

УДК 575.22

# Генетическое разнообразие и структура неравновесия по сцеплению гена *MTHFR* в популяциях Северной Евразии

Е. А. Трифонова<sup>1</sup>, Е. Р. Еремина<sup>2</sup>, Ф. Д. Урнов<sup>3</sup>, В. А. Степанов<sup>1,4,5\*</sup><sup>1</sup>Научно-исследовательский институт медицинской генетики СО РАМН, 634050, Томск, Набережная реки Ушайки, 10<sup>2</sup>Бурятский государственный университет Министерства образования и науки РФ, 670000, Улан-Удэ, ул. Ранжурова, 6а<sup>3</sup>University of California, Berkley, 313 Richmond, CA 94720, USA<sup>4</sup>Томский государственный университет, 634050, Томск, просп. Ленина, 36<sup>5</sup>ООО «Геномная диагностика», 634055, Томск, просп. Академический, 17

\*E-mail: vadim.stefpanov@medgenetics.ru

Поступила в редакцию 13.12.2011 г.

**РЕФЕРАТ** Исследована структура гаплотипов и неравновесия по сцеплению (LD) в локусе метилентетрагидрофолатредуктазы (*MTHFR*) в девяти этнотерриториальных выборках Северной Евразии и популяциях из международного проекта НарМар. Полученные данные свидетельствуют о том, что архитектура LD в геноме человека в значительной степени определяется эволюционной историей популяций, тем не менее, результаты гаплотипического и филогенетического анализа указывают на возможность существования общего, «древнего» механизма формирования некоторых паттернов LD. Вариабельность структуры LD и уровень гаплотипического разнообразия гена *MTHFR* в исследованных выборках обуславливают определенный набор tagSNPs (tagging single nucleotide polymorphisms; полиморфизмы, аллельные варианты которых маркируют гаплотипические блоки) с установленной прогностической значимостью для каждой популяции. Полученные в настоящей работе данные представляют, по нашему мнению, значительный интерес в понимании нескольких генетических феноменов: ассоциации межпопуляционных различий в характере LD со структурой наследственной предрасположенности к многофакторным заболеваниям; функциональной значимости и плейотропного «поля действия» гена *MTHFR*. Суммируя результаты проведенного исследования, можно заключить, что анализ генетической вариабельности с акцентом на структуру LD в популяциях человека является мощным инструментом, способным внести большой вклад в такие отрасли медико-биологической науки, как эволюционная биология человека, функциональная геномика, генетика многофакторных заболеваний и фармакогеномика.

**КЛЮЧЕВЫЕ СЛОВА** геном, неравновесие по сцеплению, популяции Северной Евразии, метилентетрагидрофолатредуктаза, гаплотип.

**СПИСОК СОКРАЩЕНИЙ** МФЗ – многофакторные заболевания; LD – неравновесие по сцеплению; *MTHFR* – метилентетрагидрофолатредуктаза; SNP – однонуклеотидная замена; ГЦ – гомоцистеин.

## ВВЕДЕНИЕ

Генетическая вариабельность составляет основу фенотипической изменчивости человека и имеет огромное значение для объяснения индивидуальных различий в подверженности к многофакторным заболеваниям (МФЗ) и определения метаболических путей, вовлеченных в прогрессирование патологических процессов. Наиболее распространенным типом вариабельности генома являются однонуклеотидные замены (SNP). Усилиями международного консорциума по SNP к настоящему моменту в геноме человека выявлено около 10 млн SNPs с плотностью при-

близительно 1 полиморфизм на 300 п.н. [1]. Каждый новый аллель полиморфного варианта возникает на фоне уже существующего гаплотипа, с аллелями, составляющими который, изначально ассоциирован предковый вариант данного маркера. Новые гаплотипы формируются путем накопления новых мутаций и рекомбинации. Совместное наследование аллелей в гаплотипе на популяционном уровне проявляется как неравновесие по сцеплению (LD).

Архитектура LD в геноме человека в настоящее время является предметом оживленных дискуссий и интенсивных исследований [2–7]. Ряд работ по-

казывает, что в геноме можно выделить блоки сцепленных сайтов, не демонстрирующие свидетельств значительной рекомбинации в истории существования нашего вида, отделяемые участками с более интенсивным темпом рекомбинации, так называемыми «горячими точками» [8, 9]. Характер LD в современных популяциях человека формируется в результате комплексного эволюционного процесса, который включает как демографическую историю популяций (изменения эффективной численности, характер подразделенности, миграции), так и генспецифические факторы, например темп мутирования, рекомбинации, давление отбора. Анализ структуры LD позволяет реконструировать демографическую историю современных популяций и занимает центральное место при картировании генов МФЗ [10].

Наряду с полногеномными паттернами LD, изученными в современных широкогеномных исследованиях [11–15], существенный интерес представляет структура LD в отдельных функционально значимых участках генома, в частности в локусах генов, ассоциированных с распространенными болезнями человека. Важность анализа гаплотипической структуры этих областей генома объясняется, с одной стороны, необходимостью выявления функционально значимых вариантов этих генов, обеспечивающих их участие в общей компоненте наследственной подверженности к МФЗ, а с другой – значимостью оценки эволюционно-генетических механизмов генерации генетического разнообразия в этих локусах генома, формировавшихся, вероятно, с учетом действия естественного отбора, генетического дрейфа и миграций и сопоставления генетико-демографических сценариев, полученных при анализе тонкой структуры генов-кандидатов МФЗ с таковыми по данным условно-нейтральных генетических систем и данными полногеномных исследований.

В представленной работе в качестве локуса для изучения структуры LD в популяциях различного этнического происхождения был выбран ген метилентетрагидрофолатредуктазы (*MTHFR*), полиморфные варианты которого, по результатам многочисленных исследований, ассоциированы с развитием многих многофакторных состояний (сердечно-сосудистые, онкологические заболевания, дефекты нервной трубки, осложненное течение беременности и ряд других патологических процессов). Фермент метилентетрагидрофолатредуктаза катализирует единственную внутриклеточную реакцию образования 5-метилтетрагидрофолата, необходимого для восстановления гомоцистеина (ГЦ) до метионина. Снижение активности этого фермента, часто обусловленное мутациями в гене *MTHFR*, приводит к накоплению ГЦ и развитию умеренной гипергомоцистеинемии.

Роли генетической вариабельности локуса *MTHFR* в развитии многих социально-значимых заболеваний посвящено огромное количество работ, результаты которых указывают на небольшой вклад отдельных SNPs гена *MTHFR* в структуру наследственной предрасположенности к МФЗ и зачастую противоречивы для многих этнических групп. Поскольку одной из потенциально наиболее продуктивных стратегий выявления генетических вариантов, лежащих в основе подверженности к МФЗ, считается анализ структуры LD в области генов-кандидатов и обнаружение связанных с болезнью гаплотипов и их tagSNPs [3, 16, 17], весьма актуальной представляется оценка генетической вариабельности генов-кандидатов на уровне SNP и гаплотипов в различных популяционных выборках.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

### Популяции

Данные, представленные в работе, получены при молекулярно-генетическом анализе ДНК индивидов из девяти популяционных выборок, проживающих в различных регионах Северной Евразии и принадлежащих к семи этническим группам (табл. 1). Общая численность суммарной выборки составила 837 человек. Исследованные популяции относятся к двум основным расовым типам населения Евразии (европеоидному и монголоидному) и к четырем лингвистическим семьям (алтайской, палеоазиатской, индоевропейской и уральской). В обследование вошли индивиды неметисированные по крайней мере в трех поколениях. Национальность, родословная и в некоторых случаях принадлежность к родоплеменной группе устанавливались на основании индивидуальных опросников. Выборка тувинцев была собрана в Республике Тува (поселок Бай-Тайга). В Республике Бурятия было обследовано две популяции: буряты из г. Улан-Удэ и пос. Хуромша. Выборка якутов собрана в пос. Дюпся, расположенном на востоке Усть-Алданского улуса Якутии. Среднеазиатские популяции представлены киргизами. Две выборки киргизов представляют население севера (пос. Кегеты, пос. Талды-Су) и юга (г. Ош) Республики Киргизия и относятся к разным родоплеменным группам. Выборка русских представлена жителями г. Томска. Популяция кетов собрана в поселке Келлог Туруханского района Красноярского края, а выборка хантов – в поселке Рускинской Ханты-Мансийского автономного округа.

Кроме того, в качестве объекта для популяционных сравнений были использованы данные по популяциям европеоидов (жители штата Юта, США), китайцев (жители г. Пекин, Китай), японцев (жители

Таблица 1. Лингвистические и антропологические характеристики популяций

Этнос	Популяция, населенный пункт	N	Локализация	Языковая принадлежность, семья/ группа	Расовый и антропологический тип
Тувинцы	пос. Бай-Тайга	134	Республика Тува	Алтайская/ тюркская	Монголоидный (центральноазиатский)
Буряты	г. Улан-Удэ, пос. Хуромша	60 60	Республика Бурятия	Алтайская/ монгольская	Монголоидный (центральноазиатский)
Якуты	пос. Дюпся	81	Республика Саха	Алтайская/ тюркская	Монголоидный (центральноазиатский)
Киргизы	Север (пос. Кегеты, пос. Талды-Су), юг (г. Ош)	85 111	Республика Киргизия	Алтайская/ тюркская	Монголоидный (южносибирский)
Кеты	пос. Келлог	38	Красноярский край, Туруханский район	Палеоазиатская/ кетская	Монголоидный (североазиатский)
Ханты	пос. Рускинской	142	Сургутский район ХМАО	Уральская/ финно-угорская	Уральский (переходный)
Русские	г. Томск	126	Томская область	Индоевропейская/ славянская	Европеоидный (восточноевропейский)

г. Токио, Япония) и йоруба (население Ибадана, Нигерия), полученные в ходе реализации проекта Нар-Мар [11].

### Полиморфизмы

В качестве маркеров для изучения паттернов LD были выбраны следующие 12 SNPs гена *MTHFR*: rs3753588, rs2066470, rs17037397, rs7533315, rs4846052, rs1801133 (C677T), rs6541003, rs2066462, rs1801131 (A1298C), rs17375901, rs2274976 (G1793A) и rs1537516 (рис. 1). В табл. 2 представлена краткая характеристика изученных локусов гена *MTHFR*. Десять из 12 SNPs были результатом транзиций (3 A→G и 7 C→T), два – трансверсий (A→C). Выбранные полиморфные варианты относительно равномерно охватывают все участки гена (экзоны, интроны, 3'-нетранслируемые участки), частота минорного аллеля в большинстве локусов составляет не менее 5% (по данным базы NCBI). Генотипирование проводили согласно протоколам, описанным ранее [18–20].

### Методы статистической обработки результатов

Статистическую обработку результатов проводили с помощью пакетов статистических программ «Statistica 7.0», «ARLEQUIN» и «Haploview 4.0». Характер распределения полученных данных определяли с использованием критерия Колмогорова–Смирнова, частоты гаплотипов – с помощью EM-алгоритма. LD между парами SNPs оценивали с помощью коэффициента  $D'$ , предложенного Левонтином, и коэффициента корреляции  $r^2$  Пирсона. Блочную структуру определяли с использованием алгоритма «Solid spine of LD» [21], предусмотренного программным обеспечением «Haploview 4.1», с заданным порогом  $D' \geq 0.8$ .

Уровень генетического разнообразия и межпопуляционной дифференциации вычисляли, анализируя молекулярную вариабельность (AMOVA). Селективную нейтральность полиморфизмов изучали с помощью теста Эвенса–Ваттерсона [22]. Роль давления отбора в формировании паттернов LD и уровня генетического разнообразия в популяциях оценивали с использованием стандартных статистических тестов нейтральности Таджимы и Фу [23, 24].

### РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

#### Генетическое разнообразие и структура гаплотипов локуса *MTHFR* в популяциях

Генофонд популяций современного человека сформировался в результате последовательных эволюционно-демографических процессов: продолжительного периода эволюции генетического разнообразия в Африке и расхождения популяций при расселении современного человека с частичной изоляцией и снижением интенсивности потока генов обратно пропорционально расстоянию. Популяции в различных географических регионах имеют как общую, так и уникальную эволюционную историю, «отпечатки» которых можно наблюдать в геноме современного человека, в том числе в виде паттернов LD [3, 25–27].

Распределение генотипов, частоты аллелей, наблюдаемая гетерозиготность и уровень значимости соответствия равновесию Харди–Вайнберга исследованных SNPs гена *MTHFR* приведены в табл. 3. Все 12 локусов оказались полиморфными практически во всех изученных популяциях (за исключением rs2066470 в популяции кетов). Частота минорного ал-

Таблица 2. Характеристика исследованных SNPs гена *MTHFR*

№	SNP	Позиция на хромосоме (по данным базы NCBI)	Аллели dbSNP	Предковый аллель	Тип мутации	Локализация в гене <i>MTHFR</i> (по данным базы NCBI)
1	rs3753588	11863904	A/G	G		Интрон 1
2	rs2066470	11863057	C/T	C	Синонимичная (39 Pro/Pro)	Экзон 2
3	rs17037397	11862163	A/C	C		Интрон 2
4	rs7533315	11860683	C/T	C		Интрон 3
5	rs4846052	11857951	C/T	T		Интрон 4
6	rs1801133	11856378	C/T	C	Несинонимичная (222 Val/Ala)	Экзон 5
7	rs6541003	11855867	A/G	G		Интрон 5
8	rs2066462	11854896	C/T	C	Синонимичная (352 Ser/Ser)	Экзон 7
9	rs1801131	11854476	A/C	A	Несинонимичная (429 Ala/Glu)	Экзон 8
10	rs17375901	11852516	C/T	C		Интрон 9
11	rs2274976	11850927	A/G	G	Несинонимичная (594 Gln/Arg)	Экзон 12
12	rs1537516	11847861	C/T	C		3'-нетранслируемая область

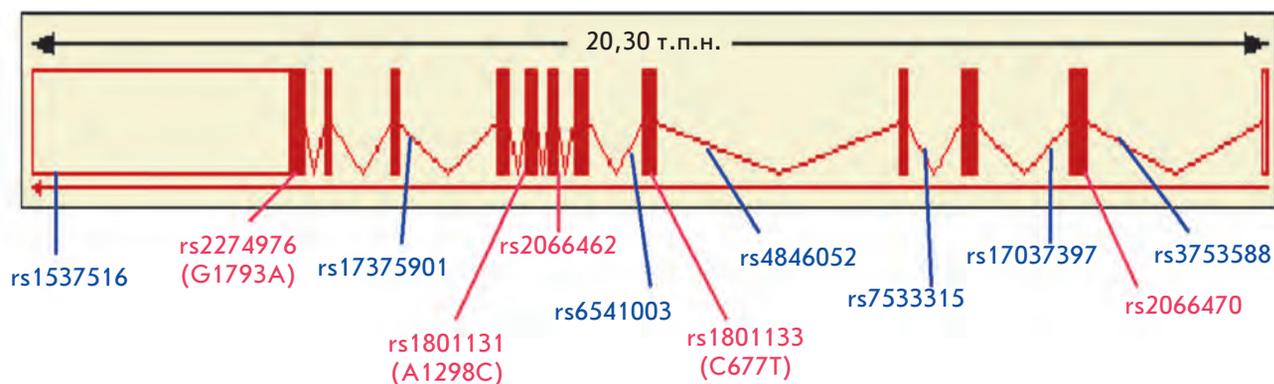


Рис. 1. Локализация исследованных SNP в гене *MTHFR*. Синим цветом показаны полиморфизмы, расположенные в интронах и 3'-нетранслируемой области, красным – в экзонах.

лея варьировала от 0 до 39%, семь SNPs (rs3753588, rs7533315, rs4846052, rs1801133, rs6541003, rs1801131 и rs1537516) встречались с частотой более 5% во всех популяциях. Полученные данные находятся в пределах вариаций частот аллелей и генотипов полиморфизмов гена *MTHFR*, опубликованных и представленных в базах данных европеоидных и монголоидных популяций. Во всех выборках распределение частот генотипов практически всех маркеров соответству-

ет равновесию Харди-Вайнберга (исключение составляют локусы rs17375901, rs2066470, rs3753588, rs2274976 и rs1537516 в популяции бурятов из пос. Хуромша). В большинстве случаев отмечается низкая и средняя гетерозиготность, что согласуется с мировыми данными. Максимальная гетерозиготность по совокупности локусов зафиксирована в популяции якутов (0.28), минимальная – у жителей поселка Хуромша (0.18). Разумеется, приведенные показате-

Таблица 3. Распределение генотипов и минорных аллелей изученных полиморфных вариантов гена *MTHFR* в исследованных выборках

№	Исследованные SNPs	Генотип, аллель	Частота, %								
			Тувицы (N=134)	Киргизы южные (N=111)	Киргизы северные (N=85)	Кеты (N=38)	Буряты г. Улан-Удэ (N=60)	Буряты пос.Хуромша (N=60)	Якуты (N=81)	Ханты (N=142)	Русские (N=126)
1	2	3	4	5	6	7	8	9	10	11	12
1	rs3753588	AA	2	1	1	0	0	3	4	2	1
		AG	13	14	22	11	17	5	20	28	18
		GG	85	85	77	89	83	92	76	70	81
		A	8	8	12	5	8	6	14	16	10
		$H_e$	0.16	0.15	0.23	0.10	0.17	0.13	0.24	0.28	0.18
		$p$	0.21	0.48	1.00	1.00	1.00	<b>0.008</b>	0.15	1.00	1.00
2	rs2066470	CC	90	85	78	100	83	92	76	79	83
		CT	10	14	18	0	17	5	19	19	16
		TT	0	1	1	0	0	3	5	2	1
		T	5	8	10	0	8	6	14	11	9
		$H_e$	0.13	0.16	0.20	0	0.17	0.13	0.25	0.20	0.16
		$p$	1.00	0.53	1.00	0	1.00	<b>0.007</b>	<b>0.05</b>	0.67	1.00
3	rs17037397	AA	0	0	0	0	0	0	3	0	0
		AC	12	11	8	16	20	5	17	26	11
		CC	88	89	92	84	80	95	80	74	89
		A	6	5	4	8	10	3	11	13	6
		$H_e$	0.12	0.11	0.09	0.17	0.20	0.07	0.20	0.23	0.11
		$p$	1.00	1.00	1.00	1.00	1.00	1.00	0.24	0.13	1.00
4	rs7533315	CC	55	69	60	68	78	63	63	75	53
		CT	38	29	39	29	19	35	37	23	42
		TT	7	2	1	3	3	2	0	2	5
		T	26	16	21	17	13	19	18	13	26
		$H_e$	0.38	0.28	0.34	0.31	0.24	0.32	0.31	0.23	0.38
		$p$	1.00	0.74	0.18	1.00	0.21	0.67	0.06	0.72	0.36
5	rs4846052	CC	46	55	40	55	58	58	40	51	30
		CT	41	42	52	39	36	35	53	39	53
		TT	13	3	8	6	6	7	7	10	17
		T	34	24	34	25	23	24	34	29	43
		$H_e$	0.45	0.37	0.46	0.38	0.35	0.38	0.45	0.41	0.50
		$p$	0.33	0.12	0.23	1.00	1.00	0.73	0.13	0.68	0.47
6	rs1801133 (C677T)	CC	67	53	53	79	72	55	61	67	50
		CT	28	37	44	18	25	42	33	29	37
		TT	5	10	3	3	3	3	6	4	13
		T	19	28	24	12	16	24	23	18	31
		$H_e$	0.32	0.41	0.38	0.24	0.27	0.37	0.36	0.29	0.44
		$p$	0.26	0.36	0.14	0.41	0.62	0.48	0.75	0.57	0.10
7	rs6541003	AA	43	54	38	53	60	58	37	49	29
		AG	46	43	52	42	35	35	52	40	56
		GG	11	3	10	5	5	7	11	11	15
		G	34	24	36	26	23	24	32	31	43
		$H_e$	0.45	0.38	0.47	0.39	0.35	0.38	0.47	0.43	0.50
		$p$	0.85	0.12	0.34	1.00	1.00	0.73	0.47	0.44	0.28

окончание таблицы 3											
1	2	3	4	5	6	7	8	9	10	11	12
8	rs2066462	CC	94	88	93	89	97	97	89	74	94
		CT	6	12	7	11	3	3	11	26	6
		T	3	6	3	5	2	2	6	13	3
		$H_e$	0.06	0.12	0.07	0.10	0.05	0.05	0.11	0.23	0.06
		$p$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.13	1.00
9	rs1801131 (A1298C)	AA	44	62	34	58	60	58	46	51	40
		AC	35	35	54	37	36	35	48	38	48
		CC	21	3	12	5	4	7	6	11	12
		C	38	20	39	24	21	24	30	30	36
		$H_e$	0.38	0.33	0.49	0.37	0.33	0.38	0.43	0.42	0.47
		$p$	0.79	0.56	0.25	1.00	1.00	0.73	0.29	0.32	0.69
10	rs17375901	CC	98	96	89	97	98	94	94	94	91
		CT	2	4	11	3	2	3	6	6	9
		TT	0	0	0	0	0	3	0	0	0
		T	1	2	5	1	1	5	3	3	4
		$H_e$	0.02	0.04	0.11	0.03	0.03	0.11	0.07	0.06	0.09
		$p$	1.00	1.00	1.00	1.00	1.00	<b>0.003</b>	1.00	1.00	1.00
11	rs2274976 (G1793A)	AA	0	0	0	0	0	3	3	1	0
		AG	10	30	11	8	17	4	17	23	13
		GG	90	70	89	92	83	93	80	76	87
		A	5	15	5	4	8	5	13	13	6
		$H_e$	0.11	0.26	0.11	0.10	0.17	0.11	0.20	0.23	0.12
		$p$	1.00	0.12	1.00	1.00	1.00	<b>0.003</b>	0.24	1.00	1.00
12	rs1537516	CC	84	85	80	89	83	92	80	69	82
		CT	16	14	20	11	17	5	17	29	17
		TT	0	1	0	0	0	3	3	2	1
		T	8	8	10	5	8	6	13	16	10
		$H_e$	0.15	0.16	0.18	0.13	0.17	0.013	0.20	0.27	0.18
		$p$	1.00	0.53	1.00	1.00	1.00	<b>0.007</b>	0.24	0.53	0.34
Средняя $H_e$ по 12 локусам			0.21	0.24	0.27	0.19	0.21	0.18	0.28	0.27	0.27
Средняя $H_e$ по 12 локусам			0.23	0.23	0.26	0.19	0.21	0.20	0.27	0.27	0.23

Примечание:  $N$  – число индивидов в выборке,  $H_e$  – ожидаемая гетерозиготность,  $H_o$  – наблюдаемая гетерозиготность,  $p$  – уровень значимости соответствия равновесию Харди–Вайнберга. Полужирным шрифтом выделены статистически значимые отличия.

ли не отражают уровень гетерозиготности изученных популяций по SNP вообще, в силу малого числа рассматриваемых локусов, а представляют интерес лишь как сведения о степени полиморфизма гена *MTHFR*. Что касается наблюдаемого в нашей работе отклонения от равновесия Харди–Вайнберга, то оно может быть обусловлено смещением оценок частот вследствие небольшого объема выборки. С другой стороны, случаи достоверного отклонения распределения от ожидаемого могут отражать специфику популяционно-генетических процессов в популяции, которая может быть связана как с характеристиками генетико-демографической структуры популяции, так и со сцеплением с функционально значимым локусом, что, на наш взгляд, более вероятно.

Из 12 рассматриваемых SNPs гена *MTHFR* наиболее хорошо изучен полиморфный вариант C677T (rs1801133). Миссенс-мутация C677T (замена цитозина на тимин в положении 677) вызывает замену аланина на валин в каталитическом домене фермента. У гомозигот по полиморфному аллелю активность фермента *in vitro* снижена на 70%, а у гетерозигот – на 35%. Частота мутантного аллеля 677T в популяциях мира колеблется от полного отсутствия у представителей племени денди до 55% у испанцев [28–30]. В России у жителей московского региона частота встречаемости аллеля 677T составляет 29%, у жителей Сибири – 32% [31, 32]. В исследованных нами популяциях частота данного аллеля варьирует от 12% у кетов до 31% в выборке русских.

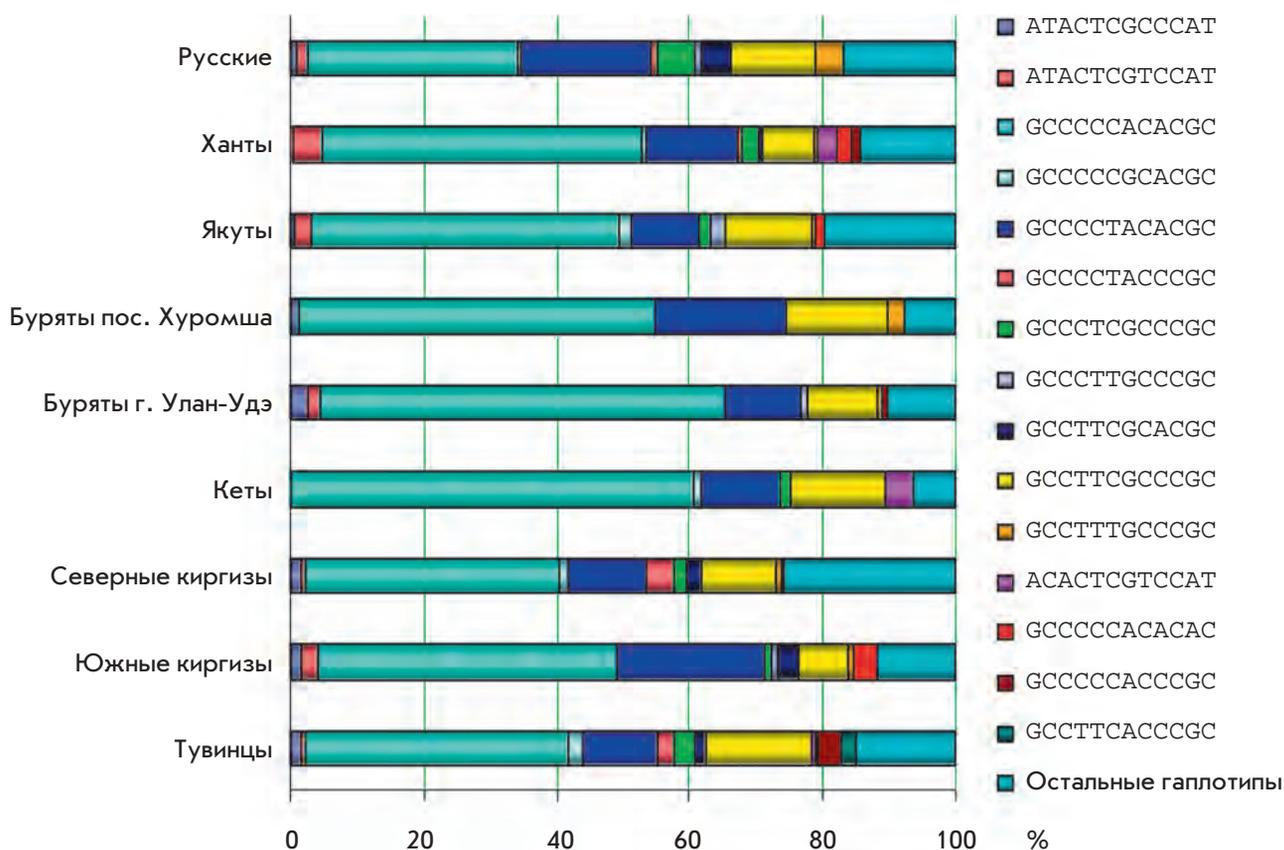


Рис. 2. Распределение гаплотипов в изученных популяциях.

Второй распространенный полиморфизм в гене *MTHFR* – транзигия А1298С (rs1801131), приводящая к замене остатка глутаминовой кислоты на аланин в регуляторном домене фермента. У носителей аллеля 1298С также снижена активность фермента, хотя и не столь значительно, как у носителей аллеля 677Т. У индивидов, являющихся компаунд-гетерозиготами по маркерам С677Т и А1298С, согласно некоторым исследованиям, наблюдается уменьшение активности *MTHFR* на 40–50% и биохимический профиль, схожий с профилем у гомозиготных носителей аллеля 677Т [33]. Минимальная частота аллеля 1298С зафиксирована у жителей Сенегала (4%), максимальная – в популяциях Израиля и Новой Гвинеи (41%) [34, 35]. У жителей России частота данного аллеля варьирует от 24 до 38% [31]. В изученных нами популяциях аллель С встречается с частотой от 20% (у южных киргизов) до 39% (у северных киргизов).

К настоящему времени установлено, что на уровень ГЦ влияет еще один SNP гена *MTHFR* – rs2274976 (G1793A). Этот локус менее изучен, чем С677Т и А1298С. Известно, что частота миссенс-мутации G1793A варьирует от 1.3% у евреев ашкенази до 26.6% у индонезийцев острова Ява [36]. Показано, что гете-

розиготность по сайту G1793A приводит к повышению уровня ГЦ в крови на 40% [37, 38]. В наших популяциях минимальная частота аллеля А составляет 4% в выборке кетов, наибольшая частота данного аллеля (15%) наблюдается в популяции южных киргизов.

В исследованных выборках обнаружено 160 гаплотипов из 4096 теоретически возможных. Максимальное число гаплотипов выявлено в популяции тувинцев (47), минимальное – у кетов (11). В большинстве выборок, за исключением кетов, бурят и южных киргизов, отмечен высокий уровень гаплотипического разнообразия. Если считать каждую мутацию, приводящую к образованию закрепившегося в популяциях полиморфного варианта, уникальным событием, а скорость образования мутаций пренебрежимо малой, то следует постулировать возможность формирования вследствие мутационного процесса лишь 12 гаплотипов. В этом случае существенная часть гаплотипического разнообразия даже в столь физически небольшом участке генома, как локус *MTHFR*, длиной около 20 т.п.н., должна быть сформирована за счет рекомбинационных событий (см. далее).

На рис. 2 представлено распределение гаплотипов, встречающихся с частотой более 2% в изученных по-



**Рис. 3.** Структура неравновесия по сцеплению в гене *MTHFR* в исследованных популяциях. Цветовая гамма отображает силу сцепления между SNP: ярко-красный – сильное сцепление ( $D' = 1$ ,  $LOD > 2$ ), красный и розовый – значительное сцепление ( $D' < 1$ ,  $LOD > 2$ ), белый – слабое сцепление ( $D' < 1$ ,  $LOD < 2$ ). Ячейка сиреневого цвета обозначает невозможность расчета неравновесия по сцеплению вследствие низкой частоты минорного аллеля полиморфизма ( $D' = 1$ ,  $LOD < 2$ ). Расположение SNP в каждой популяции следующее (слева направо): rs3753588, rs2066470, rs17037397, rs7533315, rs4846052, rs1801133, s6541003, rs2066462, rs1801131, rs17375901, rs2274976, rs1537516, за исключением популяций из проекта ХарМар, где отсутствует полиморфизм rs3753588. В популяции йоруба маркер rs17375901 исключен из анализа, так как является мономорфным.

пуляциях. Гаплотипы с частотой более 5% обозначены как основные. Во всех обследованных популяциях, за исключением русских, выявлены три основных гаплотипа – GCCCCACACGC, GCCCCTACACGC, GCCTTCGCCCGC, частоты которых в сумме составляют более 83% наблюдаемых хромосом в популяциях кетов и бурят, более 61% в остальных выборках.

В исследованных нами популяциях обнаружена различная степень гаплотипического разнообразия, тем не менее, все выборки содержали одинаковые основные гаплотипы, что указывает на возможность общего механизма формирования данных паттернов LD.

#### Архитектура неравновесия по сцеплению гена *MTHFR* в некоторых популяциях Евразии

На рис. 3 показана структура LD между исследованными локусами гена *MTHFR* в 13 популяционных выборках. Максимальное сцепление между изученными SNP показано для южных киргизов, кетов, китайцев и японцев – в этих выборках все анализируемые аллельные варианты гена *MTHFR* входят в состав одного гаплотипического блока. У европеоидов из проекта ХарМар также наблюдается

один блок, однако он не включает маркеры rs2274976 и rs1537516. Два блока обнаружено у северных киргизов: первый включает три SNP (rs2066470, rs17037397 и rs7533315), второй охватывает участок длиной 10 т.п.н. и содержит восемь SNP. В популяции хантов также выявлено два блока: первый содержит rs3753588, rs2066470 и rs17037397, второй идентичен блоку № 2 у северных киргизов. Сильное сцепление между первыми девятью SNP, составляющими первый блок размером 9 т.п.н., и между rs2274976 и rs1537516, образующими второй небольшой блок, выявлено у бурятос пос. Хуромша. В популяции бурятос г. Улан-Удэ также обнаружено значительное сцепление между многими полиморфными вариантами, но в структуре LD в этой популяции представлены два других блока (3 и 8 т.п.н.). У русских показаны три блока: первый – из двух близко расположенных SNP (rs3753588 и rs2066470), второй включает пять полиморфизмов – rs7533315, rs4846052, rs1801133, rs6541003 и rs2066462, третий состоит из четырех SNP – rs1801131, rs17375901, rs2274976 и rs1537516.

В популяциях тувинцев и якутов можно выделить четыре небольших гаплотипических блока, состоя-

щих из двух или трех соседних SNPs. В популяции йоруба представлены два блока: первый состоит из 2 т.п.н. и содержит rs2066470, rs17037397 и rs7533315, второй блок достаточно протяженный (7 т.п.н.), он включает четыре локуса (rs2066462, rs1801131, rs2274976 и rs1537516). Таким образом, в исследованных популяциях гаплотипические блоки длиной от 847 п.н. до 16 т.п.н. представлены несколькими (3–6) основными гаплотипами, которые в сумме составляют более 90% наблюдаемых хромосом. Состав и размер блоков варьируют в зависимости от популяционной структуры. Примечательно, что функционально значимые полиморфизмы C677T и A1298C во всех популяциях, кроме южных киргизов, кетов и популяций из проекта НарМар, не сцеплены. Необходимо также отметить, что во многих популяциях наблюдается тесное сцепление между SNPs, образующими второй блок у йоруба.

На сегодняшний день степень гаплотипического разнообразия и протяженность блоков сцепления в разнообразных участках генома определены во многих популяциях человека [39, 40]. Обнаружена большая изменчивость гаплотипической структуры, которая значительно отличается при переходе от одного геномного региона к другому, часто перемежаясь регионами низкого уровня LD. Размеры гаплотипических блоков варьируют от 1 до 100 и более т.п.н. [9]. Некоторые авторы сообщают о согласованности пространственного размещения гаплотипических блоков в нескольких регионах генома человека, указывая на возможность общего механизма образования этих блоков в различных популяциях как вероятной причины данного феномена [3, 41]. Описана также дихотомия африканцы/неафриканцы в нескольких сегментах генома человека [42, 43], где гаплотипические блоки имеют большую протяженность в неафриканских популяциях (> 44 т.п.н.), чем у населения Африки (>22 т.п.н.) [9]. Обнаружены паттерны мейотической рекомбинации, которые подтверждают, что регионы низкого уровня LD, соответствующие регионам с высокими показателями рекомбинации, ограничивают гаплотипические блоки [10, 44]. В рамках интенсивного изучения рекомбинации в геноме человека оценивали также взаимосвязь между частотой рекомбинаций, структурой хроматина и различными структурными и функциональными компонентами ядра. Этот анализ показал, что различные потенциальные факторы как локальные (нуклеотидная последовательность ДНК, структура хроматина), так и несвязанные непосредственно со структурой точки рекомбинации, могут повлиять на показатели рекомбинации специфического хромосомного сегмента. Существует предположение, что рекомбинация является основной причиной об-

разования блоков сцепления вследствие разрушения протяженных гаплотипов в хромосомных регионах со сравнительно высокой рекомбинацией. Доказательством этой гипотезы служит похожая структура LD в нескольких регионах генома в различных популяциях [45–47]. Наряду с этим существуют данные, свидетельствующие о значимых межпопуляционных различиях в степени и характере LD в одном и том же участке генома [10, 48–50]. Эти результаты указывают на то, что характер LD, выявленный в конкретной популяции или выборке, вероятно, не может автоматически экстраполироваться на другие популяции, по крайней мере, в некоторых участках генома. Маловероятно, что одна общая карта неравновесия по сцеплению в геноме окажется полезной при выборе генетических маркеров для ассоциативных исследований во многих популяциях, поскольку нельзя игнорировать взаимодействие между различными популяционно-специфическими факторами и геном-специфическими механизмами при формировании структуры LD.

В настоящей работе получено подтверждение популяционно-специфического характера формирования паттернов LD. Наиболее сильное сцепление (все SNPs входят в состав одного блока) между исследованными локусами обнаружено в популяциях кетов, южных киргизов, китайцев и японцев. Тесное сцепление наблюдалось также у бурят. Можно предложить два объяснения наблюдаемому сохранению структуры и протяженности блока: недавнее происхождение от общих предков или позитивный отбор, который часто приводит к увеличению размера блока, содержащего полезный аллель [51]. Поскольку у кетов и бурят выявлено минимальное число гаплотипов среди всех исследованных популяций, есть вероятность, что в данном случае имел место эффект основателя. Однако на увеличение структуры LD влияет и ряд других факторов [49] (генетическая изоляция, подразделенность или смешение популяций, балансирующий отбор, эффект «горлышка бутылки», небольшой размер популяции и другие причины), действие которых на некоторые популяции также нельзя исключить.

Показано, что протяженность паттернов LD в геноме человека определяет потенциал и дизайн ассоциативных исследований, использующих SNP для картирования генов, лежащих в основе сложных признаков. Текущая оценка числа маркеров, необходимых для основанного на LD геномного сканирования различных популяций, варьирует от 120000 до нескольких миллионов и свидетельствует об огромной стоимости генотипирования и возможных проблемах статистических выводов. Предполагается, что в популяциях с высокой степенью LD ко-

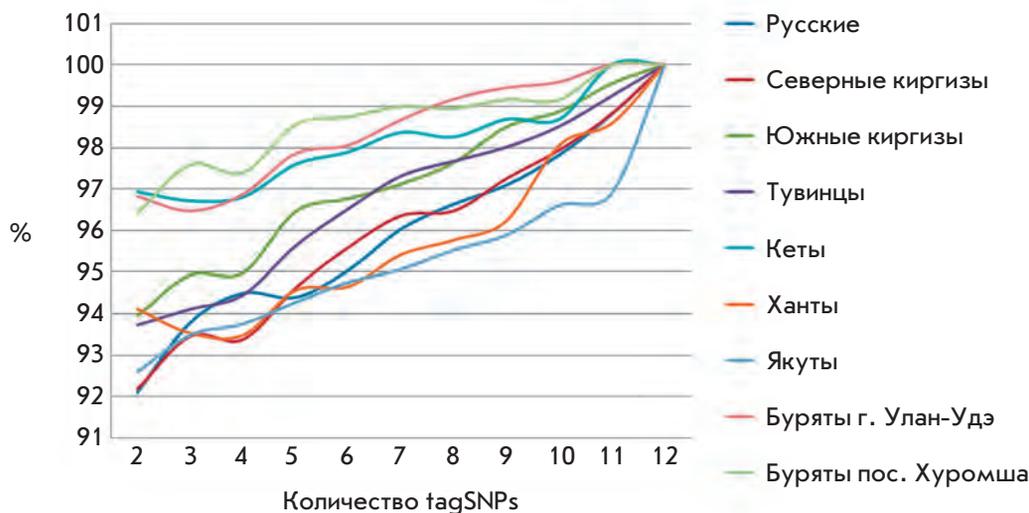


Рис. 4. Вариабельность точности прогноза набора tagSNP гена *MTHFR* в зависимости от их количества и популяционной выборки.

личество маркеров, необходимых для картирования МФЗ, существенно снизится [52].

#### Выбор tagSNPs гена *MTHFR*

По оценкам различных ученых геном человека содержит более 7.5 млн распространенных SNPs с частотой минорного аллеля (MAF) не менее 5% [21, 53], которые объясняют часть наследственного риска развития многих МФЗ. В настоящее время одной из наиболее продуктивных стратегий создания панели генетических маркеров для анализа ассоциаций с МФЗ является выбор tagSNPs с целью расширения «генетического охвата» [54]. В данном случае «генетический охват» определяется как доля набора всех распространенных SNPs с MAF 5%, находящихся в пороговой корреляции, по крайней мере, с одним SNP из заданного массива полиморфизмов [55]. Подход с использованием tagSNPs имеет преимущества в связи с известными на сегодняшний день знаниями о блочной структуре LD организации генома человека. Кроме того, данная стратегия существенно редуцирует силы и средства, необходимые для генотипирования [53, 56]. Исходя из несомненной потенциальной пользы выбора tagSNPs для ассоциативных исследований, предлагается эффективно идентифицировать их посредством различных алгоритмов.

В данной работе при установлении tagSNPs использовали два метода – «STAMPA» и «Tagger». «STAMPA» использует для анализа информацию о генотипах и не требует сведений о гаплотипах и блочной архитектуре исследуемого геномного региона [57]. Этот алгоритм основан на гипотезе о том, что корреляция между SNPs имеет тенденцию к снижению с увеличением физического расстояния между ними, а аллельный вариант SNP может быть

определен на основе знания аллельных вариантов ближайших tagSNPs с каждой стороны. Во всех популяциях уже при двух выбранных tagSNPs гена *MTHFR* достигается прогностическая ценность более 90% (рис. 4). Возможно, это обусловлено небольшим размером гена и физически близким расположением рассматриваемых полиморфных сайтов. Однако для достижения точности прогноза 99–100% требуется значительно увеличить количество tagSNPs, что, на наш взгляд, является существенным недостатком данного алгоритма.

Также в представленной работе tagSNPs гена *MTHFR* определяли с помощью приложения «Tagger», предусмотренного программным обеспечением «HaploView» [21]. В данном методе использовался алгоритм «Aggressive tagging», основанный на выявлении tagSNP в двух и трех маркерных гаплотипах, в пределах которых все входящие полиморфные варианты должны находиться в состоянии прочного сцепления ( $LOD > 3$ ) друг с другом [26]. TagSNPs, выявленные при помощи описанного алгоритма, представлены в табл. 4.

Предсказательная способность данного набора tag-меток относительно всего массива исследованных полиморфизмов составляет 100%. Согласно полученным результатам, при заданном пороге прогностической ценности в различных популяциях наблюдаются неодинаковые наборы tagSNPs, что, вероятно, обусловлено варьированием структуры LD и гаплотипического разнообразия гена *MTHFR* в выборках. Подтверждением этого служит статистически значимая корреляция между гаплотипическим разнообразием и вариабельностью числа tagSNPs ( $r^2 = 0.85$ ;  $p < 0.01$ ). Вопрос о возможности «переноса» tagSNPs с одной популяции на другую доволь-

Таблица 4. TagSNPs гена *MTHFR*, идентифицированные в программе «Tagger»

Популяция	Число tagSNP	SNP гена <i>MTHFR</i>											
		1	2	3	4	5	6	7	8	9	10	11	12
Русские	9	1	2	3	4	5	6	7	8	9	10	11	12
Северные киргизы	11	1	2	3	4	5	6	7	8	9	10	11	12
Южные киргизы	8	1	2	3	4	5	6	7	8	9	10	11	12
Тувинцы	11	1	2	3	4	5	6	7	8	9	10	11	12
Кеты	8	1	2	3	4	5	6	7	8	9	10	11	12
Ханты	9	1	2	3	4	5	6	7	8	9	10	11	12
Якуты	10	1	2	3	4	5	6	7	8	9	10	11	12
Буряты г. Улан-Удэ	8	1	2	3	4	5	6	7	8	9	10	11	12
Буряты пос. Хуромша	7	1	2	3	4	5	6	7	8	9	10	11	12

Примечание: использована следующая нумерация SNP: 1 – rs3753588, 2 – rs2066470, 3 – rs17037397, 4 – rs7533315, 5 – rs4846052, 6 – rs1801133, 7 – rs6541003, 8 – rs2066462, 9 – rs1801131, 10 – rs17375901, 11 – rs2274976, 12 – rs1537516. Серым фоном выделены tagSNP гена *MTHFR*.

но важен, так как анализ ассоциаций с использованием tagSNPs, идентифицированных на основании данных проекта НарМар, весьма актуален. В серии недавно выполненных исследований показана достаточно высокая прогностическая ценность наборов tagSNPs в нескольких регионах генома, выбранных для CEU, CHB и JPT, применительно к ряду европеоидных и монголоидных популяций [58–60]. Тем не менее обнаружено, что уровень экстраполяции tagSNP снижается при использовании набора tagSNPs, найденных для CEU, в ассоциативных исследованиях в африканских и некоторых изолированных европеоидных популяциях [26, 61, 62]. Наряду с этим показано, что наиболее универсальные tagSNPs, обеспечивающие максимальный «генетический охват» в других популяциях, идентифицируются у йоруба вследствие минимальной силы LD в этой выборке [53].

В рамках нашей работы проведен сравнительный анализ мощности алгоритмов «STAMPA» и «Tagger» в зависимости от уровня прогностической оценки. Как видно из табл. 5, при уровне прогноза 90–95% минимальное число tagSNPs определяется посредством алгоритма «STAMPA», а при предсказательной точности 98–100% более эффективен метод «Tagger».

Использование tagSNPs как средства, эффективно отражающего генетическое разнообразие, на сегодняшний день широко распространено в различных генетических исследованиях. Тем не менее качество выбранных tagSNPs зависит от исходного массива, в котором они охарактеризованы. Если изначально плотность маркеров слишком редкая, то выбранный tagSNP «захватит» меньше информации, чем требу-

ется для анализа. Необходимая плотность маркеров в исходном массиве данных варьирует в различных участках генома в зависимости от ряда показателей, таких, например, как уровень рекомбинации, структура LD, частоты SNP, характер мутации и демографическая история населения [17].

#### Филогенетический анализ взаимоотношений гаплотипов локуса *MTHFR* и оценка селективной нейтральности изученных полиморфизмов

В данной работе филогенетический анализ взаимоотношений гаплотипов, определяемых по диаллельным маркерам и встречающихся с частотой более 0.1% в суммарной выборке, проводили, конструируя филогенетические деревья (сети) гаплотипов по алгоритму медианных сетей, реализованному в программе «Network». В качестве предкового гаплотипа рассматривали гаплотип, состоящий из предковых аллелей (информация взята из базы данных NCBI). Полученные результаты свидетельствуют о схождении всех гаплотипов, наблюдаемых в изученных популяциях человека, к одному общему предковому варианту, встречающемуся с частотой около 12% у йоруба и 1% в популяциях русских и северных киргизов (гаплотип 10 на рис. 5), что свидетельствует в пользу теории недавнего африканского происхождения современного человека. Подтверждает эту гипотезу и значительная частота у йоруба гаплотипов № 4 и 7, наиболее близких к предковому варианту.

Все наблюдаемые гаплотипы находятся в пределах шести мутационных шагов от общего предка и разбиваются на три основных кластера – А, В и С, формирующихся из гаплотипов ССТТСГСАСГС, СССТСГСССГС и СССТСАСАСГС (№ 4, 5 и 7 со-

Таблица 5. Сравнительная характеристика алгоритмов определения tagSNPs «STAMPA» и «Aggressive tagging»

Популяция	Точность прогноза 90%		Точность прогноза 95%		Точность прогноза 98%		Точность прогноза 100%		Гаплогенетическое разнообразие	Число блоков
	STAMPA	Tagger	STAMPA	Tagger	STAMPA	Tagger	STAMPA	Tagger		
	Число tagSNP гена <i>MTHFR</i>									
Русские	2	8	6	9	12	9	10	9	0.69	3
Северные киргизы	2	10	6	11	12	11	10	11	0.77	2
Южные киргизы	2	7	4	8	12	8	9	8	0.62	1
Тувинцы	2	10	5	11	12	11	9	11	0.82	4
Кеты	2	7	2	8	12	8	10	8	0.49	1
Ханты	2	8	7	9	12	9	10	9	0.78	3
Якуты	2	9	7	10	12	10	12	10	0.72	4
Буряты г. Улан-Удэ	2	7	2	8	12	8	5	8	0.59	3
Буряты пос. Хуромша	2	6	2	7	12	7	5	7	0.55	2

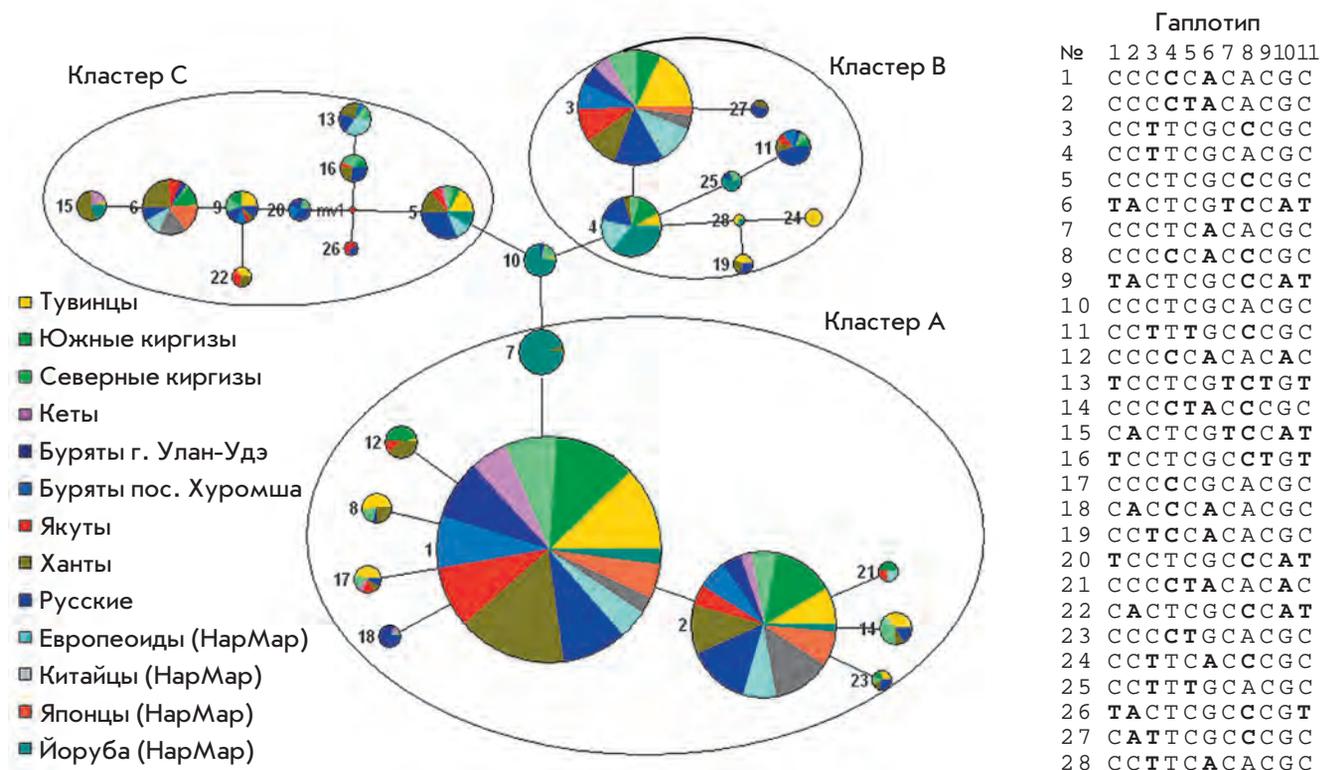


Рис. 5. Медианное дерево гаплотипов, встречающихся с частотой более 0.1% в суммарной выборке. Жирным выделены мутантные аллели, предковый гаплотип обозначен № 10. Диаметр узла отражает частоту гаплотипа в суммарной выборке. Нумерация SNP в гаплотипах: 1 – rs2066470, 2 – rs17037397, 3 – rs7533315, 4 – rs4846052, 5 – rs1801133, 6 – rs6541003, 7 – rs2066462, 8 – rs1801131, 9 – rs17375901, 10 – rs2274976, 11 – rs1537516.

ответственно, рис. 3). Кластер А представлен 10 гаплотипами, два из которых (№ 1, 2) наиболее распространенные (их суммарная частота составляет более 50% от частоты всех гаплотипов в суммарной выборке) и выявляются во всех изученных популяциях. Необходимо отметить, что структура дерева в данном кластере носит ярко выраженный звездообразный характер, что, очевидно, указывает на резкую экспансию численности в демографической истории населения. Кластер В включает восемь гаплотипов, в том числе гаплотип № 3, третий по частоте встречаемости и представленный во всех популяциях, за исключением йоруба. Примечательно, что гаплотипы № 8, 12, 17, 18, 21, 14, 23, 27, 11, 24, 19 кластеров А и В, расположенные на концах ветвей филогенетического дерева, встречаются только в популяциях Северной Евразии и предположительно возникли недавно. Гаплотипы кластера С содержат большое число мутантных аллелей, они образовались, вероятно, с участием рекомбинационных событий.

Поскольку темп мутирования SNP и их разнообразие, наблюдаемое в современных популяциях, поддаются оценке, можно рассчитать время происхождения данной гаплотипической линии. Известно, что оценки разнообразия, основанные на филогении нуклеотидных последовательностей ДНК, не зависят от демографических процессов [63] и пригодны для выявления эволюционного «возраста» генетических линий. При этом под возрастом понимают время коалесценции («схождения» к общему предку), или, другими словами, время генерации наблюдаемого разнообразия. Учитывая эти соображения, в качестве мутационного шага для оценки времени коалесценции гаплотипов рассматривали изменения аллелей одного SNP в одном локусе. Всем полиморфным вариантам задавали одинаковый темп мутирования, составляющий  $1 \times 10^{-8}$  на локус на поколение [64]. Время одного поколения считали равным 20 годам. Время коалесценции гаплотипов оценивали посредством программы «Network». В целом возраст генерации разнообразия, определенный по 12 SNPs гена *MTHFR*, составил  $314000 \pm 135000$  лет. Для оценки точности данного результата были проанализированы медианные деревья гаплотипов, полученные при «разбиении» гена *MTHFR* на два блока (первый включает rs3753588, rs2066470, rs17037397, rs7533315, rs4846052, rs1801133, rs6541003, второй – rs2066462, rs1801131, rs17375901, rs2274976, rs1537516). Такой анализ был проведен в связи с тем, что в большинстве популяций, а также у йоруба указанные участки гена *MTHFR* входят в состав разных блоков LD. Тем самым независимый филогенетический анализ двух блоков гена *MTHFR* может быть более устойчивым к возможным ошибкам, привносимым за счет рекомбинации. В итоге время коалесценции

первого блока оказалось равным  $350000 \pm 188000$  лет, возраст предкового гаплотипа второго блока оценивается в  $306000 \pm 188000$  лет.

Хотя проведенный филогенетический анализ является достаточно мощным и продуктивным для характеристики эволюционных взаимоотношений между гаплотипами, следует отметить, что абсолютные оценки времени коалесценции (т.е. времени, выраженному в годах) следует интерпретировать с осторожностью, поскольку ключевые параметры, на которых строятся эти оценки, – темп мутирования и отсутствие рекомбинации в анализируемом участке генома.

Согласно результатам ряда исследований [65, 66], сравнительно недавняя и быстрая экспансия человека из Африки оставила существенный «отпечаток» в нашем геноме, сформировав структуру генетических различий в популяциях человека, обладающую, в том числе и биомедицинской значимостью. Необходимо отметить, что вариабельность генома, обуславливающая фенотипические различия между двумя индивидами, составляет всего 0.1%. Фактически, большинство этих изменений в ДНК должно быть эволюционно нейтральным, но, тем не менее, выявлено большое число полиморфизмов, влияющих на фенотип, которые могут быть объектами отбора или подвергнуться ему в дальнейшем [65].

При анализе селективной нейтральности полиморфизмов гена *MTHFR* с помощью теста Эвенса–Ваттерсона отклонение от нейтральности обнаружено только у rs4846052 и rs6541003 в выборках русских (наблюдаемый критерий  $F$  для теста Эвенса–Ваттерсона составил 0.99 ( $p = 0.039$ ) и 0.84 ( $p = 0.041$ ) соответственно) и европеоидов из проекта НарМар ( $F = 0.79$  ( $p = 0.021$ ) и  $F = 0.98$  ( $p = 0.030$ ) соответственно). Все три функционально значимых SNPs, обуславливающие повышение уровня ГЦ в крови, оказались селективно нейтральными. Вероятно, это объясняется тем, что даже некоторое изменение фенотипа может быть селективно нейтральным, если не влияет на репродуктивную эффективность [67]. Тем не менее существуют данные о селекции аллеля 677Т в испанской популяции, основанные на изучении изменения распределения частот генотипов и аллелей полиморфизма С677Т в течение XX века. В популяции отмечен рост числа индивидов с генотипом 677ТТ, обусловленный повышением жизнеспособности носителей аллеля 677Т на ранних стадиях эмбриогенеза, вследствие приема многими женщинами фолиевой кислоты в первом и втором триместрах беременности [68]. Кроме того, селективную значимость аллеля Т подтверждают результаты анализа распределения частот аллелей, генотипов и гаплотипов полиморфизмов С677Т и А1298С гена *MTHFR* в популяциях Израиля, Японии и Африки, согласно

которым аллель 677Т обнаруживается в гаплотипах, имеющих селективное преимущество [69].

Известно, что частота аллеля 677Т в мировых популяциях очень гетерогенна: от полного отсутствия у представителей африканских племен до 55% у испанцев [28–30]. Более того, наблюдается градиент частоты данного аллеля с севера на юг Европы [70]. Показано, что в Северной Америке частота гомозигот 677ТТ увеличивается в направлении от Западной Канады (Альберта) до юго-востока США (Атланта) и достигает своего пика в Мексике [71]. Механизмы генерации этого градиента достоверно не установлены, однако, существуют по крайней мере три гипотезы, которые базируются на предположении, что высокая частота аллеля 677Т обусловлена действием естественного отбора. Первая гипотеза постулирует, что во время голода уменьшение активности МТНFR приводит к снижению реметилирования гомоцистеина, и таким образом сберегает одноуглеродные радикалы тетрагидрофолатного метаболизма для жизненно важной синтеза ДНК и РНК. Согласно другой гипотезе, у носителей мутантного аллеля меньше вероятность заболеть раком толстой кишки, в результате чего частота мутации в популяции может постепенно возрастать [18]. Третья гипотеза в качестве основного фактора накопления гомозигот 677ТТ в популяции рассматривает ген-средовые взаимодействия между МТНFR и содержанием фолиевой кислоты. Последняя теория нашла свое подтверждение в ряде экспериментальных работ [34, 70, 71].

Роль давления отбора в формировании паттернов LD и уровня генетического разнообразия в популя-

циях оценивали с использованием стандартных статистических тестов нейтральности Таджимы и Фу [23, 24]. Значение критерия *D* Таджимы оказалось отрицательным во всех исследованных популяциях, но статистически незначимым. Значение критерия *F<sub>s</sub>* теста Фу было отрицательным и статистически значимым в популяциях тувинцев (*F<sub>s</sub>* = -11.28, *p* < 0.01), северных киргизов (*F<sub>s</sub>* = -24.15; *p* < 0.00001), якутов (*F<sub>s</sub>* = -19.76, *p* < 0.00001) и хантов (*F<sub>s</sub>* = -10.31, *p* < 0.01), что свидетельствует о возможном действии отрицательного отбора на данный участок генома в этих популяциях либо о популяционной экспансии. Однако смешение данных популяций с соседними также могло привести к увеличению разнообразия ДНК, а вместе с этим и значения критерия *F<sub>s</sub>*.

В целом полученные данные свидетельствуют о вероятном действии стабилизирующего отбора на локусы rs4846052 и rs6541003 у европеоидов из проекта Нар-Мар и русских г. Томска и о возможном действии отрицательного отбора на определенные гаплотипы гена *MTHFR* в популяциях тувинцев, северных киргизов, якутов и хантов. Необходимо отметить, что эти четыре популяции характеризуются наиболее высоким уровнем гаплотипического разнообразия (более 70%) и низким уровнем LD среди всех исследованных групп.

### Генетическая дифференциация и взаимоотношения между исследованными популяциями

Данные о степени генной дифференциации (*F<sub>st</sub>*) в суммарной выборке по каждому из выбранных маркеров представлены на рис. 6. Необходимо отметить, что все

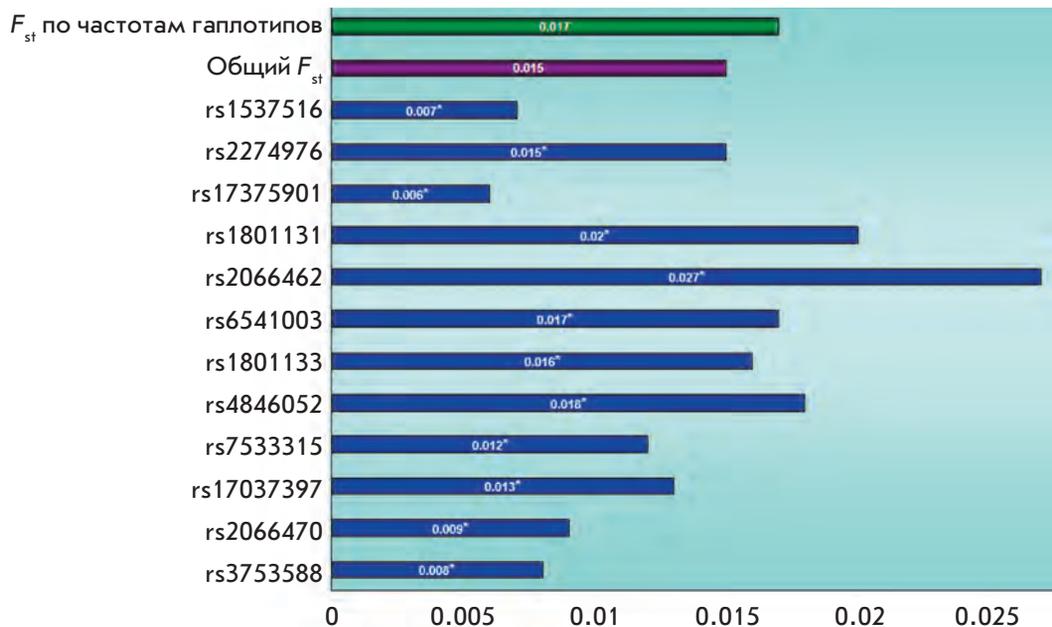


Рис. 6. Общая генетическая дифференциация суммарной выборки по исследованным полиморфизмам гена *MTHFR*. \*Статистически значимые различия (*p* < 0.05).

изученные полиморфные варианты показали статистически значимую дифференциацию. Показано, что наибольший вклад в межпопуляционное разнообразие вносят различия в частотах аллелей локусов rs4846052, rs1801133, rs6541003, rs2066462, rs1801131 и rs2274976. Наименьшая же степень межпопуляционного разнообразия характерна для rs17375901. Уровень генетической дифференциации изученных популяций по частотам аллелей 12 исследованных SNPs гена *MTHFR* составил 0.015, а по частотам гаплотипов 0.017. Оценка проводилась с помощью коэффициента  $F_{st}$ .

Филогенетический анализ генетического родства популяций проводили на массиве данных для 13 популяций. Полученная дендрограмма генетических взаимоотношений популяций показана на рис. 7. На представленном дереве можно выделить два основных кластера. Первый включает йоруба, европеоидов НарМар, русских, северных киргизов, тувинцев и якутов; второй – хантов, кетов, бурятов, южных киргизов, японцев, китайцев. Данный подход выявляет существенные различия генофондов европеоидных (русские, европеоиды НарМар) и монголоидных (японцы, китайцы, буряты, южные киргизы) популяций, а также тесное генетическое родство между йоруба и европеоидами НарМар, китайцами и японцами. Интересным представляется факт расположения северных и южных киргизов в разных кластерах на значительном генетическом расстоянии. Разрешающая способность дерева, построенного всего по 11 локусам, недостаточна, чтобы делать какие-то определенные выводы о деталях генетических взаимоотношений между исследованными популяциями, а свидетельствует лишь о степени генетических различий между ними по локусу *MTHFR*.

### ЗАКЛЮЧЕНИЕ

Изучение архитектуры неравновесия по сцеплению (LD) локуса *MTHFR* в девяти популяционных выборках, проживающих на территории Северной Евразии, базировалось на концепции блочной структуры генома человека. В качестве объекта для популяционных сравнений использованы данные по популяциям европеоидов, китайцев, японцев и йоруба, полученные в ходе реализации проекта НарМар.

В результате проведенной работы выявлена популяционная специфичность структуры LD гена *MTHFR* в различных этнотерриториальных группах Северной Евразии. Наряду с этим обнаружено сходство в архитектуре LD среди некоторых популяций, свидетельствующее о роли эволюционной истории в организации блочной структуры LD.

Нами показана различная степень гаплотипического разнообразия в исследованных популяциях,

тем не менее, во всех выборках, за исключением йоруба, найдены идентичные основные гаплотипы, что указывает на возможность общего механизма формирования паттернов LD гена *MTHFR*. Проведенный филогенетический анализ гаплотипов показывает схождение всех гаплотипов, наблюдаемых в изученных популяциях, к одному общему предковому варианту, свидетельствует о значительной роли рекомбинации в генерации генетического разнообразия локуса *MTHFR* и возможности резкой экспансии численности населения. Возраст генерации разнообразия по 12 SNP гена *MTHFR* составил  $314000 \pm 135000$  лет.

Также в представленной работе получены данные, подтверждающие действие стабилизирующего отбора на локусы rs4846052 и rs6541003 у европеоидов НарМар и влияние отрицательного отбора на определенные гаплотипы гена *MTHFR* в популяциях тувинцев, северных киргизов, якутов и хантов, характеризующихся наиболее высоким уровнем гаплотипического разнообразия (более 70%) и низким уровнем LD среди всех исследованных групп. Все изученные SNPs гена *MTHFR* показали достоверную дифференциацию. Данные по отдельным локусам демонстрируют, что наибольший вклад в межпопуляционное разнообразие вносят различия по частоте



Рис. 7. Дендрограмма генетических взаимоотношений между исследованными популяциями. Длина ветвей филогенетического дерева соответствует генетическим расстояниям.

там аллелей локусов rs4846052, rs1801133, rs6541003, rs2066462, rs1801131 и rs2274976. Наименьшая же степень межпопуляционного разнообразия характерна для маркера rs17375901, являющегося мономорфным у йоруба.

Таким образом, архитектура LD в геноме человека, в частности в локусе *MTHFR*, носит популяционно-специфический характер и в значительной степени определяется эволюционной историей популяции. Очевидно, что при анализе структуры LD и выборе tagSNP при проведении генетического картирования широко распространенных заболеваний как в полногеномном масштабе, так и при ассоциативных исследованиях, когда вариант, связанный с болезнью, выявляется по сцеплению с близлежащими сайтами

в относительно узком регионе генома, необходимо учитывать значительную этноспецифическую вариабельность гаплотипических блоков. Дальнейшее изучение характера генетического разнообразия и неравновесия по сцеплению в геноме в конкретных географических, этнических или популяционных группах позволит реконструировать генетическую историю популяций, выявить следы действия естественного отбора, связанного с адаптивной изменчивостью. ●

*Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» (ГК № П321).*

#### СПИСОК ЛИТЕРАТУРЫ

- Gamazon E.R., Zhang W., Dolan M.E., Cox N.J. // PLoS One. 2010. V. 5. № 2. e9366.
- Zhao H., Nettleton D., Dekkers J.C.M. // Genet. Res. 2007. V. 89. P. 1–6.
- Slatkin M. // Genetics. 2008. V. 9. P. 477–485.
- Altshuler D., Daly M.J., Lander E.S. // Science. 2008. V. 322. P. 881–888.
- Lambert C.A., Tishkoff S.A. // Cold Spring Harbor Symp. Quant. Biol. 2009. V. 74. P. 395–402.
- Neale B.M. // Cold Spring Harbor Protoc. 2010. V. 3. P. 74
- Huang Y.T., Chang C.J., Chao K.M. // Curr. Drug. Metab. 2011. V. 12. № 5. P. 498–506.
- Daly M.J., Rioux J.D., Schaffner S.F., Hudson T.J., Lander E.S. // Nature. 2001. V. 29. P. 229–232.
- Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., et al. // Science. 2002. V. 296. P. 2225–2229.
- Jeffreys A.J., Kauppi L., Neumann R. // Nat. Genet. 2001. V. 29. P. 217–222.
- The International HapMap Consortium. // Nature. 2007. V. 449. P. 851–862.
- Shriner D., Adeyemo A., Gerry N.P., Herbert A., Chen G., Doumatey A., Huang H., Zhou J., Christman M.F., Rotimi C.N. // PLoS One. 2009. V. 4. № 12. e8398.
- Joubert B.R., North K.E., Wang Y., Mwapasa V., Franceschini N., Meshnick S.R., Lange E.M. // J. Hum. Genet. 2010. V. 55. № 6. P. 366–374.
- O'Dushlaine C.T., Morris D., Moskvina V., Kirov G., Consortium I.S., Gill M., Corvin A., Wilson J.F., Cavalleri G.L. // Eur. J. Hum. Genet. 2010. V. 18. № 11. P. 1248–1254.
- Dumitrescu L., Carty C.L., Taylor K., Schumacher F.R., Hindorf L.A., Ambite J.L., Anderson G., Best L.G., Brown-Gentry K., Bůžková P., et al. // PLoS Genet. 2011. V. 7. № 6. e1002138.
- Crawford D.C., Nickerson D.A. // Annu. Rev. Med. 2005. V. 56. P. 303–320.
- Zhang R., Li X., Jiang Y., Liu G., Li C., Zhang F., Xiao Y., Gong B. // Science in China Ser. 2009. V. 52. № 2. P. 163–172.
- Frosst P., Blom H.J., Milos R., Goyette P., Sheppard C.A., Matthews R.G., Boers G.J., den Heijer M., Kluijtmans L.A., van den Heuvel L.P., et al. // Nat. Genet. 1995. V. 10. P. 111–113.
- Friedman G., Goldschmidt N., Friedlander Y. // J. Nutr. 1999. № 129. P. 1656–1661.
- Трифоновна Е.А., Спиридонова М.Г., Пузырёв В.П., Степанов В.А. // Мед. генетика. 2009. № 1. С. 39–47.
- Barrett J.C., Fry B., Maller J., Daly M.J. // Bioinformatics. 2005. № 21. P. 263–265.
- Watterson G.A. // Theor. Popul. Biol. 1975. V. 7. P. 256–276.
- Tajima F. // Genetics. 1989. V. 123. P. 585–595.
- Fu Y.X. // Genetics. 1997. V. 147. P. 915–925.
- Schaid D.J. // Genetics. 2004. V. 166. P. 505–512.
- de Bakker P.I., Yelensky R., Pe'er I., Gabriel S.B., Daly M.J., Altshuler D. // Nat. Genet. 2005. V. 7. № 11. P. 1217–1223.
- Zhao H., Nettleton D., Soller M., Dekkers J.C. // Genet. Res. 2005. V. 86. P. 77–78.
- Fodinger M., Horl W.H., Sunder-Plassmann G. // J. Nephrol. 2000. V. 13. P. 20–33.
- Спиридонова М.Г., Степанов В.А., Пузырёв В.П. // Клини. медицина. 2001. № 2. С. 10–16.
- Jääskeläinen E., Keski-Nisula L., Toivonen S., Romppanen E.L., Helisalmi S., Punnonen K., Heinonen S. // Hypertens. Pregnancy. 2006. V. 25. P. 73–80.
- Назаренко М.С., Пузырев В.П., Лебедев И.Н. // Генетика. 2006. Т. 42. № 5. С. 711–717.
- Фетисова И.Н., Добролюбов А.С., Липин М.А., Поляков А.В. // Вест. нов. мед. технологий. 2007. Т. 10. № 1. С. 12–17.
- Weisberg I., Tran P., Christensen B., Sibani S., Rozen R. // Mol. Genet. Metab. 1998. V. 64. P. 169–172.
- Callejón G., Mayor-Olea A., Jiménez A.J., Gaitán M.J., Palomares A.R., Martínez F., Ruiz M., Reyes-Engel A. // Hum. Reprod. 2007. V. 22. P. 3249–3254.
- Branco C.C., Pereirinha T., Cabral R. // Thromb. J. 2009. V. 7. P. 4–9.
- Rady P.L., Szucs S., Grady J., Hudnall S.D., Kellner L.H., Nitowsky H., Tyring S.K., Matalon R.K. // Am. J. Med. Genet. 2002. V. 107. P. 162–168.
- Melo S.S., Persuhn D.C., Meirelles M.S., Jordao A.A., Vannucchi H. // Mol. Nutr. Food Res. 2006. V. 50. P. 769–774.
- Mao R., Fan Y., Chen F., Sun D., Bai J., Fu S. // Cell Biochem. Funct. 2008. V. 26. P. 352–358.
- Evans D., Cardon L.R. // Am. J. Hum. Genet. 2005. V. 76. P. 681–687.
- Collins A. // Mol. Biotechnol. 2009. V. 41. P. 83–89.
- Phillips M.S. // Nat. Genet. 2003. V. 33. P. 382–387.

42. Wang N., Akey J.M., Zhang K., Chakraborty R., Jin L. // *Am. J. Hum. Genet.* 2002. V. 71. P. 1227–1234.
43. Wall J.D., Pritchard J.K. // *Nat. Rev. Genet.* 2003. V. 4. P. 587–597.
44. Templeton A.R., Clark A.G., Weiss K.M., Nickerson D.A., Boerwinkle E., Sing C.F. // *Am. J. Hum. Genet.* 2000. V. 66. P. 69–83.
45. Barrett W.A., Doshi J.M., Hacker C.R., Kautzer C.R., Lee D.H., Marjoribanks C., McDonough D.P., et al. // *Science.* 2001. V. 294. P. 1719–1723.
46. Rana N.A., Ebenezer N.D., Webster A.R., Linares A.R., Whitehouse D.B., Povey S., Hardcastle A.J. // *Hum. Mol. Genet.* 2004. V. 24. P. 3089–3102.
47. Oota H., Pakendorf B., Weiss G., von Haeseler A., Pookajorn S., Settheetham-Ishida W., Tiwawech D., Ishida T., Stoneking M. // *PLoS Biol.* 2005. V. 3. P. 536–542.
48. Reich D.E., Cargill M., Bolk S., Ireland J., Sabeti P.C., Richter D.J., Lavery T., Kouyoumjian R., Farhadian S.F., Ward R., Lander E.S. // *Nature.* 2001. V. 411. P. 199–204.
49. De La Vega F.M., Su X., Avi-itzhak H. // *Am. J. Hum. Genet.* 2002. V. 71. P. 205–213.
50. Liu N., Sawyer S.L., Mukherjee N., Pakstis A.J., Kidd J.R., Kidd K.K., Brookes A.J., Zhao H. // *Genetic Epidemiol.* 2004. V. 27. P. 385–400.
51. Zhang K., Qin Z., Liu J., Chen T., Waterman M.S., Sun F. // *Genome Res.* 2004. V. 14. P. 908–916.
52. Service S., Sabatti C., Freimer N. // *Genetic Epidemiol.* 2007. V. 31. P. 189–194.
53. Gonzalez-Neira A., Ke X., Lao O., Lao O., Calafell F., Navarro A., Comas D., Cann H., Bumpstead S., Ghori J., et al. // *Genome Res.* 2006. V. 16. P. 323–330.
54. Barrett J.C., Cardon L.R. // *Nat. Genet.* 2006. V. 38. P. 659–662.
55. Hao K. // *Bioinformatics.* 2007. V. 23. № 23. P. 3178–3184.
56. Carlson C.S., Eberle M.A., Rieder M.J., Smith J.D., Kruglyak L., Nickerson D.A. // *Nat. Genet.* 2003. V. 33. P. 518–521.
57. Halperin E., Kimmel G., Shamir R. // *Bioinformatics.* 2005. V. 21. P. 195–203.
58. Montpetit A., Nelis M., Laflamme P., Magi R., Ke X., Remm M., Cardon L., Hudson T.J., Metspalu A. // *PLoS Genet.* 2006. V. 2. № 3. e27.
59. Gu S., Pakstis A.J., Li H., Speed W.C., Kidd J.R., Kidd K.K. // *Eur. J. Human Genet.* 2007. V. 15. P. 302–312.
60. Lundmark P.E., Liljedahl U., Boomsma D.I., Mannila H., Martin N.G., Palotie A., Peltonen L., Perola M., Spector T.D., Syvänen A.C. // *Eur. J. Hum. Genet.* 2008. V. 16. № 9. P. 1142–1150.
61. de Bakker P.I., Burt N.P., Graham R.R., Guiducci C., Yelensky R., Drake J.A., Bersaglieri T., Penney K.L., Butler J., Young S., et al. // *Nat. Genet.* 2006. V. 38. P. 1298–1303.
62. Hu C., Jia W., Zhang W., Wang C., Zhang R., Wang J., Ma X. // *BMC Genet.* 2008. V. 27. P. 9–19.
63. Степанов В.А. Этногеномика населения Северной Евразии. Томск: Печатная мануфактура, 2002. 244 с.
64. Горлов И.П., Горлова О.Ю. // *Вестник ВОГиС.* 2007. Т. 11. № 2. С. 363–372.
65. Kidd K.K., Kidd J.R. *Human genetic variation of medical significance. Evolution in Health and Disease.* N.Y.: Oxford Univ. Press, 2008. 363 p.
66. Tenesa A., Navarro P., Hayes B.J., Duffy D.L., Clarke G.M., Goddard M.E., Visscher P.M. // *Genome Res.* 2007. V. 17. № 4. P. 520–526.
67. Spiroski I., Kedev S., Antov S., Arsov T., Krstevska M., Dzhekova-Stojkova S., Kostovska S., Trajkov D., Petlichkovski A., Strezova A., et al. // *Croat Med. J.* 2008. V. 49. P. 39–49.
68. Mayor-Olea A., Callejon G., Palomares A.R., Jiménez A.J., Gaitán M.J., Rodríguez A., Ruiz M., Reyes-Engel A. // *BMC Med. Genet.* 2008. V. 9. P. 104.
69. Haggarty P., Campbell D.M., Duthie S., Andrews K., Hoad G., Piyathilake C., Fraser I., McNeill G. // *BJOG.* 2008. V. 115. № 7. P. 851–856.
70. Guéant-Rodriguez R.M., Guéant J.L., Debard R., Thirion S., Hong L.X., Bronowicki J.P., Namour F., Chabi N.W., Sanni A., Anello G., et al. // *Am. J. Clin. Nutr.* 2006. V. 83. № 3. P. 701–707.
71. Wilcken B., Bamforth F., Li Z., Zhu H., Ritvanen A., Renlund M., Stoll C., Alembik Y., Dott B., Czeizel A.E., et al. // *J. Med. Genet.* 2004. V. 41. № 8. P. 619–625.