

ГЕНОМИКА.
ТРАНСКРИПТОМИКА

УДК 577.29

УНИВЕРСАЛЬНАЯ ПАНЕЛЬ ИНСЕРЦИОННО-ДЕЛЕЦИОННЫХ
ПОЛИМОРФИЗМОВ *ChipID106* ДЛЯ ГЕНЕТИЧЕСКОЙ
ИДЕНТИФИКАЦИИ ЛИЧНОСТИ И БИОЧИП НА ЕЕ ОСНОВЕ

© 2023 г. Д. О. Фесенко^{а,*}, И. Д. Ивановский^б, П. Л. Иванов^с, Е. Ю. Земскова^с, С. А. Поляков^а,
О. Е. Фесенко^д, М. А. Филиппова^а, А. С. Заседателей^а

^аИнститут молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991 Россия

^бООО “Центр исследования ДНК”, Химки, 141402 Россия

^сРоссийский центр судебно-медицинской экспертизы Министерства здравоохранения Российской Федерации,
Москва, 125284 Россия

^дНаучно-исследовательский институт физики Южного федерального университета,
Ростов-на-Дону, 344090 Россия

*e-mail: deferos@yandex.ru

Поступила в редакцию 26.12.2022 г.

После доработки 26.12.2022 г.

Принята к публикации 25.01.2023 г.

Предложен новый подход к генетической идентификации личности, включающий панель из 106 инсерционно-делеционных (InDel) полиморфизмов и метод их генотипирования на биологических микрочипах. Такие фундаментальные свойства InDel-маркеров, как малая длина и низкая частота мутирования, обеспечивают им существенные преимущества в экспертной практике перед широко используемыми тандемными повторами (STR). В настоящей работе проанализировано распределение аллелей всех известных InDel-полиморфизмов в пяти крупнейших мировых популяциях (европейской, восточно-азиатской, южно-азиатской, африканской и американской), отобраны маркеры, отвечающие следующим критериям: частота минорного аллеля более 0.30, физическое расстояние между маркерами более 3 млн.п.н., отсутствие во фланкирующих последовательностях полиморфизмов, тандемных повторов и палиндромов, соотношение AT/GC близкое к 1. Сформирована панель из 106 полиморфизмов со средней частотой минорного аллеля 0.396 во всех пяти популяциях. Разработан метод генотипирования панели, включающий одностадийную мультиплексную ПЦР с последующей гибридизацией на биологическом микрочипе. Средняя длина ампликона составила 72 п.н. На выборке из 201 жителя Москвы и Санкт-Петербурга определены основные характеристики сформированной панели: вероятность случайного совпадения, равная 1.89×10^{-43} , потенциал исключения отцовства (CPE) 0.99999999063. Предлагаемый метод представляет собой альтернативу молекулярно-генетической идентификации личности, основанной на вариациях длин STR-локусов.

Ключевые слова: идентификация личности, биочип, криминалистика, InDel-полиморфизмы

DOI: 10.31857/S0026898423040067, **EDN:** QLARZQ

ВВЕДЕНИЕ

Широкое внедрение методов генетической идентификации личности произвело революцию в расследовании преступлений и стало “золотым стандартом” доказательной базы в суде. В частности, выявление преступников с использованием национальной базы генетических данных (вклю-

чая нераскрытые дела прошлых лет) снижает нагрузку на следственные органы и сокращает долю преступников, находящихся на свободе и продолжающих причинять вред обществу, а реабилитация ошибочно осужденных позволяет вернуть свободу невиновным. Предоставляя веские доказательства вины подозреваемых, современные методы генетической идентификации существенно упрощают принятие верных судебных решений.

Согласно принятой в США судебной практике, идентификация считается доказательной, если генотип подозреваемого по выбранному перечню маркеров уникален в популяции, численность

Сокращения: InDel – инсерционно-делеционный полиморфизм; SNP – однонуклеотидный полиморфизм; STR – короткий тандемный повтор; AFR – африканская популяция; AMR – смешанная американская популяция; EAS – восточно-азиатская популяция; EUR – европейская популяция; SAS – южно-азиатская популяция.

которой на порядок превышает население Земли (10^{11}) [1]. В геноме человека в избытке присутствуют полиморфные элементы, комбинация которых позволяет достоверно отличить одного человека от другого. Количественным показателем идентифицирующей способности единичного полиморфизма является вероятность случайного совпадения (MP, *matching probability*), которая вычисляется как сумма квадратов всех возможных его генотипов. Таким образом, чтобы обеспечить указанные судебные критерии, необходимо использовать набор полиморфизмов, комбинированная вероятность случайного совпадения (CMP, *combined matching probability*, вычисляется как произведение MP всех маркеров, входящих в идентифицирующую систему) при использовании которого не превышает 10^{-11} .

В подавляющем большинстве случаев генетические экспертизы проводят с целью идентификации личности и определения родства (чаще всего отцовства). И, если для доказательства идентичности двух образцов достаточно набора полиморфизмов, имеющего $CMP \approx 10^{-11}$, то для достоверного установления родства требуется на 10–20 порядков более высокий идентифицирующий потенциал. Кроме того, чем выше идентифицирующий потенциал, тем больше вероятность найти в национальной генетической базе родственников разыскиваемого лица, что крайне важно для следствия. В силу этого требования к CMP задаются именно экспертизами по установлению родства.

Более 30 лет доминирующим подходом к генетической идентификации личности было определение длины локусов, содержащих короткие tandemные повторы — STR (*short tandem repeats*), например (AATG)_n, где аллели отличаются числом повторяющихся мотивов (*n*). Одна из наиболее распространенных STR-систем, используемая в национальной генетической базе США CODIS, гарантирует определение по 13 основным локусам уникального генетического профиля в масштабах мировой популяции, имея $CMP \approx 10^{-15}$ [2].

Выбор геномной дактилоскопии в пользу STR-маркеров сложился исторически и обусловлен, в основном, ограниченной методической базой в период формирования этого направления криминалистики. Различия между аллелями STR-локусов обусловлены вариациями длины этих локусов, а определение длины фрагментов ДНК по их электрофоретической подвижности является одним из наиболее ранних и широко распространенных методов изучения ДНК, при помощи которого tandemные повторы и были впервые обнаружены [3]. Таким образом, развитие методов идентификации личности по этому пути до определенного момента видится вполне закономерным. Со временем был выявлен суще-

ственный недостаток STR-маркеров — их большая протяженность. В экспертной практике нередко встречаются образцы деградированной ДНК, которые фрагментированы настолько, что длинные локусы не амплифицируются, приводя либо к полному выпадению информации о данном локусе, либо к потере более длинного аллеля и ложной гомозиготности в гетерозиготных образцах. Острота этой проблемы немного снизилась при смене типа tandemных повторов: локусы с минисателлитами заменили на микросателлиты, что позволило сократить среднюю длину локуса [3].

Другой недостаток обусловлен высокой скоростью мутирования в STR-локусах, что создает риск ошибок при определении родства: в каждой десятой–сотой экспертизе истинного отцовства ребенок будет носителем аллеля, отличающегося от родительского. И эту проблему невозможно решить с помощью каких-либо методических приемов, так как она носит фундаментальный характер, связанный с высокой мутабельностью STR: 10^{-2} – 10^{-3} [5, 6]. Оптимальный выход в этой ситуации нам видится в замене используемых полиморфизмов более стабильными.

В качестве альтернативы STR-маркерам для идентификации личности в 1993 году были предложены однонуклеотидные полиморфизмы — SNP [7] — самый распространенный и самый короткий тип геномного полиморфизма. SNP на порядки более консервативны, чем STR: скорость мутирования 2.5×10^{-8} [8], но в силу биаллельности для получения сопоставимого дискриминирующего потенциала количество SNP должно быть в 2.5–4 раза больше, чем STR [9]. Однако это требовало сложных методических решений: высокочувствительной >50-плексной ПЦР и подходящих методов генотипирования, поэтому развитие этот подход на тот момент не получил. Тем не менее, основатель направления А.Д. Jeffreys (Сэр Алек Джеффрис) в заключении к своей обзорной публикации, посвященной 20-летию геномной дактилоскопии, возлагал надежды именно на SNP [4]. Идея использовать биаллельные полиморфизмы была положительно воспринята и, благодаря развитию новых методов генотипирования, в начале 2000-х появилась потенциальная возможность ее практической реализации.

Для создания экспертного инструмента на базе SNP или InDel необходимо решить две задачи: из десятков миллионов полиморфизмов собрать панель с требуемыми характеристиками и разработать метод генотипирования, отвечающий практическим задачам криминалистики. В последние два десятилетия опубликован ряд работ, посвященных как отбору маркеров-кандидатов [10–13], так и апробации методов их генотипирования [14], включая оценку применимости в экспертной практике [15–17].

Интересующий нас класс полиморфизмов представлен SNP и короткими InDel, практически равнозначными с точки зрения размеров, информативности, распространенности в геноме и мутабельности. Однако, принимая во внимание, что наиболее перспективным методом их генотипирования является гибридизационный анализ амплифицированных фрагментов, предпочтительнее использовать короткие InDel, так как в сравнении с SNP дискриминация аллелей InDel надежнее в силу более явных различий в температуре плавления совершенных и несовершенных дуплексов.

С 2009 г. предприняты десятки попыток сформировать InDel-панель для решения экспертных задач в криминалистике. В первых публикациях, посвященных использованию этого типа полиморфизмов в судебно-генетической экспертизе, предложены панели из 38 [18], 33 InDel [19], 49 InDel для глобальной панели [20]; за последние годы китайскими исследователями предложены панели из 12 [21], 32 [22], 35 [23], 47 [24], 60 [17] InDel-полиморфизмов.

Из всех опубликованных подходов к решению данной проблемы только один завершился созданием коммерческого продукта. Он получил название DIPplex® (“Qiagen”, США) и представляет собой набор реагентов для генотипирования панели из 30 InDel-полиморфизмов посредством капиллярного электрофореза [25]. Около сотни публикаций, посвященных апробации DIPplex® на различных популяциях мира, свидетельствуют о востребованности этого направления в криминалистике, в первую очередь, благодаря возможности генотипирования деградированных образцов (средняя длина ампликона в DIPplex® – 114 п.н.). DIPplex® был разработан как расходный материал под имеющееся в большинстве экспертных лабораторий оборудование для STR-типирования. Это должно было облегчить внедрение метода в практику, однако он не получил широкого распространения, и ни одна страна мира не использует его для создания национальной базы генетических данных. По-видимому, это связано с тем, что, несмотря на достаточность идентифицирующего потенциала ($CMP \leq 2.83 \times 10^{-13}$), достоверность установления родства крайне низка: CPE (combined probability of exclusion – вероятность исключения, основной показатель, используемый при определении родства) составляет всего 0.998, иначе говоря, кровное родство первой степени с исследуемым будет допускаться для каждого пятисотого случайного индивида.

Причина, по которой ни одна из разработок не доведена до внедрения на национальном уровне ни в одной стране, на наш взгляд, комплексная. Она состоит в отсутствии как универсальной панели биаллельных полиморфизмов, имеющей до-

статочный идентифицирующий потенциал во всех мировых популяциях, так и метода ее мультиплексного генотипирования, обладающего требуемыми в экспертной практике характеристиками. В настоящей работе нами предпринята попытка решить эту задачу. При формировании панели маркеров мы учли опыт предыдущих работ и требования, предъявляемые современным экспертно-криминалистическим научным сообществом к аналитическим возможностям систем генетической идентификации. В качестве метода генотипирования сформированной панели маркеров мы использовали развитый в ИМБ РАН оригинальный подход, представляющий собой мультиплексную ПЦР с последующим гибридизационным анализом на гидрогелевых биологических микрочипах. Для предварительной оценки аналитических характеристик разработанного инструмента, который мы назвали *ChipID106*, исследована выборка образцов ДНК российской популяции.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Отбор InDel-полиморфизмов. Для первичного отбора кандидатов написан макрос для MS Excel, анализирующий файлы gvfc исходными данными *1000 Genomes Project*, размещенными на FTP-сервере геномного браузера Ensemble [http://ftp.ensembl.org/pub/grch37/release-90/variation/gvf/homo_sapiens]. Заданы следующие критерии отбора: полиморфизм должен быть инсерцией или делецией и иметь частоту минорного аллеля ≥ 0.30 во всех пяти популяциях (AFR, AMR, EUR, SAS, EAS). В результате был сформирован пул из 12375 InDel-полиморфизмов. При вторичном отборе исключили InDel, входящие в состав тандемных повторов и полинуклеотидов, что сократило пул до 1395. На третьем этапе для снижения проблем при дизайне специфичных праймеров из списка были удалены кандидаты, содержащие тандемные повторы и полинуклеотиды во фланкирующей области (± 50 п.н. от полиморфизма), либо представляющие собой AT- или GC-богатые регионы. Оставшиеся претенденты были разбиты на кластеры, исходя из их расположения на хромосоме, и на финальном этапе отбор проводили с учетом проверки уникальности фланкирующих последовательностей по BLAST и минимального расстояния до следующего полиморфизма более 3 млн.п.н. Окончательный перечень составил 106 полиморфизмов (табл. 1). Номенклатура полиморфизмов и их координаты указаны согласно сборке GRCh38.p13.

Образцы ДНК. В работе использовали 201 образец ДНК от неродственных совершеннолетних доноров (100 мужчин и 101 женщина), проживающих в Москве, Московской области и Санкт-Петербурге. Концентрацию ДНК измеряли спек-

Таблица 1. Инсерционно-делеционные полиморфизмы, вошедшие в панель *ChipID106*, частоты аллелей и генетическое разнообразие в российской популяции по результатам настоящего исследования и в пяти основных популяциях по данным *1000 Genome Project*

№	ID полиморфизма	Название в системе <i>ChipID</i>	Положение	Алель	MA*	L**, п.н.	Частоты генотипов и аллелей в российской популяции (N = 201), F ₀				MP	Частота минорного аллеля в основных мировых макропопуляциях, F ₀					
							In/In	In/Del	Del/Del	In		Del	EAS	EUR	AMR	SAS	AFR
1	rs10617833	CID.01-042	1:41899280	TCC/-	Del	55	0.443	0.388	0.169	0.637	0.363	0.375	0.398	0.374	0.401	0.331	0.476
2	rs11394524	CID.01-120	1:119994503	G/-	Del	66	0.179	0.408	0.413	0.383	0.617	0.369	0.414	0.412	0.496	0.318	0.316
3	rs72082809	CID.01-155	1:154776010	GTA/-	Del	65	0.210	0.580	0.210	0.500	0.500	0.425	0.336	0.498	0.326	0.404	0.495
4	rs35379245	CID.01-171	1:170771419	AG/-	In	62	0.144	0.448	0.408	0.368	0.632	0.388	0.495	0.327	0.376	0.482	0.398
5	rs71107464	CID.01-192	1:191712777	AAAT/-	Del	68	0.353	0.483	0.164	0.595	0.405	0.385	0.340	0.408	0.421	0.473	0.489
6	rs66765111	CID.01-208	1:207842739	ATTTCAGTTTCTT CGAG/-	Del	65	0.455	0.440	0.105	0.675	0.325	0.412	0.332	0.381	0.318	0.320	0.473
7	rs35126101	CID.01-223	1:223222802	TG/-	Del	64	0.458	0.433	0.109	0.674	0.326	0.409	0.369	0.319	0.349	0.410	0.473
8	rs34450453	CID.01-240	1:239878518	TGG/-	In	84	0.191	0.437	0.372	0.410	0.590	0.366	0.400	0.426	0.454	0.445	0.325
9	rs10693974	CID.02-012	2:11733545	ACTC/-	In	61	0.095	0.433	0.473	0.311	0.689	0.420	0.436	0.388	0.484	0.387	0.471
10	rs66610049	CID.02-036	2:36026644	CTGT/-	Del	74	0.338	0.473	0.189	0.575	0.425	0.374	0.448	0.373	0.388	0.343	0.320
11	rs57616890	CID.02-043	2:42762877	ATA/-	Del	69	0.483	0.388	0.129	0.677	0.323	0.400	0.346	0.374	0.324	0.454	0.405
12	rs72339906	CID.02-080	2:80191212	AG/-	Del	58	0.328	0.517	0.154	0.587	0.413	0.399	0.431	0.435	0.454	0.373	0.394
13	rs832955	CID.02-102	2:101846803	G/-	In	71	0.217	0.455	0.328	0.444	0.556	0.362	0.405	0.427	0.440	0.483	0.309
14	rs111665673	CID.02-216	2:215632035	AGAG/-	In	76	0.210	0.490	0.300	0.455	0.545	0.374	0.343	0.398	0.308	0.427	0.375
15	rs57993635	CID.02-232	2:232275593	C/-	In	76	0.189	0.393	0.418	0.386	0.614	0.365	0.419	0.423	0.347	0.390	0.320
16	rs35926392	CID.03-016	3:16145615	GC/-	Del	66	0.338	0.493	0.169	0.585	0.415	0.386	0.362	0.474	0.441	0.385	0.372
17	rs139738417	CID.03-031	3:30759018	CATAAGC/-	In	75	0.194	0.507	0.299	0.448	0.552	0.384	0.321	0.354	0.349	0.402	0.451
18	rs7495895	CID.03-082	3:82011691	G/-	Del	70	0.368	0.458	0.174	0.597	0.403	0.375	0.456	0.393	0.406	0.310	0.343
19	rs35291341	CID.03-100	3:99843746	CT/-	In	64	0.154	0.438	0.408	0.373	0.627	0.382	0.350	0.367	0.432	0.465	0.492
20	rs35464243	CID.03-107	3:106930411	CAC/-	In	72	0.109	0.388	0.502	0.303	0.697	0.415	0.477	0.330	0.363	0.445	0.359
21	rs10540628	CID.03-124	3:123992597	AC/-	Del	67	0.313	0.488	0.199	0.557	0.443	0.376	0.330	0.424	0.455	0.308	0.465
22	rs59170274	CID.04-027	4:26821887	C/-	Del	65	0.338	0.507	0.154	0.592	0.408	0.396	0.371	0.405	0.483	0.468	0.350
23	rs34184705	CID.04-045	4:44871992	G/-	Del	59	0.398	0.463	0.139	0.629	0.371	0.392	0.310	0.469	0.334	0.492	0.433
24	rs56281469	CID.04-057	4:57360344	CT/-	In	66	0.085	0.338	0.577	0.254	0.746	0.455	0.476	0.312	0.484	0.368	0.387
25	rs35932180	CID.04-086	4:85828903	CT/-	In	47	0.129	0.428	0.443	0.343	0.657	0.396	0.366	0.351	0.473	0.353	0.470
26	rs72273695	CID.04-113	4:112718708	ATCTGCATTG/-	Del	106	0.537	0.393	0.070	0.734	0.266	0.448	0.421	0.302	0.326	0.456	0.306
27	rs70938781	CID.04-138	4:138358907	CCTACCTTG- TAGCT/-	In	62	0.205	0.500	0.295	0.455	0.545	0.379	0.418	0.490	0.448	0.403	0.409
28	rs3067059	CID.04-178	4:178345624	ACTTG/-	In	77	0.090	0.512	0.398	0.346	0.654	0.429	0.368	0.331	0.316	0.373	0.328
29	rs61235135	CID.05-003	5:3358309	C/-	Del	62	0.423	0.448	0.129	0.647	0.353	0.396	0.396	0.379	0.395	0.348	0.382

Таблица 1. Продолжение

№	ID полиморфизма	Название в системе <i>ChipID</i>	Положение	Аллель	МА*	L**, п.н.	Частоты генотипов и аллелей в российской популяции (N = 201), F ₀				MP	Частота минорного аллеля в основных мировых макропопуляциях, F ₀					
							In/In	In/Del	Del/Del	In		Del	EAS	EUR	AMR	SAS	AFR
30	rs34159280	CID.05-014	5:13705943	CTTCAACAAGA-	Del	72	0.448	0.478	0.075	0.687	0.313	0.434	0.393	0.393	0.405	0.452	0.330
31	rs142490175	CID.05-064	5:63882726	GAAG/-	Del	96	0.403	0.458	0.139	0.632	0.368	0.391	0.336	0.340	0.380	0.404	0.413
32	rs138124389	CID.05-087	5:86744866	TAGACACTG-CAGACTAC/-	In	70	0.179	0.507	0.313	0.433	0.567	0.388	0.465	0.416	0.412	0.449	0.433
33	rs34990478	CID.05-095	5:95002440	CAT/-	Del	65	0.378	0.453	0.169	0.604	0.396	0.377	0.319	0.382	0.370	0.365	0.435
34	rs6149192	CID.05-115	5:114944029	GACCCTAAACA-GAG/-	Del	72	0.234	0.473	0.294	0.470	0.530	0.364	0.478	0.489	0.402	0.454	0.415
35	rs71621395	CID.05-119	5:118963126	AAC/-	Del	85	0.408	0.453	0.139	0.634	0.366	0.391	0.487	0.441	0.379	0.491	0.327
36	rs66881681	CID.05-132	5:131813412	AGG-	Del	77	0.423	0.483	0.095	0.664	0.336	0.421	0.384	0.333	0.353	0.431	0.470
37	rs10701589	CID.05-140	5:140036116	TGAT/-	In	71	0.184	0.522	0.294	0.445	0.555	0.393	0.452	0.431	0.363	0.451	0.384
38	rs3832355	CID.05-150	5:150038428	ACTC/-	Del	65	0.353	0.502	0.144	0.604	0.396	0.398	0.331	0.417	0.494	0.406	0.473
39	rs10560231	CID.05-180	5:179630655	ATG/-	Del	92	0.308	0.463	0.229	0.540	0.460	0.362	0.471	0.412	0.460	0.411	0.396
40	rs35948562	CID.06-007	6:7387821	CA/-	In	73	0.095	0.488	0.418	0.338	0.662	0.421	0.450	0.385	0.463	0.366	0.497
41	rs56071028	CID.06-016	6:15988568	C/-	In	73	0.159	0.483	0.358	0.400	0.600	0.387	0.455	0.404	0.415	0.363	0.426
42	rs10529292	CID.06-031	6:31409184	ACA/-	In	64	0.114	0.403	0.483	0.316	0.684	0.408	0.304	0.330	0.427	0.312	0.470
43	rs68133212	CID.06-072	6:71745181	TCT/-	Del	76	0.463	0.418	0.119	0.672	0.328	0.403	0.384	0.347	0.402	0.408	0.346
44	rs11267328	CID.06-078	6:77750524	ACT-CAGTTCATGA/-	Del	69	0.388	0.488	0.124	0.632	0.368	0.404	0.325	0.447	0.346	0.344	0.465
45	rs3831872	CID.06-088	6:88183335	CAGTTAGAAAG-GAACACT/-	In	61	0.159	0.403	0.438	0.361	0.639	0.379	0.306	0.346	0.487	0.390	0.435
46	rs56343259	CID.06-096	6:96279205	GCCC/-	In	56	0.119	0.468	0.413	0.353	0.647	0.403	0.347	0.354	0.366	0.397	0.427
47	rs141141409	CID.06-125	6:124946472	GTT/-	In	64	0.090	0.485	0.425	0.333	0.668	0.424	0.411	0.358	0.318	0.303	0.414
48	rs7057239	CID.06-135	6:135024728	GAAT/-	Del	95	0.358	0.522	0.119	0.619	0.381	0.415	0.442	0.359	0.356	0.447	0.356
49	rs6149835	CID.06-144	6:143819409	TTGAA-CAAAGTCC/-	Del	76	0.468	0.413	0.119	0.674	0.326	0.403	0.482	0.339	0.440	0.359	0.478
50	rs11278298	CID.06-159	6:158930905	TCAITTCAG/-	Del	71	0.448	0.433	0.119	0.664	0.336	0.402	0.364	0.366	0.405	0.491	0.487
51	rs35345873	CID.07-010	7:103522781	G/-	In	85	0.149	0.483	0.368	0.391	0.609	0.391	0.379	0.407	0.334	0.408	0.356
52	rs68111476	CID.07-023	7:23027696	TTTGCTGATA-ACC/-	Del	66	0.420	0.445	0.135	0.643	0.358	0.393	0.468	0.363	0.329	0.347	0.358
53	rs1610906	CID.07-082	7:81725951	TAT/-	In	72	0.219	0.517	0.264	0.478	0.522	0.385	0.393	0.437	0.450	0.380	0.346
54	rs7943214	CID.07-097	7:96765416	GTAITTTG/-	In	82	0.100	0.428	0.473	0.313	0.687	0.416	0.468	0.351	0.412	0.401	0.470
55	rs3216282	CID.07-126	7:126081845	AGAT/-	Del	74	0.482	0.437	0.081	0.701	0.299	0.430	0.379	0.318	0.392	0.369	0.340
56	rs35575707	CID.07-140	7:140101500	AG/-	Del	68	0.418	0.393	0.189	0.614	0.386	0.365	0.495	0.399	0.370	0.406	0.486

Таблица 1. Продолжение

№	ID полиморфизма	Название в системе <i>ChipID</i>	Положение	Аллель	MA*	L**, п.н.	Частоты генотипов и аллелей в российской популяции (N = 201), F ₀				MP	Частота минорного аллеля в основных мировых макропопуляциях, F ₀					
							In/In	In/Del	Del/Del	In		Del	EAS	EUR	AMR	SAS	AFR
57	rs71520104	CID.07-153	7:153064801	TTTTCTGTA-CAGTG/-	In	93	0.139	0.431	0.426	0.356	0.644	0.387	0.389	0.425	0.350	0.395	0.424
58	rs35146459	CID.08-002	8:1788862	G/-	Del	55	0.323	0.537	0.139	0.592	0.408	0.413	0.377	0.385	0.383	0.332	0.338
59	rs35604660	CID.08-059	8:59055597	GT/-	In	81	0.194	0.502	0.303	0.445	0.555	0.382	0.342	0.411	0.458	0.355	0.378
60	rs34067160	CID.08-065	8:65072454	AA/-	In	96	0.189	0.438	0.373	0.408	0.592	0.367	0.471	0.427	0.435	0.418	0.392
61	rs35679778	CID.08-104	8:103906808	C/-	In	71	0.144	0.453	0.403	0.371	0.629	0.388	0.432	0.398	0.486	0.473	0.456
62	rs11323264	CID.08-127	8:126654202	G/-	Del	44	0.373	0.448	0.179	0.597	0.403	0.372	0.475	0.414	0.365	0.443	0.341
63	rs6150855	CID.08-141	8:140579696	AGGCTGAGCAG/-	Del	62	0.428	0.443	0.129	0.649	0.351	0.396	0.357	0.356	0.326	0.391	0.414
64	rs56962291	CID.09-072	9:72269223	CTCCT/-	In	81	0.104	0.473	0.423	0.341	0.659	0.413	0.486	0.385	0.432	0.460	0.315
65	rs34499887	CID.09-087	9:86660018	G/-	Del	50	0.234	0.557	0.209	0.512	0.488	0.409	0.493	0.463	0.352	0.465	0.365
66	rs59923204	CID.09-109	9:108837145	C/-	Del	54	0.473	0.368	0.159	0.657	0.343	0.384	0.409	0.347	0.373	0.457	0.362
67	rs5592736	CID.09-117	9:117187806	TAGC/-	In	76	0.095	0.687	0.219	0.438	0.562	0.528	0.329	0.335	0.378	0.301	0.379
68	rs71953876	CID.09-135	9:135380181	CCAC/-	Del	63	0.418	0.458	0.124	0.647	0.353	0.400	0.352	0.305	0.330	0.319	0.469
69	rs10552811	CID.10-071	10:70563084	GC/-	Del	79	0.323	0.473	0.204	0.560	0.440	0.370	0.302	0.467	0.320	0.336	0.405
70	rs786663	CID.10-085	10:84613642	TA/-	Del	70	0.443	0.403	0.154	0.644	0.356	0.382	0.391	0.394	0.377	0.398	0.320
71	rs140576359	CID.10-118	10:117733203	CCGATATT/-	Del	71	0.313	0.542	0.144	0.585	0.415	0.413	0.409	0.412	0.414	0.373	0.429
72	rs10544160	CID.10-128	10:128080091	TA/-	Del	74	0.343	0.483	0.174	0.585	0.415	0.381	0.349	0.410	0.428	0.351	0.303
73	rs10583916	CID.11-006	11:5870839	TG/-	Del	94	0.404	0.481	0.115	0.644	0.356	0.408	0.460	0.369	0.314	0.372	0.362
74	rs113764748	CID.11-020	11:19528202	TAGC/-	Del	71	0.264	0.542	0.194	0.535	0.465	0.401	0.435	0.452	0.415	0.411	0.327
75	rs74933486	CID.11-083	11:83237025	CA/-	Del	86	0.333	0.517	0.149	0.592	0.408	0.401	0.458	0.458	0.313	0.461	0.437
76	rs35847449	CID.11-094	11:94191207	C/-	In	57	0.100	0.502	0.398	0.351	0.649	0.421	0.403	0.373	0.406	0.412	0.335
77	rs56221618	CID.11-111	11:111065696	AAG/-	Del	61	0.443	0.483	0.075	0.684	0.316	0.435	0.403	0.356	0.346	0.471	0.499
78	rs11394480	CID.11-135	11:134778510	G/-	In	47	0.179	0.403	0.418	0.381	0.619	0.369	0.465	0.349	0.463	0.301	0.408
79	rs10533848	CID.12-013	12:12861586	TC/-	Del	77	0.224	0.512	0.264	0.480	0.520	0.382	0.450	0.445	0.437	0.465	0.312
80	rs10689649	CID.12-090	12:89969919	ACTC/-	In	93	0.104	0.473	0.423	0.341	0.659	0.413	0.362	0.380	0.303	0.433	0.424
81	rs56392226	CID.13-040	13:40028153	AATGC/-	Del	81	0.383	0.498	0.119	0.632	0.368	0.409	0.365	0.412	0.424	0.442	0.359
82	rs59502417	CID.13-087	13:87305513	TGTC/-	Del	90	0.443	0.448	0.109	0.667	0.333	0.409	0.313	0.406	0.431	0.346	0.313
83	rs5812371	CID.14-056	14:56159616	ACATGTTCCA/-	In	69	0.095	0.458	0.448	0.323	0.677	0.419	0.367	0.343	0.310	0.427	0.307
84	rs3085775	CID.14-091	14:91332415	AGTG/-	In	87	0.174	0.498	0.328	0.423	0.577	0.386	0.407	0.437	0.474	0.368	0.329
85	rs67320356	CID.15-067	15:66925590	GAGTTT/-	Del	70	0.423	0.448	0.129	0.647	0.353	0.396	0.349	0.359	0.398	0.477	0.368
86	rs3995812	CID.15-098	15:97502107	TTAC/-	In	61	0.100	0.488	0.413	0.343	0.657	0.418	0.401	0.334	0.401	0.307	0.307

Таблица 1. Окончание

№	ID полиморфизма	Название в системе <i>ChipID</i>	Положение	Аллель	МА*	L**, п.н.	Частоты генотипов и аллелей в российской популяции (N=201), F ₀				MP	Частота минорного аллеля в основных мировых макропопуляциях, F ₀					
							In/In	In/Del	Del/Del	In		Del	EAS	EUR	AMR	SAS	AFR
87	rs34436424	CID.16-012	16:11820079	GC/-	Del	80	0.328	0.478	0.194	0.567	0.374	0.319	0.499	0.304	0.356	0.467	
88	rs3841767	CID.16-024	16:24136639	CC/-	Del	68	0.338	0.537	0.124	0.607	0.419	0.433	0.415	0.455	0.340	0.428	
89	rs34816330	CID.16-076	16:76041094	C/-	Del	60	0.289	0.562	0.149	0.570	0.422	0.330	0.485	0.424	0.433	0.421	
90	rs5821134	CID.17-057	17:566529684	ACTC/-	In	78	0.219	0.493	0.289	0.465	0.374	0.308	0.454	0.318	0.446	0.478	
91	rs141730340	CID.17-081	17:80625313	CCTGTGCAGCG/-	Del	95	0.408	0.498	0.095	0.657	0.423	0.419	0.311	0.304	0.348	0.363	
92	rs35452393	CID.18-028	18:27684881	AT/-	In	56	0.343	0.463	0.194	0.575	0.370	0.472	0.490	0.318	0.458	0.432	
93	rs33990768	CID.18-048	18:48155140	G/-	Del	58	0.244	0.517	0.239	0.502	0.384	0.362	0.481	0.375	0.410	0.441	
94	rs112877961	CID.19-029	19:29495432	C/-	Del	59	0.393	0.507	0.100	0.647	0.422	0.445	0.393	0.411	0.468	0.492	
95	rs35894782	CID.20-020	20:20384749	AATAG/-	Del	78	0.428	0.478	0.095	0.667	0.420	0.366	0.410	0.369	0.348	0.440	
96	rs73622051	CID.20-038	20:37610825	TTTA/-	Del	65	0.438	0.453	0.109	0.664	0.409	0.430	0.347	0.370	0.340	0.347	
97	rs34198404	CID.20-056	20:56104372	GTG/-	Del	76	0.338	0.478	0.184	0.577	0.376	0.446	0.451	0.336	0.362	0.478	
98	rs10583984	CID.21-039	21:39042013	AG/-	Del	85	0.368	0.507	0.124	0.622	0.409	0.415	0.384	0.441	0.318	0.315	
99	rs3282131	CID.21-044	21:43822529	ACAT/-	Del	67	0.328	0.498	0.174	0.577	0.386	0.457	0.370	0.344	0.389	0.466	
100	rs6028580	CID.22-033	22:33384984	G/-	In	76	0.169	0.488	0.343	0.413	0.384	0.358	0.401	0.411	0.424	0.425	
101	rs35997669	CID.X-005	X:4674479	C/-	Del	67	0.356	0.248	0.391	0.483	0.341	0.328	0.366	0.342	0.390	0.346	
102	rs57522085	CID.X-029	X:28920529	C/-	Del	75	0.510	—	0.490	0.510	0.490	—	—	—	—	—	
							Ж	0.208	0.495	0.297	0.455	0.545	—	—	—	—	
							Σ	0.446	0.252	0.297	0.575	0.425	0.335	0.336	0.391	0.348	0.315
103	rs202172626	CID.X-082	X:82244300	CTTA/-	Del	109	0.590	—	0.410	0.590	0.410	—	—	—	—	—	
							Ж	0.307	0.505	0.188	0.559	0.441	—	—	—	—	
							Σ	0.584	0.208	0.203	0.692	0.308	0.377	0.486	0.450	0.317	0.458
104	rs4991528	CID.X-094	X:94398625	C/-	In	80	0.680	—	0.320	0.680	0.320	—	—	—	—	—	
							Ж	0.495	0.416	0.089	0.703	0.297	—	—	—	—	
							Σ	0.257	0.262	0.475	0.391	0.609	0.404	0.419	0.438	0.444	0.311
105	rs35020323	CID.X-137	X:136803649	TATT/-	In	64	0.400	—	0.600	0.400	0.600	—	—	—	—	—	
							Ж	0.119	0.525	0.356	0.381	0.619	—	—	—	—	
							Σ	0.361	0.257	0.376	0.493	0.507	0.338	0.415	0.478	0.441	0.395
106	rs10689407	CID.X-140	I39879947	AT/-	Del	68	0.430	—	0.570	0.430	0.570	—	—	—	—	—	
							Ж	0.297	0.515	0.188	0.554	0.446	—	—	—	—	
							Σ	0.500	0.238	0.252	0.625	0.375	0.415	0.400	0.370	0.410	0.328
							М	0.610	—	0.390	0.610	0.390	—	—	—	—	
							Ж	0.400	0.480	0.120	0.640	0.360	—	—	—	—	

*МА — аллель, минорный во всех популяциях.

**L — длина ампликона; указана для ПЦР-продукта, содержащего аллель Del.

трофотометрически (5–500 нг/мкл), но не приводили к единым значениям.

Мультиплексная праймерная система. На стадии подбора ПЦР-праймеров основная задача состояла в минимизации размера ампликона. В табл. 1 приведены значения длин каждого локуса, среднее значение составило 72 п.н. В ходе экспериментальной оценки работоспособности праймеров и подбора условий ПЦР использовали несколько образцов свежeweделенной ДНК в концентрации 0.5 и 0.05 нг/мкл каждый. После достижения стабильной работы мультиплексной ПЦР проведено генотипирование образцов исследуемой выборки из российской популяции.

Аmplификация. ПЦР проводили на амплификаторе SpeedCycler (“AnalytikJena”, Германия) в 25 мкл смеси следующего состава: ПЦР-буфер с полимеразой HotTaqMulti, 4 ед. акт. (“Асфоген”, Россия), 5 мМ MgSO₄, 0.2 мМ каждого из dNTP (“Сибэнзим”, Россия), смесь праймеров, 200 пмоль универсального праймера Су5-ТСАТТGGATCT-САТТА, 0.05–500 нг геномной ДНК. Полимеразу активировали (2 мин при 95°C) и проводили на первом этапе 50 циклов амплификации (95°C – 20 с, 65°C – 30 с, 66°C – 30 с, 69°C – 40 с), затем на втором этапе проводили 40 циклов амплификации (95°C – 20 с, 56°C – 30 с, 72°C – 30 с).

Схема биочипа. Биочип изготовлен согласно схеме, приведенной на рис. 1а. Аллель-специфичные олигонуклеотидные зонды нанесены слева направо в порядке расположения в геноме соответствующих им InDel-полиморфизмов (от начала первой хромосомы до конца X-хромосомы). Каждому полиморфизму на биочипе соответствует пара вертикально расположенных ячеек: верхняя ячейка содержит зонд для аллеля с инсерцией, нижняя – с делецией. ПЦР-праймеры и ДНК-зонды произведены компанией “Lumiprobe” (Россия). Гидрогелевые биочипы изготавливали по оригинальной технологии ИМБ РАН в беспылевых помещениях с использованием пинового робота Qarray2 (“Genetix”, Великобритания), согласно ранее описанной методике [26].

Гибридизация. Гибридизационную камеру биочипа заполняли смесью следующего состава: 25% формамида, 5× SSPE, 50 об. % ПЦР-продукта. После инкубации (10 ч, 37°C) и отмывки (10 мин в 1× SSPE при комнатной температуре) биочипы промывали дистиллированной водой, высушивали сжатым воздухом, помещали в портативный анализатор “Пикодетект” (ООО “БИОЧИП-ИМБ”, Москва, Россия), флуоресценцию регистрировали с экспозицией 0.5–2 с в формате SPE. На рис. 1 приведены примеры таких флуоресцентных изображений четырех образцов ДНК (рис. 1б–1д). Изображения анализировали с ис-

пользованием программ ImaGel 2.0 и ImaGelStudio (ИМБ РАН). Результатом автоматической обработки каждого биочипа был текстовый файл с генотипом образца.

Статистический анализ. Частоты аллелей и соответствие равновесию Харди–Вайнберга оценивали при помощи онлайн-программы <https://genecalc.pl/hardy-weinberg-page>. Основные экспертные характеристики сформированной панели (MP, PD, CPD, PE, SPE) вычисляли по общеизвестным формулам в программе PowerStats v12.xls (“Promega Corp.”, США).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Стартовой задачей при создании инструмента для генетической идентификации личности, является выбор маркеров, которые обеспечат требуемый дискриминирующий потенциал всей системы. Добавление каждого маркера увеличивает мультиплексность ПЦР, что негативно сказывается на чувствительности и стабильности работы тест-системы. Чтобы минимизировать количество маркеров, обеспечивающих достижение цели, следует задать оптимальные параметры их первичного отбора.

Параметры отбора маркеров

Базовым параметром, определяющим дискриминирующий потенциал биаллельного полиморфизма, является частота минорного аллеля (MAF), идеальное значение MAF составляет 0.5. В связи с отсутствием в геноме нужного количества полиморфизмов, обладающих такими характеристиками, причем во всех мировых макропопуляциях, требуется найти компромиссный диапазон MAF, которому будет соответствовать некий пул, из которого можно будет провести дальнейший отбор по критериям второго порядка. Чем менее строго задано значение MAF, тем больший выбор маркеров возникает у исследователя, но тем меньшим итоговим дискриминирующим потенциалом будет обладать панель. В других аналогичных работах задавалось значение $MAF \geq 0.20$ [20, 27] и менее [28]. Мы попытались провести отбор маркеров, отвечающих более жестким критериям: $MAF \geq 0.30$.

При отборе маркеров-кандидатов мы использовали данные международного исследования *1000 Genomes Project* о частотах аллелей и генотипов в популяциях AFR ($n = 661$), AMR ($n = 347$), EUR ($n = 503$), SAS ($n = 489$) и EAS ($n = 504$). Частота минорного аллеля маркеров (табл. 1) в пяти рассматриваемых популяциях составила 0.301–0.499. Усредненная частота всех минорных аллелей в каждой популяции составила 0.394–0.402. Замысел планируемой работы по выбору InDel-полиморфизмов состоял в том, что при соблюде-

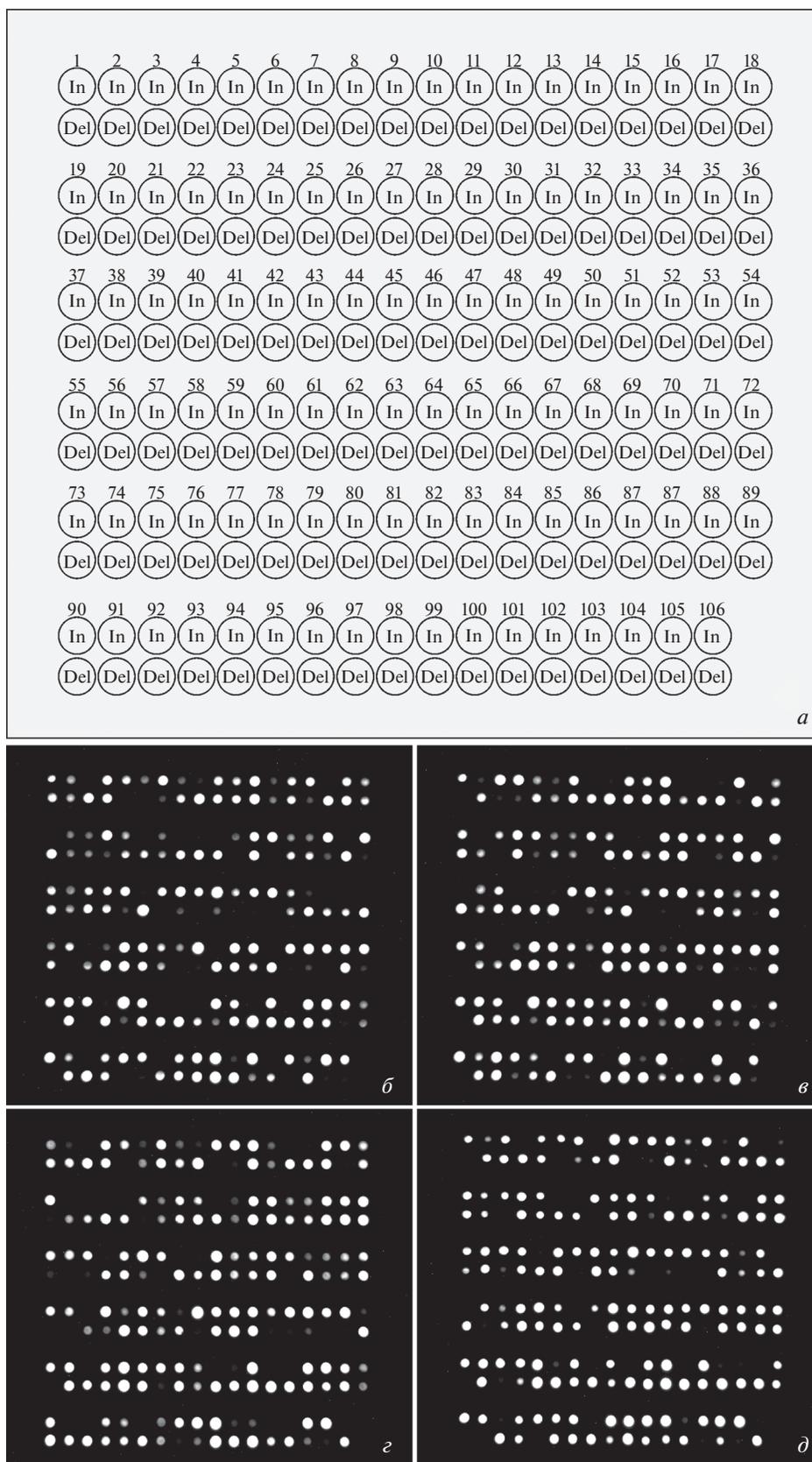


Рис. 1. Схема расположения ДНК-зондов на биологическом микрочипе *ChipID106* (а) и флуоресцентные изображения биочипа, полученные в результате генотипирования четырех неродственных образцов ДНК (б–д). Цифры над каждой парой ячеек соответствуют номеру полиморфизма из табл. 1. Обозначения внутри ячеек соответствуют аллелям полиморфизмов.

нии заданной частоты минорного аллеля во всех пяти макропопуляциях, дискриминирующий потенциал существенно не уменьшится и для более локальных субпопуляций, к которым потенциально может быть применен разрабатываемый метод. Это придаст методу универсальность в глобальном масштабе.

После формирования первичного пула из 12375 InDel, отобранных по критерию $MAF \geq 0.30$, в несколько этапов проведен отсев полиморфизмов, генотипирование которых посредством гибридационного анализа будет технически затруднительно, либо невозможно. В результате последовательного отбора по указанным критериям создана панель из 106 InDel-полиморфизмов.

Характеристики маркеров

Панель состоит из 106 InDel, 100 из которых расположены на хромосомах 1–22 и шесть – на X-хромосоме. Физическое расстояние между соседними маркерами варьирует от 3 до 113.8 млн.п.н., среднее расстояние 24 млн.п.н., что исключает возможное сцепленное наследование маркеров. Наличие шести маркеров на X-хромосоме дает дополнительные возможности определения родства в некоторых сложных случаях [29].

Размер InDel-полиморфизмов варьирует от 1 до 18 п.н. (среднее значение 4.2, медианное 3). Длина амплифицируемых локусов составила от 44 до 109 п.н., среднее и медианное значения совпадают и составляют 72 п.н. для локусов, содержащих аллель Del, и 76 п.н. для локусов, содержащих аллель In. Минимальная длина фрагмента, который можно получить в ПЦР с последующей гибридизацией, зависит от нуклеотидного состава локуса и в идеальном случае составляет 40–50 п.н. Нам удалось добиться средней длины локуса 72 п.н., что обеспечивает методу фундаментальное преимущество при исследовании деградированной ДНК.

Популяционные данные и дискриминирующие характеристики метода

Несмотря на то, что ряд авторов полагают достаточным опубликовать перечень маркеров, отобранных по определенным критериям *in silico*, мы считаем необходимым их верифицировать, так как результаты современных полногеномных исследований нередко содержат ошибки. Целесообразность включения в панель каждого из выбранных полиморфизмов подтверждена путем экспериментальной оценки их генетического разнообразия в российской популяции.

В исследованных в настоящей работе образцах ($n = 201$), взятых от жителей Москвы, Москов-

ской области и Санкт-Петербурга, усредненная частота минорных аллелей составила 0.385, средняя наблюдаемая гетерозиготность, $H_o = 0.471$ (0.338–0.687). У двух полиморфизмов частота минорного аллеля оказалась незначительно меньше заданного значения: rs56281469 – 0.254 и rs72273695 – 0.266 (табл. 1). Результаты проверки сформированной панели по частотам аллелей в российской популяции свидетельствуют о соответствии выбранных полиморфизмов критериям, заданным при отборе.

Опираясь на полученные данные, мы определили значение MP для каждого полиморфизма (табл. 1), а также оценили основные дискриминирующие характеристики всей панели для российской популяции: совокупная вероятность случайного совпадения, CMP , составила 1.89×10^{-43} , потенциал исключения отцовства $CPE = 0.99999999063$.

Сравнивая полученный показатель вероятности случайного совпадения, $CMP = 1.89 \times 10^{-43}$, с показателем у наиболее известных STR-систем: 13-локусный CODIS – 2.34×10^{-15} ; Identifiler – 5.93×10^{-18} ; PowerPlex16 – 2.43×10^{-18} ; NGM – 1.12×10^{-19} ; New FBI core – 6.28×10^{-30} [2], можно видеть, что предлагаемая панель обладает на 13 порядков более высоким дискриминирующим потенциалом, чем New FBI core – наиболее информативная панель, применяемая на практике. Эта избыточность дискриминирующего потенциала обеспечивает не только преимущество при экспертизах сложных образцов, для которых получены неполные профили, но и более высокую достоверность при определении кровного родства.

Отдельный важный аспект для любой системы генетической идентификации – информация о том, с какой частотой встречаются ее аллели в популяции, к которой относится исследуемая ДНК. Это необходимо для предоставления в суд экспертиз, содержащих обоснованные вероятностные расчеты. К примеру, если обнаруженный аллель встречается с частотой 0.01 в европейской популяции и 0.99 в азиатской, то без информации о популяционной принадлежности ее носителя при расчете будет допущена грубая ошибка. В свете этого, оптимальным решением представляется создание универсальной панели с полиморфизмами, не имеющими значимой популяционной специфичности в распределении аллелей. Этот подход в сочетании с избыточностью дискриминирующего потенциала позволит полностью исключить критическую недостоверность в расчетах вероятности при проведении экспертизы.

*Практические характеристики
метода генотипирования*

Протокол исследования содержит четыре этапа: выделение ДНК, амплификация, гибридизация и регистрация результатов. Выделение ДНК может быть выполнено любым подходящим для биоматериала способом. Метод протестирован в широком диапазоне концентраций и, в отличие от STR-типирования на капиллярном электрофоретическом оборудовании, не требует приведения к допустимым значениям.

Особые усилия были приложены к созданию условий внедрения метода в практику и снижению риска контаминации. Ранее тест-системы, созданные на базе биочип-технологии, обладали рядом недостатков, ограничивающих их широкое применение. В частности, в связи с необходимостью получения одноцепочечного ампликона для гибридизации, использовали двухстадийную “гнездную” ПЦР с переносом ампликонов из симметричной стадии ПЦР в асимметричную, что требовало отдельной лабораторной зоны и создавало серьезный риск контаминации. Благодаря применению ранее опубликованной модификации (LATE)-PCR [30], на стадии амплификации удалось объединить симметричный и асимметричный этап ПЦР.

Кроме того, ограничения в мультиплексности создавали необходимость в постановке нескольких ПЦР с разными пулами праймеров для одного образца. Это не только повышало трудоемкость, но и снижало такой критически важный параметр в криминалистике, как аналитическую чувствительность. Известны публикации, в которых заявляется успешная амплификация сотен локусов, однако это декларируется для высоких концентраций геномной ДНК, которые нечасто встречаются в экспертной практике [31]. Нам удалось подобрать условия амплификации всех 106 локусов в одной мультиплексной ПЦР с сохранением чувствительности, значимой для экспертной работы.

Таким образом, протокол исследования был максимально упрощен, сведен в схему “один образец – одна пробирка – один биочип” и избавлен от промежуточных стадий переноса продукта.

Помимо дискриминирующих возможностей, к важнейшим качествам методов практической криминалистики относятся способность генотипирования деградированных образцов и чувствительность. Залогом успеха при работе с фрагментированной ДНК является минимизация ампликонов при проектировании праймеров. В нашем случае средний размер ампликона доведен до 72 п.н., однако экспериментальная проверка на деградированных образцах в план этой работы не входила.

Что касается чувствительности, то ее типичный нижний предел для современных систем генетической идентификации составляет 50–100 пг ДНК на ПЦР (что соответствует 8–16 геном-эквивалентам). В задачи настоящей работы не входили специальные мероприятия по повышению чувствительности метода, однако, учитывая критическую практическую важность этого параметра, мы провели его предварительную экспериментальную оценку. С этой целью проведено девятикратное генотипирование одного из образцов (50 пг на ПЦР). В среднем наблюдалось выпадение 9 локусов (от 5 до 16), причем не случайных, а вполне определенных, что, вероятнее всего, связано с эффективностью отжига праймеров и может быть исправлено в дальнейшем. Тем не менее, пул из 90–101 сработавших локусов с избытком обеспечивает достаточную дискриминирующую силу не только в идентифицирующих генетических экспертизах, но и для определения отцовства в трио и дуэте.

Графические примеры результатов генотипирования приведены на рис. 1б–1д. При внимательном изучении рисунков могут возникнуть несколько вопросов. Во-первых, некоторые пары ячеек в гетерозиготном состоянии светятся слабее других. Это связано с тем, что не все локусы амплифицируются с одинаковой эффективностью. Однако надежность генотипирования обеспечивается высокой чувствительностью камеры и ее широким динамическим диапазоном (0–65528). Это позволяет одинаково успешно анализировать как слабые, так и сильные пары. Во-вторых, в некоторых парах, очевидно находящихся в гетерозиготном состоянии, можно видеть неравнозначность по яркости верхней (In) и нижней (Del) ячеек. При визуальной интерпретации результатов это могло бы трактоваться неоднозначно, однако обработка выполняется автоматически путем оценки дискриминации (дискриминационного отношения сигнала) в каждой паре In-Del. Дискриминация представляет собой частное от деления интенсивности сигналов от ячеек с ДНК-зондами, комплементарными In- и Del-аллелю, соответственно, $D = \frac{I_{in}}{I_{del}}$.

Этот показатель позволяет разделить генотипы биаллельных маркеров на три пула $D_{In/In} \gg 1$, $D_{In/Del} \approx 1$ и $D_{Del/Del} \ll 1$. Для достоверного генотипирования необходимо, чтобы эти пулы не только не пересекались, но и имели достаточно широкий коридор значений между собой.

На рис. 2а–2г приведены примеры такой обработки сигналов. Для удобства восприятия результаты расположены по мере возрастания значения дискриминационного отношения сигналов. Кроме того, при такой подаче легко найти экстремальные значения в каждом пуле, чтобы по ним

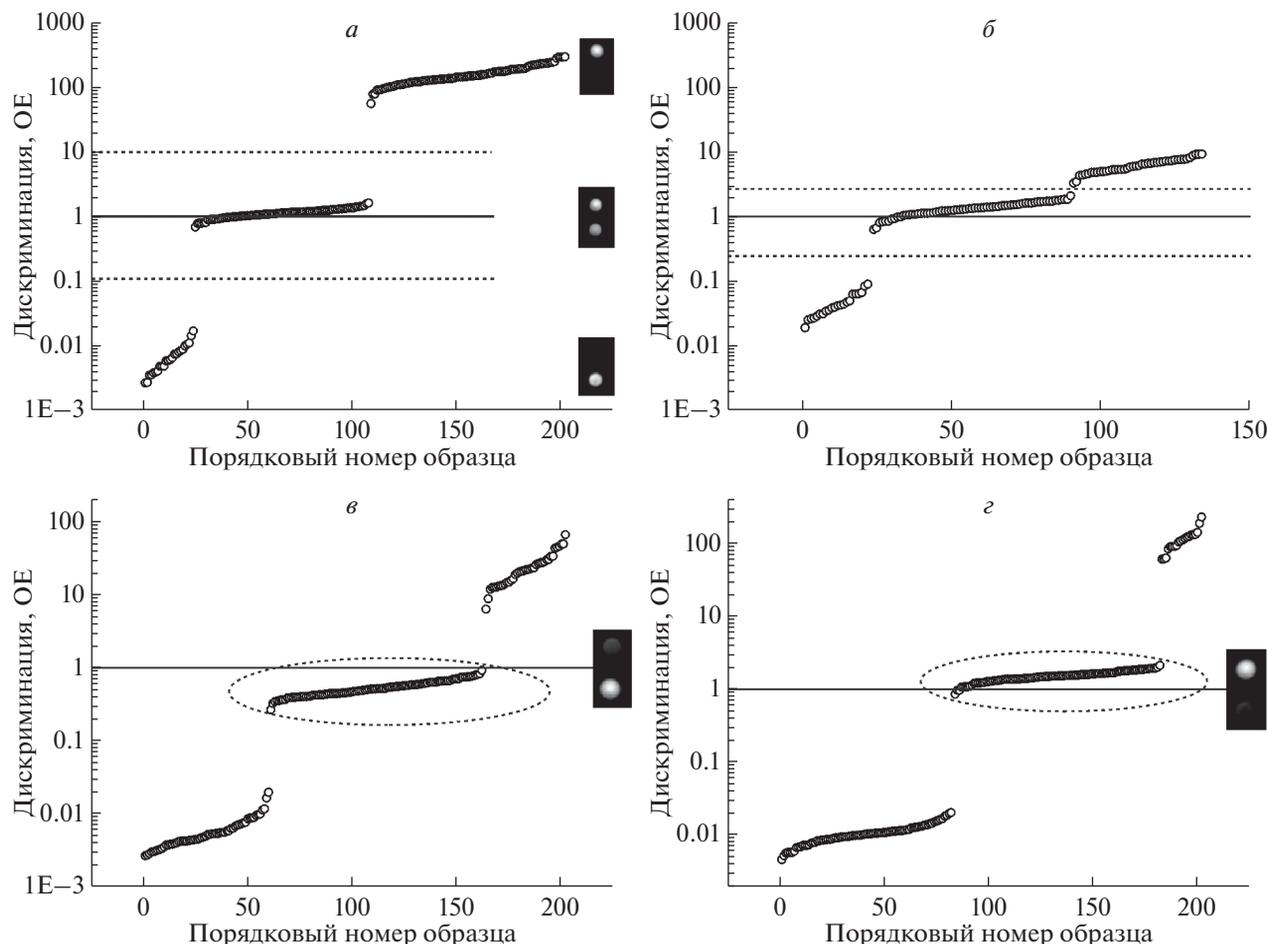


Рис. 2. Примеры распределения дискриминационных значений D , полученных при обработке гибридационных изображений. *a, в, г* – Графики, полученные при генотипировании образцов коллекции и относящиеся к маркерам CID.06-144, CID.03-031 и CID.15-098 соответственно. График, приведенный на части *б*, относится к однонуклеотидному полиморфизму rs4959270, использованы данные [30]. Пунктирные прямые на *a* и *б* – пороговое значение, используемое для автоматизированного генотипирования в программе ImageWare. Пунктирные овалы на *в* и *г* – смещение значений D для гетерозиготных пулов относительно 1. По оси абсцисс отложены порядковые номера образцов при нумерации их по мере возрастания значения дискриминации; OE – относительные единицы.

определить границы коридора и правильно установить пороговые значения, используемые программой ImageWare для автоматической обработки результатов и формирования отчета о генотипе. На рис. 2*a* приведено образцовое распределение сигналов: гетерозиготный пул расположен в области единицы, а полярные пулы In/In и Del/Del отстоят от него более чем на полтора порядка. Это позволяет с существенным запасом установить верхний и нижний порог значений (пунктирные линии на рис. 2*a*).

Чтобы проиллюстрировать преимущества In-Del-полиморфизмов перед SNP и обосновать наш выбор в пользу первых, на рис. 2*б* приведен пример распределения дискриминационных значений для однонуклеотидного полиморфизма rs4959270. Как видно из графика, коридор между пулами существенно уже, что повышает

риск ошибок при генотипировании. Это обусловлено сравнительно большей близостью температур плавления совершенных и несовершенных дуплексов соответствующих ДНК-зондов у SNP, что создает трудности при дизайне ДНК-зондов. Безусловно, если для решения поставленной задачи необходимо генотипировать SNP (как в большинстве практических решений на основе биочип-технологии), то приходится находить оптимальный дизайн олигонуклеотидов. Однако в нашем случае SNP и InDel равнозначны по критически важным для экспертной практики свойствам (перечислены во введении), поэтому мы сделали выбор в пользу InDel.

Графики на рис. 2*в* и 2*г* указывают на возможность визуальных различий в яркости ячеек In и Del при гетерозиготном генотипе, что не влияет на результат. При этом верное генотипирование

обеспечивается смещением пороговых значений в соответствующую сторону при применении автоматизированного подхода. В дальнейшем такие пары ДНК-зондов могут быть скорректированы путем соответствующего изменения длины олигонуклеотида, либо его концентрации в ячейке биочипа.

ЗАКЛЮЧЕНИЕ

На настоящий момент генетическая идентификация личности целиком базируется на исследовании STR: на этот тип маркеров ориентировано и оборудование, и базы данных, и методика интерпретации результатов. Однако, как упомянуто во введении, стратегически STR проигрывают SNP и InDel по протяженности и мутабельности, и это фундаментальный проигрыш, его нельзя преодолеть техническими средствами. Запрос на метод идентификации, использующий короткие биаллельные маркеры, возник давно, но не был реализован, в первую очередь, в связи с отсутствием подходящего инструмента для генотипирования. В настоящей работе нам удалось продвинуться в создании практического инструмента для решения задач судебно-криминалистической экспертизы на новом уровне. Создана универсальная панель биаллельных маркеров, которая может стать основой для формирования национальной базы генетической информации не только в России, но и в любой стране мира. Выбор метода в пользу отечественной технологии биологических микрочипов низкой плотности в наибольшей степени соответствует поставленным задачам, так как обеспечивает размещение необходимого количества зондов и не требует сложных протоколов пробоподготовки, дорогостоящего оборудования и расходных материалов, что создает предпосылки для его успешного внедрения.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 года и ее последующим изменениям или сопоставимым нормам этики. От каждого включенного в исследование участника получено информированное добровольное согласие.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Животовский Л.А. (2001) ДНК в суде. *Химия и жизнь*. № 12, 23–27.
2. Ge J., Eisenberg A., Budowle B. (2012) Developing criteria and data to determine best options for expanding the core CODIS loci. *Investigative Genetics*. 3(1), 1–14.

3. Wyman A.R., White R. (1980) A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA*. 77(11), 6754–6758.
4. Tamaki K., Jeffreys A.J. (2005) Human tandem repeat sequences in forensic DNA typing. *Legal Medicine*. 7(4), 244–250.
5. Aşcıoğlu F., Oguz-Savran F., Ozbek U. (2004). Mutation rate at commonly used forensic STR loci: paternity testing experience. *Disease Markers*. 20(6), 313–315.
6. Hohoff C., Dewa K., Sibbing U., Hoppe K., Forster P., Brinkmann B. (2007) Y-chromosomal microsatellite mutation rates in a population sample from Northwestern Germany. *Int. J. Legal Med.* 121(5), 359–363.
7. Syvänen A.C., Sajantila A., Lukka M. (1993) Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am. J. Hum. Genet.* 52(1), 46.
8. Reich D.E., Schaffner S.F., Daly M.J., McVean G., Mullikin J.C., Higgins J.M., Richter D.J., Lander E.S., Altshuler D. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat. Genet.* 32(1), 135–142.
9. Lee H.J., Lee J.W., Jeong S.J., Park M. (2017) How many single nucleotide polymorphisms (SNPs) are needed to replace short tandem repeats (STRs) in forensic applications? *Int. J. Legal Med.* 131(5), 1203–1210.
10. Lee H.Y., Park M.J., Yoo J.E., Chung U., Han G.R., Shin K.J. (2005) Selection of twenty-four highly informative SNP markers for human identification and paternity analysis in Koreans. *Forensic Sci. Int.* 148(2–3), 107–112.
11. Freire-Aradas A., Fondevila M., Kriegel A.K., Phillips C., Gill P., Prieto L., Schneider P.M., Carracedo A., Lareu M.V. (2012) A new SNP assay for identification of highly degraded human DNA. *Forensic Sci. Int.: Genetics*. 6(3), 341–349.
12. Zeng Z., Wang L., Feng Q., Zhang L., Lee L., Wang L., Yue Y., Fang Y., Yang W., Qiu H., Dong Z. (2012) Evaluation of 96 SNPs in 14 populations for worldwide individual identification. *J. Forensic Sci.* 57(4), 1031–1035.
13. Mo S.K., Ren Z.L., Yang Y.R., Liu Y.C., Zhang J.J., Wu H.J., Li Z., Bo X.C., Wang S.Q., Yan J.W., Ni M. (2018) A 472-SNP panel for pairwise kinship testing of second-degree relatives. *Forensic Sci. Int.: Genet.* 34, 178–185.
14. Krjtskov K., Viltrop T., Palta P., Metspalu E., Tamm E., Suvi S., Sak K., Merilo A., Sork H., Teek R., Nikopenius T., Kivisild T., Metspalu A. (2009) Evaluation of the 124-plex SNP typing microarray for forensic testing. *Forensic science international*. *Genetics*. 4(1), 43–48.
15. Musgrave-Brown E., Ballard D., Balogh K., Bender K., Berger B., Bogus M., Børsting C., Brion M., Fondevila M., Harrison C., Oguzturun C., Parson W., Phillips C., Proff C., Ramos-Luis E., Sanchez J.J., Sánchez Diz P., Sobrino Rey B., Stradmann-Bellinghausen B., Thacker C., Carracedo A., Morling N., Scheithauer R., Schnei-

- der P.M., Syndercombe Court D. (2007) Forensic validation of the SNPforID 52-plex assay. *Forensic Sci. Int. Genet.* **1**(2), 186–190.
16. Fondevila M., Børsting C., Phillips C., de la Puente M., Consortium E.N, Carracedo A., Morling N., Lareu M.V. (2017) Forensic SNP genotyping with SNaPshot: technical considerations for the development and optimization of multiplexed SNP assays. *Forensic Sci. Rev.* **29**(1), 57–76.
 17. Liu J., Du W., Jiang L., Liu C., Chen L., Zheng Y., Hou Y., Liu C., Wang Z. (2022) Development and validation of a forensic multiplex InDel assay: the AGCU InDel 60 kit. *Electrophoresis.* **43**(18–19), 1871–1881.
 18. Pereira R., Phillips C., Alves C., Amorim A., Carracedo A., Gusmão L. (2009) Insertion/deletion polymorphisms: a multiplex assay and forensic applications. *Forensic Sci. Int.: Genet. Suppl. Ser.* **2**(1), 513–515.
 19. Freitas N.S., Resque R.L., Ribeiro-Rodrigues E.M., Guerreiro J.F., Santos N.P., Ribeiro-dos-Santos A., Santos S. (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. *Int. J. Legal Med.* **124**(6), 589–593.
 20. LaRue B.L., Lagacé R., Chang C.W., Holt A., Hennessy L., Ge J., King J.L., Chakraborty R., Budowle B. (2014) Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification. *Legal Med.* **16**(1), 26–32.
 21. Sun K., Yun L., Zhang C., Shao C., Gao T., Zhao Z., Hou Y., Xie J., Tang Q. (2019) Evaluation of 12 Multi-InDel markers for forensic ancestry prediction in Asian populations. *Forensic Sci. Int. Genet.* **43**, 102155.
 22. Huang Y., Liu C., Xiao C., Chen X. Yi S., Huang D. (2020) Development of a new 32-plex InDels panel for forensic purpose. *Forensic Sci. Int. Genet.* **44**, 102171.
 23. Zhang W., Jin X., Wang Y., Kong T., Cui W., Chen C., Guo Y., Zhu B. (2020) Genetic polymorphisms and forensic efficiencies of a set of novel autosomal InDel markers in a Chinese Mongolian group. *Biomed. Res. Int.* **2020**, 3925189.
 24. Lan Q., Zhao C., Wei C., Xu H., Shen C., Zhu B. (2022) Genetic insights and evaluation of forensic features in Mongolian and Ewenki groups using the InDel variations. *Front. Biosci. (Landmark Ed.)*. **27**(2), 67.
 25. LaRue B.L., Ge J., King J.L., Budowle B. (2012) A validation study of the Qiagen Investigator DIPplex® kit: an INDEL-based assay for human identification. *Int. J. Legal. Med.* **126**(4), 533–540.
 26. Fesenko D.O., Chudinov A.V., Surzhikov S.A., Zasedatelev A.S. (2014) Biochip for genotyping SNPs defining core Y-chromosome haplogroups in Russian population groups. *BioChip J.* **8**, 171–178.
 27. Li C., Zhang S., Li L., Chen J., Liu Y. (2012) Selection of 29 highly informative InDel markers for human identification and paternity analysis in Chinese Han population by the SNPlex genotyping system. *Mol. Biol. Reports.* **39**(3), 3143–3152.
 28. Huang J., Luo H., Wei W., Hou Y. (2014) A novel method for the analysis of 20 multi-I ndel polymorphisms and its forensic application. *Electrophoresis.* **35**(4), 487–493.
 29. Степанов В.А., Вагайцева К.В., Харьков В.Н., Чердниченко А.А., Бочарова А.В. (2016) Панель однонуклеотидных сцепленных с X-хромосомой полиморфных маркеров для ДНК-идентификации (XSNPid) на основе мультиплексного генотипирования с использованием методов многофокусной ПЦР и масс-спектрометрии MALDI-TOF. *Молекуляр. биология.* **50**(3), 445–456.
 30. Фесенко Д.О., Ивановский И.Д., Иванов П.Л., Земскова Е.Ю., Агапитова А.С., Поляков С.А., Фесенко О.Е., Филиппова М.А., Заседателев А.С. (2022) Биочип для генотипирования полиморфизмов ассоциированных с цветом глаз, волос, кожи, группой крови, половой принадлежностью, основной гаплогруппой Y-хромосомы и результаты исследования славянской популяции с его помощью. *Молекуляр. биология.* **56**(5), 860–880.
 31. Křjutskov K., Andreson R., Mägi R., Nikopensius T., Khrunin A., Mihailov E., Tammekivi V., Sork H., Remm M., Metspalu A. (2008) Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. *Nucl. Acids Res.* **36**(12), e75.

Universal Panel of Insertion–Deletion Polymorphism for Human Genetic Identification and a Biochip-Based Kit *ChipID106* for This Purpose

D. O. Fesenko^{1, *}, I. D. Ivanovsky², P. L. Ivanov³, E. Yu. Zemskova³, S. A. Polyakov¹, O. E. Fesenko⁴, M. A. Filippova¹, and A. S. Zasedatelev¹

¹Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

²DNA Research Center, LLC, Hkimki, 141402 Russia

³Russian Center of Forensic Medical Expertise, Ministry of Health of the Russian Federation, Moscow, 125284 Russia

⁴Research Institute of Physics, Southern Federal University, Rostov-on-Don, 344090 Russia

*e-mail: deferos@yandex.ru

The paper proposes a panel of 106 insertion–deletion (InDel) polymorphisms and practical testing of the method of their genotyping on biochips. Such fundamental properties of InDel markers as short length and low mutation rate provide them with significant advantages in expert practice over widely used tandem repeats (STRs).

In this work, we analyzed the allele distribution of InDel polymorphisms in the five largest world populations (European, East Asian, South Asian, African and American), and selected markers that meet the following criteria: minor allele frequency of more than 0.30, physical distance between markers more than 3 million bp, the absence of polymorphisms, tandem repeats and palindromes in the flanking sequences, the proximity of the AT/GC ratio to 1. As a result, a panel of 106 polymorphisms was formed with an average frequency of the minor allele for all five populations of 0.396. For panel genotyping, a method was developed that included one-step multiplex PCR followed by hybridization on a biological microarray. The average amplicon length was 72 bp. On a sample of 201 residents of Moscow and St. Petersburg, the main characteristics of the formed panel were determined: the probability of a random match (MP) 1.89×10^{-43} , the paternity exclusion potential (CPE) 0.99999999063. The proposed method is an alternative to molecular genetic identification of a person based on variations in the lengths of STR loci.

Keywords: human identification, microarray, forensics, InDel polymorphisms