

*На правах рукописи*

ФИЛИППОВА ИРИНА НИКОЛАЕВНА

ПОПУЛЯЦИОННОЕ РАЗНООБРАЗИЕ ГЕНОМНЫХ КЛАСТЕРОВ  
ГЛУТАТИОН-S ТРАНСФЕРАЗНЫХ ГЕНОВ ЧЕЛОВЕКА

03.01.03 - Молекулярная биология

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва

2015

Работа выполнена в отделе молекулярных основ генетики человека Федерального государственного бюджетного учреждения науки Института молекулярной генетики Российской академии наук (ИМГ РАН)

**Научный руководитель:** Хрунин Андрей Владимирович, кандидат биологических наук, ФГБУН ИМГ РАН.

**Официальные оппоненты:** Спицын Виктор Алексеевич, д.б.н., профессор, заведующий лабораторией экологической генетики, Федеральное государственное бюджетное научное учреждение Медико-генетический научный центр.

Лебедев Юрий Борисович, д.б.н., профессор, заведующий лабораторией сравнительной и функциональной геномики, Федеральное государственное бюджетное учреждение науки Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук.

**Ведущая организация:** Федеральное государственное бюджетное учреждение науки Институт биологии гена Российской академии наук

Защита состоится «\_\_» \_\_\_\_\_ 2015г. на заседании Диссертационного совета Д217.013.01 при ФГУП «Государственный научно-исследовательский институт генетики и селекции промышленных микроорганизмов» по адресу: 117545, Москва, 1-й Дорожный проезд, д.1.

С диссертацией можно ознакомиться в библиотеке ФГУП «ГосНИИгенетика» и на сайте [www.genetika.ru](http://www.genetika.ru)

Автореферат разослан «\_\_» \_\_\_\_\_ 2015г.

Ученый секретарь

диссертационного совета,

кандидат химических наук, доцент

Воюшина Т.Л.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Большую роль в развитии современных направлений молекулярной биологии и генетики сыграла грандиозная работа по расшифровке генома человека. По её завершению на одно из первых мест по значимости выдвинулась проблема изучения геномного разнообразия различных популяций и этнических групп. Актуальность этих исследований связана с тем, что они позволяют выявить новые аспекты структурно-функциональной организации генома, особенности его эволюционного развития, а также определить вклад полиморфных участков генома в возникновение и риски развития заболеваний и реакций на внешне средовые воздействия, включая лекарственные средства.

За последнее время важным подходом в изучении особенностей генофонда различных популяций и этнических групп стало полногеномное исследование его полиморфизма с применением технологий ДНК-микрочипов (полногеномное сканирование). Этот метод позволяет одновременно анализировать аллельный статус сотен тысяч полиморфных участков и получать глобальные сведения об изменчивости генома с высокой степенью точности и надёжности.

Кроме глобальных сведений, результаты сканирования содержат данные о популяционной изменчивости отдельных геномных регионов. Регионы могут иметь свои собственные эволюционные траектории в популяционной истории, и информацию об этом можно извлечь из результатов полногеномного генотипирования популяций. Важная роль, которую играют в организме системы генов глутатион S-трансфераз, определяет значимость этих геномных регионов для изучения их популяционной изменчивости, основываясь на анализе данных полногеномного сканирования различных популяций.

Цели и задачи исследования: Целью настоящей работы является анализ аллельного и гаплотипического разнообразия в двух кластерах генов глутатион-S трансфераз, принадлежащим к классам мю (*GSTM*) и альфа (*GSTA*) в русских популяциях.

Для достижения цели были поставлены следующие задачи:

1. С использованием данных полногеномного типирования однонуклеотидных полиморфных замен в трех русских популяциях провести анализ их аллельного разнообразия в геномных регионах кластеров генов *GSTM* и *GSTA* в сравнении с другими популяциями (популяции китайцев (CHB), японцев (JPT), нигерийцев (YRI) и европейцев по происхождению из американского штата Юта (CEU)).
2. Исследовать гаплотипическое разнообразие тех же кластеров в указанных популяциях.
3. Провести анализ гаплотипического разнообразия кластера генов *GSTM* в трех русских популяциях в подгруппах индивидуумов, различающихся по наличию или отсутствию делеции гена *GSTM1* (делеционный полиморфизм гена *GSTM1*).
4. Определить возможные взаимосвязи между аллельным статусом делеционного полиморфизма гена *GSTM1* и гаплотипическим разнообразием в кластере генов *GSTM*.

Научная новизна и практическая значимость работы. В ходе проведения диссертационной работы было исследовано гаплотипическое и аллельное разнообразие двух геномных кластеров генов глутатион-S трансфераз, принадлежащим к классам мю и альфа в русских популяциях.

Однонуклеотидный полиморфизм в кластерах генов *GSTA* и *GSTM*, расположенных на разных хромосомах, имеет характеристики, указывающие на сходные эволюционные траектории в обоих участках генома. Показано отсутствие выраженных отличий между тремя русскими популяциями и европейской по происхождению популяцией CEU из проекта HarMap. В сравнительном анализе с популяциями японцев, китайцев и нигерийцев выявлены достоверные отличия. Проведенный анализ гаплотипического разнообразия обоих кластеров *GSTs* обнаружил сходство по некоторым гаплоглокам между европейскими выборками (включая русских) и нигерийской популяцией, что

указывает на схожие траектории в эволюционном развитии этих участков генома в данных общностях. Сопоставление гаплотипического разнообразия геномного кластера *GSTM* с делецией гена *GSTM1* в трех русских популяциях показало статистически значимые различия в отдельных гапблоках между подгруппами гомозигот по делеции и остальными индивидуумами. Анализ в объединенной русской популяции показал существование достоверных характеристик гаплотипического разнообразия кластера *GSTM* у лиц, имеющих делецию гена *GSTM1*, что может указывать либо на «эффект основателя», либо на возможность предпочтительного возникновения делеции в определенных вариантах геномного окружения. Выявленный в работе ряд ассоциаций, описывается впервые.

Полученные результаты, в целом, расширяют имеющиеся представления об особенностях структурной организации кластеров генов *GSTA* и *GSTM*, белковые продукты которых имеют большое значение в процессах детоксикации. Они могут быть основой для разработки новых подходов в фармакогеномике различных лекарственных препаратов, а также для создания панелей генетических маркеров для тестирования реакций больных на лекарства.

Положения, выносимые на защиту:

1. Геномные регионы, включающие кластеры генов глутатион-S-трансфераз А и М, имеют сходную эволюционную историю.
2. Отдельные геномные блоки (гапблоки) в кластерах генов глутатион-S-трансфераз А и М могут иметь специфичные эволюционные пути в различных популяционных общностях.
3. Вероятность возникновения делеции гена *GSTM1* на конкретной хромосоме может зависеть от его геномного окружения.

Публикации и апробация работы. Диссертационная работа была представлена на заседании Ученого совета ИМГ РАН 27 октября 2014г., протокол №11 и 09 декабря 2014 года на заседании секции «Молекулярная биология» ученого совета ФГУП «ГосНИИгенетика». Основные результаты работы были представлены на международной конференции «The Biology of Geneomes» (Cold Spring Harbor, USA, 2012) и V Международной Школе молодых учёных по молекулярной

генетике «Непостоянство генома» (Звенигород, Россия, 2012). По материалам работы опубликовано 3 статьи в журналах, рекомендованных ВАК, а также тезисы в сборниках материалов конференций.

Структура и объем работы. Диссертационная работа состоит из введения, обзора литературы, материалов и методов исследования, результатов, обсуждения, выводов и списка литературы. Работа изложена на 102 страницах машинописного текста, включает 14 таблиц, 11 рисунков. Список цитируемых литературных источников включает 181 наименование.

### СОДЕРЖАНИЕ РАБОТЫ

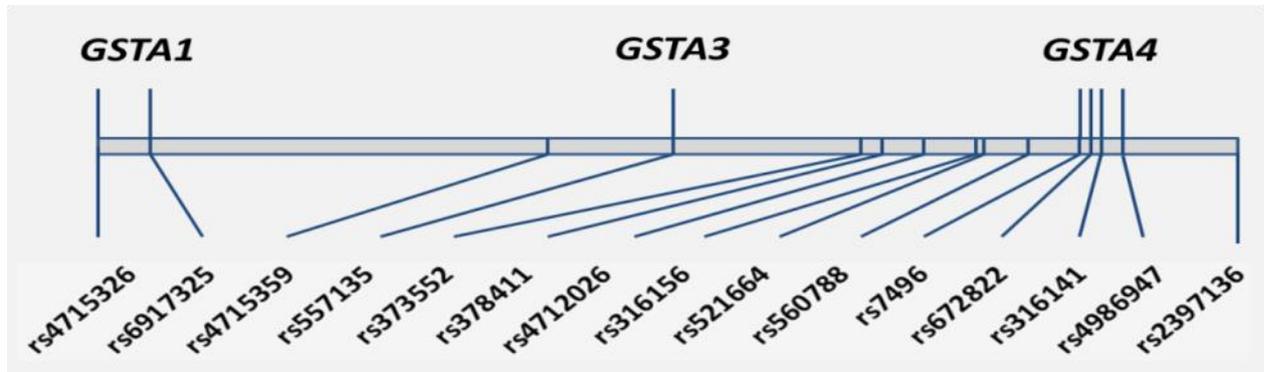
Настоящая работа посвящена изучению гаплотипического и аллельного разнообразия двух геномных кластеров генов глутатион-S трансфераз, принадлежащим к классам мю и альфа в русских популяциях.

Для проведения анализа генетического разнообразия двух геномных кластеров – *GSTA* и *GSTM*, были использованы результаты полногеномного сканирования русских популяций, полученные с использованием ДНК микрочипов фирмы Illumina (Khrunin et al., 2013). Из всего массива данных, нами были отобраны однонуклеотидные полиморфные локусы, которые располагаются в районах локализации вышеупомянутых кластеров генов. Полученный набор локусов был проанализирован в трех русских популяциях, а также в нескольких популяциях, взятых из международного проекта ХарМар. Были проанализированы три популяции русских из центральных районов европейской России – Андреапольского района Тверской области, рядом расположенных Курского и Октябрьского районов Курской области и Муромского района Владимирской области, а также четыре популяции из международного проекта ХарМар: европейская (CEU), китайская (CHB), японская (JPT) и нигерийская (YRI).

#### **1. Анализ однонуклеотидного полиморфизма в кластере генов *GSTA*.**

На рисунке 1 изображены исследованные однонуклеотидные полиморфные локусы локализуемых в состав кластера генов *GSTA*. Расположение полиморфизмов приведено в соответствии с их локализацией относительно генов

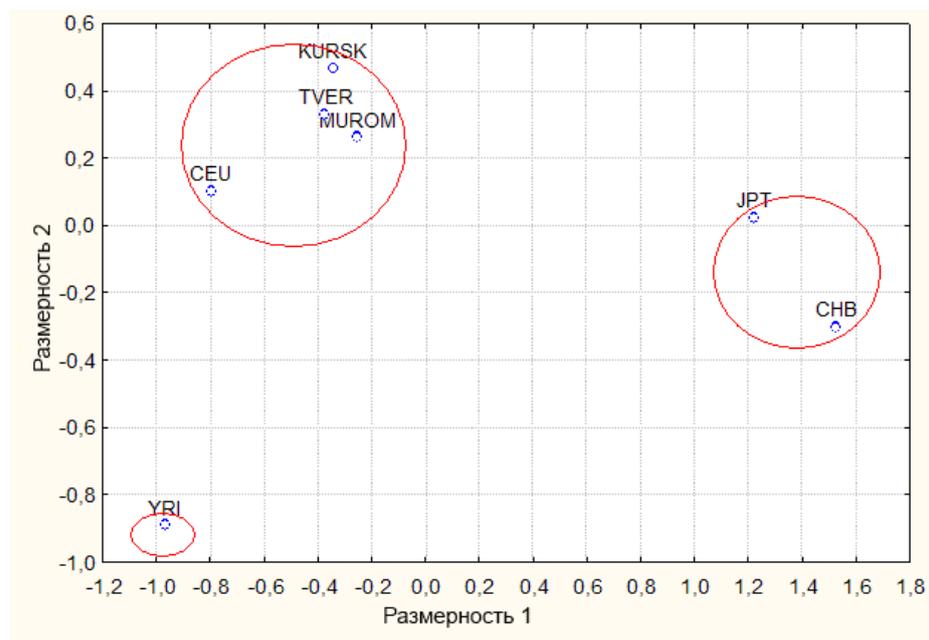
данного кластера. Всего в кластере генов *GSTA*, было проанализировано 15 однонуклеотидных полиморфных маркеров.



**Рисунок 1. Исследованные однонуклеотидные полиморфизмы в кластере генов *GSTA*.**

Сначала нами был проведен анализ частот встречаемости аллельных вариантов каждого из полиморфизмов в изучаемых популяциях. Сравнительный анализ частот показал отсутствие различий в распределении аллелей исследованных локусов в популяциях русского населения. Близкие к ним параметры частот встречаемости аллелей были отмечены в европейской по происхождению популяции CEU. Распределение же аллелей в трех других популяциях из проекта ХарМар – китайской, японской и нигерийской – в большинстве случаев существенно отличалось от наблюдаемого в выборках населения европейского происхождения.

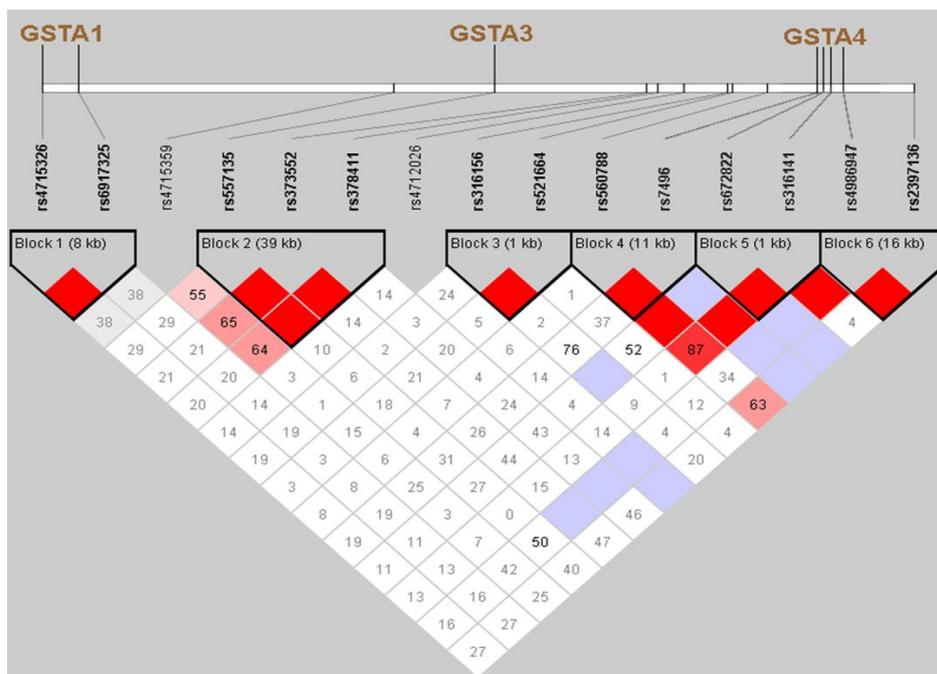
Для количественной оценки уровней межпопуляционных вариаций частот аллелей анализируемых локусов нами были рассчитаны индексы фиксации  $F_{ST}$ . На рисунке 2 представлены результаты обработки полученной матрицы линеаризованных попарных величин  $F_{ST}$  методом многомерного шкалирования. Из диаграммы видно, что русские популяции входят в состав единого кластера вместе с близкорасположенной к ним популяцией CEU, а африканская популяция YRI и азиатские CHB и JPT находятся на значительном удалении от них.



**Рисунок. 2. Двумерный график многомерного шкалирования генетических расстояний ( $F_{ST}$ ) между популяциями русских и популяциями из проекта ХарМар, полученный по результатам анализа однонуклеотидного полиморфизма в кластере *GSTA*.**

Наряду с оценкой полиморфизма отдельных локусов был также проведен анализ их полиморфизма в составе гаплотипов. Для этого нами сначала были оценены уровни неравновесия по сцеплению между всеми исследованными полиморфизмами, на основе которых далее были вычленены гаплотипические блоки. Последние представляют собой группы локусов, характеризующихся высоким уровнем сцепления между собой. В качестве порогового значения для объединения отдельных полиморфизмов в составе гапблока была использована величина  $D' \geq 0.7$  (Khrunin et al., 2009).

По результатам анализа в составе кластера *GSTA* было выделено 6 гаплотипических блоков (рисунок 3). Гапблоку кластера генов *GSTA* были идентичными для всех исследуемых популяций русского населения из центрально-европейского региона России. В качестве примера, на рисунке 3 представлен график паттернов неравновесия по сцеплению, полученный для популяции русских из Андреапольского района Тверской области.



**Рисунок. 3. Паттерны неравновесия по сцеплению и гапоблоки, вычлененные в регионе кластера генов *GSTA***

Данные, полученные для двух других русских популяций – популяции из Муромского района Владимирской области и объединенной выборки из Курского и Октябрьского районов Курской области, имели сходные значения. Частоты встречаемости гаплотипов в составе выделенных гапоблоков в анализируемых популяциях приведены в таблице 1. В отдельных популяциях ряд гапоблоков не вычленялся.

Сравнение популяций основывалось на сопоставлении частот гаплотипов, рассчитанных для каждого из блоков. Результаты сравнений в виде полученных величин вероятности «P» представлены в таблице 2. Уровни значимости «P» были рассчитаны с учетом поправки на множественность сравнений для каждого из блоков. Полученные данные свидетельствуют об отсутствии выраженных различий в частотах встречаемости гаплотипов между популяциями русских из Твери, Мурома, Курска и европейской по происхождению популяцией CEU. В случае же сравнения частот гаплотипов в популяциях русских с популяциями китайцев, японцев и нигерийцев можно говорить о достоверных различиях между ними. Величины большинства значений вероятности были существенно ниже установленных уровней значимости.

Таблица 1

Частоты гаплотипов в 6 вычлененных блоках кластера *GSTA*

Гаплотипы	Тверь 2N=192	Муром 2N=192	Курск 2N=192	CEU 2N=330	CHB 2N=168	JPT 2N=172	YRI 2N=332
<u>1БЛОК</u>							
TC	0.604	0.630	0.590	0.594	0.905	0.872	0.701
CT	0.396	0.370	0.410	0.406	0.095	0.128	0.299
<u>2БЛОК</u>							
AAA	0.682	0.686	0,670	0.652	0.500	0.523	0.311
GGG	0.245	0.234	0,217	0.279	0.042	0.105	0.021
GAA	0.073	0.079	0,107	0.107	0.458	0.372	0.641
GAG	0	0	0	0	0	0	0.027
<u>3БЛОК</u>							
TT	0.469	0.479	0.479	0.551	0.214	0.163	0.810
TC	0.396	0.375	0.432	0.288	0.560	0.547	0.073
GT	0.135	0.146	0.089	0.161	0.226	0.291	0.112
<u>4 БЛОК</u>							
CC	0.505	0.529	0.494	0.534	0.755	0.767	0.361
TC	0.323	0.299	0.374	0.367	0.097	0.146	0.391
TT	0.172	0.164	0.120	0.094	0.136	0.081	0.238
CT	0	0	0,011	0	0.013	0	0
<u>5БЛОК</u>							
AG	0.557	0.562	0.547	0.556	-	-	0.518
AA	0.380	0.306	0.351	0.340	-	-	0.422
GA	0.063	0.131	0.101	0.104	-	-	0.060
<u>6БЛОК</u>							
GA	0.849	0.869	0.831	0.855	-	-	-
AG	0.078	0.046	0.089	0.055	-	-	-
GG	0.073	0.079	0.074	0.091	-	-	-

Имеющиеся исключения относятся к блоку №5 и в некоторой мере к блокам №1 и №4, где значения  $P$  для попарных сравнений частот гаплотипов в популяциях русских и нигерийцев были выше величин « $P$ », определенных для соответствующих блоков.

Таблица 2.

Результаты сравнения частот гаплотипов генов кластера *GSTA* в популяциях Твери, Мурома, Курска и в популяциях проекта НарМар (CEU, CHB, JPT, YRI)

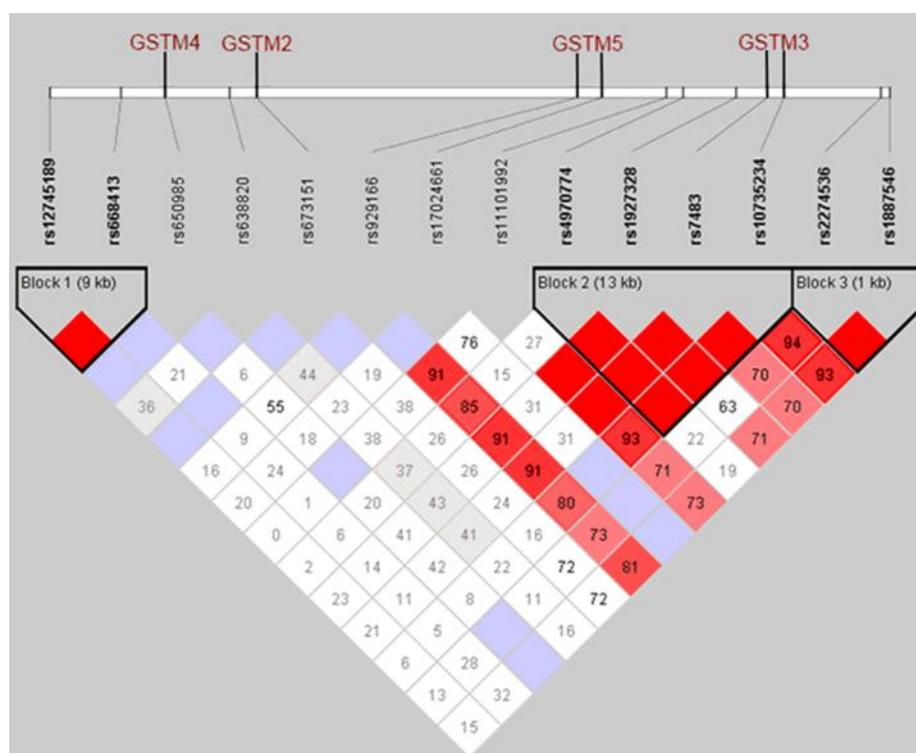
	№ ГАПЛО-БЛОКА	МУРОМ	КУРСК	CEU	CHB	JPT	YRI	$P_{\text{МНОЖ}}^*$
ТВЕРЬ	I	0,6746	0,8353	0,8534	<0,0001	<0,0001	0,0269	0,004
	II	0,9599	0,4282	0,6984	<0,0001	<0,0001	<0,0001	0,0025
	III	0,9030	0,3306	0,0403	<0,0001	<0,0001	<0,0001	0,0025
	IV	0,8238	0,2076	0,0326	<0,0001	<0,0001	0,0065	0,0019
	V	0,0484	0,3957	0,2428	—	—	0,6452	0,0025
	VI	0,4538	0,9276	0,4634	—	—	—	0,0025
МУРОМ	I		0,4641	0,4574	<0,0001	<0,0001	<0,0001	0,004
	II		0,5661	0,5307	<0,0001	<0,0001	<0,0001	0,0025
	III		0,1764	0,1188	<0,0001	<0,0001	<0,0001	0,0025
	IV		0,1599	0,0408	<0,0001	<0,0001	0,0006	0,0019
	V		0,5052	0,5524	—	—	0,0031	0,0025
	VI		0,2664	0,8195	—	—	—	0,0025
КУРСК	I			0,9266	<0,0001	<0,0001	0,0097	0,004
	II			0,1321	<0,0001	<0,0001	<0,0001	0,0025
	III			0,0013	<0,0001	<0,0001	<0,0001	0,0025
	IV			0,2053	<0,0001	<0,0001	0,0006	0,0019
	V			0,9682	—	—	0,1181	0,0025

\* $P_{\text{МНОЖ}}$  – пороговые значения вероятности, рассчитанные с учетом поправки на множественность сравнений для каждого из блоков

## 2. Анализ однонуклеотидного полиморфизма в кластере генов *GSTM*.

В пределах участка генома, на котором располагается кластер генов *GSTM*, было проанализировано 14 маркерных локусов (рисунок 4). Как и в случае кластера генов *GSTA*, в распределении частот аллелей кластера *GSTM* также прослеживается сходство между исследованными популяциями русского населения, а также между ними и европейской по происхождению популяцией CEU. Различия в частотах аллелей наблюдаются при сравнении их с выборками из Азии (CHB и JPT) и Африки (YRI). Двумерный график многомерного шкалирования матрицы рассчитанных для данного кластера  $F_{ST}$  демонстрирует высокое сходство с приводившимся ранее графиком кластера *GSTA*.

По результатам анализа паттернов неравновесия по сцеплению между исследованными локусами в составе кластера *GSTM* было выделено 3 гаплотипических блока (рисунок 4).



**Рисунок 4.** Исследованные однонуклеотидные полиморфизмы и паттерны неравновесия по сцеплению внутри кластера генов *GSTM*.

Данные о частотах гаплотипов для каждого из гапоблоков в кластере *GSTM* представлены в таблице 3. Как и в случае кластера *GSTA*, внутри блоков

наблюдалось различное количество гаплотипов. Основываясь на данных таблицы 4, где приведены значения «P» для попарных сравнений, можно говорить о том, что сколько-нибудь выраженные различия в частотах встречаемости гаплотипов между популяциями русских из Твери, Муром, Курска, а также CEU отсутствуют. В то же время статистически значимые отличия наблюдались в большинстве случаев сравнений частот гаплотипов между популяциями русских и популяциями китайцев, японцев и нигерийцев. Имеющиеся исключения относятся лишь к блоку №1, значения «P» которого для попарных сравнений частот гаплотипов между популяциями русских и нигерийцев был значительно выше установленного уровня значимости.

Таблица 3

Частоты гаплотипов в гапблоках кластера *GSTM*

Гаплотипы	Тверь 2N=192	Муром 2N=192	Курск 2N=192	CEU 2N=330	CHB 2N=168	JPT 2N=172	YRI 2N=332
<u>1 БЛОК</u>							
TC	0.443	0.458	0.495	0.395	0,628	0,654	0,451
CA	0.458	0.411	0.400	0.456	0,314	0,297	0,414
CC	0.099	0.130	0.105	0.142	0,058	0,036	0,136
TA	0	0	0	0	0	0,012	0
<u>2 БЛОК</u>							
AACG	0.391	0.365	0.378	0.430	0.150	0.179	0.092
CGTA	0.266	0.339	0.287	0.297	0.743	0.690	0.035
AACA	0.198	0.167	0.213	0.130	0.083	0.071	0.377
CACA	0.141	0.120	0.122	0.142	0	0	0.494
CGTG	0	0	0	0	0.013	0	0
AATA	0	0	0	0	0.012	0	0
AGTA	0	0	0	0	0	0.060	0
<u>3 БЛОК</u>							
AG	0.693	0.687	0.658	0.712	0.895	0.881	0.364
GT	0.302	0.313	0.337	0.288	0.105	0.095	0.601
GG	0	0	0	0	0	0.024	0.035

Таблица 4

**Результаты сравнения частот гаплотипов кластера генов *GSTM* популяций Твери, Муром, Курска и популяций проекта ХарМар (CEU, CHB, JPT, YRI)**

	№ ГАПЛО- БЛОКА	МУРОМ	КУРСК	CEU	CHB	JPT	YRI	$P_{\text{МНОЖ}}$
ТВЕРЬ	I	0,5078	0,5183	0,2862	0,0021	<0,0001	0,4485	0,0025
	II	0,3202	0,8943	0,2160	<0,0001	<0,0001	<0,0001	0,0019
	III	0,9120	0,9206	0,7647	<0,0001	<0,0001	<0,0001	0,0025
МУРОМ	I		0,6541	0,3994	0,0027	<0,0001	0,5856	0,0025
	II		0,4052	0,1369	<0,0001	<0,0001	<0,0001	0,0019
	III		0,9392	0,5527	<0,0001	<0,0001	<0,0001	0,0025
КУРСК	I			0,0811	0,0313	0,0018	0,1701	0,0025
	II			0,0906	<0,0001	<0,0001	<0,0001	0,0019
	III			0,8592	<0,0001	<0,0001	<0,0001	0,0025

Таким образом, хотя гены подсемейств *GSTA* и *GSTM* находятся на разных хромосомах, проведенный нами анализ полиморфизма этих двух генных кластеров продемонстрировал значительное сходство соответствующих им паттернов вариации аллельных и гаплотипических частот, зафиксированных в исследованных популяциях. Спектры гаплотипов в русских популяциях из Твери, Муром и Курска были похожи. Также не было выявлено заметных различий между ними и популяцией CEU, что, очевидно, является следствием их общего европейского происхождения. В этом контексте особенно интересным было обнаружить определенное сходство в частотах гаплотипов некоторых блоков (блок 5 в кластере *GSTA* и блок 1 в кластере *GSTM*) между русскими популяциями и популяцией YRI из Нигерии. Так как европейские популяции кардинально отличались от популяций из Китая и Японии по частотам гаплотипов всех гапоблоков, можно предположить, что отмеченное сходство может быть следствием каких-либо особых черт этих гапоблоков в микроэволюционной

истории популяций. В то же время русские и нигерийская популяции статистически значимо отличались по всем остальными гапблокам обоих генных кластеров.

### **3. Анализ делеционного полиморфизма гена *GSTM1* в контексте гаплотипического разнообразия геномного кластера *GSTM*.**

Особенностью кластера генов *GSTM*, в сравнении с другим исследованным нами кластером *GSTA*, является делеция его значительной части – около 16kb, захватывающая целиком ген *GSTM1* и встречающаяся с высокой частотой во многих популяциях человека. Представляло интерес изучить, есть ли взаимосвязи между делецией гена *GSTM1* и представленностью тех или иных гаплотипов в составе гаплотипических блоков кластера *GSTM1*.

Классический метод ПЦР по анализу делеционного полиморфизма гена *GSTM1*, основывается на одновременной амплификации участка данного гена и используемого в качестве контроля участка какого-либо другого гена. В нашем исследовании в качестве такового был взят участок гена ГТР-циклогидролазы [Khrunin A. et. al., 2008]. В результате удалось выделить в составе каждой популяции две подгруппы индивидуумов, у которых ген *GSTM1* либо отсутствовал вовсе (подгруппа *GSTM1* «0», гомозиготы по делеции), либо имелся хотя бы один его функциональный вариант (подгруппа *GSTM1* «+»: нормальные гомозиготы и гетерозиготы по делеции). По результатам анализа неравновесия по сцеплению в анализируемом регионе было выделено три гаплотипических блока (рисунок 4).

В каждой из подгрупп нами был проведен анализ паттернов неравновесия по сцеплению и вычленены гаплотипические блоки. Частоты встречаемости конкретных гаплотипов в подгруппах *GSTM1* «0» и *GSTM1* «+» в каждой из исследованных популяций отображены в таблице 5.

Таблица 5.

Частоты гаплотипов в подгруппах индивидуумов, несущих функциональные (0/+ и +/+; *GSTM1* «+») и нулевые (0/0; *GSTM1* «0») варианты гена *GSTM1* в трех русских популяциях

№ гапто-блока	Гаплотип	Тверь		<i>P</i>	Муром		<i>P</i>	Курск		<i>P</i>
		<i>GSTM1</i> «0» 2N=86	<i>GSTM1</i> «+» 2N=106		<i>GSTM1</i> «0» 2N=90	<i>GSTM1</i> «+» 2N=102		<i>GSTM1</i> «0» 2N=80	<i>GSTM1</i> «+» 2N=106	
1	TC	0,326	0,538	0,0120	0,422	0,490	0,0851	0,423	0,566	0,1149
	CA	0,558	0,377		0,489	0,343		0,474	0,330	
	CC	0,116	0,085		0,089	0,167		0,103	0,104	
2	AACG	0,337	0,434	0,0009	0,411	0,324	0,0079	0,423	0,346	0,4665
	AACA	0,186	0,208		0,122	0,206		0,154	0,240	
	CACA	0,070	0,198		0,056	0,176		0,115	0,135	
	CGTA	0,395	0,160		0,411	0,275		0,308	0,279	
	CGCA	0,012	0		0	0		0	0	
	AGCG	0	0		0	0,020		0	0	
3	AG	0,779	0,628	0,0191	0,811	0,578	0,0006	0,731	0,623	0,0785
	GT	0,209	0,377		0,189	0,422		0,256	0,377	
	GG	0,012	0		0	0		0,013	0	

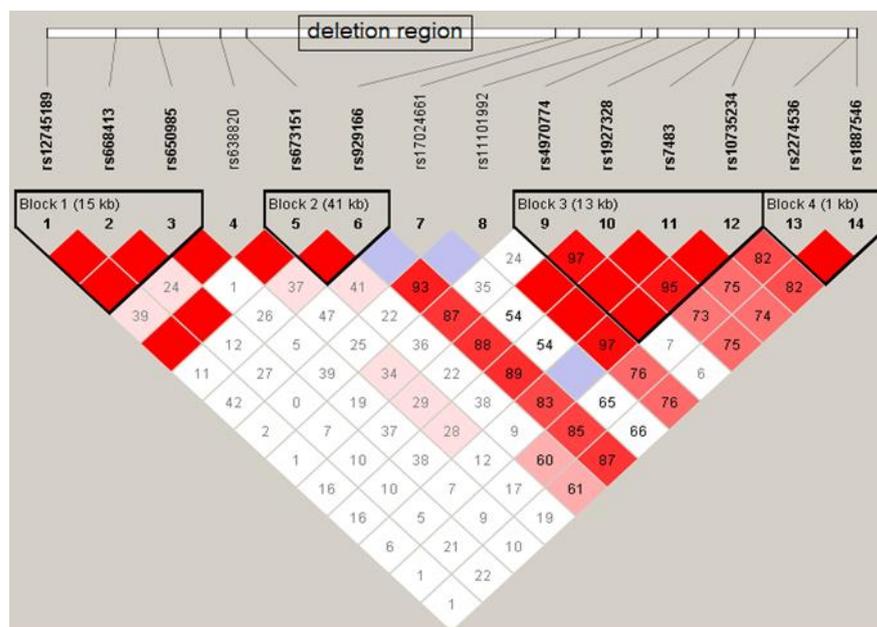
Из представленных в таблице 5 данных видно, что в первом гаплоглоке частоты гаплотипов в подгруппах *GSTM1*«0» и *GSTM1*«+» во всех 3-х анализируемых популяциях русского населения имеют схожие значения. По данному гаплоглоку достоверных различий ( $P < 0.0028$ ) между изучаемыми подгруппами *GSTM1*«0» и *GSTM1*«+» не было выявлено ни в одной из популяций.

Несколько иная ситуация наблюдается при рассмотрении второго гаплоглока. Здесь обращает на себя внимание факт существования различий в количестве выделенных гаплотипов как между анализируемыми популяциями, так и между выделенными подгруппами индивидуумов внутри каждой популяции. Статистически значимыми в данном гаплоглоке являются различия в распределении частот гаплотипов между подгруппами *GSTM1*«0» и *GSTM1*«+» в тверской популяции. Третий гаплоглок характеризуется разницей по числу гаплотипов. Здесь достоверные различия между изучаемыми подгруппами наблюдаются лишь в популяции из Муром

В целом, в данном контексте популяция из Курской области выглядит более гомогенной, в сравнении с популяциями из Тверской и Владимирской областей. В двух последних различия в частотах встречаемости гаплотипов между подгруппами *GSTM1* «0» и *GSTM1* «+» сохраняются по отдельным блокам даже после введения поправки на множественность сравнений ( $P$ ), равной 0.0028. Это относится ко второму блоку популяции из Тверской области и третьему – в случае популяции из Владимирской области.

Мы предположили, что неоднозначность результатов сравнений распределений гаплотипов между подгруппами в исследованных популяциях могла быть следствием малого размера самих этих подгрупп, недостаточных для эффективного расчета частот гаплотипов. Поэтому на следующем этапе мы провели такого же рода анализ в объединенной популяции, включающей выборки из популяций Твери, Муром и Курска. Правомочность данного объединения основывалась на данных проведенного нами полногеномного анализа однонуклеотидного полиморфизма в популяциях Европейской части

России (Khrunin et al., 2013), где было продемонстрирована практически полная перекрываемость этнических облаков, соответствующих курской, тверской и муромской популяциям, и их дифференция от остальных популяций. Изменения в схеме анализа, т. е. переход от анализа разнообразия гаплотипов в индивидуальных популяциях к его анализу в составе объединенной популяции, привели к изменениям в количестве вычлняемых в кластере *GSTM* гапблоков, число которых возросло до четырех (рисунок 5). Кроме того, улучшившаяся статистика параметров неравновесия по сцеплению между локусами rs12745189 и rs650985 позволила нам увеличить размеры 1-го гапблока и, соответственно, длину анализируемых гаплотипов. Частоты гаплотипов в выявленных гапблоках в объединенной популяции русских, включающей выборки русского населения из Тверской, Курской и Владимирской областей, приведены в таблице 6. Из представленных данных видно, что распределения гаплотипов в двух изучаемых подгруппах имеют значительные различия. Так, в первом гапблоке в подгруппе *GSTM1* «+» наиболее частым гаплотипом является – ТСТ, а в подгруппе *GSTM1* «0» – САТ.



**Рисунок 5. Паттерны неравновесия по сцеплению и вычлненные гапблоки в кластере генов *GSTM* в объединенной русской популяции.**

Полученные для каждого из блоков значения вероятности «Р» свидетельствуют о том, что различия в частотах гаплотипов между выделенными подгруппами индивидуумов по делеции являются существенными. Во всех четырех вариантах сравнений их величины были значительно меньше порогового уровня значимости, рассчитанного с учетом поправки на множественность сравнений ( $P=0.006$ ). Результаты сравнений позволили предположить существование взаимосвязи между гаплотипическим разнообразием кластера *GSTM* и делеционным полиморфизмом гена *GSTM1*, в частности, его «нулевым» вариантом (*GSTM1* «0»). Однако оставалась непонятными детали этой связи применительно к взаимоотношениям в подгруппе *GSTM1*«+». Для вычленения из подгруппы *GSTM1*«+» носителей с двумя функциональными вариантами гена *GSTM1* и индивидуумов, гетерозиготных по данному гену, нами был использован метод ПЦР в режиме реального времени в технологии *TaqMan*.

В таблице 7 частоты гаплотипов представлены уже во всех трех подгруппах: гомозиготах по делеции, гетерозиготах и гомозиготах с обеими функциональными вариантами (копиями) гена *GSTM1*. Анализ сравнения частот гаплотипов был нами начат со второго по порядку, блока. Он представляет наибольший интерес в связи с тем, что два SNPs, из которых он состоит, располагаются с обеих сторон от делезируемого региона, причем не на очень большом расстоянии. Из данных таблицы 7 видно, что частоты гаплотипов второго блока различаются между подгруппами по делеции. Так, частота гаплотипа СТ в подгруппе нормальных гомозигот составляет 0,736, а в группе – гомозигот по делеции - лишь 0,583. Для гаплотипа СG этого же гапблока наблюдается противоположная ситуация, его частота оказалась выше в подгруппе гомозигот по делеции.

Сравнения частот встречаемости гаплотипов в подгруппах индивидов показало, что статистически достоверно они различаются в парах нормальные гомозиготы – гомозиготы по делеции и гомозиготы по делеции –

гетерозиготы, но не в паре нормальные гомозиготы – гетерозиготы (таблица 8).

**Таблица 6.**  
**Частоты гаплотипов в двух подгруппах в объединенной выборке 3-х русских популяций**

	Гаплотипы	<i>GSTM1</i> «0» 2N=256	<i>GSTM1</i> «+» 2N=314	<i>P</i>
1 БЛОК	САТ	0.508	0.350	0.0021
	ТСТ	0.390	0.532	
	ССС	0.075	0.082	
	ССТ	0.028	0.036	
2 БЛОК	СТ	0.583	0.710	<0.0001
	СG	0.406	0.217	
	ТТ	0.012	0.074	
3 БЛОК	ААСG	0.390	0.369	0.00008
	СGТА	0.374	0.237	
	ААСА	0.154	0.218	
	САСА	0.079	0.170	
	СGСА	0.004	0	
	АGСG	0	0.006	
4 БЛОК	GТ	0.217	0.392	<0.0001
	АG	0.776	0.608	
	GГ	0.008	0	

Далее мы проанализировали распределение частот гаплотипов в тех блоках, которые находятся на значительно большем удалении от делетируемого региона (блоки 1, 3 и 4). Данные таблицы 7 частот встречаемости гаплотипов показывают, что отмеченное выше разнонаправленное изменение частот гаплотипов здесь также сохраняется.

При этом частоты гаплотипов в подгруппе гетерозигот во всех случаях занимают промежуточное положение. Статистически значимыми, даже с учетом поправки на множественность сравнений ( $P < 0.004$ ), всегда были различия в частотах встречаемости гаплотипов в паре нормальные гомозиготы – гомозиготы по делеции. Несколько меньшее число статистически значимых различий в распределении гаплотипов было выявлено в паре гомозиготы по делеции – гетерозиготы и лишь одно – в паре нормальные гомозиготы – гетерозиготы (таблица 8). Молекулярные основы возникновения делеции в кластере *GSTM* были исследованы в работе Shi-jie Xu et al. [1998], в которой было показано, что у индивидуумов с функциональными вариантами гена *GSTM1* кластер генов *GSTM* содержал два высоко гомологичных (идентичных более чем на 98%) региона, которые с обеих сторон фланкируют ген *GSTM1*.

Таблица 7.

Частоты гаплотипов в 4 гапблоках кластера генов *GSTM*.

	Гаплотипы	Гомозиготы по делеции (2N=256)	Гетерозиготы (2N=242)	Нормальные гомозиготы (2N=72)
1 БЛОК	ТСТ	0.390	0,512	0.597
	САТ	0.508	0,380	0.250
	ССТ	0.028	0,069	0.125
	ССС	0.075	0,039	0.028
2 БЛОК	СТ	0.583	0.705	0.736
	СГ	0.406	0.237	0.139
	ТТ	0.012	0.055	0.125
3 БЛОК	ААСГ	0.390	0,392	0.319
	СГТА	0.374	0,254	0.167
	ААСА	0.154	0,188	0.306
	САСА	0.079	0,158	0.208
	СГСА	0.004	0	0
	АГСГ	0	0.008	0
4 БЛОК	ГТ	0.217	0.343	0.542
	АГ	0.776	0.657	0.458
	ГГ	0.008	0	0

**Таблица 8.**

**Статистика различий частот встречаемости гаплотипов кластера генов *GSTM***

	Гомозиготы по делеции – Нормальные гомозиготы	Гомозиготы по делеции – Гетерозиготы	Нормальные гомозиготы – Гетерозиготы
1 БЛОК	<b>&lt;0,0001</b>	0,0060	0,9512
2 БЛОК	<b>&lt;0,0001</b>	<b>&lt;0,0001</b>	0,0391
3 БЛОК	<b>&lt;0,0001</b>	<b>0,0038</b>	0,0700
4 БЛОК	<b>&lt;0,0001</b>	<b>0,0022</b>	<b>0,0024</b>

Такая структурная организация региона, т.е. наличие высоко гомологичных повторов, свидетельствует в пользу возможности прохождения неаллельной гомологичной рекомбинации на данном участке хромосомы и появлению делеции [Gu et al., 2008], в данном конкретном случае – к делеции гена *GSTM1*. Высокая частота встречаемости делеционных вариантов гена *GSTM1* во всех основных группах популяций человека, включая популяции русского населения, по-видимому, свидетельствует об отсутствии действия отбора против делеции (гомозигот по делеции) и поэтому может рассматриваться как результат ее возникновения в далекой предковой популяции с последующим ее распространением по типу «эффекта основателя». С другой стороны, если высоко гомологичные участки в кластере генов *GSTM1* являются горячими точками рекомбинации, то рассматриваемая делеция, в принципе, могла возникать неоднократно. Полученные нами данные по гаплотипическому разнообразию региона кластера генов *GSTM* в контексте делеционного полиморфизма гена *GSTM1* у русского населения позволяют предположить, что делеция, скорее всего, возникала в разных гаплотипах, так как не было обнаружено различий в составе и количестве гаплотипов между выделенными группами индивидуумов по наличию/отсутствию делеции. В то же время, были выявлены статистически достоверные различия по частоте

встречаемости отдельных гаплотипов между группами, индивидуумов гомозиготных по делеции и гомозиготных по наличию функционального варианта гена *GSTM1*, что указывает на возможность предпочтительного возникновения делеции в определённых вариантах геномного окружения.

## ВЫВОДЫ

1. Полиморфизм однонуклеотидных замен в кластерах генов *GSTA* и *GSTM*, расположенных на разных хромосомах, имеет характеристики, указывающие на сходные эволюционные траектории в обоих участках генома.
2. Анализ гаплотипического разнообразия обоих кластеров *GSTs* выявил наличие в некоторых гапблоках сходства в частотах встречаемости гаплотипов между европейскими популяциями, включая русских, и африканской популяцией из Нигерии, что указывает на схожие эволюционные пути этих геномных блоков в исследуемых общностях.
3. Оценка гаплотипического разнообразия геномного кластера *GSTM* в подгруппах индивидуумов с делецией гена *GSTM1* показала наличие статистически значимых различий в отдельных гапблоках между ними и остальными индивидуумами в трех русских популяциях.
4. Объединение выборок трех русских популяций позволило выявить статистически значимые различия по всем изучаемым гаплотипическим блокам класса *GSTM* в подгруппах гомозигот по делеции гена *GSTM1* и остальными индивидуумами объединенной русской популяции. Это может указывать либо на «эффект основателя», либо на возможность предпочтительного возникновения делеции в определенных вариантах геномного окружения.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Filippova I, Khrunin A, Limborska S.** Analysis of DNA variations in *GSTA* and *GSTM* gene clusters based on the results of genomewide data from three Russian populations taken as an example. **BMC Genetics. 2012;** 13:89.
2. **Khrunin AV, Khokhrin DV, Filippova IN, Esko T, Nelis M, Bebyakova NA, Bolotova NL, Klovins J, Nikitina-Zake L, Rehnström K, Ripatti S, Schreiber S, Franke, Macek M, Krulišová V, Lubinski J, Metspalu A, Limborska SA.** A genome-wide analysis of populations from European Russia reveals a new pole of genetic diversity in northern Europe. **PlosOne. 2013;**8(3):e58552.
3. **Филиппова И.Н., Хрунин А.В., Лимборская С.А.** Анализ делеции гена *GSTM1* в контексте гаплотипического разнообразия геномного кластера *GSTM* в трех русских популяциях. Молекулярная генетика, микробиология и вирусология. 2014;2:8-12.

Материалы конференций:

1. **Limborska S, Khrunin A, Filippova I, Khokhrin D, Bebyakova N, Bolotova N, Esko E, Metspalu A.** Genomic variations in populations from the Far North East corner of Europe. Abstracts of papers presented at the 2012 meeting on "The Biology of Genomes", Cold Spring Harbor, NY, USA, 08 - 12 May 2012, p.167.
2. **Филиппова И.Н.** Геномный полиморфизм регионов *GSTA* и *GSTM* в русских популяциях. Материалы V Международной Школы молодых учёных по молекулярной генетике «Непостоянство генома». 3-7 декабря 2012 г., Звенигород. С. 66. № 117.