариотических организмов самая низкая, как было показано, в каждом поколении возникает *de novo* около 80 вариантов нуклеотидной последовательности (Jakob M. Goldmann и др., 2016). Эти процессы являются источниками как популяционного разнообразия, так и причиной возникновения широкого спектра заболеваний.

Список литературы

1. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // *Proceedings of the National Academy of Sciences, USA*. 1977. Vol. 74. P. 5463–5467.
2. Sanger F., Air G.M., Barrell B.G. et al. Nucleotide sequence of bacteriophage phi X174 DNA // *Nature*. 1977. Vol. 265, N 5596. P. 687–695.
3. Fleischmann R.D., Adams M.D., White O. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd // *Science*. 1995. Vol. 269, N 5223. P. 496–512.
4. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology // *Science*. 1998. Vol. 282, N 5396. P. 2012–2018.
5. Adams M.D., Celniker S.E., Holt R.A. et al. The genome sequence of *Drosophila melanogaster* // *Science*. 2000. Vol. 287, N 5461. P. 2185–2195.
6. Venter J.C. et al. The sequence of the human genome // *Science*. 2001. Vol. 291, N 5507. P. 1304–1351.
7. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome // *Nature*. 2001. Vol. 409. P. 860–921.
8. Levy S., Sutton G., Ng P.C., Feuk L. et al. The diploid genome sequence of an individual human // *PLoS Biol*. 2007. Vol. 5. P. e254.
9. Auton A., Abecasis G.R., Altshuler D.M. et al. A global reference for human genetic variation // *Nature*. 2015. Vol. 526, N 7571. P. 68–74.
10. Wheeler D.A., Srinivasan M., Egholm M. et al. The complete genome of an individual by massively parallel DNA sequencing // *Nature*. 2008. Vol. 452, N 7189. P. 872–876.
11. Goldmann J.M., Wong W.S.W., Pinelli M. et al. Parent-of-origin-specific signatures of *de novo* mutations // *Nature Genetics*. 2016. Vol. 48. P. 935–939.