

МУЛЬТИПЛЕКСНОЕ ГЕНОТИПИРОВАНИЕ ОДНОНУКЛЕОТИДНЫХ ПОЛИМОРФНЫХ МАРКЕРОВ МЕТОДОМ МАСС-СПЕКТРОМЕТРИИ MALDI-TOF: ЧАСТОТЫ 56 SNP В ГЕНАХ ИММУННОГО ОТВЕТА В ПОПУЛЯЦИЯХ ЧЕЛОВЕКА

© 2013 г. В. А. Степанов*, Е. А. Трифонова

Научно-исследовательский институт медицинской генетики Сибирское отделение Российской академии медицинских наук, 634050, Томск

Поступила в редакцию 13.05.2013 г.

Принята к печати 27.06.2013 г.

Однонуклеотидные полиморфные маркеры (SNP) — наиболее распространенный тип генетического полиморфизма. Несмотря на прогресс в технологиях секвенирования и постгеномных технологиях, точечное генотипирование SNP остается самым востребованным подходом в генетике человека и медицинской генетике. В работе описан опыт применения мультиплексного генотипирования SNP с помощью масс-спектрометрии MALDI-TOF для анализа генетического разнообразия генов иммунного ответа в популяциях человека. Показано, что масс-спектрометрия MALDI-TOF — это быстрый, точный и эффективный метод для среднemasштабного генотипирования SNP. В четырех исследованных российских популяциях (русские, коми, ханты и буряты) выявлен близкий спектр аллельных частот 56 SNP в 41 гене, продукты которых участвуют в регуляции иммунного ответа. В изученных популяциях зафиксирован сходный уровень генетического разнообразия и обнаружена кластеризация популяций в соответствии с их географической локализацией. Оценена экономическая эффективность генотипирования с помощью масс-спектрометрии MALDI-TOF в сравнении с технологией ПЦР в реальном времени.

Ключевые слова: однонуклеотидные полиморфные маркеры (SNP), масс-спектрометрия MALDI-TOF, мультиплексное генотипирование, ПЦР в реальном времени, генетическое разнообразие, популяции человека, гены иммунного ответа.

MULTIPLEX GENOTYPING OF SINGLE NUCLEOTIDE POLYMORPHISMS BY MALDI-TOF MASS-SPECTROMETRY: FREQUENCIES OF 56 SNP IN IMMUNE RESPONSE GENES IN HUMAN POPULATIONS, by V. A. Stepanov*, E. A. Trifonova (Research Institute for Medical Genetics, Russia Academy of Medical Sciences, Tomsk, 634050 Russia; *e-mail: vadim.stepanov@medgenetics.ru). Single nucleotide polymorphism (SNP) are the most common type of genetic polymorphisms. Despite the progress in sequencing and postgenomic technologies, point genotyping of SNPs remains the most demanded approach in human and medical genetics. In this paper, the experience in the application of multiple SNP genotyping using MALDI-TOF mass-spectrometry for the analysis of genetic diversity of immune response genes in human populations has been described. It is demonstrated that MALDI-TOF mass-spectrometry is rapid, accurate and efficient method for medium-scale SNP genotyping. In four populations (Russian, Komi, Khant, and Buryat) the similar spectrum of allelic frequencies for 56 SNPs in 41 genes implicated in the regulation of immune response was found. Populations under study have the similar level of genetic diversity and are clustered according their geographic location. Economic effectiveness of MALDI-TOF mass-spectrometry in comparison with a real time PCR technology is estimated.

Keywords: single nucleotide polymorphisms (SNPs), MALDI-TOF mass-spectrometry, multiplex genotyping, real time PCR, genetic diversity, human populations, immune response genes.

DOI: 10.7868/S0026898413060153

Однонуклеотидные полиморфные маркеры (SNP, single nucleotide polymorphism) представляют собой наиболее распространенный тип генетического полиморфизма. В текущей версии базы

данных dbSNP (build 137), аккумулирующей информацию об однонуклеотидной вариабельности геномов, описано около 38 млн. валидированных SNP в геноме человека [1]. Тем самым, в среднем

*Эл. почта: vadim.stepanov@medgenetics.ru



один из каждых 100 нуклеотидов в геноме человека вариателен. Прогрессу в накоплении информации о вариателности генома человека способствует появление технологий секвенирования ДНК следующего поколения (NGS) и работа международных консорциумов НарМар и “1000 геномов” [2, 3]. В каждом индивидуальном геноме при NGS-секвенировании обнаруживается 3–5 млн. SNP [4–7]. Считается, что геномы двух неродственных индивидов различаются, в среднем, по одному нуклеотиду из каждой тысячи.

Несмотря на прогресс в технологиях секвенирования и постгеномных технологиях, точечное генотипирование SNP остается самым востребованным подходом в генетике человека и медицинской генетике. Анализ вариателности SNP используется в генетическом картировании болезней человека, в ассоциативных исследованиях (в том числе полногеномных) распространенных болезней, в популяционной и эволюционной генетике человека, в ДНК-диагностике моногенных заболеваний, при ДНК-идентификации в криминалистике и судебной медицине и во множестве других фундаментальных направлений и практических приложений [8].

Технологии генотипирования SNP развивались в течение последних трех десятилетий от «ручного» генотипирования единичных маркеров методами гибридизации нуклеиновых кислот, ПЦР и выявления различий в ампликонах (ПЦР-ПДРФ, аллель-специфической ПЦР и др.) до автоматизированного мультиплексного анализа сотен тысяч–нескольких миллионов SNP на высокоплотных микроматрицах ДНК (биочипах) [8–11].

На сегодня самый распространенный метод генотипирования SNP в лабораторной и исследовательской практике — ПЦР в реальном времени с использованием флуоресцентно-меченых зондов (TaqMan-проб) [12]. Для более масштабных исследований применяют генотипирование на высокоплотных биочипах [9, 13, 14]. Первый из этих подходов позволяет в рамках одного эксперимента получать данные о единичных маркерах в сотнях или тысячах образцов, второй — о десятках и сотнях тысяч SNP в единицах или десятках образцов. При этом ниша промежуточного масштаба, самого востребованного при решении большинства задач — генотипирование десятков или сотен маркеров в сотнях образцов, — остается слабо заполненной современными технологическими решениями. Подходы, основанные на низкоплотных микроматрицах ДНК [15–19] или на SNaPShot-анализе с помощью капиллярного гель-электрофореза [20, 21], остаются мало востребованными из-за ограничений в возможности мультиплексирования и лимитированной разрешающей способности в разделении ампликонов.

Одна из технологий, “работающей” в описываемой нише задач, — масс-спектрометрия биологических макромолекул, в том числе ДНК [22]. В генетике и молекулярной биологии технологии на основе масс-спектрометрии используются для генотипирования, анализа экспрессии и метилирования ДНК [23–25]. В данной работе представлен опыт применения мультиплексного генотипирования SNP с помощью масс-спектрометрии на примере анализа 56 маркеров в генах иммунного ответа в популяциях человека, а также приводится оценка экономической эффективности разработанного подхода в сравнении с технологией ПЦР в реальном времени.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Принцип метода масс-спектрометрии MALDI-TOF. Метод масс-спектрометрии заключается в измерении отношения массы заряженных частиц вещества к их заряду. Существуют различные способы ионизации органических веществ в зависимости от природы и фазы вещества. Метод матрично-активированной лазерной десорбции/ионизации (MALDI, *matrix assisted laser desorption/ionisation*) позволяет ионизировать крупные биомолекулы без их деградации и заключается в облучении короткими лазерными импульсами образца, представляющего собой твердый раствор анализируемого соединения в органической матрице. Матрица выбирается таким образом, чтобы ее молекулы активно поглощали фотоны, эмитируемые УФ- или ИК-лазером. Над поверхностью образца создается плотная высокотемпературная плазма, в которой наряду с молекулами и ионами матрицы оказываются и молекулы анализируемого соединения. Ионизация последних путем поглощения энергии фотонов или в результате ионно-молекулярных реакций приводит к образованию положительных и отрицательных ионов, которые выталкиваются высоким потенциалом из области ионизации и направляются в анализатор. В анализаторе заряженные частицы разделяются по времени пролета определенного расстояния (TOF, *time of flight*), при этом время пролета пропорционально соотношению массы частицы к ее заряду. В настоящее время методом MALDI-TOF успешно анализируют пептиды, белки, нуклеотиды, полисахариды, синтетические полимеры, гуминовые кислоты, органические комплексные соединения и другие биомолекулы с молекулярными массами до нескольких сотен тысяч дальтон [22].

Основные этапы эксперимента. Мультиплексное генотипирование проводили на масс-спектрометре MassARRAY Analyzer 4 (“Sequenom”, США). Процесс подготовки и анализа образцов состоит из ряда последовательных этапов:

- 1) мультиплексная ПЦР;

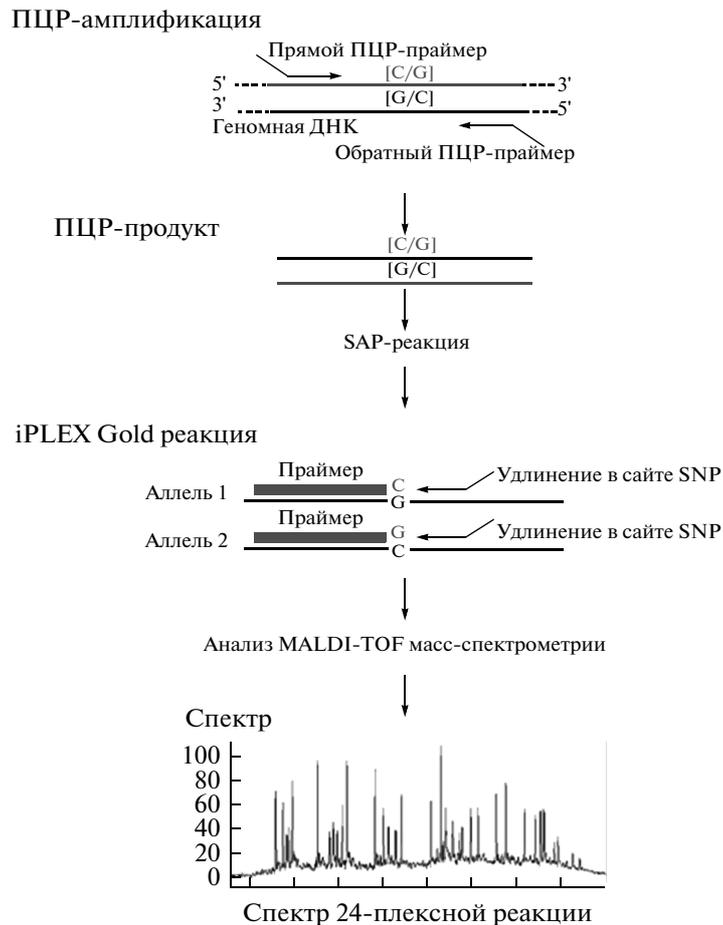


Рис. 1. Стадии подготовки и анализа образцов при генотипировании методом масс-спектрометрии MALDI-TOF.

- 2) SAP-реакция;
- 3) iPLEX-реакция;
- 4) ионизация и анализ спектров.

После iPLEX-реакции образцы обессоливают, наносят на SpectroCHIP и анализируют с помощью масс-спектрометрии MALDI-TOF. На рисунке 1 представлены основные этапы генотипирования с помощью масс-спектрометра MassARRAY Analyzer 4.

Образцы ДНК. В работе использовали образцы ДНК, выделенные из лимфоцитов периферической крови стандартными способами. Всего проанализировано 192 образца крови, взятой у индивидов четырех различных популяций: русские из г. Томска ($N = 48$), коми-зыряне из г. Сыктывкар Республики Коми ($N = 48$), ханты из пос. Казым Ханты-Мансийского автономного округа Тюменской области ($N = 48$) и буряты из пос. Курумкан Республики Бурятия ($N = 48$).

Гены и SNP. Исследование проведено на 58 SNP в генах, продукты которых участвуют в регуляции иммунного ответа. На основании анализа базы данных по полногеномным ассоциативным

исследованиям (GWAS) Национального института исследования генома человека США (NHGRI) [26] выбраны маркеры, ассоциированные с болезнями аутоиммунной природы (системная красная волчанка, рассеянный склероз, ревматоидный артрит), либо с продукцией цитокинов и иммуноглобулина E. Маркеры отбирали по следующим критериям: высокая степень достоверности ассоциации с фенотипом, полученная в GWAS ($p \leq (5 \times 10^{-8})$); повторяемость ассоциации с геном/маркером в нескольких GWAS; локализация маркера в кодирующей или регуляторной областях генома; подтверждение функциональной значимости полиморфизма по литературным данным.

Дизайн праймеров. Выбранные SNP объединяли в два мультиплекса: первый включал 34 полиморфных варианта, второй – 24. Состав мультиплексов приведен в табл. 1. Дизайн праймеров осуществляли с помощью программного обеспечения Sequenom Assay Design (“Sequenom”) доступного в он-лайн режиме на сайте www.sequenom.com. К каждому маркеру подобрали по два ПЦР-праймера (прямой и обратный) и пролон-

Таблица 1. Состав мультиплексов для ПЦР и масс-спектрометрии

| Мультиплекс 1 | | | Мультиплекс 2 | | |
|---------------|-------------|------------------|---------------|------------|----------------|
| № | SNP (rs) | Ген | № | SNP (rs) | Ген |
| 1 | rs1042713 | <i>ADRB2</i> | 1 | rs1024161 | <i>CTLA4</i> |
| 2 | rs11150610 | <i>ITGAM</i> | 2 | rs11865121 | <i>CLEC16A</i> |
| 3 | rs1128334 | <i>ETS1</i> | 3 | rs1800875 | <i>CMA1</i> |
| 4 | rs12722489 | <i>IL2RA</i> | 4 | rs20541 | <i>IL13</i> |
| 5 | rs1295685 | <i>IL13</i> | 5 | rs2227306 | <i>IL8</i> |
| 6 | rs1327474 | <i>IFNgR1</i> | 6 | rs2230926 | <i>TNFAIP3</i> |
| 7 | rs13277113 | <i>BLK</i> | 7 | rs2234711 | <i>IFNgR1</i> |
| 8 | rs1335532 | <i>CD58</i> | 8 | rs2240335 | <i>PADI4</i> |
| 9 | rs144651842 | <i>IL4RA</i> | 9 | rs2300747 | <i>CD58</i> |
| 10 | rs1800693 | <i>TNFRSF1A</i> | 10 | rs2305480 | <i>GSDMB</i> |
| 11 | rs1800896 | <i>IL10</i> | 11 | rs231735 | <i>CTLA4</i> |
| 12 | rs1801275 | <i>IL4RA</i> | 12 | rs2381416 | <i>IL33</i> |
| 13 | rs1805015 | <i>IL4RA</i> | 13 | rs2476601 | <i>PTPN22</i> |
| 14 | rs1837253 | <i>TSLP</i> | 14 | rs2569190 | <i>CD14</i> |
| 15 | rs2056626 | <i>CD247</i> | 15 | rs3087243 | <i>CTLA4</i> |
| 16 | rs2070874 | <i>IL4</i> | 16 | rs324015 | <i>STAT6</i> |
| 17 | rs2104286 | <i>IL2RA</i> | 17 | rs3890745 | <i>STAT4</i> |
| 18 | rs2227284 | <i>IL4</i> | 18 | rs485499 | <i>IL12A</i> |
| 19 | rs2430561 | <i>IFNg</i> | 19 | rs5744455 | <i>CD14</i> |
| 20 | rs2546890 | <i>LOC285626</i> | 20 | rs6441286 | <i>IL12A</i> |
| 21 | rs2618476 | <i>BLK</i> | 21 | rs7097397 | <i>WDFY4</i> |
| 22 | rs2736340 | <i>BLK</i> | 22 | rs907092 | <i>IKZF3</i> |
| 23 | rs3761847 | <i>TRAF1</i> | 23 | rs9303277 | <i>IKZF3</i> |
| 24 | rs3790567 | <i>IL12RB2</i> | 24 | rs9888739 | <i>ITGAM</i> |
| 25 | rs3821236 | <i>STAT4</i> | | | |
| 26 | rs4810485 | <i>CD40</i> | | | |
| 27 | rs4986790 | <i>TLR4</i> | | | |
| 28 | rs5743708 | <i>TLR2</i> | | | |
| 29 | rs6604026 | <i>RPL5</i> | | | |
| 30 | rs6679677 | <i>PTPN22</i> | | | |
| 31 | rs6897932 | <i>IL7R</i> | | | |
| 32 | rs6920220 | <i>TNFAIP3</i> | | | |
| 33 | rs743777 | <i>IL2RB</i> | | | |
| 34 | rs7574865 | <i>STAT4</i> | | | |

гирующий праймер для iPLEX-реакции. В масс-спектрометрическом анализе продуктов мультиплексной ПЦР для успешной идентификации маркеров и аллелей необходимо, чтобы масса пролонгирующих праймеров различалась, по крайней мере, на 30 Да для различных SNP и на 5 Да между ампликонами альтернативных аллелей одного маркера. Праймеры синтезированы компанией “Metabion” (Германия). Последовательности праймеров доступны по запросу у авторов.

Мультиплексная ПЦР. Цель данного этапа – наработка ампликонов, которые содержат исследуемые SNP. ПЦР проводили в 96-луночных планшетах в объеме 5 мкл на амплификаторах “Thermo Scientific” (Германия) и “Applied Biosystems” (США) в присутствии 10 нг геномной ДНК-матрицы и смеси ПЦР-праймеров соответствующего мультиплекса. Схема ПЦР: начальная денатурация при 94°C (5 мин), затем 42 цикла амплификации в следующих условиях:

денатурация при 94°C (20 с), отжиг при 56°C (40 с), элонгация при 72°C (60 с), после чего пробы инкубировали 5 мин при 72°C.

SAP-реакция. SAP-реакция – дефосфорилирование не включенных в ампликоны дезоксинуклеотидтрифосфатов (dNTPs) щелочной фосфатазой (SAP). Данный этап необходим для удаления из продуктов амплификации не прореагировавших dNTPs. Процедура выполняется на 96-луночных ПЦР-планшетах после реакции амплификации. В продукты амплификации, полученные на первом этапе, добавляют SAP-смесь (“Sequenom”), состоящую из однократного SAP-буфера и 0.5 единиц на образец щелочной фосфатазы креветок. Далее ПЦР-планшеты центрифугируют и инкубируют при 37°C в течение 40–50 мин, а затем, для инактивации SAP, прогревают при 85°C в течение 5–20 мин.

iPLEX-реакция. На этом этапе происходит мультиплексное удлинение пролонгирующих праймеров путем включения модифицированного дидезокситрифосфата (ddNTP) с модифицированной массой, комплементарного нуклеотида, находящегося в полиморфной позиции каждого из SNP. В результате образуется смесь коротких ПЦР-продуктов, соответствующих аллелям представленных в мультиплексе полиморфных маркеров. После SAP-реакции в образец добавляют раствор, содержащий десятикратный iPLEX-буфер (“Sequenom”), ДНК-полимеразу (iPLEX-фермент, “Sequenom”), четыре ddNTPs с измененной молекулярной массой, и смесь пролонгирующих праймеров. Необходимо отметить, что вследствие зависимости между детектируемой интенсивностью и массой анализируемого вещества в масс-спектрометрии MALDI-TOF, требуются различные концентрации пролонгирующих праймеров в реакционной смеси (концентрация праймеров с низкой молекулярной массой должна быть меньше концентрации праймеров с большей массой). Программа амплификации включает 1 цикл начальной денатурации при 94°C (30 с), далее 40 циклов, состоящих из денатурации при 94°C (5 с), отжига при 52°C (5 с), элонгации при 80°C (5 с), после чего образцы инкубируют 5 мин при температуре 72°C. Реакция проводится в том же планшете, что и предыдущие стадии эксперимента.

Очистка ПЦР-продуктов (обессоливание образцов). Чтобы избежать повышенного фонового шума в масс-спектрометрическом анализе, необходимо очистить реакционную смесь от солей, накопившихся при мультиплексной ПЦР и iPLEX-реакции. Для удаления ионов Na⁺, K⁺, и Mg²⁺ продукты реакции обрабатывают смолой Spectro-CLEAN (“Sequenom”). В каждую лунку 96-луночного планшета добавляют 15 мг смолы, планшет заклеивают пленкой, вращают на ротаторе 15 мин

и затем центрифугируют при 3200 × g в течение 5 мин.

Перенос образцов на спектро-чип. Перед ионизацией образцов и анализом масс-спектров образцы необходимо перенести на спектро-чип. Спектро-чип (“Sequenom”) представляет собой 96- или 384-луночный носитель с MALDI-матрицей для ионизации. Перенос образцов из ПЦР-планшетов на спектро-чип осуществляется в автоматическом режиме на рабочей станции Nanodispenser RS1000 (“Sequenom”). В каждую из лунок спектро-чипа наносится 8–12 нл образца.

Ионизация образцов и анализ масс-спектров. Спектро-чип помещается в рабочую камеру масс-спектрометра, где происходит последовательная ионизация образца в каждой из лунок чипа и разделение ионизированных частиц по массе. Процесс ионизации и получения масс-спектров 96-луночного спектро-чипа занимает примерно 20 минут. Анализ масс-спектров в реальном времени поводят с помощью программного обеспечения MassARRAY TYPER 4.0 (“Sequenom”). Этот же пакет программ используют для дальнейшего анализа и документирования результатов эксперимента.

Статистическая обработка результатов. Для анализа полученных данных применяли стандартные методы популяционной генетики и математической статистики. Соответствие распределения генотипов равновесию Харди–Вайнберга оценивали, используя точный тест Гуо и Томсона, а межпопуляционные различия – по коэффициенту геной дифференциации (Gst). Многомерный анализ частот аллелей проводили методом главных компонент в пакете программ Statistica (“StatSoft”, США).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ масс-спектров и генотипирование

На рисунке 2 приведен пример масс-спектра, полученного для одного из образцов по мультиплексу 2 на 24 SNP. Анализируемые молекулы распределены по массе от 5000 до 9000 Да; при этом в указанном интервале возможна детекция 72 продуктов – по 3 на каждый SNP, включая исходный пролонгирующий праймер и 2 продукта iPLEX-реакции, состоящие из пролонгирующего праймера с присоединенным нуклеотидом, соответствующим одному из двух альтернативных аллелей. Позиции возможных пиков масс-спектра для 72 продуктов показаны вертикальными прерывистыми линиями. В данном образце успешно прошла амплификация и элонгация всех 24 маркеров, и мы видим 24 пика пролонгирующих праймеров и 29 пиков специфических продуктов iPLEX-реакции (17 маркеров в гомозиготном состоянии и 6 – в гетерозиготном). Заметим, что при этом пики для непролонгированных прайме-

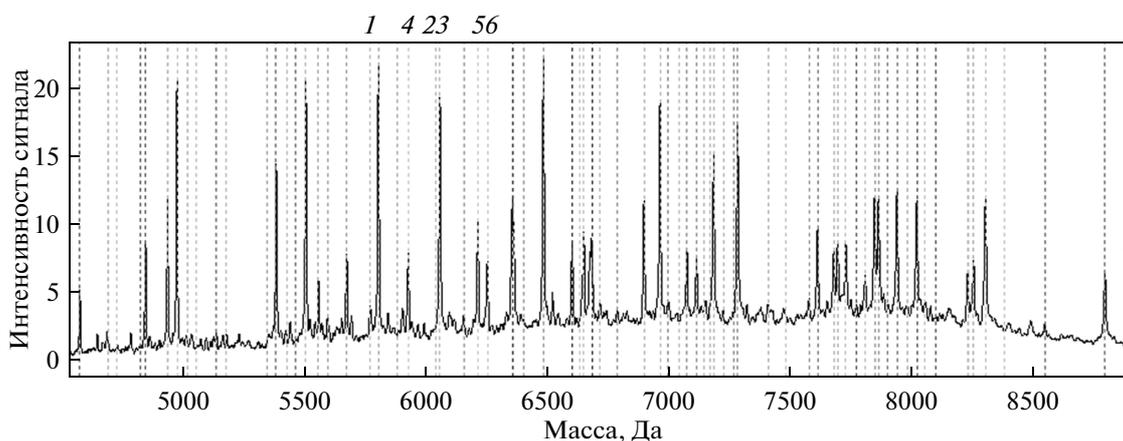


Рис. 2. Масс-спектр образца 24-плексной реакции (мультиплекс 2). Пример гомозиготного генотипа (rs20541): 1 – позиция пика нативного пролонгирующегося праймера без включенных ddNTP; 2 – позиция пика олигонуклеотида для аллеля Т (пик отсутствует); 3 – позиция пика олигонуклеотида для аллеля С (пик присутствует). Пример гетерозиготного генотипа (rs6441286): 4 – позиция пика нативного пролонгирующегося праймера без включенных ddNTP; 5 – позиция пика олигонуклеотида для аллеля G (пик присутствует); 6 – позиция пика олигонуклеотида для аллеля Т (пик присутствует).

ров не превышают 50 единиц интенсивности и всегда ниже, чем пики продуктов iPLEX-реакции для соответствующего SNP, что свидетельствует о высокой эффективности iPLEX-реакции и включении большей части молекул праймеров в специфичный iPLEX-продукт.

На рис. 3 представлен фрагмент масс-спектра 24-плексной реакции для трех образцов с тремя различными генотипами по rs20541. Разница между специфичными для аллелей продуктами iPLEX-реакции в 16 Да позволяет легко идентифицировать аллели и генотипы даже при анализе масс-спектра вручную.

Автоматическое чтение генотипов на основе мультиплексного масс-спектра в пакете MassARRAY TYPER 4.0 основано на анализе соотношения высоты пиков продуктов iPLEX-реакции двух аллелей. Пример кластеризации образцов в соответствии с генотипами для rs20541 в 24-плексной реакции приведен на рис. 4. Несмотря на существенный разброс высоты пиков между образцами, в данном примере успешно определены генотипы всех 95 образцов. Один образец без пиков (аллели

не найдены), находящийся в нулевой точке осей координат, соответствует контрольному образцу без ДНК.

Эффективность генотипирования

В рамках настоящей работы нами проведено 4 эксперимента по анализу эффективности и воспроизводимости метода – генотипирование двух разработанных мультиплексов (34-плекс и 24-плекс) в двух планшетах. Один из планшетов включал образцы коми и русских, второй – бурятов и хантов. В каждом из 96-луночных планшетов анализировали 95 образцов из соответствующих популяций и один контрольный образец без ДНК. Некоторые образцы дублировали для проверки воспроизводимости генотипов.

Общая эффективность генотипирования в 24-плексной реакции оказалось очень высокой (табл. 2). Генотипы прочитаны практически во всех образцах для всех SNP. Суммарное значение параметра “call rate” (доля определенных генотипов из всех возможных) составило 98.64%.

Таблица 2. Доля выявленных генотипов (call rate) в мультиплексных реакциях

| Эксперимент | Мультиплекс 1 (34-плекс)* | | | Мультиплекс 2 (24-плекс) | | | Call rate, суммарно (%) |
|-------------------------|---------------------------|----------------------|---------------|--------------------------|----------------------|---------------|-------------------------|
| | всего генотипов | генотип не определен | Call rate (%) | всего генотипов | генотип не определен | Call rate (%) | |
| Планшет 1 | 3040 | 113 | 96.29 | 2280 | 26 | 98.85 | 97.39 |
| Планшет 2 | 3040 | 16 | 99.47 | 2280 | 36 | 98.42 | 99.02 |
| Call rate, суммарно (%) | 6080 | 129 | 97.87 | 4560 | 62 | 98.64 | 98.20 |

* Без учета rs4810485 и rs3761847 (32-плекс).

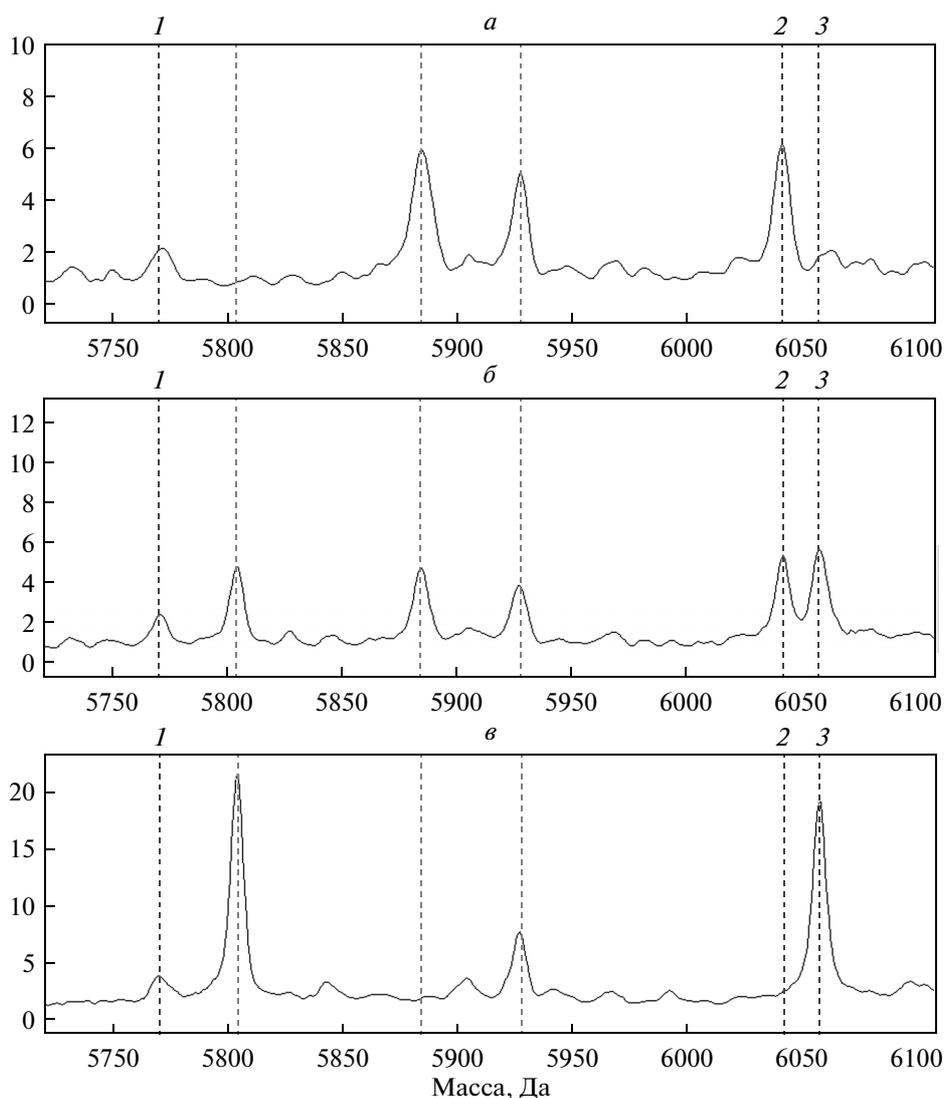


Рис. 3. Фрагмент масс-спектра трех образцов с различными генотипами rs20541: ТТ (а), ТС (б), СС (в). 1 – Позиция пика нативного пролонгирующегося праймера без включенных ddNTPs (5 771 Да); 2 – позиция пика олигонуклеотида для аллеля Т (6 042 Да); 3 – позиция пика олигонуклеотида для С (6 058 Да).

Эффективность 34-плекса оказалась несколько ниже. Два из 34 SNP в составе этого мультиплекса не сработали. Так, генотипы rs4810485 не удалось определить ни в одном образце, значение call rate для rs3761847 составило 42.1%, поэтому эти два маркера мы не рассматривали в последующем анализе. Без них эффективность генотипирования 34-плекса составила 97.87%. Для большинства SNP в обоих мультиплексах (52 из 56) значение call rate варьировало в пределах 98–100%. Самая низкая эффективность генотипирования зафиксирована для rs6441286, rs22403350 (по 96.84%), rs6604026 (93.15%) и rs6920220 (87.89%). Суммарное значение call rate для 56 SNP составило 98.2% (определено 10449 генотипов из 10640 возможных, табл. 2).

Популяционно-генетический анализ частот аллелей

В таблице 3 приведены частоты аллелей 56 изученных SNP, показатели генетического разнообразия и генетической дифференциации популяций. (Распределение генотипов по каждому из маркеров можно получить по запросу у авторов.) В проанализированном массиве данных, из 224 распределений генотипов (56 маркеров в 4 популяциях) в 12 наблюдалось отклонение от равновесия Харди-Вайнберга (РХВ), что не превышает ожидаемого числа случайных отклонений. Не наблюдалось накопления отклонений от РХВ по отдельным маркерам или популяциям.

У четырех исследованных популяций выявлен сходный спектр аллельных частот. Лишь по

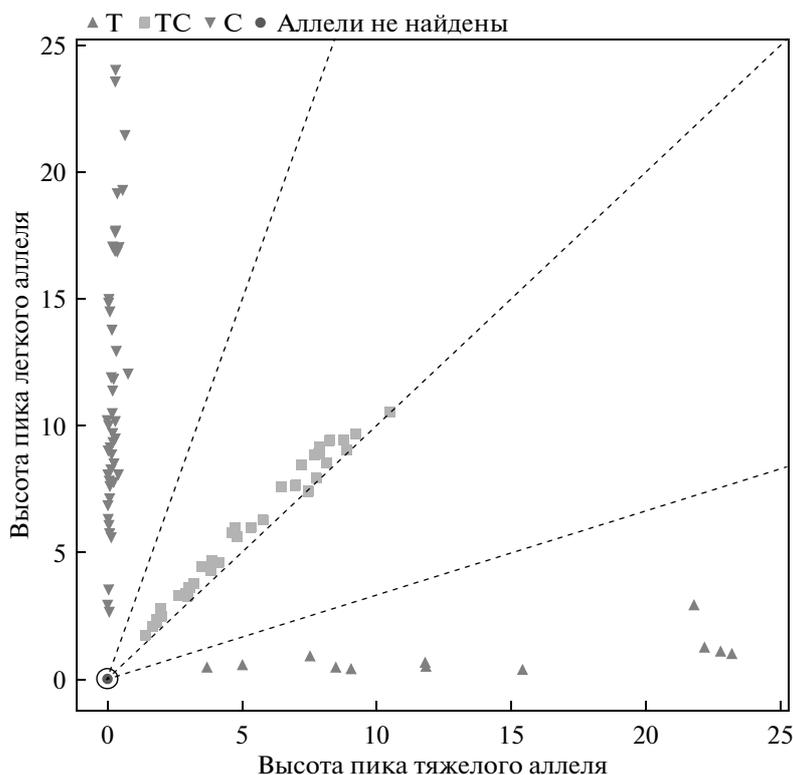


Рис. 4. Кластеризация образцов по генотипам в программе MassARRAY TYPER 4.0 на примере rs20541.

rs6604026 в гене *RPL5* и rs1128334 в гене *ETS1* аллель, редкий для русских и коми, часто встречается у бурят и хантов. Общий уровень генетического разнообразия популяций по исследованным маркерам также близок. Средняя ожидаемая гетерозиготность варьирует от 0.3522 у бурят до 0.3698 у коми.

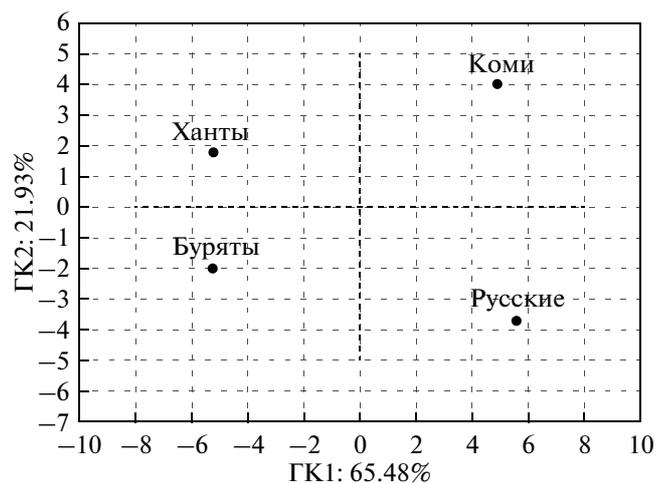


Рис. 5. Пространство двух первых главных компонент аллельных частот 56 SNP в исследованных популяциях.

Полокусные значения коэффициента генетической дифференциации для большинства маркеров (30 из 56) не превышают 2% (табл. 3). Лишь по трем SNP изученные популяции дифференцированы более чем на 10%, включая уже упоминавшиеся rs6604026 и rs1128334, для которых показан исключительно высокий уровень дифференциации ($G_{st} = 34.9\%$). Средний же уровень дифференциации по 56 маркерам составляет 4.2%, что соответствует уровню генетических различий между исследованными популяциями по различным аутосомным маркерным системам, включая широкогеномные наборы SNP [7, 14].

Многомерный анализ генетического разнообразия четырех изученных популяций по 56 маркерам проведен методом главных компонент (ГК). Две первые ГК объясняют в сумме 87.5% вариабельности аллельных частот в популяциях. Пространство ГК1 и ГК2 приведено на рисунке 5. Положение исследованных популяций в пространстве ГК1–ГК2 соответствует их географической принадлежности, что ожидалось, исходя из данных, полученных в предыдущих исследованиях [27, 28]. Популяции европейской части России кластеризуются в правой части графика, популяции сибирского региона – в левой. Тем самым, ГК1, вероятно, соответствует клинальной изменчивости частот аллелей по долготе, наблюдаемой по аутосомным маркерам, линиям Y-хромосомы

Таблица 3. Частоты аллелей, генетическое разнообразие и генетическая дифференциация исследованных популяций

| SNP | аллель | Частоты аллелей | | | | He* | Gst** |
|-------------|--------|-----------------|--------|--------|--------|--------|--------|
| | | русские | коми | буряты | ханты | | |
| rs1024161 | T | 0.50 | 0.58 | 0.70 | 0.72 | 0.4536 | 0.0337 |
| rs1042713 | G | 0.66 | 0.73 | 0.52 | 0.62 | 0.4532 | 0.0244 |
| rs11150610 | C | 0.61 | 0.63 | 0.41 | 0.49 | 0.4818 | 0.0320 |
| rs1128334 | G | 0.86 | 0.81 | 0.28 | 0.20 | 0.3183 | 0.3598 |
| rs11865121 | A | 0.32 | 0.30 | 0.22 | 0.24 | 0.3894 | 0.0085 |
| rs12722489 | G | 0.84 | 0.89 | 0.89 | 0.94 | 0.1939 | 0.0148 |
| rs1295685 | C | 0.67 | 0.68 | 0.69 | 0.77 | 0.4148 | 0.0066 |
| rs1327474 | A | 0.50 | 0.63 | 0.84 | 0.84 | 0.3751 | 0.1029 |
| rs13277113 | G | 0.79 | 0.73 | 0.46 | 0.52 | 0.4323 | 0.0787 |
| rs1335532 | T | 0.78 | 0.73 | 0.42 | 0.48 | 0.4328 | 0.0983 |
| rs144651842 | G | 0.97 | 0.92 | 0.81 | 0.77 | 0.2181 | 0.0573 |
| rs1800693 | A | 0.86 | 1.00 | 0.97 | 0.97 | 0.0865 | 0.0592 |
| rs1800875 | G | 0.48 | 0.49 | 0.50 | 0.53 | 0.4991 | 0.0018 |
| rs1800896 | A | 0.51 | 0.53 | 0.80 | 0.73 | 0.4271 | 0.0680 |
| rs1801275 | G | 0.17 | 0.14 | 0.16 | 0.16 | 0.2635 | 0.0013 |
| rs1805015 | T | 0.87 | 0.89 | 0.91 | 0.97 | 0.1650 | 0.0166 |
| rs1837253 | T | 0.24 | 0.27 | 0.46 | 0.45 | 0.4397 | 0.0433 |
| rs20541 | C | 0.67 | 0.69 | 0.68 | 0.75 | 0.4197 | 0.0044 |
| rs2056626 | G | 0.37 | 0.39 | 0.15 | 0.27 | 0.3963 | 0.0427 |
| rs2070874 | T | 0.32 | 0.22 | 0.50 | 0.29 | 0.4205 | 0.0495 |
| rs2104286 | A | 0.79 | 0.81 | 0.87 | 0.92 | 0.2545 | 0.0222 |
| rs2227284 | A | 0.38 | 0.28 | 0.54 | 0.38 | 0.4609 | 0.0355 |
| rs2227306 | C | 0.58 | 0.51 | 0.61 | 0.61 | 0.4841 | 0.0070 |
| rs2230926 | T | 0.97 | 0.94 | 0.98 | 0.99 | 0.0565 | 0.0096 |
| rs2234711 | C | 0.38 | 0.49 | 0.46 | 0.47 | 0.4917 | 0.0064 |
| rs2240335 | G | 0.64 | 0.66 | 0.52 | 0.49 | 0.4785 | 0.0211 |
| rs2300747 | A | 0.81 | 0.73 | 0.42 | 0.47 | 0.4221 | 0.1157 |
| rs2305480 | C | 0.56 | 0.50 | 0.52 | 0.53 | 0.4973 | 0.0020 |
| rs231735 | G | 0.35 | 0.24 | 0.23 | 0.18 | 0.3686 | 0.0193 |
| rs2381416 | A | 0.86 | 0.81 | 0.94 | 0.91 | 0.2074 | 0.0229 |
| rs2430561 | T | 0.62 | 0.57 | 0.73 | 0.76 | 0.4311 | 0.0266 |
| rs2476601 | G | 0.93 | 0.89 | 0.96 | 0.98 | 0.1129 | 0.0197 |
| rs2546890 | A | 0.53 | 0.53 | 0.46 | 0.50 | 0.4981 | 0.0036 |
| rs2569190 | G | 0.55 | 0.49 | 0.43 | 0.38 | 0.4884 | 0.0173 |
| rs2618476 | T | 0.79 | 0.73 | 0.42 | 0.52 | 0.4279 | 0.0968 |
| rs2736340 | C | 0.78 | 0.73 | 0.40 | 0.50 | 0.4312 | 0.1011 |
| rs3087243 | G | 0.69 | 0.76 | 0.77 | 0.81 | 0.3635 | 0.0091 |
| rs324015 | A | 0.27 | 0.40 | 0.40 | 0.43 | 0.4610 | 0.0164 |
| rs3790567 | A | 0.28 | 0.18 | 0.24 | 0.23 | 0.3561 | 0.0064 |
| rs3821236 | G | 0.78 | 0.77 | 0.64 | 0.72 | 0.3888 | 0.0160 |
| rs3890745 | G | 0.42 | 0.38 | 0.46 | 0.40 | 0.4830 | 0.0041 |
| rs485499 | T | 0.56 | 0.56 | 0.88 | 0.74 | 0.3959 | 0.0844 |
| rs4986790 | A | 0.94 | 0.89 | 0.97 | 0.99 | 0.0991 | 0.0295 |
| rs5743708 | G | 0.95 | 0.97 | 0.96 | 0.98 | 0.0711 | 0.0031 |
| rs5744455 | C | 0.78 | 0.74 | 0.76 | 0.69 | 0.3799 | 0.0054 |
| rs6441286 | T | 0.62 | 0.65 | 0.53 | 0.56 | 0.4789 | 0.0095 |
| rs6604026 | T | 0.30 | 0.34 | 0.92 | 0.91 | 0.2948 | 0.3754 |
| rs6679677 | C | 0.94 | 0.88 | 0.96 | 0.98 | 0.1135 | 0.0259 |
| rs6897932 | C | 0.74 | 0.67 | 0.77 | 0.76 | 0.3883 | 0.0078 |
| rs6920220 | G | 0.87 | 0.89 | 0.91 | 0.90 | 0.1894 | 0.0027 |
| rs7097397 | G | 0.66 | 0.64 | 0.53 | 0.50 | 0.4765 | 0.0198 |
| rs743777 | G | 0.34 | 0.20 | 0.06 | 0.11 | 0.2709 | 0.0749 |
| rs7574865 | T | 0.22 | 0.28 | 0.33 | 0.27 | 0.3944 | 0.0079 |
| rs907092 | G | 0.49 | 0.56 | 0.48 | 0.44 | 0.4965 | 0.0067 |
| rs9303277 | T | 0.53 | 0.45 | 0.58 | 0.57 | 0.4930 | 0.0097 |
| rs9888739 | T | 0.08 | 0.08 | 0.07 | 0.09 | 0.1473 | 0.0007 |
| Средняя He | | 0.3690 | 0.3698 | 0.3522 | 0.3471 | 0.3595 | 0.0419 |

* He – ожидаемая гетерозиготность. ** Gst – коэффициент генной дифференциации.

Таблица 4. Финансовые и временные затраты на генотипирование SNP методом MALDI-TOF на масс-спектрометре Sequenom MassARRAY 4 в сравнении с ПЦР в реальном времени (ПЦР-PB)

| Показатель | MALDI-TOF | ПЦР-PB |
|--|-----------|---------|
| Реагенты и расходные материалы на генотипирование 30 маркеров в 960 образцах (руб.)* | 347908 | 1228027 |
| В том числе (руб.): | | |
| – Комплект реагентов | 263 093 | 608 304 |
| – Олигонуклеотидные праймеры | 82758 | – |
| – TaqMan-пробы | – | 558010 |
| – Пластик (планшеты и наконечники) | 2057 | 61713 |
| Стоимость генотипирования 1 маркера в 1 образце (руб.)** | 12.08 | 42.64 |
| Число генотипов, получаемое 1 оператором за 1 день работы*** | 5760 | 384 |

* Для 96-луночного формата. Расчет на основе цен официальных российских дистрибуторов на апрель 2013 года.

** Только реактивы и расходные материалы, без учета стоимости и амортизации оборудования, стоимости труда оператора, накладных и прочих расходов.

*** Для 96-луночного формата: MALDI-TOF – 2 планшета на 96 образцов, по 30 маркеров на образец; MALDI-TOF – 4 планшета на 96 образцов по 1 маркеру на образец.

и мтДНК [7, 14, 27, 28]. ГК2, по-видимому, соответствует оси Север–Юг.

Оценка временных и финансовых затрат на генотипирование SNP методом MALDI-TOF в сравнении с ПЦР в реальном времени

Наиболее востребованным методом мелкомасштабного генотипирования SNP на сегодняшний день остается ПЦР в реальном времени. Генотипирование с помощью масс-спектрометрии способно решать как задачи мелкомасштабного (единичные маркеры в сотнях или тысячах образцов), так и среднемасштабного (десятки и сотни маркеров в сотнях или тысячах образцов) генотипирования. Давайте рассмотрим гипотетический пример решения задачи по генотипированию 30 маркеров в 960 образцах с помощью масс-спектрометрии MALDI-TOF и методом ПЦР в реальном времени в 96-луночном формате для обоих подходов. ПЦР с флуоресцентно-мечеными TaqMan-зондами потребует отдельного генотипирования каждого маркера. В MALDI-TOF-варианте масс-спектрометрии 30 маркеров могут быть объединены в один мультиплекс. Расчеты затрат на расходные материалы и реагенты для обоих методов приведены в таблице 4. Расчеты сделаны на основе цен официальных российских дистрибуторов на апрель 2013 г. При этом учтена возможность уменьшения объемов реакционной смеси для ПЦР в реальном времени в 3–4 раза по сравнению с инструкциями производителя TaqMan-зондов. Расчеты показывают, что стоимость генотипирования одного маркера в одном образце методом MALDI-TOF в 3.5 раза ниже, чем ПЦР в реальном времени. При этом 30-маркерный мультиплекс не есть максимально возможный – разрешающая способность использованного масс-

спектрометра позволяет мультиплексировать до 40 SNP. Немаловажным нам представляется и большая производительность масс-спектрометрии – за один рабочий день на одном приборе можно проанализировать в 15 раз больше генотипов, чем при тех же условиях методом ПЦР в реальном времени (табл. 4). Преимущества масс-спектрометрии в себестоимости и производительности генотипирования несколько нивелируются тем, что далеко не все маркеры поддаются мультиплексированию, а также высокой стоимостью оборудования для MALDI-TOF. Разница в стоимости приборов между ПЦР и масс-спектрометрией окупится лишь при относительно больших масштабах генотипирования – при указанной разнице себестоимости одного маркера потребуется примерно 400 тыс. генотипов, чтобы перекрыть разрыв в цене оборудования. По-видимому, масс-спектрометрию имеет смысл рассматривать как оптимальную технологию генотипирования для больших лабораторий или центров коллективного пользования, тогда как в случае небольших научных подразделений оптимальными останутся ПЦР в реальном времени и другие методы мелкомасштабного генотипирования или генотипирование на заказ в сервисных компаниях.

В данной работе мы описали опыт применения мультиплексного генотипирования SNP с помощью тандемной масс-спектрометрии для анализа генетического разнообразия генов иммунного ответа в популяциях человека. Показано, что масс-спектрометрия MALDI-TOF – это быстрый, точный и эффективный метод для среднемасштабного генотипирования SNP. В исследованных российских популяциях обнаружен близкий спектр аллельных частот 56 SNP в 41 гене, продукты которых участвуют в регуляции иммунного

ответа. В популяциях русских, коми, хантов и бурятов зафиксирован идентичный уровень генетического разнообразия и выявлена кластеризация популяций в соответствии с их географической локализацией. Оценена экономическая эффективность генотипирования с помощью масс-спектрометрии MALDI-TOF в сравнении с технологией ПЦР в реальном времени.

Работа частично финансировалась ФЦП “Исследования и разработки по приоритетным направлениям развития науки и техники” (Госконтракт № 11.519.11.2036), ФЦП “Научные и научно-педагогические кадры инновационной России” (соглашение № 8042), Российским фондом фундаментальных исследований (грант № 12-04-00595а).

СПИСОК ЛИТЕРАТУРЫ

1. http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi
2. The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1092 human genomes. *Nature*. **491**, 56–65.
3. The International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature*. **467**, 52–58.
4. Levy S., Sutton G., Ng P.C., et al. 2007. The diploid genome sequence of an individual human. *PLoS Biology*. **5**, 2113–2144.
5. Schuster S.C., Miller W., Ratan A., et al. 2007. Complete Khoisan and Bantu genomes from southern Africa. *Nature*. **463**, 963–947.
6. Скрыбин Г.К., Прохорчук Е.Б., Мазур А.М., и др. 2009. Комбинирование двух технологических платформ для полногеномного секвенирования человека. *Acta Naturae*. **1**, 113–119.
7. Степанов В.А. 2010. Геномы, популяции, болезни: этническая геномика и персонифицированная медицина. *Acta Naturae*. **2**, 18–34.
8. Kim S., Misra A. 2007. SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.* **9**, 289–320.
9. Steemers F.J., Gunderson K.L. 2007. Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J.* **2**, 41–49.
10. Kwok P.Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* **2**, 235–258.
11. Kumar S., Banks T.W., Cloutier S. 2012. SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genomics*. doi: 10.1155/2012/831460.
12. Shen G.Q., Abdullah K.G., Wang Q.K. 2009. The TaqMan method for SNP genotyping. In: *Single Nucleotide Polymorphisms: Methods and Protocols. Methods in Molecular Biology*, vol. 578. Ed. A.A. Komar. Totowa, N.J: Humana Press. pp. 293–306.
13. Карунас А.С., Юнусбаев Б.Б., Федорова Ю.Ю., и др. 2011. Полногеномный анализ ассоциации бронхиальной астмы в Волго-Уральском регионе России. *Молекуляр. биология*. **45**, 992–1003.
14. Степанов В.А., Прохорчук Е.Б., Хуснутдинова Э.К., и др. 2010. Евразийский генетический ландшафт: геогеография в зеркале полногеномной вариабельности. Материалы VI Российского съезда общества медицинских генетиков. Ростов-на-Дону. 14–18 мая 2010 г. *Медицинская генетика*. С. 171.
15. Tebbutt S.J., He J.Q., Burkett K.M. et al. 2004. Microarray genotyping resource to determine population stratification in genetic association studies of complex disease. *Biotechniques*. **37**, 977–985.
16. Федорова Ю.Ю., Гра О.А., Карунас А.С. и др. 2009. Ассоциация полиморфных вариантов генов системы биотрансформации с atopическими заболеваниями у русских детей из Республики Башкортостан. *Молекуляр. биология*. **43**, 1032–1039.
17. Грядунов Д.А., Зименков Д.В., Михайлович В.М., и др. 2009. Технология гидрогелевых биочипов и ее применение в медицинской лабораторной диагностике. *Лаборатория*. **3**, 10–14.
18. Гра О.А., Кожекбаева Ж.М., Гра Д.В., и др. 2009. Биочиповые технологии в диагностике социально-значимых многофакторных заболеваний. *Вестник биотехнологии и физико-химической биологии*. **5**, 5–25.
19. Павлов А.Е., Апалько С.В., Воробьев Е.В. 2012. Молекулярно-генетическая диагностика муковисцидоза в формате микрочипа. *Лаборатория*. **4**, 16–19.
20. Pati N., Schowinsky V., Kokanovic A., et al. 2004. A comparison between SNaPshot, pyrosequencing, and biplex invader SNP genotyping methods: accuracy, cost and throughput. *Biomed. Biophys. Methods*. **60**, 1–12.
21. Esteves L.M., Bulhoes S.M., Brilhante M.J., Mota-Vieira L. 2013. The multiplex snapshot assays for SNP genotyping in candidate innate immune genes. *BMC Res. Notes*. **6**, 54. doi: 10.1186/1756-0500-6-54.
22. Cole R.B. 2012. *Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications (2nd Edition)*. John Wiley & Sons.
23. Ratnayake M., Reynard L.N., Raine E., et al. 2012. Allelic expression analysis of the osteoarthritis susceptibility locus maps to MICAL3. *BMC Medical Genetics*. **13**, 12. doi: 10.1186/1471-2350-13-12.
24. Bradic M., Costa J, Chelo I.V. 2011. Genotyping with Sequenom. In: *Molecular Methods for Evolutionary Genetics. Methods in Molecular Biology*, 772, pp. 193–210.
25. Gabriel S., Ziaugra L., Tabbaa D. 2009. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr. Protoc. Hum. Genet.* **60**, 2.12.1–2.12.18.
26. Hindorff L.A., Sethupathy P., Junkins H.A., et al. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA*. **106**, 9362–9367.
27. Степанов В.А. 2002. *Этногеномика населения Сибири и Средней Азии*. Томск: Печатная Мануфактура. 242 с.
28. Степанов В.А., Харьков В.Н., Пузырев В.П. Эволюция и филогеография линий Y-хромосомы человека. *Вестник ВОГиС*. **10**, 57–73.