

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
ИНСТИТУТ МОЛЕКУЛЯРНОЙ ГЕНЕТИКИ РОССИЙСКОЙ АКАДЕМИИ НАУК

---

*На правах рукописи*

Филиппова Ирина Николаевна

**ПОПУЛЯЦИОННОЕ РАЗНООБРАЗИЕ ГЕНОМНЫХ КЛАСТЕРОВ  
ГЛУТАТИОН-S ТРАНСФЕРАЗНЫХ ГЕНОВ ЧЕЛОВЕКА**

03.01.03 - Молекулярная биология

Диссертация на соискание ученой степени

кандидата биологических наук

Научный руководитель:

к.б.н., с.н.с. А.В. Хрунин

Москва, 2015г.

## ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ.....	4
ВВЕДЕНИЕ.....	5
Актуальность работы.....	5
Цель и задачи исследования.....	6
Научно-практическая значимость работы.....	6
Положения, выносимые на защиту.....	7
Глава 1. ЛИТЕРАТУРНЫЙ ОБЗОР.....	8
1. Система генов глутатион-S трансфераз и кодируемые ею ферменты.....	8
1.2. Полиморфизм системы генов глутатион-S трансфераз.....	17
1.2.1 Явление полиморфизма в геноме человека.....	17
1.2.2. Варианты полиморфизма системы генов глутатион-S-трансфераз.....	21
1.3. Феномен неравновесия по сцеплению. Понятие о гаплотипе.....	30
Глава 2.МАТЕРИАЛЫ И МЕТОДЫ.....	38
Глава 3. РЕЗУЛЬТАТЫ.....	47
3.1 Анализ изменчивости двух кластеров генов <i>GSTA</i> и <i>GSTM</i> .....	47
3.1.1 Анализ изменчивости в кластере генов <i>GSTA</i> .....	47
3.1.2 Анализ изменчивости в кластере генов <i>GSTM</i> .....	56
3.2 Анализ делеционного полиморфизма гена <i>GSTM1</i> в контексте гаплотипического разнообразия геномного кластера <i>GSTM</i> в трех русских популяциях.....	63
3.2.1 Анализ гаплотипического разнообразия геномного кластера <i>GSTM</i> в трех русских популяциях в зависимости от делеционного полиморфизма гена <i>GSTM1</i> .....	63
3.2.2 Анализ гаплотипического разнообразия кластера <i>GSTM</i> в двух подгруппах в объединенной выборке трех русских популяций.....	68
Глава 4. ОБСУЖДЕНИЕ.....	73
4.1 Исследование изменчивости двух кластеров генов <i>GSTA</i> и <i>GSTM</i> , основанном на результатах полногеномного сканирования.....	73

4.2 Исследование делеции гена <i>GSTM1</i> в контексте гаплотипического разнообразия геномного кластера <i>GSTM</i> в трех русских популяциях.....	76
ЗАКЛЮЧЕНИЕ.....	79
ВЫВОДЫ.....	83
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ.....	84

## СПИСОК СОКРАЩЕНИЙ

- GSTs (glutathione S-transferase) –глутатион-S трансфераза;
- GSH ( $\gamma$ -глутамилцистеинилглицин) – молекула глутатиона;
- SNP (single nucleotide polymorphism) – однонуклеотидный полиморфизм;
- MAF (minimum allele frequency) – частота минимального аллеля в популяции;
- LD (linkage disequilibrium) – неравновесие по сцеплению;
- CEU (individuals of northern and western European ancestry living in Utah from the Centre d'Etude du Polymorphisme Humain collection) – популяция европейского происхождения из штата Юта, США;
- CHB (Han Chinese individuals from Beijing, China) – популяция из Китая, Пекин;
- JPT (individuals from Tokyo, Japan) – популяция из Японии, Токио;
- YRI (individuals from the Yoruba in Ibadan, Nigeria – популяция из Африки, Нигерия;
- GSTM1 (0/0) – индивидуумы, являющиеся гомозиготами по делеции гена GSTM1;
- GSTM1 (+/+) – индивидуумы, являющиеся гомозиготами по наличию полнофункциональных копий гена GSTM1;
- GSTM1 (0/+) – индивидуумы, являющиеся гетерозиготами по делеции гена GSTM1;
- CNV (copy number variation) – полиморфизм числа копий участков ДНК.

## ВВЕДЕНИЕ

### 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 сведения об этом можно извлечь из результатов полногеномного генотипирования популяций. Важная роль, которую играют в организме системы генов глутатион S-трансфераз, дает повод начать исследования в данной области именно с анализа этих геномных регионов для изучения их популяционной изменчивости, основываясь на анализе данных полногеномного сканирования различных популяций.

## **Цель и задачи исследования**

Целью настоящей работы является анализ аллельного и гаплотипического разнообразия в двух кластерах генов глутатион-S трансфераз, принадлежащим к классам мю и альфа 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, японцев, нигерийцев и европейцев по происхождению из штата Юта).

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*.

## **Научно-практическая значимость работы**

Результаты работы могут быть использованы в дальнейших популяционно-генетических исследованиях, в частности, для оценки динамики изменения генофонда русских популяций, проживающих на Европейской территории России. Кроме того, они могут быть использованы при изучении эпидемиологии различных заболеваний, предварительной оценке эффективности лекарственной терапии, а также для анализа влияния климато-экологических условий на

генетические особенности популяций, в том числе связанные и с распространенностью многофакторных заболеваний у населения.

**Положения, выносимые на защиту:**

1. Геномные регионы, включающие кластеры генов глутатион-S-трансфераз A1 и M1, имеют сходную эволюционную историю.
2. Отдельные геномные блоки (гапоблоки) в кластерах генов глутатион-S-трансфераз A1 и M1 могут иметь специфичные эволюционные пути в различных популяционных общностях.
3. Вероятность возникновения делеции гена *GSTM1* на конкретной хромосоме может зависеть от его геномного окружения.

## ГЛАВА 1. ЛИТЕРАТУРНЫЙ ОБЗОР

### 1.1. Система генов глутатион-S трансфераз и кодируемые ею ферменты

Система генов глутатион-S трансфераз играет важную роль в процессах обеспечения и поддержания нормальной жизнедеятельности всех живых организмов. Это связано, прежде всего, с тем, что гены, входящие в состав данной системы, кодируют ферменты, катализирующие различные типы химических превращений в клетке. Именно данная система генов играет первостепенную роль, принимая участие во второй фазе детоксикации (обезвреживания) ксенобиотиков и в защите клеток против окислительного стресса [Leiers B. et. al. 2003; An J. et. al. 2003]. Кроме того, белковые продукты данных генов, участвуют в метаболизме стероидных гормонов, простагландинов, билирубина, желчных кислот, различных лекарств (участие в процессах противоастматической и противоопухолевой лекарственной терапии), инсектицидов и гербицидов [Frova C. 2006; Desikan R. et. al. 2001; Kenneth D. et. al. 2011]. Ферменты генов семейства глутатион-S трансфераз участвуют в формировании конъюгатов, образующихся при взаимодействии глутатиона с различными субстратами. Для дальнейшего выведения данных конъюгатов из клеток, последние под действием глутатиона превращаются в малотоксичные меркаптуровые кислоты [Di Pietro G. et. al. 2010].

Процесс детоксикации в клетках живых организмов проходит в две фазы. I фаза – модификация исходных соединений, катализируемая в основном ферментами семейства цитохрома P450 (CYPs). Обезвреживание ксенобиотиков на данном этапе, заключается в трансформации чужеродных веществ в более полярные соединения (увеличение гидрофильности) и связано с внедрением или образованием функциональных групп типа –OH, –NH<sub>2</sub>, –SH, –COOH. Данный процесс, включает в себя реакции O-, N- и S-деалкилирования, алифатическое и ароматическое гидроксильное, N- и S- окисление и дезаминирование [Guengerich, 2001; Iyanagi, 2007].

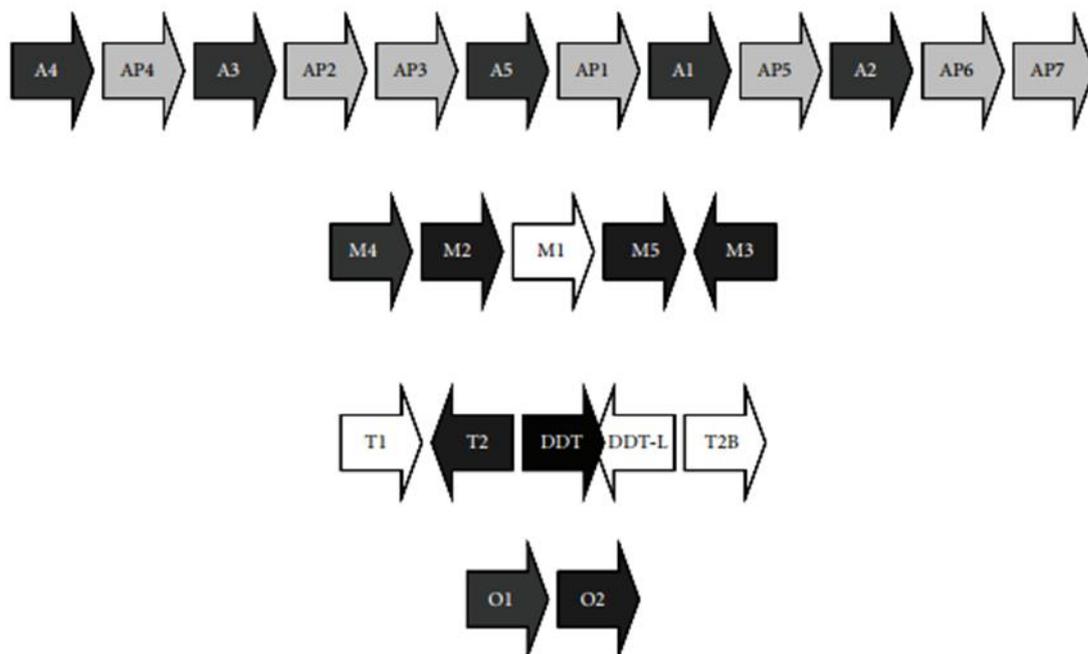
II фаза – конъюгация. Ферменты II фазы играют важную роль в биотрансформации эндогенных веществ и ксенобиотиков, в метаболической инактивации активных веществ: конъюгация с глутатионом, глюкуроновой кислотой, сульфатация, метилирование, ацетилирование, и аминокислотная конъюгация. Ферменты, принимающие участие во второй фазе: UDP-глюкороносульфотрансферазы (UGTs), сульфотрансферазы (SULTs), N-ацетилтрансферазы (NATs), глутатион S-трансферазы (GSTs), различные метилтрансферазы (главным образом, тиопурин S-метилтрансферазы (TPMT) и катехол O-метилтрансферазы (COMT) [Jancovaa P. et. al. 2010]. В результате серии реакций чужеродные вещества начинают обладать гидрофильными свойствами, что способствует снижению их активности и токсичности, с последующим удалением конъюгата из клетки. Все вещества, полученные в итоге данных процессов, выводятся с жёлчью в кишечник посредством транспортеров, таких, как белок множественной лекарственной устойчивости MDR1, и преобразуются в меркаптуровые кислоты (N-ацетилцистеин), которые затем удаляются. Система обезвреживания включает множество разнообразных ферментов, под действием которых практически любой ксенобиотик может быть модифицирован [Kenneth D. et. al. 2011]. Всего в метаболизме ксенобиотиков участвуют более 200 различных ферментов, действующих на разных стадиях детоксикации [Oakley A. 2012].

Важная роль в процессах II фазы отводится непосредственно ферментам, которые кодируются генами глутатион-S трансферазами. Химизм катализируемых ими реакций основан на присоединении глутатиона к электрофильному центру разнообразных соединений. GSTs принимают участие в различных глутатион зависимых процессах, включая восстановление гидропероксидов и изомеризацию различных ненасыщенных производных, а также вовлечены в синтез простагландинов и лейкотриенов. Суперсемейство GSTs разделяется на несколько классов – альфа (A1-A4), каппа (K1), мю (M1-M5), пи (P1), сигма ( $\Sigma$ ), тета (T1, T2), зета (Z) и омега ( $\Omega$ ) [Mannervik B. et. al. 1992].

Историческое название глутатион трансфераз — глутатион-S трансферазы, и именно это название дало начало широко распространенной в научной литературе аббревиатуре *GSTs* [Hayes J. et. al. 2005]. Эволюционно данная группа генов появилась около 2 млрд. лет назад [Nebert D. et. al. 2004]. Система генов *GSTs* представляет собой обширную и разнообразную группу генов, встречающихся практически во всех живых организмах. К примеру, у нематоды *Caenorhabditis elegans* – 48 различных генов *GSTs* [Lindblom T. et. al. 2006], у *Aedes aegypti* (комар желтолихорадочный) – 26 [Lumjuan N. et. al. 2007], и более 70 генов *GSTs* у черного тополя *Populus trichocarpa* [Lan T. et. al. 2009]. Система генов глутатион-S трансфераз человека включает в себя около 18 генов [Mitchell A. et. al. 2005; Попова С.Н. и др. 2005; Ginsberg G. et. al. 2009]. Разделение *GSTs* на классы производится на основании гомологии нуклеотидных последовательностей, а также иммунореактивности [Singh S. et. al. 2010]. Обозначаются данные классы заглавными римскими буквами с арабскими цифрами, обозначающими определенный ген [Laborde E. 2010]. Однако существует некоторая неясность в номенклатуре *GSTs* человека. К примеру, фермент, кодируемый геном *GSTA5*, класса альфа, был исключен из таблиц перечня человеческих *GSTs*, так как продукт экспрессии данного гена не был обнаружен в клетках человека. Однако кодирующая последовательность гена является не поврежденной, и экспрессия ферментативно-активного рекомбинантного белка происходит. Таким образом, вероятно, что фермент на самом деле экспрессируется, может быть в условиях, которые пока еще не были обнаружены [Josephy D. et. al. 2010].

Гены, входящие в состав системы *GSTs*, располагаются в геноме человека на разных хромосомах, причем располагаются они не поодиночке, а группами, образуя генные кластеры (Рисунок 1). Последнее, очевидно, связано с особенностями их эволюционного происхождения: гены, образующие кластер, являются гомологами, возникшими в результате дупликаций одного и того же гена, т.е. паралогами. К примеру, кластер генов *GSTA* локализуется на 6 хромосоме в районе 6p12 [Coles B. et. al. 2005; Morel F. et. al. 2002], кластеры генов *GSTM* и *GSTT* – на 1(1q13) [Tan K-L. et. al. 1995; Sies H. et. al. 2005], и на 22

(22q11.23) [Coggan M. et. al. 1998; Zhao Y. et. al. 2009], и класс омега – хромосома 10q24.3 [Whitbread A. et. al. 2003; Whitbread A. et. al. 2005], гены классов *GSTP*, *GSTK* на 11 (11q13) и 7 (7q34) хромосомах соответственно.



**Рисунок 1. Хромосомная организация некоторых генных кластеров семейства глутатион-S трансфераз.** Буквы А, М, Т и Q обозначают сокращенное название классов генов *GSTs* (альфа, мю, тета, омега). Значок “P” в названии гена указывает, что данный ген является псевдогеном (например, AP3), такие ячейки показаны серым цветом. Гены *GSTs* указаны черным цветом. Гены *GSTM1*, *GSTT2B*, и *GSTT1* – подверженные делециям, показаны белым цветом.

В клетках человека кодируемые генами *GSTs* ферменты наиболее представлены в цитозоле, однако имеются определенные ферменты, встречающиеся в ядрах клеток и митохондриях. Цитозольные *GSTs* представляют собой самое многочисленное подсемейство из всех генов глутатион-S трансфераз. Например, в геноме человека и других видов млекопитающих присутствует около 15-20 различных типов цитозольных *GSTs* [Hayes J. et. al. 2005], у растений 40-60 [McGonigle B. et. al. 2000; Soranzo N. et. al. 2004], у бактерий 10-15 [Ranson H. et.

al. 1998; Wagner U. et. al. 2002] и около 10 типов у насекомых [Morel F. et. al. 2002]. Цитозольные GSTs представлены практически во всех клетках организма, однако имеют место значительные межтканевые особенности уровней экспрессии различных изоформ GSTs. Так, например, концентрация ферментов, кодируемых генами альфа-класса, особенно велика в печени, концентрация ферментов пи-класса выше в плаценте, легких, эритроцитах [Mannervik B. 1985; Pilo D. et. al. 1990; Armstrong R. 1997]. Для всех ферментов данной системы, также характерно проявление глутатион-пероксидазной активности по отношению к гидропероксиду кумола (CuOOH).

Главный компонент катализируемых GSTs реакций – глутатион (GSH) – это трипептид  $\gamma$ -глутамилцистеинилглицин (g-Glu-Cys-Gly), содержащий одну пептидную связь между карбокси-группой боковой цепи глутамата и аминогруппой цистеина, и вторую пептидную связь, соединяющую цистеин с глицином [Antomasi T. et. al. 1997]. Особенности строения атома серы, входящей в состав молекулы GSH, обуславливают способность данного трипептида принимать участие в различных химических реакциях в качестве восстановителя. Синтез молекулы GSH *de novo* осуществляется главным образом в печени [Allocati N. et. al. 2003; Veal E. et. al. 2002]. Окислительно-восстановительные свойства GSH играют важную роль в различных процессах клеточного метаболизма: выведение токсинов и метаболитов процессов распада, наркотиков и лекарственных препаратов, а также регуляция процессов апоптоза, участие в трансмембранном транспорте растворенных органических веществ [William H. et. al. 1994; Tsuchida S. et. al. 1992]. Также есть данные о его участии в процессах активации транскрипции генов. Глутатион своими гидрофобными центрами может нековалентно связывать огромное количество липофильных соединений (физическое обезвреживание), предотвращая их внедрение в липидный слой клеточных мембран и нарушение функций самой клетки. Поэтому глутатион иногда называют внутриклеточным альбумином. Также глутатион способен ковалентно связывать ксенобиотики, являющиеся сильными электролитами. Присоединение таких веществ - «самоубийство» для глутатиона, но это свойство

является дополнительным защитным механизмом для клетки [Amicarelli F. et. al. 2004; Grundy J. et. al. 1998; Kobayashi M. et. al. 2002; Wang M. et. al. 2003]. Молекулы глутатиона полифункциональны, и принимают участие в таких процессах как: ответ клеток на тепловой шок, детоксикация электрофилов, карциногенез, и иммуномодуляторные процессы [Jefferies H. et. al., 2003].

В ходе аэробного метаболизма в клетках постоянно происходят реакции с образованием активных форм кислорода, окислительно-восстановительные свойства GSH помогают клеткам поддерживать жизнеспособность. Активные формы кислорода представлены в клетках в виде гидроксильных радикалов, пероксида водорода и окиси азота. В связи с высокой реакционной способностью данные соединения имеют крайне неустойчивую структуру, что приводит к переоислению липидов, окислению ДНК и белков. Установлено, что окислительный стресс участвует в системе клеточных сигналов и регуляции систем генов [Jefferies H. et. al., 2003].

Одним из основных и особенно важных аспектов ферментов GST является их способность принимать участие в реакциях с большим количеством различных по структуре и химической природе субстратов, общее число которых превышает 3000. Их субстраты включают галогенонитробензолы, оксиды аренов, хиноны,  $\alpha$ ,  $\beta$ -ненасыщенные карбонильные соединения, алкил-и арилгалогениды, лактоны, эпоксиды, сложные эфиры и активированные алкены [Armstrong R. 1997]. Субстрат-специфические особенности характерны для каждого из классов глутатион-S трансфераз, например, ферменты альфа-класса связываются с гидроперекисью кумола, мю-класс – эпоксиды, пи-класс – этакриновая кистота [Hayes J. et. al., 1991]. Глутатион-S-трансферазы классов альфа, мю и пи обладают повышенной аффинностью и связываются с такими матрицами как: S-гексилглутатион-сефароза или S-гексилглутатион—агароза [Jefferies H. et. al., 2003; Flagg E. et. al., 1994], тогда как ферменты, кодируемые другими классами, например, тета, либо не связываются с ними вовсе, либо связываются очень слабо. Из-за относительно высокого уровня экспрессии и легкости очистки, GSTs альфа, мю и пи классов исследовались более основательно, нежели другие классы

человеческих GSTs. Специфичности GSTs в отношении электрофильных субстратов значительно перекрываются. Так, к примеру, 1-хлоро-2,4-динитробензол, часто используемый для спектрофотометрических анализов активности GSTs, является субстратом для большинства человеческих GSTs (кроме GST-T1-1 или GST-T2-2). С другой стороны, некоторые субстраты относительно специфичны для отдельных GSTs. Субстрат-специфические особенности характерны для каждого из классов глутатион-S трансфераз, например, ферменты альфа-класса связываются с гидроперекисью кумола, мю-класс – эпоксидами, пи-класс – этакриновой кислотой [Hayes J. et. al., 1991].

Растворимые, или цитозольные глутатион-S трансферазы, к числу которых относятся большинство рассмотренных выше ферментов – это димерные ферменты, в состав которых входит приблизительно 199-244 аминокислот [Bolt H. et. al., 2006].

Помимо участия цитозольных GSTs в реакциях конъюгации и изомеризации, они могут ковалентно и нековалентно связывать гидрофобные несубстратные лиганды. Данный тип активности требуется для внутриклеточного транспорта, изоляции и ликвидации ксенобиотиков и гормонов. Эти соединения включают билирубин, азокрасители, гем, полиароматические углеводороды, стероиды и гормоны щитовидной железы. [Rajaud J. et. al., 2012]. Цитозольные GSTs могут быть представлены как в гомодимерной, так и в гетеродимерной формах. Так представители классов альфа и мю образуют гетеродимеры, их субъединицы могут формировать значительно большее количество изоферментов. Субъединицы, из которых состоят ферменты, являются уникальными и с характерными особенностями для каждого из классов. Разнообразие количества субъединиц варьирует от двух (тета класс) до шести и более (класс мю) [Ginsberg G. et. al., 2009; Sheehan D. et. al., 2001]. Между ферментами, кодируемыми генами одного класса, наблюдается высокое сходство аминокислотных последовательностей (около 60-80%), схожие размеры мономеров (от 24 до 28 kDa) [Board P. et. al., 1990]. Гетеродимеризация является очень выгодным свойством, которым обладают представители данных классов. Это свойство

заключается в том, что в определённых изоферментах между субъединицами находятся два разных сайта связывания с ксенобиотиками. Ферменты растворимых, или цитозольных GSTs, являются димерами по 25КДа на субъединицу [Ommen B. et. al., 1990]. Были определены кристаллические структуры для многих растворимых GSTs, зачастую со связанными субстратами или продуктами. Каждая субъединица димерных GSTs каталитически является независимой, т.к. имеет специфические сайты в структуре белковой молекулы, называемые – G-сайт и H-сайт. Третичная структура растворимой GSTs субъединицы, имеет N-концевой-альфа/бетта домен, формирующий сайт связывания глутатиона (G-сайт), и второй альфа спиральный домен, формирующий большую часть H сайта, который связывает электрофильные субстраты. Все выявленные в природе ферменты, кодируемые генами *GSTs*, представляют собой каталитически активные димеры. Совсем недавно некоторые исследователи пытались дать ответ на вопрос о молекулярной основе димеризации и искать объяснения, почему димерная организация оставалась столь консервативной в ходе эволюции. Во всех димерных GSTs две субъединицы связаны осью второго порядка. Основное взаимодействие между субъединицами происходит между доменом I одной субъединицы и доменом II ее партнера. Было выявлено, что H-сайт, в гетеродимерах заметно отличается от такового в гомодимерах. Отличительной особенностью данных ферментов, является то, что они принимают участие в процессе катаболизме фенилаланина и тирозина; также, демонстрируют тиол-трансферазную активность. В то же время, данные ферменты являются участниками в процессах превращения тринитроглицерина в глицерил-1,3-динитрат, с освобождением азотистой кислоты, снижении уровня дегидроаскорбиновой и монометиларсоновой кислот [Adler V. et. al., 1999; Axarli I. et. al., 2004; Cho S. et. al., 2001; Dixon D. et. al., 2000; Hayes J. et. al., 1999; Smith A. et. al., 2004; Keen J. et. al., 1978; Khojasteh-Bakht S. et. al., 1999; Mueller L. et. al., 2000].

Во многих исследованиях было выявлено существование прямой взаимосвязи между наличием различных типов полиморфизмов в данной системе генов и активностью ферментов, кодируемыми данными генами.

Рядом авторов была показана связь полиморфизма изолейцин/валин в 105-м положении с пониженной активностью конъюгирования 1-хлоро-2,4-динитробензола (CDNB). Так, в работе Watson M. A. [Watson M. A. 1998] было показано значительное различие активности ферментов, кодируемых геном *GSTP1* в тканях легких, у носителей полиморфизма изолейцин/валин в 105-м положении в зависимости от гомозиготного или гетерозиготного генотипа [Ishii T. et. al., 1999]. Ранее 30% снижение активности рассматриваемого фермента в отношении конъюгации CDNB у носителей аллеля 105 Val было показано в работе Zimniak et al [Zimniak P. et. al., 1994; Pandya U. et. al., 2000] и подтверждено Lo, Ali-Osman и их коллегами [Lo H., et. al., 1997]. Интересно, что по данным Zimniak et. al., содержание 105Val в *GSTP1*-1 приводило к повышению активности кодируемого фермента в отношении других субстратов – этакриновой кислоты и бромсульфопфталеина.

Система обезвреживания с участием GSTs и глутатиона играет важную роль в формировании резистентности организма к самым различным воздействиям и является надежным защитным механизмом клетки. Многие противораковые химиотерапевтические агенты являются электрофильными соединениями, (либо их метаболитами предшественниками). И катализируемые GSTs реакции являются важным путем инактивации и выведения многих таких препаратов, например, таких как 1,3-бис(2хлор-этил)-1-нитрозомочевина, циклофосфамид, элфалан. Препараты из других терапевтических классов могут быть метаболизированы до электрофилов, которые являются субстратами для конъюгации глутатиона, к примеру, фенилпропеновый метаболит противоэпилептического препарата фелбамат [Kenneth D. et. al., 2011]. Многие другие широко используемые препараты, включая ацетаминофен, клозапин и фуросемид метаболизируются до конъюгатов глутатиона, хотя реакция с глутатионом не во всех случаях катализируется GSTs [Hayes J. et. al., 2005].

Таким образом, во второй фазе детоксикации, ферменты семейства GSTs играют ключевую роль в процессах обезвреживания чужеродных веществ в клетках всех живых существ [Di Pietro G. et. al., 2010]. Отсюда понятна значимость изучения генетических систем, кодирующих ферменты GSTs. Полиморфизм, выявленный в генах данного генного супер-семейства, может оказывать свое непосредственное влияние на активность кодируемых ферментов, что в конечном итоге способно приводить к изменениям в клеточном метаболизме. Значимость изучения данной группы генов, обусловлена к тому же широким использованием их полиморфных характеристик в популяционно-генетических исследованиях, что будет рассмотрено далее.

## **1.2. Полиморфизм системы генов глутатион-S трансфераз**

### **1.2.1 Явление полиморфизма в геноме человека**

В результате многочисленных исследований было обнаружено важное свойство генома – его полиморфизм. Геномный полиморфизм обычно обозначает нейтральные изменения в строении генома у разных людей. Нейтральные вариации – это те вариации, которые не проявляются фенотипически. Исследование генетического полиморфизма – важнейший инструмент популяционной генетики, и один из основных способов изучения структурной организации и идентификации функциональных последовательностей генома [Chakravarti A. 1999]. Огромное количество полиморфных маркеров, выявленных в геноме человека, является мощным инструментом изучения генофонда и его основных характеристик, таких как динамика, история и география. Генетические маркеры, широко применяются в медико-биологических исследованиях, в судебной медицине и криминалистике, а также для создания карт генома человека и локализации различных генов. Исследование популяционной изменчивости генетических маркеров позволяет анализировать роль микроэволюционных процессов – отбора, миграций, мутационного давления – в формировании

генетической структуры популяций. В настоящее время изучение полиморфизма ДНК проводится во многих популяциях мира. Подобные исследования позволяют выявлять значительные внутри- и межпопуляционные различия в частотах полиморфных вариантов ДНК во многих географических регионах мира, что стало одной из важнейших характеристик генетической структуры популяций человека [Лимборская С.А. и др., 2002; Лимборская С.А. и др., 2003; Хуснутдинова Э.К. и др., 2005].

История открытия систем генетических маркеров отражает историю развития и становления популяционной генетики. Первым изученным генетическим полиморфным признаком был дальтонизм – признак, сцепленный с X-хромосомой, позволяющий таким образом, напрямую оценивать частоту соответствующего гена. В нашей стране популяционные исследования генетического полиморфизма человека впервые были проведены в конце позапрошлого века А.В. Люблинским [Люблинский А.В. 1885]. Значительную роль в становлении медицинской и популяционной генетики в дальнейшем сыграло открытие Ландштейнером в 1900г. групп крови АВ0. Широкое применение системы АВ0 в медико-генетических исследованиях обусловлено тем, что фенотипы АВ0 определяют предрасположенность или устойчивость к различным заболеваниям. По частотам встречаемости групп крови выявлены резкие различия между популяциями коренного населения в различных регионах мира. Согласно некоторым гипотезам, неравномерность распределения групп крови на земном шаре является результатом прошедших эпидемий тяжелых инфекционных болезней, которые выступали как факторы отбора: чумы, оспы и холеры [Vogel F. 1970].

Непосредственно в геноме выявлено несколько типов полиморфизма. Это – полиморфные мини- и микросателлиты, инсерционно-делеционный полиморфизм, однонуклеотидные замены. [Лимборская С.А. и др., 2002; Лимборская С.А. и др., 2003; Gray I. et. al., 2000].

Под понятием делеционно-инсерционного полиморфизма подразумевается наличие или отсутствие тех или иных участков последовательности ДНК, размер

которых может составлять от одной до нескольких тысяч пар оснований и более. Как правило, такие изменения возникают на фоне активных рекомбинационных процессов, либо репарации двуцепочечных разрывов ДНК [Алтухов Ю.П. и др., 2002; Гречко В. 2002]. Инсерционно-делеционный полиморфизм может оказывать сильное влияние на фенотип в случае отсутствия компенсирующего действия других генов. Данный вид полиморфизма является характерным, например, для семейства глутатион-S трансфераз. Так, делеционный полиморфизм в 6 интроне гена *GSTM3* (отсутствует последовательность AGG) приводит к изменению уровня экспрессии данного гена. Так как фермент, кодируемый этим геном, участвует в метаболизме ксенобиотиков, это изменение может затронуть ответную реакцию организма на воздействие многих веществ, включая определенные компоненты пищи и лекарственные препараты [Feng X. et. al., 2012; Tan X. et. al., 2013; Gray I. et. al., 2000].

Другой тип полиморфизма, встречающийся в геноме человека – полиморфные однонуклеотидные замены или однонуклеотидный полиморфизм (SNPs), выражающийся в единичных (точковых) заменах нуклеотидов в последовательности ДНК. Этот полиморфизм оказался наиболее представленным в геноме [Hayes J. et. al., 2005; Smolenski S. et. al., 2009]. Общее известное на данный момент число вариабельных точек в геноме составляет более 40 млн. [The 1000 Genomes Project Consortium, 2012; National Center for Biotechnology Information dbSNP, build 141]. Подавляющее большинство SNPs имеет два аллеля, представляющих собой замену одного основания другим. В каждом SNPs один аллель является мажорным, т.е. наиболее представленным в популяции, другой обозначается как минорный. Основываясь на данных, полученных о частоте каждого из аллелей, индивидуумы можно разделить на следующие категории по встречающимся генотипам: гомозиготный по мажорному аллелю, гетерозиготный и гомозиготный в отношении минорного аллеля. Однонуклеотидные замены в экзонах могут приводить к заменам кодируемой аминокислоты – это так называемые несинонимичные замены, которые составляют около 65% от всех экзонных SNPs, что зачастую приводит к изменению функций соответствующих

белков. Замены, не приводящие к изменению кодируемой аминокислоты, называются синонимичными.

Практически повсеместная встречаемость SNPs обеспечивает этому виду полиморфизма ряд преимуществ, по сравнению с другими полиморфизма [Kimura-Kataoka K. et. al., 2014]. Исследования ДНК семей показали, что SNPs расположенные рядом на одной хромосоме, могут наследоваться совместно. При этом аллельные варианты образуют сцепленные структуры, называемые гаплотипом. В последнее время при анализе популяций именно структура гаплотипов является одним из ключевых объектов изучения. В связи с этим SNPs, являются подходящим объектом как ДНК-маркеры для различных ассоциативных генетических исследований [Manichaikul A. et. al., 2014]. Следует отметить, что этот тип полиморфизма характеризуется низкой частотой возникновения. Одна точковая замена может происходить очень редко, например, 1 раз в течение нескольких тысяч лет. Чем меньше скорость возникновения полиморфизма, тем более отдаленные события он маркирует. Также немаловажную роль играет изучение генетического полиморфизма в популяциях при поиске ассоциации генотипа с такими признаками, как риск развития различных заболеваний и ответ организма на действие определенных факторов среды [Dawson E. et. al., 2002; Gray I. et. al., 2000; Major J. et. al., 2014; Resler A. et. al., 2014].

Все маркеры ДНК с позиций популяционных исследований можно разделить на три группы: маркеры митохондриальной ДНК, аутосомные маркеры, и маркеры Y-хромосомы.

Полиморфизм митохондриальной ДНК используется в популяционных исследованиях сравнительно давно. Основной особенностью этого полиморфизма является отсутствие рекомбинаций, высокий уровень изменчивости и материнский тип наследования [Ward R. et. al., 1991]. Y-хромосомный полиморфизм является как бы комплементарным митохондриальному – имеет отцовское наследование, но также обнаруживает отсутствие рекомбинации (за исключением псевдоаутосомного региона) [Semino O. et. al., 2000]. Использование этих двух маркерных систем дало новые, не существовавшие ранее возможности

в популяционных исследованиях – проследить и сопоставить историю женской и мужской части популяции и оценить их вклад в популяционный генофонд. Ядерные же аутосомные маркеры ДНК характеризуют сообщества в целом, не акцентируя внимание на особенности генетического вклада различных полов. Использование определенных типов ядерного полиморфизма, как полагают многие исследователи, позволяет оценить те или иные временные события, происходившие в истории популяции [Gray I. et. al., 2000; Jorde L. et. al., 1998].

### **1.2.2. Варианты полиморфизма системы генов глутатион-S-трансфераз**

Как и для большинства других генов человека, для генов *GSTs* характерно явление полиморфизма и как следствие – существование полиморфных вариантов. Наличие различных вариантов аллелей способно приводить к изменению функций ферментов. Помимо одной из важных функций в организме – участие в детоксикации – глутатион-S трансферазы также участвуют в воспалительных процессах. Наличие полиморфизма в данных генах может стать причиной развития различных заболеваний, именно поэтому семейство данных генов является одним из основных объектов молекулярно-эпидемиологического изучения исследований уже на протяжении многих лет [Nebert D. et. al., 2004]. Исследования показывают, что аллельные варианты *GSTs* могут играть роль предрасполагающих факторов в процессах канцерогенеза, развитии воспалительных заболеваний, а также оказывать влияние на эффективность химиотерапевтических воздействий. Представленность же тех или иных аллельных вариантов многих из этих генов в популяциях существенно зависит от расовой и этнической принадлежности данной популяции. Поэтому получение информации об этнических/популяционных вариациях полиморфизма генов в целом может иметь решающее значение, как для оценки риска развития определённых патологических состояний у групп индивидуумов, так и эффективности тех или иных лекарственных воздействий [Хуснутдинова Э.К. и др., 2005; Daly A. 2003; Evans W. et. al., 2003; Garte S. et. al., 2001; McIlwain C. et.

al., 2006]. В конечном итоге, общепризнанная роль генов *GSTs* в обезвреживании токсических соединений и защите клеток организма от окислительного стресса [Hayes J. et. al., 1995] может указывать на то, что наличие различных вариантов полиморфизма в системе данных генов, скорее всего, связано с риском развития онкологических и дегенеративных заболеваний

В группе генов цитозольных *GSTs* наблюдается полиморфизм определенных генов, который может определять межиндивидуальные различия в ответе клеток на воздействие ксенобиотиков. Самые ранние исследования в этой области начинались с ответа на вопрос, приводит ли полиморфизм некоторых генов данного семейства к повышенной частоте случаев развития рака. После открытия аллельных вариантов в гене *GSTP1*, которые кодируют ферменты с пониженной каталитической активностью – SNPs в 5-м (105-й кодон, A>G, rs1695) и 6-м экзонах (114-й кодон, C>T, rs1138272) – была выдвинута гипотеза о том, что существует взаимосвязь между наличием полиморфизма в данной системе генов с некоторыми компонентами из окружающей среды, и это может способствовать возникновению различных онкологических заболеваний [Watson et al., 1998]. Кроме связи с риском развития рака как такового, полиморфизмы в гене *GSTP1*, как полагают, могут выступать модификаторами ответа на химиотерапию у пациентов уже имеющих те или иные онкологические заболевания [McIlwain et al., 2006], а так же влиять на риск развития вторичных новообразований, например, острого миелоидного лейкоза, обусловленного терапией у пациентов, успешно излеченных от рака груди, неходжкинской лимфомы, рака яичников, болезни Ходжкина, карциномы пищевода и синдрома Барретта.

В дополнение к аллельным вариантам в гене *GSTP1*, также были идентифицированы полиморфизмы в генах всех других классах цитозольных *GSTs*. Для ряда генов, в частности *GSTA1* и *GSTA2* *GSTM4* и *GSTT2* был особо отмечен потенциальный вклад полиморфизмов в промоторной области. Такой полиморфизм способен оказывать влияние на количество синтезируемого белка. В модельном эксперименте отдельные аллельные варианты вышеуказанных генов

*GSTs* демонстрировали способность в 1.4-1.5 раз увеличивать транскрипционную активность промотора гена люциферазы, что позволило авторам высказать предположение о взаимосвязи между полиморфизмом в промоторной области данных генов и восприимчивостью несущих их индивидуумов к различным болезням и эффективность воздействия на их организм лекарственных препаратов [Guyet al., et. al., 2004].

Наряду с однонуклеотидным полиморфизмом для генов *GSTs* характерен и делеционный полиморфизм. И здесь наиболее интересен его вариант, связанный с отсутствием целых генов. Делеционный полиморфизм генов классов мю и тета наиболее обычный вариант полиморфизма у человека [Board P. 1981]. Распространенность *GSTM1(0/0)* гомозигот по делеции в европейских и азиатских популяциях составляет примерно 50%, при значительно меньшей частоте среди африканцев и афроамериканцев. Данные по гомозиготам *GSTT1(0/0)* составляют около 20% для европейцев и афроамериканцев, и 50% для азиатов [Bolt, et. al., 2006; Tatewaki N. et. al., 2009].

Взаимосвязь между полиморфизмами по делеции и риском заболеваний изучалась во многих исследованиях, начиная с 1990 г. И хотя эти исследования сами по себе не являются фармакогенетическими, они могут иметь отношение к более общему вопросу: существует ли взаимосвязь между полиморфизмом генов глутатион-S трансфераз и ответной реакцией организма на поступление ксенобиотиков, так как воздействие окружающей среды вносит свой вклад в риск развития определенных онкологических заболеваний и других широко распространенных заболеваний [Moyer A. et. al., 2007]. Поиск в базе данных PubMed по данному вопросу, выдает большое количество ссылок на различные исследования, в которых изучалась взаимосвязь между полиморфизмами генов данного семейства и риском развития неонкологических заболеваний, к примеру, болезнь Паркинсона, болезнь Альцгеймера, астма, ишемическая болезнь сердца и ревматоидные артриты [Bolt H. et. al., 2006].

Как уже говорилось выше, гены классов альфа, мю, тета и омега в геноме человека, располагаются группами, образуя кластеры генов, локализованные на разных хромосомах (Рисунок 1). Гены, образующие кластер, являются гомологами, возникшими в результате дупликаций одного и того же гена, т.е. паралогами. Делеционный полиморфизм генов классов мю и тета наиболее обычный вариант полиморфизма у человека [Board P. 1981], более детально данные делеции были описаны в работе Н. М. Volt и R. Thier в 2006 году [Nebert D. et. al., 2004; Volt H. et. al., 2006]. Авторы опубликовали обширный обзор по молекулярной эпидемиологии *GSTs*. С того времени было опубликовано еще несколько новых работ по взаимосвязи эпидемиологических аспектов с определенными заболеваниями [Khoury M. et. al., 2000].

Самые ранние исследования начинались с ответа на вопрос, приводит ли делеция генов *GSTM1* и/или *GSTT1*, а именно гомозиготность по делеции (*GSTM1(0/0)* и/или *GSTT1(0/0)*), к повышенной частоте случаев развития рака мочевого пузыря, груди, рака головы/шеи и рака легких. Было обнаружено, что наличие у индивидуумов «нулевого» генотипа *GSTM1(0/0)* может оказывать влияние на развитие рака легких, а *GSTM1(0/0)* и *GSTT1(0/0)* – дают в совокупности увеличение частоты случаев возникновения рака головы и шеи, и *GSTP1(0/0)* влияет на риск развития синдрома Барретта и карциномы пищевода [Rebbeck T. 1997]. Стоит отметить, что возможный недостаток многих исследований биологического действия *GSTM1(0/0)* и *GSTT1(0/0)* генотипов заключается в том, что с помощью наиболее популярного (простого) метода ПЦР идентифицировались только индивидуумы гомозиготные по делеции этих генов. Индивидуумы гетерозиготные (0/+) и гомозиготные (+/+) по наличию функционального аллеля не различались этим методом и анализировались вместе в составе единой группы. Как следствие, преимущества, которые дает гомозиготность дикого типа по *GSTM1(+/+)* и *GSTT(+/+)* редко рассматриваются. Как отмечено Minelli и соавт. [Minelli C. et. al., 2010], классификация генотипов «только как «нулевой» и «не нулевой», может не отражать действительную генетическую модель, лежащую в основе, достоверной и точной оценки

генетического риска». Сейчас, когда стали доступны аналитические методы, которые оценивают число копий, (ПЦР в режиме реального времени [Girault I. et al., 2005; Roodi N. et al., 2004; Timofeeva M. et al., 2009]); более старые и менее информативные методы генотипирования постепенно теряют свои позиции [Parl F. 2009].

Имеются исследования, в которых дается оценка распределения нескольких широко встречающихся полиморфных вариантов генов подсемейства цитозольных *GSTs* в различных популяциях России. В работе Хрунина с соавт. [Хрунин и др., 2008] были проанализированы SNPs в 5-м (105-й кодон, A>G, rs1695) и 6-м экзонах (114-й кодон, C>T, rs1138272) гена *GSTP1*, SNPs в промоторной зоне *GSTA1* (-69 C>T, rs3957357, изменение уровня экспрессии гена), делеция трех нуклеотидов AGG в 6-м интроне *GSTM3* (возможность связывания транскрипционного фактора YY1 с ДНК) и делеции генов *GSTM1* и *GSTT1* в выборках русского населения, проживающего в географически различных регионах европейской части России, а также взятых для сравнения в выборках коми и якутов. Анализ полиморфизма гена *GSTP1* не выявил значимых различий в распределении аллелей исследованных локусов в большинстве русских популяций и между популяциями коми. Распределение же аллелей в якутской популяции статистически достоверно отличалось от наблюдаемого в выборках русских и коми. Похожие результаты были получены и при анализе распределения в исследованных популяциях аллелей полиморфного локуса гена *GSTM3*: достоверные различия в каждом из попарных сравнений были отмечены лишь для якутской выборки.

Анализ распределения гомозигот делеционного генотипа гена *GSTM1* не продемонстрировал значимых различий в частотах их встречаемости не только в русских популяциях, но и между русскими, якутами и коми. В то же время аналогичное сопоставление для гомозигот *GSTT1(0/0)* выявило заметные различия в частоте их встречаемости между русской популяцией из Мезени и всеми остальными популяциями. Частота нулевого генотипа *GSTT1(0/0)* в мезенской выборке была существенно ниже (9%), чем в других русских популяциях (17-

24%) и тем более в выборке ижемских коми (29%). Тест на гомогенность, проведенный для всего пула исследованных русских популяций, продемонстрировал неоднородность такого группирования ( $P=0,056$ ). В случае же, когда мезенская популяция исключалась из группы, получаемая совокупность демонстрировала выраженную гомогенность ( $P=0,752$ ). Для распределения аллелей локуса гена *GSTA1* было характерно сочетание обеих вышеуказанных тенденций. С одной стороны, это достоверное отличие якутской популяции от всех других, а с другой – особое положение выборки русских из Мезени, частота встречаемости минорного аллеля T в которой существенно ниже, чем в остальных русских популяциях.

Таким образом, проведенный анализ полиморфизма *GSTs* позволил вычленил две особенности в распределении аллелей в исследованных популяциях России, одна из которых связана с якутской популяцией, а вторая – с выборкой русских из Мезени. И если для якутов такой результат был высоковероятным и основанным на их принадлежности к другой расе, то во втором случае он может быть обусловлен взаимодействием целого ряда факторов. С одной стороны, здесь определенную роль могут играть исторически обусловленные особенности становления русского населения северных территорий (низкая плотность населения и удаленность от центральных районов) [Balanovsky O. et. al., 2008; Khrunin A. et. al., 2008; Khrunin A. et. al., 2005]. С другой стороны, можно предположить, что это может быть и следствием отбора под действием факторов среды обитания, например, более экстремальных климатических условий. Так, пониженная частота встречаемости делеционных генотипов *GSTT1* характерна для популяций Северной Европы (финнов, шведов, датчан) [Tasa G. et. al., 1996; Xu S. et. al., 1998]. Мезенский район в этом отношении является одной из наиболее северных территорий исторического проживания русскоговорящего населения. Важным, с точки зрения результатов возможных селективных воздействий, является несоответствие зафиксированного в популяции Мезени распределения генотипов гена *GSTA1* равновесию Харди-Вайнберга: в выборке наблюдается повышенное содержание СС-гомозигот,

характеризующихся нормальным (не сниженным) уровнем экспрессии белка GSTA1. С позиции известного фактора перекрытия спектров активности GSTA1 и GSTT1 [Bolt, et. al., 2006] такое сочетание активных вариантов может рассматриваться как причинно обусловленное, так как в целом повышает суммарную детоксикационную (функциональную) активность системы GSTs.

В связи с тем, что одной из задач в настоящей работе было изучение влияния делеционного полиморфизма гена *GSTM1*, следует подробнее охарактеризовать сам ген *GSTM1*, а также влияние этого делеционного полиморфизма на частоту возникновения онкологических заболеваний у человека.

В отличие от других систем *GSTs* в кластере генов *GSTM*, наряду с небольшими по размеру однонуклеотидными полиморфизмами, имеется протяженная делеция, удаляющая из хромосомы один из генов кластера – ген *GSTM1*. Встречаемость такой делеции в популяциях достаточно высока: применительно к европеоидам, до половины населения являются гомозиготами по делеции гена *GSTM1* [Xu S. et. al., 1998]. В мировых популяциях частота данной делеции также высока. Так, в азиатских популяциях ее частота будет около 58% для китайских популяций, около 45% для популяции из Японии [Tatewaki N. et. al., 2009] и около 55% для популяций из Кореи. В популяциях Африки делеция гена *GSTM1* достигает частоты 20-25% [Piacentini S. et. al., 2011].

Данный кластер генов в геноме человека располагается на р-плече первой хромосомы в локусе 13.3. В нормальном состоянии ген *GSTM1* кодирует функционально активный фермент – GSTM1\*А. В то же время мутация в 7 экзоне данного гена ведет к выработке другого варианта фермента – GSTM1\*В. Данная мутация не сказывается на функциональных изменениях в активности фермента. Частота встречаемости делеции данного гена различается между популяциями людей в мире [Board P. et. al., 1990; Pemble S. et. al., 1994]. Ген *GSTM1* является высоко полиморфным, наличие его делеционного варианта представляет особый интерес, так как большое количество исследований указывают на наличие взаимосвязи между данным видом полиморфизма и реакциями организма на

различные виды токсических веществ, устойчивостью при лечении методом химиотерапии, и ответной реакцией на введение различных лекарственных препаратов и наркотиков, а также проявление ряда заболеваний, приводящих к развитию рака. Процент идентичных последовательностей между генами *GSTM1* и его ближайшим гомологом *GSTM2* составляет приблизительно 75-99 %, в то время как процент гомологии между различными классами (например, *GSTθ* и *GSTπ*) составляет примерно 25-30%. Результаты сопоставлений трехмерной структуры белка гена *GSTM1* с ферментами, принадлежащими к другим генам данного семейства отчетливо продемонстрировали, что фермент гена *GSTM2* имеет самую высокую степень сходства с ферментом гена *GSTM1* [Bhattacharjee P. et. al., 2013]. Наиболее высокий уровень экспрессии данного гена наблюдается в печени. В легких экспрессия гена *GSTM1* также осуществляется, но протекает на более низком уровне. Недавние исследования показали, что у индивидуумов, обладающих нулевым генотипом *GSTM1(0/0)*, увеличивается уровень экспрессии гена *GSTM2* в 2 раза. Ген *GSTM2* может эффективно компенсировать ферментативную активность в клетках, в случае отсутствия у индивидуума гена *GSTM1*. В то же время уровень экспрессии генов *GSTM3*, *GSTM5* и *GSTM4* остается неизменным [Pearson W. et. al., 1993].

Множественные исследования показали, что различные полиморфизмы в классах мю и тета могут оказывать непосредственное влияние на онкогенез. Потеря этих генов повышает подверженность воспалительным заболеваниям, таким, как астма и аллергии, атеросклероз, ревматоидный артрит и системный склероз. Большое количество публикаций посвящено изучению взаимосвязи делеции гена *GSTM1* и развитием различных онкологических заболеваний. Практически во всех работах было показано, что индивидуумы, являющиеся гомозиготами по делеции данного гена (*GSTM1(0/0)*), входят в группу риска развития различных раковых заболеваний [Zhao Y. et. al., 2013]. Неоднократно было показано, что гомозиготный генотип по делеции гена *GSTM1* связан с повышенным риском развития рака легких [Ford J. et. al., 2000; Shi X. et. al., 2008]. Многочисленные исследования подтверждают связь между отсутствием

активности *GSTM1* и риском развития рака мочевого пузыря [Habdous M. et. al., 2004], аденокарциномой желудка и прямой кишки. Было показано, что риск заболеваемости раком кишечника и желудка почти в 3 раза выше у людей, имеющих генотип *GSTM1(0/0)* [Economopoulos K. et. al., Eur J Cancer, 2010; Economopoulos K. et. al., Int J Gynecol Cancer, 2010; Zhao Y. et. al., 2013]. Делеция в этом гене также ассоциирована с повышенным риском развития астмы и других респираторных заболеваний [Tamer L. et. al., 2004]. В работе Sata F. 2003, выявлена ассоциация между наличием делеции и невынашиваемостью беременности среди женщин, гомозиготных по данному гену [Sata F. et. al., 2003].

Промоторная область данного гена также активно исследуется. Различные геномные преобразования в данном регионе могут приводить к разнообразным заболеваниям в популяциях людей. Dhillon et al. [2007] показали, что гиперметилирование промоторной области *GSTM1* приводит к развитию бесплодия у мужчин [Dhillon V. et. al., 2007]. В то же время выявлена ассоциация между определенными полиморфными локусами (rs412543, rs412302, rs4147563) и риском развития рака головы и шеи. Образуемые варианты гаплотипов могут свидетельствовать об увеличении (гаплотип CAT) или снижении (гаплотип GGC) риска развития, данного онкозаболевания [Zhang X. et. al., 2012, Zhang X. et. al., 2013].

Несмотря на большое количество исследований, связывающих полиморфизм *GSTM1* с риском тех или иных заболеваний, результаты мета-анализов не смогли продемонстрировать однозначности ассоциаций. Высокая частота делеции также свидетельствует о том, что этот ген не находится под сильным давлением отбора со стороны окружающей среды в течение эволюции. Это, по-видимому, может быть следствием того, что другие ферменты, могут быть вовлечены в схожие процессы химической детоксикации.

В последнее время в популяционных исследованиях особое место уделяется изучению однонуклеотидных полиморфизмов в составе гаплотипов. Анализ характеристик SNPs способствует развитию концепции паттернов неравновесия по сцеплению и связанной с ней представлениями о блочной структуре генома

человека. Оценка гаплотипического разнообразия активно используется в современных исследованиях и считается мощным методологическим инструментом. Сравнительный анализ гаплотипического разнообразия способен выявить межпопуляционные различия, не идентифицируемые на уровне отдельных SNPs.

### **1.3. Феномен неравновесия по сцеплению. Понятие о гаплотипе**

Изучение характера изменчивости в последовательности генома человека привело к заключению о его сегментарно-блочном строении, обусловленном особенностями взаимосвязей между аллельными вариантами отдельных полиморфных локусов и существованием их устойчивых комбинаций – гаплотипов (Daly et al., 2001; Gabriel et al., 2002).

Гаплотип – это сочетание близкорасположенных локусов одной хромосомы, следующих друг за другом и наследуемых единым блоком в ряду поколений. Гаплотип представляет собой конкретную комбинацию аллелей различных маркеров, расположенных в определенном участке хромосомы. По существу, в том или ином участке генома каждый индивидуум имеет два гаплотипа, соответствующими участками материнской и отцовской хромосомы. Основанием для объединения аллелей конкретных SNPs в составе гаплотипа является оценка уровня неравновесия по сцеплению (linkage disequilibrium (LD) между парами отдельных SNPs. LD представляет собой феномен, заключающийся в том, что присутствие на хромосоме одного аллеля может означать высокую вероятность того, что другой конкретный аллель будет находиться в рядом расположенной точке той же самой хромосомы. Знание генотипа, определенного SNP может позволить предсказывать генотип другого расположенного рядом SNP, если ассоциация (неравновесие по сцеплению) между этими двумя локусами является высокой. В этом случае частота встречаемости такого сочетания будет выше той, которую можно было бы ожидать в случае независимой сегрегации аллелей локусов.

Математически, в случае двухлокусной системы и с использованием одной из наиболее ранних и широко употребляемых до последнего времени мер неравновесия по сцеплению ( $D$ ), это можно представить следующим образом:

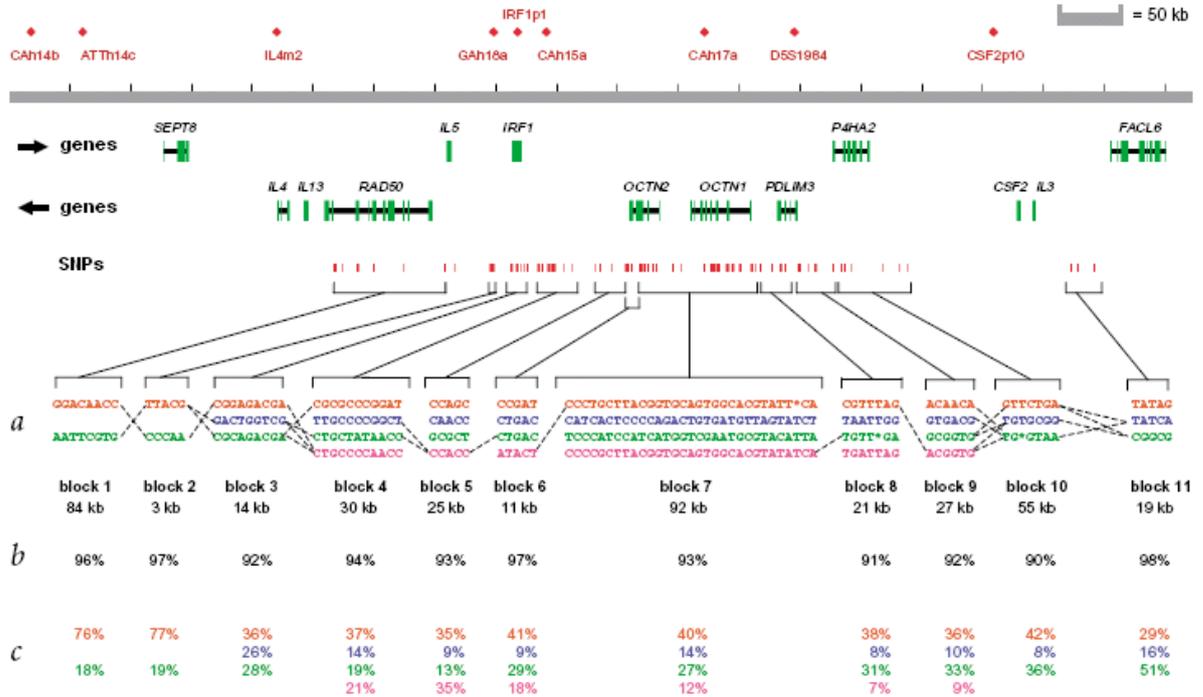
$$D = p_{AB} - p_A \times p_B, \text{ где:}$$

$p_{AB}$  – наблюдаемая частота гаплотипа, составленного из аллелей A и B;  $p_A \times p_B$  – ожидаемая частота гаплотипа AB, которая могла бы наблюдаться в случае независимой сегрегации аллелей локусов [Ardlie K. et. al., 2002].

В целом, чем выше эта величина, тем в большем сцеплении находятся локусы между собой, следовательно, увеличивается вероятность того, что данные локусы входят в состав одного гаплотипа [Kidd K. et. al., 1996]. Сами же группы же соседствующих SNPs будут в этом случае демонстрировать отчетливые паттерны LD.

Паттерны LD – предмет интенсивных исследований современной генетики, с анализом распределения которых в геноме связывают надежды на успех в изучении эволюционной истории человечества и идентификации генетических вариантов, определяющих признаки многих распространенных заболеваний человека. Решающим фактом в осознании эффективности LD-концепции стали данные о результатах секвенирования генома человека, в частности о количестве и плотности SNPs, а также последовавшая за этим возможность оперирования в популяционных исследованиях полногеномными материалами. Массированное изучение аллельных ассоциаций SNPs (групп аллелей SNPs, потенциально способных формировать паттерны LD, гаплотипы) на протяженных участках хромосом позволило получить удивительно «простую» схему архитектуры генома человека: блоки генетического материала (паттернов LD, гаплотипов) различной длины перемежаются участками с интенсивной рекомбинацией (рекомбинационные процессы происходят в основном в так называемых «горячих точках» рекомбинации — участках генома, характеризующихся повышенной частотой подобных событий [Лихтен М. и др., 1995; Lohmueller K. et. al., 2011]). При этом оказалось, что реальные размеры таких «гаплотипических» блоков (в среднем 5-20 kb) много больше предсказанных ранее (около 3 kb), а их

максимальная протяженность может достигать нескольких сотен тысяч оснований. Для каждого из таких блоков характерным является низкий уровень гаплотипического разнообразия, когда очень небольшое число гаплотипов (от 3 до 5) охватывает большую часть (до 90%) наблюдаемой генетической вариации (рисунок 2).



**Рисунок 2. Блокоподобные структуры и гаплотипическое разнообразие района хромосомы 5q31 (по Daly et al., 2001). а,** общие гаплотипические паттерны в пределах каждого из блоков; пунктиром показаны районы рекомбинации. **б,** процент от общего числа хромосом, для которых был характерен один из общих паттернов. **с,** процент каждого из паттернов в исследованной выборке хромосом.

При этом размеры блоков (протяженность паттернов LD) могут заметно различаться между популяциями людей, из различных регионов мира (принадлежащих к разным расам). Например, для гаплотипических блоков африканских популяций характерны минимальные размеры, составляющие, в среднем, около 9 kb. В то же время средняя протяженность паттернов LD в европейских и азиатских популяциях составляет около 20-25 kb. Большая протяженность паттернов неравновесия по сцеплению в неафриканских популяциях, по-видимому, отражает факт более резкого снижения численности в

истории популяций, связанного с эффектом основателя [Daly A. 2003; Gabriel S. et. al., 2002; Hinds D. et. al., 2005; Wall J. et. al., 2003].

Данный факт также свидетельствует о большей «молодости» неафриканских популяций. Для генома популяций из Африки характерным признаком является меньшая протяженность гаплотипических блоков и их большая раздробленность, по сравнению с гапблоками в неафриканских популяциях, которые могут состоять из десятков и сотен SNPs, что подтверждает факт африканского происхождения всех современных людей.

На протяженность и распределение паттернов неравновесия по сцеплению могут оказывать влияние целый ряд факторов разного уровня воздействия. Возможность одних факторов, таких как генетический дрейф, смешение и инбридинг, генерировать вариации LD реализуется на уровне популяции. Действие же других, среди которых уровень рекомбинации, геновая конверсия и естественный отбор, является специфичным в отношении геномных областей [Shifman S. et. al., 2003]. Исходно любая мутация возникает в определенном гаплотипическом окружении. Частота встречаемости такого гаплотипа, несущего новую мутацию, спустя какое-то время, могла бы, например, в условиях генетического дрейфа дорасти до высоких величин, а впоследствии, он вообще бы мог быть расщеплен на отдельные сегменты посредством рекомбинации. Такие комбинации мутирования, дрейфа, селекции, миграции/ смешения популяций и рекомбинации выливаются в генетическую вариацию, имеющую выраженную сегментарно-гаплотипическую структуру. Популяционные генетики имеют устоявшиеся виды на детерминанты неравновесия по сцеплению, но прогнозы по гаплотипам выходят за пределы выводов оценок попарного неравновесия по сцеплению. И хотя, любой из факторов, способных размыть пары неравновесия по сцеплению будет нарушать целостность гаплотипической структуры, можно отметить два неотделимых друг от друга момента: 1) гаплотипическая структура формируется из многих сайтов, имеющих общее происхождение и 2) гаплотипы возникают как внутренне объективный атрибут популяционно-генетической вариации. Именно поэтому гаплотипы «... могут быть чрезвычайно полезными

для объяснения не только существующего распределения вариации, но и быть также важными причинами такого распределения» (Kidd et. al., 1996; Clark, 2004).

Наглядным примером полезности изучения гаплотипического разнообразия в популяциях в целях изучения истории индивидуальных локусов/аллелей может служить обнаружение протяженных паттернов неравновесия по сцеплению, окружающих аллели генов глюкозо-6-фосфат дегидрогеназы (*G-6-ФДГ*) и лиганда *CD40 (TNFSF5)*, определяющих устойчивость к малярии. Гаплотипы блоков, ассоциированные с аллелями устойчивости, имели максимально-возможную протяженность: 413 и 506 kb, соответственно. Для этих гаплотипов также была характерна высокая частота встречаемости в районах распространения данного заболевания. Расчеты показали, что найденные резистентные варианты (гаплотипы) могли возникнуть около 2500 лет назад в случае гена *G-6-ФДГ* и около 6500 лет назад в случае *TNFSF5* [Sabeti P. et. al., 2002]. Анализ районов других генов человека, несущих резистентные аллели (ген гемоглобина E, ген лактазы и т.д.), также выявил протяженные паттерны неравновесия по сцеплению (протяженные гаплотипы) [Ohashi J. et. al., 2004; Bersaglieri T. et. al., 2004]. Таким образом, выявление в популяции необычно длинных, часто встречающихся, общих гаплотипов (гаплотипа) может указывать на селективную значимость такого паттерна неравновесия и его возможное недавнее историческое происхождение. Другой важный аспект – низкое разнообразие гаплотипов в пределах разных паттернов LD. Это может быть следствием либо генетического дрейфа, либо неких исторических событий, сопровождавшихся резким снижением численности популяции и последующим ростом на фоне ее изолированного проживания [Shifman S. et. al., 2003].

Особый интерес к изучению паттернов LD связан с перспективой обнаружения ассоциаций между их вариантами и наиболее распространенными заболеваниями человека [Cardon L. et. al., 2003]. Гаплотипы используются исследователями для определения локусов, вызывающих различные заболевания. Такие возможности появились сразу же после получения первых результатов по массивному типированию SNPs и формированию представлений о блочной

структуре генома, когда было выявлено, что [Gabriel S. et. al., 2002] описание гаплотипической вариации каждого блока может быть осуществлено с использованием небольшого числа конкретных маркеров (tagSNPs). В силу того, что геном человека может быть описан в терминах гаплотипических блоков, Национальный институт исследований генома человека, на базе Национальных институтов здоровья, инициировал проект под названием «НарМар». Основная цель этого проекта – определение общих паттернов вариации последовательности ДНК в геноме человека путем характеристики конкретных вариантов, их частоты и корреляций между ними в 270 образцах ДНК популяций из Нигерии, Европы, Китая и Японии [The International NarMap Consortium, 2003]. Однако с накоплением новых данных с момента начала осуществления проекта НарМар становится понятным, что представления о блочной структуре генома должны интерпретироваться с учетом многовариантности путей формирования самой этой структуры. Так, например, хотя во многих исследованиях и отмечены блокоподобные структуры на протяжении целых участков генома, однако, показана возможность возникновения таких блоков и в отсутствие горячих точек рекомбинации, а само наличие таких структур может быть обусловлено историей формирования популяции и являться результатом действия как индивидуального, так и множественного воздействия комплекса эволюционных факторов (мутаций, рекомбинации, демографической истории) [Wang N. et. al., 2002; Phillips M. et. al., 2003]. И эта идея о том, что другие силы могут оказывать воздействие на границы блоков и их размеры, находит подтверждение в исследованиях отдельных регионов генома не только в популяциях, принадлежащих к различным расам, но и в популяциях одного географического региона. Число найденных блоков и их размеры также существенно зависят и от плотности использованных маркерных SNPs [Ke X. et. al., 2004]. Последнее может оказывать существенное влияние на результативность применения найденных в подобных исследованиях tag SNPs для сравнения различных популяций [Mueller J. et. al., 2005].

Для более точного определения механизмов наследования заболеваний используются различные молекулярные методы в комбинации с

генотипированием индивидуальных образцов. Они основываются на том, что различные локусы, располагающиеся близко на одной хромосоме, будут наследоваться сообща, как неделимая единица до тех пор, пока они не будут разделены рекомбинационным событием [Lohmueller K. et. al., 2011]. Получение информации о ходе процесса рекомбинации является важным фактором, помогающим в определении присутствия взаимосвязей между полиморфными точками на молекуле ДНК и развитием различных заболеваний. Выявляются достоверные статистические ассоциации между заболеваемостью в популяциях и наличием определенных аллелей в полиморфных локусах в человеческом геноме. Вероятность рекомбинации частично определяется расстоянием между исследуемыми маркерами. Те маркеры, которые расположены физически близко друг к другу, имеют более высокую вероятность оказаться сцепленными. Локусы являются сцепленными, если их аллели передаются совместно (косегрегация) от родителей к потомству, как единый гаплотип. В процессе рекомбинации возможно образование двух новых гаплотипов, которые потенциально могут быть переданы новому поколению. При анализе ассоциаций сравнивают частоты маркеров при заболевании и в контрольной группе и анализируют одновременную встречаемость маркера и болезни. Заметная ассоциация между маркером и болезнью может подразумевать участие гена-кандидата в развитии болезни. С другой стороны, ассоциация может быть следствием сцепления маркерного аллеля с геном предрасположенности к болезни, особенно если имеется место физически близкое расположение маркера и гена болезни. Чем ближе расположены два маркера, тем больше вероятность сохранения LD между ними в популяции [Nothnagel M. et. al., 2005; Zhao Y. et. al., 2009].

Исследования болезни Крона (деструктивного воспалительного заболевания кишечника) показывает, как феномен LD может быть использован для выявления локусов на хромосоме, связанных с развитием болезни. Исследователи выяснили, что локус восприимчивости к данному заболеванию находится на 16 хромосоме. Этот участок был просеквенирован, и результат анализа выявил 3 SNPs, которые

демонстрировали сильное LD с болезнью Крона [Elding H. et. al., 2011; Sheridan J. et. al., 2013].

В целом, исследования последних лет показали, что гаплотипы являются весьма эффективным инструментом изучения характеристик генофонда. Было показано, что многие участки генома представляют собой сочетания дискретных гаплотипических блоков разного размера, границами которых являются горячие точки рекомбинации [Eslahchi S. et. al., 2011; Gibson J. et. al., 2005; Gibson J. et. al., 2013]. Изучение особенностей структуры гаплотипов, их частот встречаемости в различных популяциях чрезвычайно важны для понимания многих явлений в генетике человека. Использование представлений о гаплотипическом устройстве генома является актуальным как анализа эволюции популяций, так и для ассоциативных исследований мультифакториальных заболеваний [Zhao Y. et. al., 2009; Crawford D. et. al., 2005; Clark A. et. al., 2004].

## Глава 2. МАТЕРИАЛЫ И МЕТОДЫ

### I. Популяционные выборки

В работе использованы препараты ДНК, выделенные из образцов венозной крови русского населения, проживающего на территории центральных районов европейской части России, а именно: Андреапольском районе Тверской области (96 образцов), Курском и Октябрьском районах Курской области (93 образца) и Муромского района Владимирской области (96 образцов). Образцы крови были собраны в ходе экспедиций за период 1999-2006 гг.

Этническая принадлежность определялась на основе анкетирования: в исследование включались неродственные индивидуумы, предки двух предшествующих поколений которых являлись этническими русскими и проживали на территории исследуемого района. Включение в выборку проводилось на добровольной основе: все участники подписывали информированное согласие на участие в исследовании. Забор крови осуществлялся медицинскими работниками в содержащие антикоагулянт (ЭДТА) вакуумные пробирки (вакутейнеры). После заполнения вакутейнера образцы крови аккуратно перемешивали. Транспортировку и хранение (не более 7 суток) образцов крови осуществляли при температуре +4°C, не допуская замораживания.

### II. Экспериментальные методы

#### II.1.Выделение ДНК

Выделение и очистку образцов ДНК осуществляли стандартным методом, основанным на использовании протеиназы К с последующей фенол-хлороформной экстракцией [Mathew C. 1984]. Процесс выделения проводился в два последовательных этапа.

##### II.1.1. Осаждение ядерной фракции лейкоцитов

Всю процедуру проводили на холоде. Посуду и растворы предварительно охлаждали до +4°C.

Растворы:

а) Лизирующий буфер («буфер А») (хранится при -20°C)

- 0.32 М сахараза
- 10 мМ трисНСl (исх. 1 М, рН 7.5-8.0)
- 1% тритон Х-100
- 5 мМ MgCl<sub>2</sub>

б) Буфер 2 (хранится при комнатной температуре):

- 25 мМ ЭДТА (исх.0.5М, рН8.0)
- 10 мМ трисНСl (исх. 1 М, рН 7.5-8.0)
- 75 мМ NaCl
- 1% SDS

1) К 7-9 мл крови добавляли 8-10 объемов лизирующего буфера. Тщательно перемешивали. Если были сгустки, материал гомогенизировали в пестиковом гомогенизаторе. Оставляли во льду на 20-30мин. Затем центрифугировали 25мин, при 1600 g при +4°C.

2) Аккуратно сливали супернатант, а осадок ресуспендировали в буфере 2 объемом, равным исходному объему крови.

3) Полученные лейкоцитарные осадки хранили при -20°C.

### II.1.2. Экстракция смесью фенол-хлороформ

Свежеперегнаный фенол, содержащий в качестве антиоксиданта 0,1% оксихинолин, предварительно уравнивали 0.1% буфером трис-НСl (рН=8,0), затем к нему добавляли равный объем смеси хлороформа с изопропанолом (24:1, по объему) (хранится при +4°C).

1) К полученным на этапе 1 лейкоцитарным осадкам добавляли протеиназу К до концентрации 100 мкг/мл и оставляли на ночь (не менее 16 ч) на качалке при +42°. Протеиназу К хранили на +4°C в сухом виде, разводили перед применением до концентрации 30 мг/мл, в растворе хранили при -20°C.

2) Затем визуально проверяли гомогенность гидролизуемых образцов, при наличии сгустков повторно добавляли протеиназу К и выдерживали на качалке при +42°C до получения гомогенного раствора.

3) Добавляли хлорид натрия до концентрации 1 М, аккуратно размешивали.

4) Добавляли равный объем смеси фенол-хлороформа. Образцы помещали на 30 мин на качалку при комнатной температуре, затем центрифугировали 20-25 мин при 1600 g при +4°C и отбирали верхний водный слой, не затрагивая интерфазу.

5) Повторяли стадию 4.

6) К отобранной водной фазе добавляли 2,5-3 объема, охлажденного до -70°C 96% этанола и аккуратно перемешивали. Конденсирующуюся на дне колбы ДНК извлекали с помощью капилляра, последовательно ополаскивая сначала 70%, а затем 96% этанолом (весь используемый этанол заранее охлаждали до -20°C и следили, чтобы в процессе выделения спирт постоянно оставался холодным); подсушивали ДНК на воздухе и растворяли в 1000-1200 мкл деионизированной воды. Полученные растворы ДНК хранили при -20°C.

### II.1.3. Контроль качества ДНК с помощью гель-электрофореза

Для оценки качества полученных препаратов ДНК проводили гель-электрофорез в 0,8% агарозном геле в трис-боратном буфере (pH 8,0), содержащем 45 mM Трис-борат, 1 mM ЭДТА и 0,01% EtBr, нанося в лунку по 2 мкл предварительно смешанного с буфером нанесения образца ДНК. После окончания электрофореза, гель анализировали при УФ-освещении. Хороший образец ДНК должен быть представлен компактной не размытой полосой и не иметь выраженной остаточной светимости в лунке.

### II.2 Проведение полимеразной цепной реакции

Определение делеционного полиморфизма в гене *GSTM1*, основывалось на анализе продуктов амплификации, полученных методом ПЦР. Амплификацию фрагмента гена *GSTM1* проводили методом мультиплексной ПЦР, в которой в

качестве внутреннего контроля использовалась амплификация участка 4-го экзона гена *GCHI* (GTP-циклогидролаза 1). Используемые для амплификации праймеры и условия ПЦР указаны в таблице 1.

**Таблица 1. Последовательности праймеров и условия амплификации участков генов *GSTM1* и *GCHI*.**

Ген	Последовательность праймеров	Режим амплификации	Длина продукта
<i>GSTM1</i>	5'-gaa ctc cct gaa aag cta aag c-3' 5'-gtt ggg ctc aaa tat acg gtg g-3'	95°(45сек), 65°(30сек), 72°(40сек) 30 циклов	215
<i>GCHI</i>	5'- gtc ctt ttt gtt tta tga gga agg c -3' 5'- ggt gat gca ctc tta taa tct cag c -3'		288

ПЦР проводили на амплификаторе «Терцик» фирмы "ДНК-технология" (Россия) в пробирках " Eppendorf " объемом 0,5 мл.

Для ПЦР готовили реакционную смесь, объемом 10 мл, содержащую:

- 1 мкл 10х ПЦР-буфера “МВІ Fermentas” (500 мМ ТРИС-НСІ (рН 8,8 при 25°С, 500мМ КСІ, 0,8% Nonidet Р40);
- 0,5 мкл эквимолярной смеси 100хdNTPs “МВІ Fermentas” (по 4пмоль dАТР, dСТР, dGТР, dТТР);
- 2 мкл 25 mМ MgCl<sub>2</sub>;
- 0,5 мкл по 5 пмоль каждого праймера («Синтол», Россия);
- 0,5 мкг исследуемой ДНК;
- 5 мкл. бидеионизованной воды.

Для предотвращения испарения во время реакции на реакционную смесь наслаивали вазелиновое масло.

Для денатурации ДНК проводили предварительную инкубацию пробирок в течение 5 минут при 95<sup>0</sup>С, затем проводили 30 циклов амплификации в соответствии с программой, описанной в таблице.

После завершения ПЦР проводили анализ продуктов амплификации гена *GSTM1* методом электрофореза в 2.5% агарозном геле. Электрофорез проводили при комнатной температуре в 0,5х ТВЕ-буфере (45 мМ Трис-борат рН 8.1, 2 мМ ЭДТА), содержащем 0,5 мкг/мл бромистого этидия, при напряжении электрического поля 5 В/см в течение 20-40 мин. Распределение фрагментов ДНК наблюдали и документировали в длинноволновом ультрафиолетовом свете с использованием прибора «BioDocAnalyse» и программного обеспечения фирмы «Biometra GmbH» (Goettingen, Germany).

#### II.4. Проведение Real-Time PCR

Так как классический метод анализа делеционного полиморфизма гена *GSTM1* позволяет вычлениить из общей выборки лишь гомозигот по делеции, то для оценки количества копий самого гена *GSTM1* был использован метод ПЦР в реальном времени с использованием TaqMan зондов.

Данный метод состоял в постановке мультиплексной ПЦР, в которой в качестве внутреннего контроля использовался ген – *B2M* ( $\beta$ -2-microglobulin).

**Таблица 2. Последовательности праймеров и зондов, использованных для постановки ПЦР в реальном времени.**

Ген	Нуклеотидные последовательности	Длина продукта
<i>GSTM1</i>	Праймер 5'- tga-gcc-ctg-ctc-ggt-tta-g-3'	59
	Праймер 5'- atg-ggc-atg-gtg-ctg-gtt-3'	
	Зонд FAM 5'- ctg-tct-gcg-gaa-tc-3' BHQ1	
<i>B2M</i>	Праймер 5'- ttg-ttt-cac-tgt-cct-gag-gac-tat-tta-t-3'	83

	Праймер 5'- atg-tta-ctc-tgt-caa-tgt-tct-cca-cat-3' Зонд ROX 5'-ctc-taa-cat-gat-aac-cct-cac-3' BHQ2	
--	---	--

FAM и ROX – флуоресцентные краситель (*GSTM1*), – флуоресцентный краситель (*B<sub>2</sub>M*), BHQ1 и BHQ2 – тушители флуоресценции.

Последовательности праймеров и зондов приведены в табл. 2. В состав реакционной смеси входили следующие реактивы фирмы «Синтол» (Москва, Россия) (на одну реакцию):

HotTaq-rescue-ДНК-полимераза (5 е.а./мкл) – 0,25 мкл

буферный раствор (x10) – 2.5 мкл

MgCl<sub>2</sub> (25 мМ) – 2.5 мкл

dNTPs (по 25 мМ) – 2.5 мкл

Праймеры (20 пмоль/мкл) – по 1 мкл

Зонд FAM (10 пмоль/мкл) – по 1 мкл

Зонд ROX (5 пмоль/мкл) – по 1 мкл

Геномная ДНК (20 нг/мкл) – 3 мкл

Бидеионизованная вода – до 25 мкл

Условия ПЦР:

1. 60°C – 1 мин.

2. 95°C – 10 мин.

3. 95°C – 15 сек.

4. 60 °C – 60 сек.

5. 25 °C – ∞

} 40 циклов

Реакция амплификации и измерение интенсивности флуоресценции ПЦР-продуктов проводились с использованием амплификатора StepOnePlus фирмы «Applied Biosystems», США.

### II.5. Генотипирование однонуклеотидных полиморфных замен

Проведенный в данной работе популяционный анализ полиморфизма регионов кластеров генов *GSTA* и *GSTM* основывался на данных полногеномного исследования однонуклеотидного полиморфизма, полученных с использованием ДНК-микрочипов фирмы Иллюмина (CNV370-Duo и Human 660W-Quad). Генотипирование образцов ДНК производилось в Биоцентре г. Тарту, (Эстония). Из всего объема данных, были отобраны только те однонуклеотидные полиморфизмы, которые располагаются в области локализации кластеров генов *GSTA* и *GSTM*. Набор полиморфных локусов был определен, исходя из границ кластеров генов *GSTA* и *GSTM* (Human genome reference assembly GRCh37.p5).

## **III. Методы анализа результатов**

### III.1. Математический метод обсчета данных Real-Time PCR

Определение количества копий гена *GSTM1* в анализируемых образцах определяли в значениях показателя (коэффициента) R по формуле:

$$R = 2^{-\Delta\Delta Ct}, \text{ где}$$

$\Delta\Delta Ct = (Ct_{B2M} \text{ контрольный образец} - Ct_{GSTM1} \text{ контрольный образец}) - (Ct_{B2M} \text{ анализируемый образец} - Ct_{GSTM1} \text{ анализируемый образец})$ .

В качестве контрольного образца использовался образец – известная гетерозигота по делеции.

Значения коэффициента R больше 1,4 ( $R > 1,4$ ) интерпретировались как указывающие на наличие в исследуемом образце двух копий гена *GSTM1*, т.е. исследуемый образец являлся нормальной гомозиготой. Значения R в диапазоне от 0,7 до 1,3 ( $0,7 < R < 1,3$ ) рассматривались как свидетельствующие о присутствии в образце лишь одной копии гена *GSTM1* (гетерозигота по делеции).

### III.2. Статистические методы анализа

Статистический анализ полученных данных проводили с использованием программы PowerMarker 3.25 [Liu K. et. al., 2005].

Достоверность различий частот встречаемости индивидуальных аллелей и гаплотипов между популяциями оценивали с помощью программ GraphPad InStat (версия 3.00, GraphPad Software, Inc, San Diego, CA, USA) и «RxC» [Ward R. et. al., 1991].

Для количественной оценки вариабельности частот встречаемости аллелей между популяциями использовали индексы фиксации Райта  $F_{ST}$ , также известные как коэффициенты взаимного родства. Расчет коэффициентов осуществлялся с использованием алгоритма Вейра-Кокерхама [Weir, et. al., 1984].

Для получения картин взаиморасположения популяций в пространстве матрицы попарных величин  $F_{ST}$  подвергались многомерному шкалированию с использованием программы STATISTICA (version 6.0, StatSoft, Inc, Tulsa, OK, USA).

Вероятность существования аллелей в составе гаплотипов оценивалась на основе анализа статистик неравновесия по сцеплению. Данные по неравновесию по сцеплению в парах полиморфных локусов были получены с использованием программы Haploview 4.2 [Barrett J. et. al., 2005]. Основываясь на величинах показателей неравновесия по сцеплению между локусами ( $D' \geq 0.7$ ), Полиморфизмы со значениями показателя неравновесия по сцеплению  $D' \geq 0.7$  объединялись в гаплотипические блоки.

Для анализа полногеномных данных полиморфизма популяций был применен метод главных компонент (Price et al., et. al., 2006). Анализ был осуществлен с использованием соответствующего модуля (the Genotypic Principal Components Analysis module) программного пакета SNP and Variation Suite (version 7.4.0, (Golden Helix, Inc., Vozeman, MT). Для того, чтобы не получить завышенных значений эйгенвекторов (главных компонент) первоначальный запуск программы осуществлялся с включенной опцией «outlier removal procedure» [Luca et al., 2008], во время которого из анализируемой выборки исключались образцы (индивидуумы) со значениями эйгенвекторов,

превышающими 6 стандартных отклонений от средних значений любого из первых 10 векторов.

Все различия рассматривались как статистически значимые при  $p < 0.05$  (соответствующие пороговые уровни значимости рассчитывались с учетом поправок на множественность сравнений).

## ГЛАВА 3. РЕЗУЛЬТАТЫ

### 3.1 Анализ популяционной изменчивости двух кластеров генов *GSTA* и *GSTM*

#### 3.1.1 Анализ изменчивости в кластере генов *GSTA*

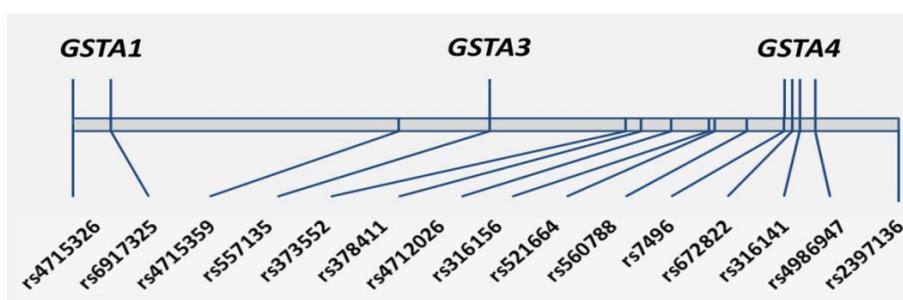
В настоящей работе, для проведения анализа генетического разнообразия двух геномных кластеров – *GSTA* и *GSTM*, были использованы результаты полногеномного сканирования русских популяций, полученные с использованием ДНК микрочипов фирмы Illumina (Khrunin et al., 2013). Из всего массива данных, нами 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 НарМар. 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 НарМар: европейская CEU, китайская – CHB, японская – JPT и нигерийская – YRI.

На рисунке 3 изображены исследованные однонуклеотидные полиморфные локусы локализуемых в состав кластера генов *GSTA*. Данная система генов находится на шестой хромосоме в районе p12.1 и имеет протяженность около 250 kb. Расположение полиморфизмов приведено в соответствии с их локализацией относительно генов данного кластера. Всего в кластере генов *GSTA*, было проанализировано 15 однонуклеотидных полиморфных маркеров.

Сначала нами был проведен анализ частоты встречаемости каждого из полиморфизмов в популяциях. Соответствующие данные представлены в таблице 3. 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. В приведенных данных таблицы 3 об аллельных частотах анализируемых локусов, обращают на себя внимание локусы rs672822 и rs4986947. Предковые варианты обоих сайтов представлены в популяциях человекообразных обезьян в форме мономорфных локусов, в которых зафиксирован аллель А, со 100% частотой встречаемости. Наиболее интересным является распределение вариантов аллелей rs4986947. В популяциях африканского и азиатского происхождения предковый аллель отсутствует – в данной позиции со 100% частотой встречается «молодой» аллель G. Однако в популяциях европейского происхождения предковый вариант встречается в виде минорного аллеля – А, с частотой встречаемости 6% [[http://www.ensembl.org/Homo\\_sapiens/](http://www.ensembl.org/Homo_sapiens/)], в анализируемых русских популяциях частота данного аллеля равна 9%.

Что касается локуса rs672822, то в азиатских популяциях он является мономорфным, причем по предковому аллелю. Однако в популяциях из Африки и Европы (включая популяции из России), он представлен обоими аллельными вариантами. Это может быть связано с особенностями микроэволюционных процессов, приведших к фиксации предкового аллеля в азиатских популяциях, и наличием обоих вариантов в Африке и Европе.



**Рисунок 3. Исследованные однонуклеотидные полиморфизмы в кластере генов *GSTA*.**

Частоты минорных аллелей полиморфных локусов кластера *GSTA*

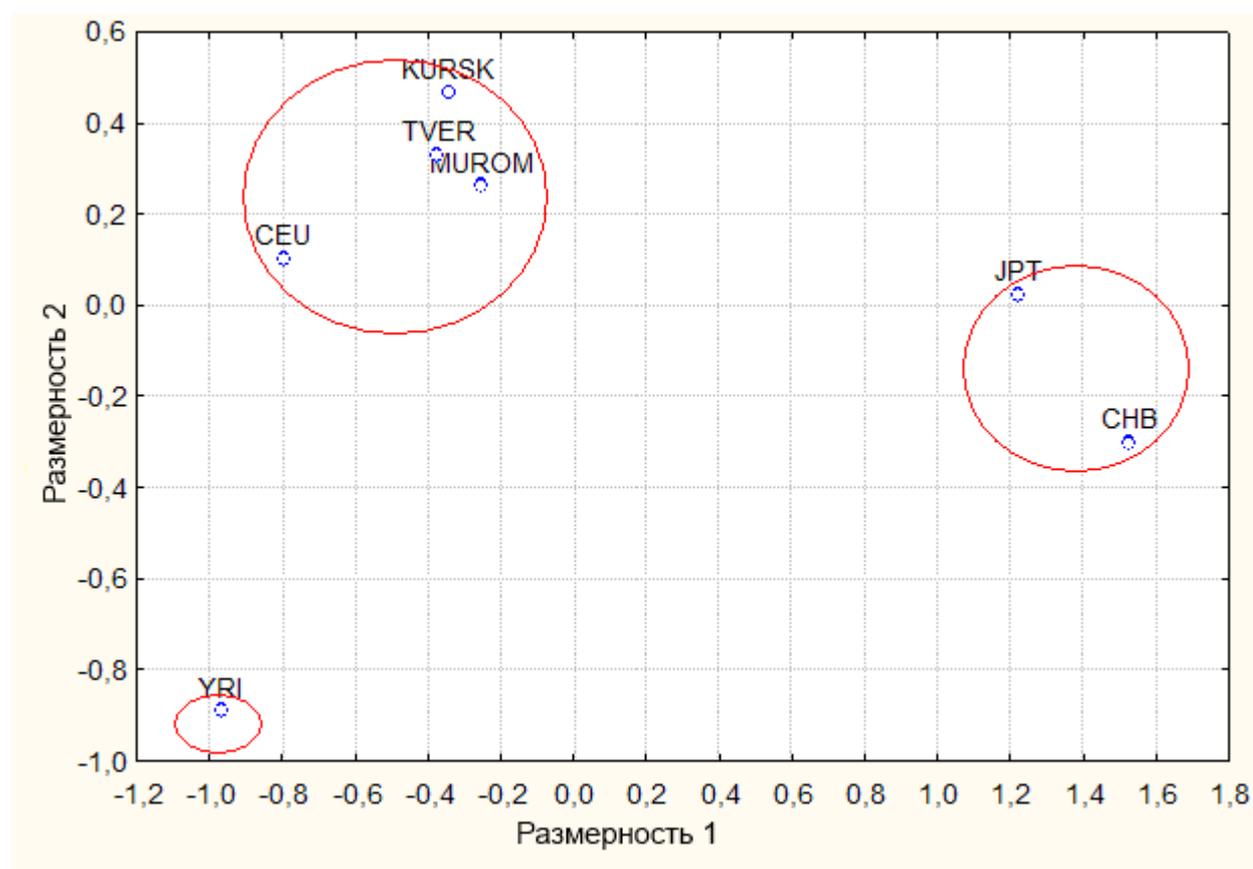
Маркер	MAF	Тверь 2N=192	Муром 2N=192	Курск 2N=192	CEU 2N=330	CHB 2N=168	JPT 2N=172	YRI 2N=332
rs4715326	C	0.396	0.400	0.403	0.406	0.095	0.128	0.299
rs6917325	T	0.396	0.370	0.416	0.409	0.096	0.128	0.298
rs4715359	G	0.339	0.292	0.346	0.312	0.133	0.151	0.260
rs557135	G	0.318	0.313	0.321	0.348	0.500	0.477	0.689
rs373552	G	0.245	0.234	0.223	0.279	0.042	0.105	0.022
rs378411	G	0.240	0.234	0.223	0.279	0.042	0.105	0.048
rs4712026	G	0.141	0.094	0.121	0.128	0.337	0.244	0.111
rs316156	G	0.135	0.146	0.089	0.161	0.226	0.291	0.117
rs521664	C	0.396	0.375	0.432	0.288	0.559	0.547	0.078
rs560788	T	0.495	0.464	0.500	0.461	0.232	0.227	0.629
rs7496	T	0.172	0.172	0.132	0.1	0.149	0.088	0.247
rs672822	G	0.063	0.125	0.101	0.104	0	0	0.060
rs316141	A	0.443	0.438	0.458	0.445	0.253	0.250	0.482
rs4986947	A	0.078	0.052	0.104	0.055	0	0	0
rs2397136	G	0.151	0.125	0.161	0.145	0.048	0.111	0.299

Таблица 4

Межпопуляционные индексы фиксации ( $F_{ST}$ ) для кластера *GSTA*.

	Тверь	Муром	Курск	CEU	CHB	JPT	YRI
Муром	0.002	0.000					
Курск	0.001	0.003	0.000				
CEU	0.008	0.010	0.012	0.000			
CHB	0.110	0.106	0.114	0.135	0.000		
JPT	0.093	0.088	0.095	0.115	0.007	0.000	
YRI	0.084	0.088	0.092	0.065	0.198	0.192	0.000

Для количественной оценки уровней межпопуляционных вариаций частот аллелей нами были рассчитаны индексы фиксации  $F_{ST}$  (таблица 4). На рисунке 4 представлены результаты обработки полученной матрицы линейризованных 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.

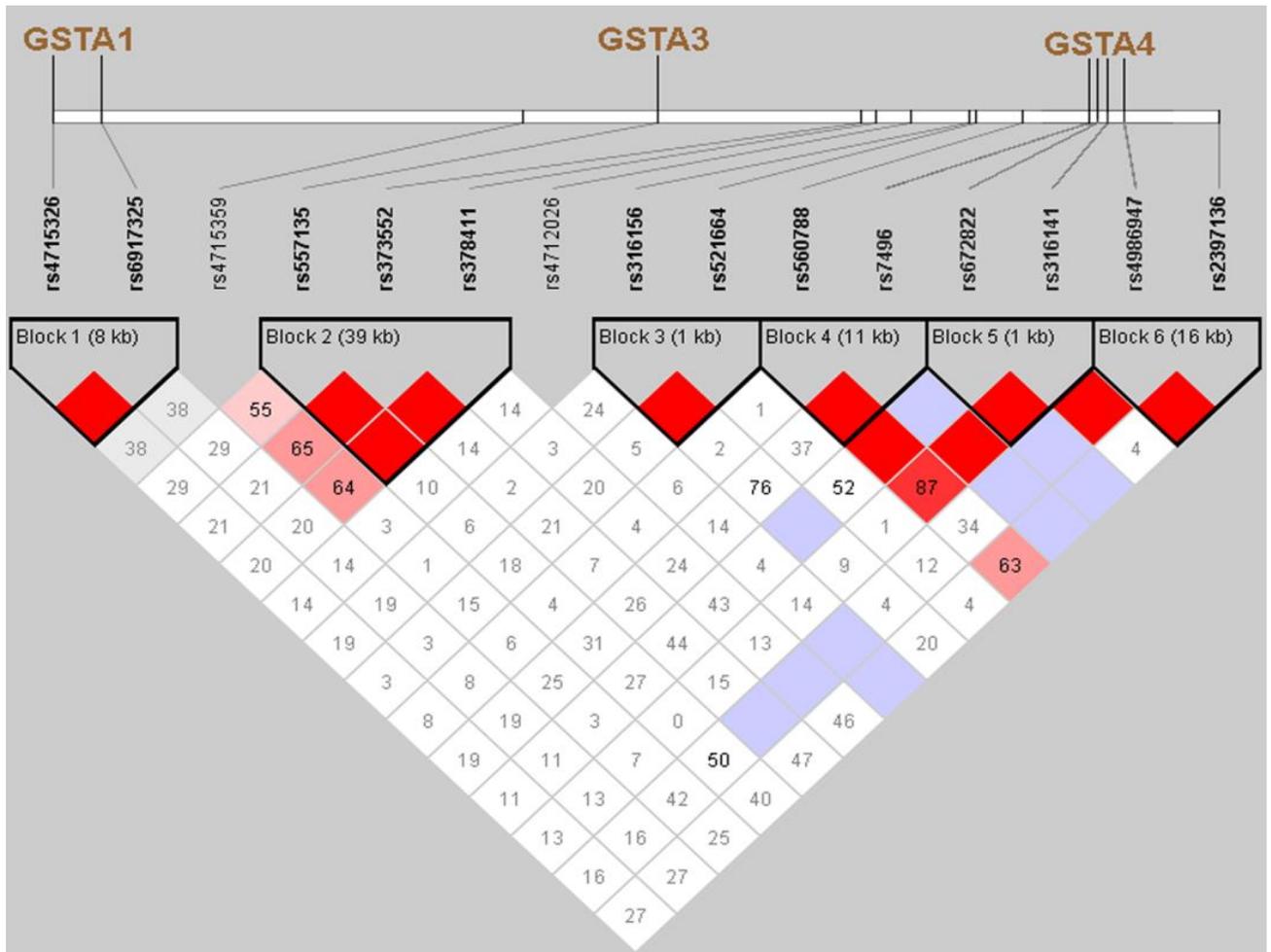


**Рисунок. 4.** Двумерный график многомерного шкалирования генетических расстояний ( $F_{ST}$ ) между популяциями русских и популяциями из проекта ХарМар, полученный по результатам анализа однонуклеотидного полиморфизма в кластере *GSTA*.

Для того чтобы сделать анализ генетической изменчивости изучаемого геномного региона более комплексным, 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. Гапблоку кластера генов *GSTA* были идентичными для всех исследуемых популяций русского населения из центрально-европейского региона России. В качестве примера, на рисунке 5 представлен график паттернов неравновесия по сцеплению, полученный для популяции русских из Андреапольского района Тверской области. 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. В тоже время, не все эти гапблоку были вычленены в популяциях из проекта HarMap (см. ниже).

Частоты встречаемости гаплотипов в составе выделенных гапблоку в анализируемых популяциях приведены в таблице 5. Видно, что в разных гапблоку имеется разное число гаплотипов, так, например, в гапблоку № 1 было выявлено только 2 гаплотипа, а в гапблоку № 4 – 4 гаплотипа. В отдельных популяциях ряд гапблоку не вычленялся, а именно, в популяциях СНВ и JPT блоки №5, №6 и в популяции YRI блок №6. Это было обусловлено тем, что взятые в анализ локусы данных блоку в этих популяциях оказались мономорфными.



**Рисунок. 5. Паттерны неравновесия по сцеплению и гаплоблоки, вычлененные в регионе кластера генов *GSTA* (на примере популяции из Андреапольского района Тверской области). Цифры в ячейках указывают величину коэффициента неравновесия по сцеплению ( $D'$ ), между каждой парой локусов. В пустых ячейках темного цвета  $D'=1$ .**

Сравнение популяций основывалось на сопоставлении частот гаплотипов, рассчитанных для каждого из блоков. Результаты сравнений в виде полученных величин вероятности «Р» представлены в таблице 6. Уровни значимости «Р» 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. 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 5

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	-	-	-

\* ввиду отсутствия некоторых гаплотипических блоков в популяциях CHB, JPT и YRI в соответствующих колонках поставлены прочерки.

Таблица 6.

Сравнение частот гаплотипов генов кластера *GSTA* в популяциях Твери, Муром, Курска и в популяциях проекта НаpMap (CEU, CHB, JPT, YRI)

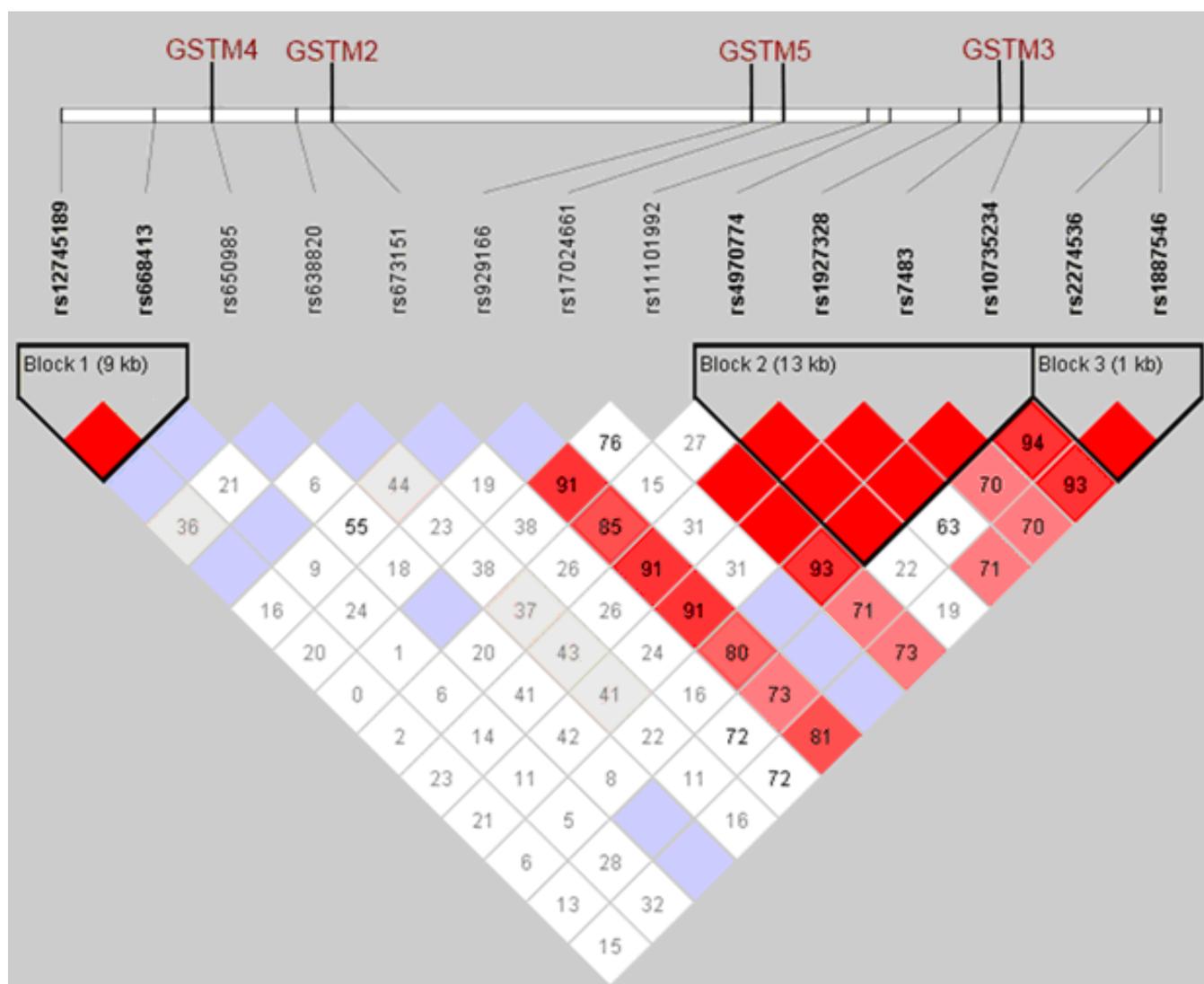
	№ ГАПЛО-БЛОКА	МУРОМ	КУРСК	CEU	CHB	JPT	YRI	$P_{\text{МНОЖ}}^*$
ТВЕРЬ	I	0,6746	0,8353	0,8534	<0,0001	<0,0001	0,0269	0,004
	II	0,9599	0,4282	0,6984	<0,0001	<0,0001	<0,0001	0,0025
	III	0,9030	0,3306	0,0403	<0,0001	<0,0001	<0,0001	0,0025
	IV	0,8238	0,2076	0,0326	<0,0001	<0,0001	0,0065	0,0019
	V	0,0484	0,3957	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

\* Пороговые значения вероятности «Р», рассчитанные с учетом поправки на множественность сравнений индивидуально для каждого из блоков по формуле:  $P=0,05/(n \times a) - 1$ , где n – число гаплотипов, а – число популяций [Glantz S. 1999]. Отсутствие некоторых гаплотипических блоков в популяциях CHB, JPT и YRI не дает возможности попарного сравнения популяций, в таких случаях в соответствующих колонках поставлены прочерки.

### 3.1.2 Анализ изменчивости кластера генов *GSTM*

Кластер генов *GSTM* располагается на хромосоме 1 в регионе p13.3, охватывая участок в 85kb. В пределах данного участка было проанализировано 14 маркерных локусов (рисунок 6). Их аллельные частоты представлены в таблице 7. 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  $F_{ST}$  (таблица 8) демонстрирует высокое сходство с приводившимся ранее графиком кластера *GSTA* (рисунок 7).

По результатам анализа паттернов неравновесия по сцеплению между исследованными локусами в составе кластера *GSTM* было выделено 3 гаплотипических блока (рисунок 6). Данные о частотах гаплотипов для каждого из гапоблоков в кластере *GSTM* представлены в таблице 9. Как и в случае кластера *GSTA*, внутри блоков наблюдалось различное количество гаплотипов. Основываясь на данных таблицы 10, 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, Мурома, 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.



**Рисунок 6. Исследованные однонуклеотидные полиморфизмы и паттерны неравновесия по сцеплению внутри кластера генов *GSTM* (на примере популяции из Андреапольского района Тверской области).**

Таблица 7

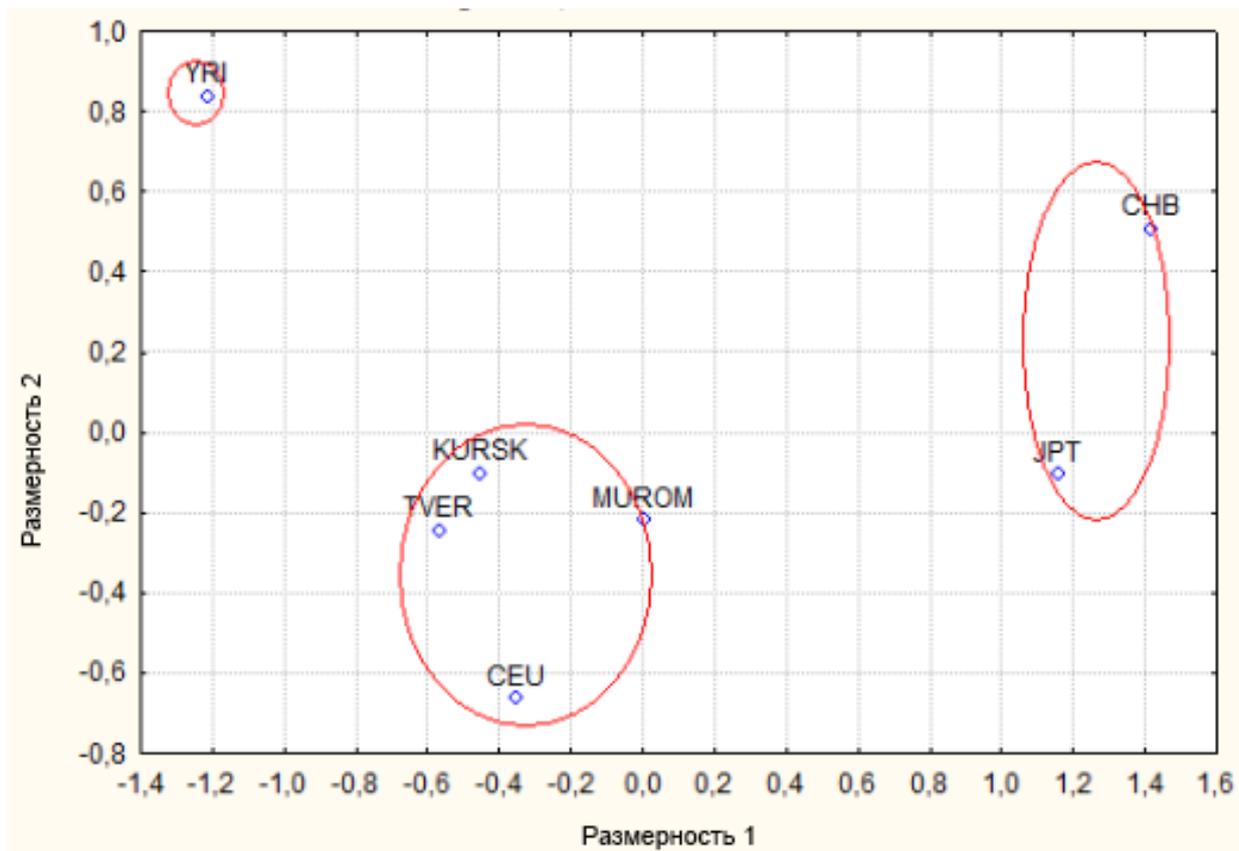
Частоты минорных аллелей полиморфных локусов кластера *GSTM*

Маркер	MAF	Тверь 2N=192	Муром 2N=192	Курск 2N=192	CEU 2N=330	CHB 2N=168	JPT 2N=172	YRI 2N=332
rs12745189	T	0,443	0,458	0,495	0,402	0,613	0,686	0,377
rs668413	A	0,458	0,411	0,400	0,463	0,317	0,291	0,467
rs650985	C	0,042	0,063	0,054	0,070	0	0	0,003
rs638820	G	0,432	0,474	0,463	0,479	0,369	0,318	0,452
rs673151	T	0,036	0,063	0,053	0,042	0,012	0	0,015
rs929166	G	0,281	0,333	0,289	0,258	0,720	0,622	0,120
rs17024661	G	0,021	0,037	0,011	0,042	0,006	0	0,039
rs11101992	C	0,307	0,266	0,307	0,250	0,133	0,140	0,809
rs4970774	C	0,411	0,458	0,414	0,439	0,732	0,674	0,521
rs1927328	G	0,279	0,349	0,287	0,297	0,756	0,721	0,036
rs7483	T	0,266	0,339	0,287	0,299	0,756	0,721	0,033
rs10735234	G	0,391	0,375	0,379	0,430	0,143	0,180	0,096
rs2274536	G	0,307	0,313	0,344	0,288	0,113	0,151	0,632
rs1887546	T	0,297	0,313	0,339	0,288	0,095	0,112	0,596

Таблица 8

Межпопуляционные индексы фиксации  $F_{ST}$  для кластера *GSTM*

	Тверь	Муром	Курск	CEU	CHB	JPT	YRI
Муром	0,005	0,000					
Курск	0,002	0,003	0,000				
CEU	0,004	0,005	0,006	0,000			
CHB	0,193	0,151	0,182	0,182	0,000		
JPT	0,165	0,125	0,153	0,154	0,007	0,000	
YRI	0,138	0,156	0,134	0,157	0,401	0,372	0,000



**Рисунок 7.** Двумерный график многомерного шкалирования матрицы генетических расстояний ( $F_{ST}$ ) между изучаемыми популяциями русских и популяциями из проекта HarMap, полученный по результатам анализа однонуклеотидного полиморфизма кластера *GSTM*.

Таблица 9

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 10

Сравнение гаплотипического разнообразия 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

### **3.2 Анализ делеционного полиморфизма гена *GSTM1* в контексте гаплотипического разнообразия геномного кластера *GSTM* в трех русских популяциях.**

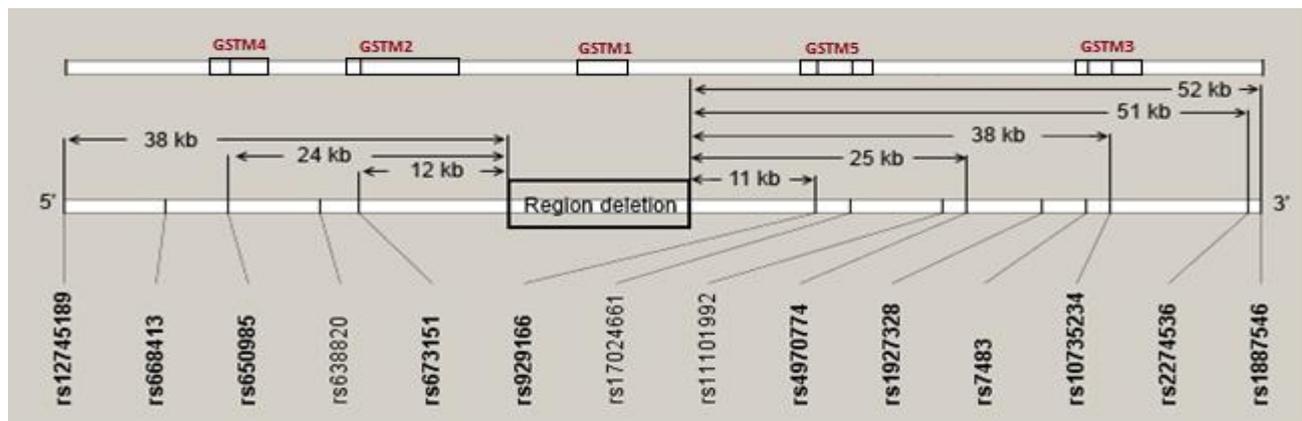
#### **3.2.1 Анализ гаплотипического разнообразия геномного кластера *GSTM* в трех русских популяциях в зависимости от делеционного полиморфизма гена *GSTM1***

Данная часть работы отражает исследования специфики гаплотипического разнообразия геномного региона *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. Целью работы было изучить, есть ли взаимосвязи между делецией гена *GSTM1* и представленностью тех или иных гаплотипов в составе гаплотипических блоков кластера *GSTM1*.

Делетируемый участок находится между исследованными локусами rs673151 и rs929166, расположенными в 5 и 1 интронах генов *GSTM2* и *GSTM5*, соответственно, и расстояние между которыми составляет около 42kb (рисунок 8). Локусы, вошедшие в состав гапблоков, выделены жирным шрифтом.

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 *GTP-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 6).



**Рисунок 8. Взаимное расположение области делеции гена *GSTM1* и анализируемых полиморфизмов.**

В каждой из подгрупп нами был произведен анализ паттернов неравновесия по сцеплению в гаплотипических блоках. Частоты встречаемости конкретных гаплотипов в подгруппах *GSTM1* «0» и *GSTM1* «+» в каждой из исследованных популяций отображены в таблице 11.

Из представленных в таблице 11 данных видно, что в первом гапоблоке частоты гаплотипов в подгруппах *GSTM1* «0» и *GSTM1* «+» во всех 3-х анализируемых популяциях русского населения имеют схожие значения. Количество выявленных гаплотипов в составе первого блока равно трем и одинаково для обеих подгрупп всех популяций, взятых в анализ. В итоге, по этому гапоблоку достоверных различий между изучаемыми подгруппами *GSTM1* «0» и *GSTM1* «+» не было выявлено ни в одной из популяций.

Несколько иная ситуация наблюдается при рассмотрении второго гапоблока. 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. В популяции из Твери в подгруппе *GSTM1* «0» выделяется 5 гаплотипов, а в подгруппе *GSTM1* «+» – 4 гаплотипа. Различия наблюдаются и в частотах гаплотипов: так для первой подгруппы наиболее частым гаплотипом является гаплотип CGTA, а для второй подгруппы – AACG. Схожая ситуация

наблюдается и при рассмотрении второго гаплоглоба в Муромской популяции. Здесь также имеется различное число гаплотипов в обеих подгруппах, с той лишь разницей, что для первой подгруппы выявлено 4 гаплотипа, а для второй – 5. В обеих подгруппах данной популяции гаплотипами с самыми высокими частотами так же, как и для Тверской популяции, являются разные гаплотипы – гаплотипы AACG и CGTA. Для популяции из Курска наблюдается отличная от предыдущих двух случаев ситуация. В обеих подгруппах второй гаплоглоб представлен только 4 гаплотипами, а наиболее частым в обеих подгруппах является гаплотип AACG. Статистически значимыми в данном гаплоглобе являются различия в распределении частот гаплотипов между подгруппами *GSTM1*«0» и *GSTM1*«+» в тверской популяции.

Третий гаплоглоб характеризуется разницей по числу гаплотипов. В двух случаях это различие обнаруживается между двумя подгруппами, а именно, в популяции из Твери и Курска в первой подгруппе 3 гаплотипа во второй – 2. В то же время, во всех случаях гаплотип с самой высокой частотой – AG, а частоты третьего гаплотипа (GG) достаточно малы. Здесь достоверные различия между изучаемыми подгруппами наблюдаются лишь в популяции из Мурома.

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. С позиций оценки внутрипопуляционного гаплотипического разнообразия полученные данные могут свидетельствовать о наличии некоторых различий между исследованными популяциями. Однако, при этом, данные различия никак не проявили себя при межпопуляционном сравнении распределения гаплотипов в каждой из рассматриваемых подгрупп *GSTM1* «0» и *GSTM1* «+».

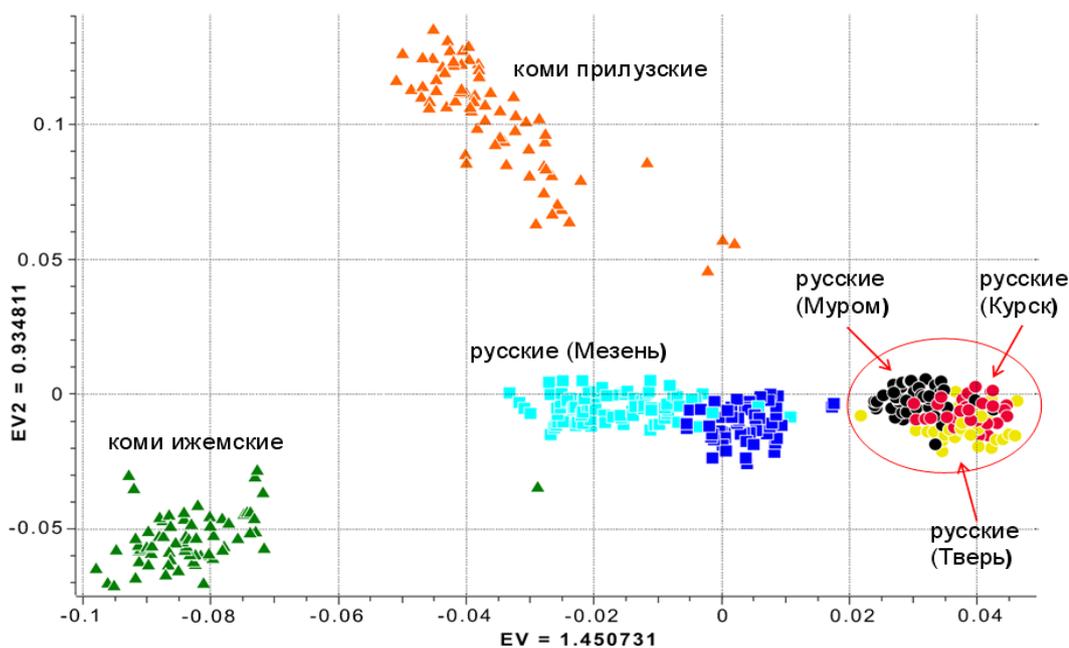
Таблица 11.

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

№ гаплогло-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	

\*С учетом поправки на множественность сравнений статистически значимыми являются различия, где  $P < 0,0028$  [Glantz S. 1999].

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, Мурома и Курска. Правомочность данного объединения основывалась на данных проведенного нами полногеномного анализа однонуклеотидного полиморфизма в популяциях Европейской части России (Khrunin et al., 2013). Результаты этого анализа представлены на рисунке 9, на котором в пространстве первой и второй главных компонент отражено взаиморасположение исследованных популяций: популяции вепсов, двух популяций коми и четырех популяций русского населения. Каждая точка, каждый отдельный значок соответствует здесь геномным характеристикам отдельного человека. И эти точки располагаются в данном виртуальном пространстве не хаотично, а формируют «этнические» облака. Видно, что облака курской, тверской и муромской популяций практически полностью перекрываются, формируя единое ядро, четко дифференцирующееся от остальных популяций.



**Рисунок 9. Популяционная структура населения Европейской части России.**

### 3.2.2 Анализ гаплотипического разнообразия кластера *GSTM* в двух подгруппах в объединенной выборке трёх русских популяций

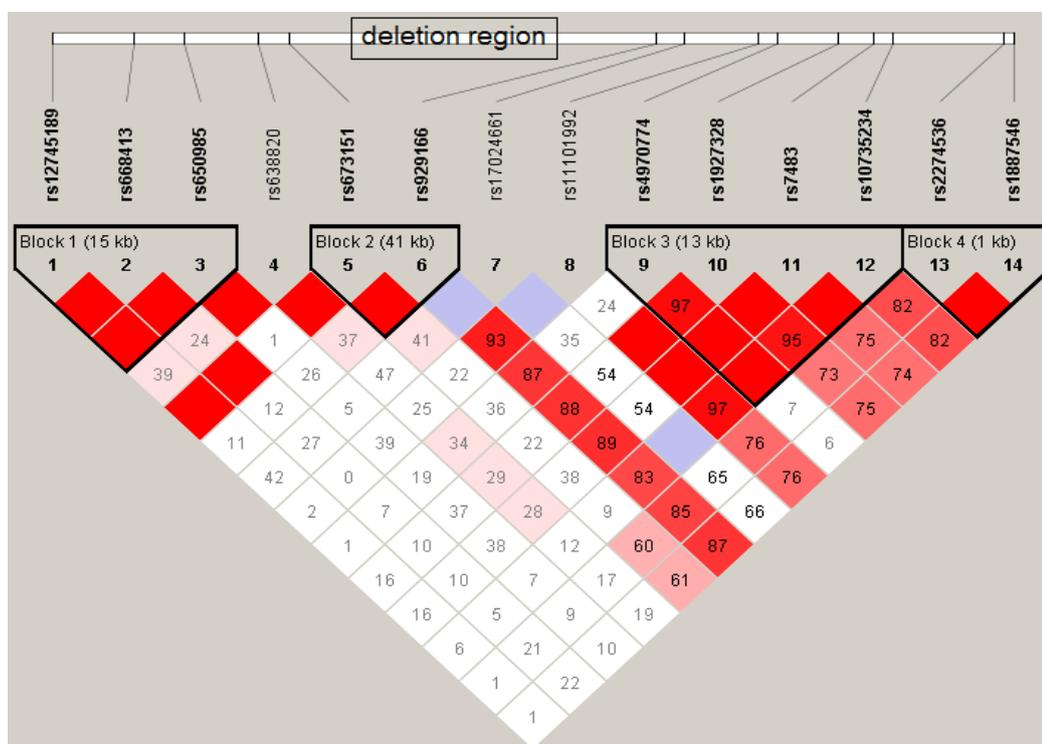
Изменения в схеме анализа, т. е. переход от анализа разнообразия гаплотипов в индивидуальных популяциях к его анализу в составе объединенной популяции, привел к изменениям в количестве вычлняемых в кластере *GSTM* гапблоков, число которых возросло до четырех (рисунок 10). Кроме того, улучшившаяся статистика 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 12. 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.

В отличие от первого блока, во втором гапблоке, лишь один гаплотип – СТ – имеет наибольшую частоту встречаемости в обеих подгруппах индивидуумов. В гапблоке 3, также как и в предыдущем случае, наиболее частый гаплотип является идентичным для обеих подгрупп по делеции (гаплотип ААСГ). Второй по частоте встречаемости гаплотип СГТА третьего блока имеет разную частоту распределения в двух подгруппах, для *GSTM1* «0» его частота равна 0.374, для *GSTM1* «+» – 0.237. Аналогичная ситуация наблюдается и для гаплотипов ААСА и САСА, с той лишь разницей, что их максимальная частота наблюдается в подгруппе *GSTM1* «+», а не в подгруппе *GSTM1* «0», как в предыдущем случае. Частоты двух оставшихся гаплотипов СГСА и АГСГ были очень малы.

В четвертом гапблоке было выявлено только три гаплотипа. Однако гаплотипом с наибольшей частотой встречаемости, в обеих подгруппах, являлся один и тот же гаплотип, а именно – АГ. Второй гаплотип ГТ с более высокой

частотой был представлен в подгруппе *GSTM1* «+». Частота гаплотипа GG была крайне низка, и он фиксировался лишь в подгруппе *GSTM1* «0».

Полученные для каждого из блоков значения вероятности «P» 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 (0.006). Результаты сравнений позволили предположить существование взаимосвязи между гаплотипическим разнообразием кластера *GSTM1* 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*.



**Рисунок 10. Паттерны неравновесия по сцеплению и вычлененные гаплоглоки в кластере генов *GSTM1* в объединенной русской популяции.**

Таблица 12.

**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	

\*Расчитанный, с учетом поправки на множественность сравнений, пороговый уровень значимости  $P=0,006$  [ Glantz S. 1999]

В таблице 13 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 делетируемого региона, причем не на очень большом расстоянии (рисунок 8). Из данных таблицы 13 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 14).

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 13 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 ( $P < 0.0024$ ), 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 14).

Таблица 13.

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 14.

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>

\* В таблице приведены значения вероятности «Р», полученные для каждой из пар сравнения. Статистически значимые величины ( $P < P_{\text{МНОЖ}} = 0.0042$ ) выделены жирным шрифтом.

## ГЛАВА 4. ОБСУЖДЕНИЕ

### 4.1 Исследование изменчивости двух кластеров генов *GSTA* и *GSTM*, основанном на результатах полногеномного сканирования

За последние годы проделана большая работа по полногеномному сканированию (полногеномному исследованию полиморфизма) многих популяций человека с помощью ДНК-микрочипов, содержащих сотни тысяч SNPs, результатом которой стали данные о глобальной изменчивости генома человека [Li et al., 2008; Auton et al., 2009; Hellenthal et al., 2014]. Однако полученная информация также может быть использована и для анализа изменчивости отдельных геномных регионов, маркирующих пути эволюционной траектории, как основных групп человека (рас), так и локальных популяций [Sabeti et al., 2007; Li et al., 2011]. В настоящей работе нами была предпринята попытка проанализировать изменчивость двух геномных участков, включающих кластеры генов *GSTA* и *GSTM*, основанная на данных полногеномного сканирования русских популяций с использованием микрочипов фирмы Illumina и их сравнение с популяциями из проекта HarMap. Интерес к исследованию популяционной изменчивости генов именно системы глутатион-S-трансфераз был обусловлен, прежде всего, их важной ролью в метаболизме широко спектра соединений как эндогенного, так и экзогенного (ксенобиотики) происхождения.

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 Yoruba из Нигерии. Так как европейские популяции кардинально отличались от популяций из Китая и Японии по частотам гаплотипов всех гапблоков, можно предположить, что отмеченное сходство может быть следствием каких-либо особых черт этих гапблоков в микроэволюционной истории популяций. В то же время русские и нигерийская популяции статистически значимо отличались по всем остальными гапблокам обоих генных кластеров.

Еще одним интересным найденным фактом, заслуживающим отдельного рассмотрения, являются аллельные статусы двух SNP rs672822 и rs4986947, входящих в состав 6 блока кластера *GSTA*, локализующегося в интроне гена *GSTA4*. У человекообразных обезьян оба данных SNPs сайтов несут лишь один нуклеотид А, характерный для обоих SNPs [Ensemble browser] и поэтому зафиксированный вариант аллеля может рассматриваться в качестве предковой формы. В двух рассмотренных популяциях из Азии – популяции из Китая и Японии, rs672822 представлен только предковым аллелем А с частотой встречаемости 100%. В популяциях из Африки (включая популяции Luhya и Maasai ) частота аллеля А в данной позиции составляет 96%, оставшиеся 4% приходятся на второй вариант – аллель G. Точно такая же картина наблюдается и в популяциях из Европы и России (частота аллеля А равна 91%, а частота аллеля G – 9%).

Особый интерес представляет локус rs4986947. В проанализированных популяциях из Азии и Африки в позиции со 100% частотой обнаруживается нуклеотид G («молодой» вариант); причем та же ситуация наблюдается и в двух других африканских популяциях из проекта HapMap – Luhya и Maasai [International HapMap Consortium, 2007; 1000 Genomes Project Consortium, 2012]. Напротив, в европейских популяциях, включая все исследованные популяции из России, этот локус представлен обоими аллельными вариантами (G/A), т.е. в этих

популяциях предковый аллель А также присутствует. Данный факт может иметь два объяснения. Первое предполагает возможность различных путей миграции из Африки как будущей европейской популяции, так и будущей азиатской популяции [McEvoy et al., 2011]. Общности, мигрирующие в Европу, могли включать индивидуумов с предковым вариантом А в локусе rs4986947, отсутствующим в современных референсных африканских популяциях. Второе объяснение подразумевает возможность обратной мутации, и ее возвращение к предковому состоянию у части европейцев. Сохранение аллеля А у европейцев может быть результатом действия естественного отбора, роль которого в становлении межэтнических вариаций в генах *GSTs* была предположена в работе Polimanti [Polimanti et al., 2001]. Наряду с популяциями русских и CEU предковый аллель А также обнаруживается с частотой около 6% в географически дистантных популяциях финнов, итальянцев и англичан [1000 Genomes Project Consortium, 2012]. Такая достаточно низкая частота встречаемости может быть следствием балансирующего отбора.

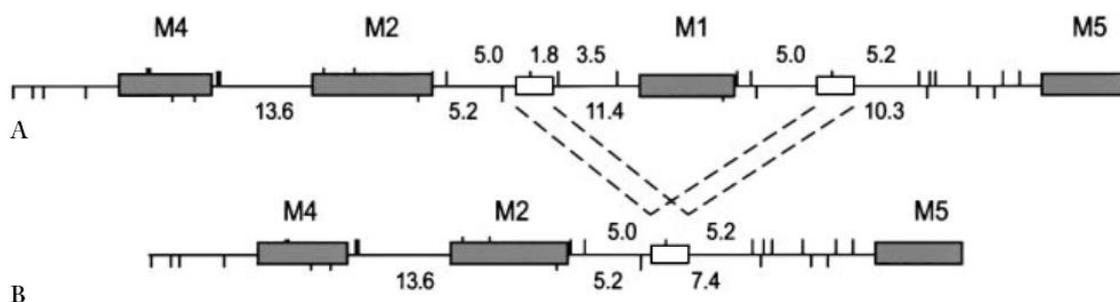
Таким образом, результаты исследования позволяют предполагать, что эволюционная история кластеров генов *GSTA* и *GSTM* имеет немало общих черт, и они, по-видимому, маркируют одни и те же события эволюционных траекторий основных групп человека.

#### **4.2 Исследование делеции гена *GSTM1* в контексте гаплотипического разнообразия геномного кластера *GSTM* в трех русских популяциях**

Особенностью кластера *GSTM*, в отличие от *GSTA*, является существование в его пределах довольно протяженной делеции. Такого рода полиморфизм, к которому, помимо делеций, относятся инсерции и дупликации и называемый, в общем случае, полиморфизмом числа копий участков ДНК (copy number variation, CNV), является важным источником геномной вариабельности. Охватываемые им участки генома варьируют от нескольких до сотен тысяч оснований и по своим

последствиям могут существенно превосходить эффекты, обуславливаемые однонуклеотидными заменами. Поэтому обнаружение CNVs и установление механизмов их появления (поведения) в геноме имеет значение для понимания молекулярных основ возникновения тех или иных заболеваний, а также особенностей эволюционной истории популяций. Однако, несмотря на бурное развитие технологий, возможности непосредственного адекватного определения вариаций числа копий, особенно, находящихся в регионах, содержащих сегментные дубликации и повторы, остаются проблематичными [Conrad, et al., 2007; Buchanan, et al., 2008; Katju, et al., 2013]. Одним из альтернативных подходов к исследованию полиморфизма числа копий является изучение структуры прилегающих к ним регионов, в частности полиморфизма однонуклеотидных замен [Conrad et al., 2010]. В настоящей работе нами было проведено исследование аллельного (гаплотипического) разнообразия SNPs региона, в котором находятся гены класса *GSTM*, включая, в том числе, делетируемый ген *GSTM1*. Молекулярные основы возникновения делеции в кластере *GSTM* были исследованы в работе Shi-jie Xu, 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* [Xu S. 1998] (рисунок 11, А). Протяженность каждого из них составляет 4.2kb. Левый регион располагается на расстоянии 5kb от 3'-конца гена *GSTM2* и 5kb от 5'-конца гена *GSTM1*, правый – на расстоянии 5kb от 3'-конца гена *GSTM1* и 10kb от 5'-конца *GSTM5*. У индивидуумов с делецией гена *GSTM1* обнаруживалась лишь последовательность, идентичная фланкирующим участкам (рисунок 11, В).

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*.



**Рисунок 11. Структура участка хромосомы в регионе кластера генов *GSTM*.**

А – хромосома с геном *GSTM1* (высокогомологичные последовательности, обрамляющие участок гена *GSTM1* обозначены прямоугольниками белого цвета),  
 В – хромосома с делецией гена *GSTM1* (приведено из работы Xu et al. (1998) с небольшими изменениями).

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, возникла 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*.

К сожалению даже в последних версиях проекта 1000 Genomes не представлены данные, касающиеся протяженных делеций. На сегодняшний день есть лишь данные по косвенному определению делеционных вариантов в геноме человека, демонстрирующие существование неравновесия по сцеплению между делециями и рядом расположенными SNPs [McCarroll et al., 2006; Conrad et al., 2005]. Есть данные, что рекомбинация может зависеть от нуклеотидного контекста прилегающих геномных регионов. Так, например, показана прямая корреляция между рекомбинацией и нуклеотидным разнообразием прилегающих областей (Tenaillon et al. 2004, Kulathinal et al. 2008, Lohmueller et al. 2011). Для точковых мутаций в работе Schaibley V.M., 2013 было показано, что геномный контекст может оказывать влияние на возникновение мутаций с последующим их закреплением и фиксацией в популяциях.

На основании вышесказанного можно полагать, что обнаруженная корреляция между делецией гена *GSTM1* и распределением частот гаплотипов в кластере *GSTM* указывает на возможность предпочтительного возникновения делеции в определённых вариантах геномного окружения. Для более точного ответа на вопрос о взаимовлиянии геномного контекста и делеционного полиморфизма было бы интересно проверить наши предположения на данных популяциях, в том числе с иными (меньшими) частотами встречаемости делеции гена *GSTM1*.

## ЗАКЛЮЧЕНИЕ

В настоящей работе было проведено популяционное исследование однонуклеотидного полиморфизма геномных регионов, включающих гены глутатион S-трансфераз из классов альфа (*GSTA*) и мю (*GSTM*). Оба класса являются одними из наиболее крупных по количеству входящих в их состав генов GSTs и занимают протяженные участки 1 и 6 хромосом. Первый этап анализа был основан на сравнении частот минорных аллелей SNPs, входящих в состав каждого из кластеров генов. Данный сравнительный анализ частот аллелей показал отсутствие различий между частотами встречаемости минорных аллелей каждого SNPs в европейских популяциях, к которым относятся три русские популяции, а также европейская по происхождению популяция – CEU. Этот же анализ показал существенные отличия в распределении частот минорных аллелей между европейскими популяциями и популяциями из Китая, Японии и Нигерии. Уровень межпопуляционных вариаций частот был оценен с использованием индексов фиксации  $F_{ST}$ . Обработка полученной линейаризованной матрицы попарных величин  $F_{ST}$  методами многомерного шкалирования позволила получить картины взаиморасположения популяций. Эти картины оказались очень похожими для каждой из групп маркеров: в обоих случаях взятые в исследование популяции формировали три отдельных кластера, в состав которых вошли популяции близкие по происхождению.

Оценка генетического разнообразия, основанная на сравнительном анализе частот гаплотипов, также показала отсутствие выраженных различий при сравнении русских популяций между собой и в парах сравнений: русские популяции – европейская по происхождению популяция CEU, что, очевидно, обусловлено их единой расовой принадлежностью. Дальнейшее сравнение показало наличие достоверных различий в распределении частот гаплотипов в парах: европейская выборка – популяция из Китая; европейская выборка – популяция из Японии; европейская выборка – популяция из Нигерии. Величины

показателя вероятности «Р» в большинстве сравнений были ниже установленного уровня значимости, определенного для каждого из гаплоблоков. В случае сравнения европейских популяций с популяциями китайцев и японцев достоверные различия между популяциями были выявлены по всем блокам в обоих кластерах генов. В то же время, в случаях попарных сравнений популяций русских с популяцией YRI из Нигерии имело место сходство распределения гаплотипов по некоторым блокам обоих кластеров. Так как по всем оставшимся гаплотипическим блокам отмечались статистически значимые различия, выявленное сходство может быть связано с какими-то особенностями микроэволюционной истории этих гаплоблоков в данных популяциях,

Из всего выше сказанного можно заключить, что, несмотря на то, что кластеры генов *GSTA* и *GSTM* находятся на разных хромосомах, результаты исследования позволяют предполагать, что их эволюционная история имеет немало общих черт. Возможно, что оба кластера маркируют одни и те же события эволюционных траекторий во многих человеческих популяциях. Они могут свидетельствовать о возможной генетической подразделенности предковой африканской популяции и существование в ней разных субпопуляций, которые приняли участие в формировании азиатских и европейских линий.

Также в нашей работе было проведено исследование делеционного полиморфизма гена *GSTM1* в контексте гаплотипического разнообразия геномного кластера *GSTM* в трех русских популяциях. Делеция захватывает значительную ( $\approx 20\%$ ) часть района кластера генов *GSTM*, находящегося в участке короткого плеча 1-й хромосомы (1p13.3) протяженностью около 85kb. Причины возникновения делеции в кластере *GSTM* обсуждались еще в работе Shi-jie Xu et.al. [1998], в которой было обнаружено, что кластер генов *GSTM* содержит два фактически одинаковых региона, фланкирующих с обеих сторон ген *GSTM1*. Рекомбинация между данными регионами может являться причиной возникновения делеции, которая захватывает по существу весь ген *GSTM1*. Однако не ясно, могут ли соседние геномные регионы благоприятствовать или,

напротив, препятствовать данному рекомбинационному событию. Для выяснения этого нами были проанализированы параметры частот встречаемости гаплотипов в контексте наличия или отсутствия делеции гена *GSTM1*.

Сравнительный внутривидовой анализ частот гаплотипов между двумя подгруппами индивидуумов – подгруппа *GSTM1* «0» и подгруппа *GSTM1* «+» – не дал однозначных результатов, о том влияет ли геномное окружение на наличие или отсутствие делеции: различия наблюдались лишь по отдельным гаплоглокам. Было предположено, 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 такого же рода анализ было решено провести в рамках объединенной популяции, включающей популяции Твери, Муром и Курска. Однако, несмотря на большую успешность такого варианта анализа, недооцененным оставался индивидуальный вклад в эти различия гетерозигот по делеции и обладателей полнофункциональных вариантов гена *GSTM1*.

С использованием метода ПЦР в режиме реального времени в рамках объединенной популяции нами было выделено уже три подгруппы – нормальных гомозигот, гетерозигот и гомозигот по делеции. Проведенный анализ распределений гаплотипов по всем гаплоглокам в парах сравнения выявил наличие высоко достоверных различий в частотах встречаемости гаплотипов между группой нормальных гомозигот и гомозигот по делеции.

Полученные в настоящей работе данные позволяют предположить, что рассматриваемая делеция могла произойти у далекого предка в конкретной хромосоме с определенными сочетаниями гаплотипов, которая затем значительно распространилась в популяциях по типу «эффекта основателя». Что определило столь значительное распространение этой хромосомы в популяциях пока не ясно; здесь могли играть роль факторы адаптации и селекции. Удивительна в этом контексте широкая распространённость данной делеции, т. к. результаты многих

работ показывают, что она может иметь отношение к развитию различных патологий, включая онкозаболевания.

В этой связи уместным будет также и другое объяснение: обнаруженная нами зависимость между делецией гена *GSTM1* и распределением частот гаплотипов в кластере *GSTM* может указывать на возможность предпочтительного возникновения делеции в определённых вариантах геномного окружения. С этим предположением, согласуются имеющиеся данные о роли геномного контекста в возникновении и последующем закреплении (фиксации) в популяциях точковых мутаций [Schaibley et al., 2013].

## 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 сходство 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 ЛИТЕРАТУРЫ**

- Алтухов Ю.П., Салменкова Е.А. Полиморфизм ДНК в популяционной генетике. Генетика, 2002. Т.38. № 9. С.1173-1195.
- Гречко В. Молекулярные маркеры ДНК в изучении филогении и систематики. Генетика, 2002. Т.38. № 8. С. 1013-1033.
- Лимборская С.А., Сломинский П.А, Молекулярная генетика человека: Медико-генетические и популяционные исследования. В кн.: Проблемы и перспективы молекулярной генетики. Том 1. Отв. ред. Свердлов Е.Д. М. Наука, 2003. с. 307-371.
- Лимборская С.А., Хуснутдинова Э.К., Балановская Е.В. Этногеномика и геногеография народов Восточной Европы. М. Наука, 2002. С.261.
- Лихтен М., Голдман А.С.Х. Ежегодник обзоров по генетике, 1995. № 29, с. 423 – 444
- Люблинский А.В. Результаты произведенных в Кронштадте в последние 5 лет исследований зрения у нижних чинов. Из медицинских прибавлений к Морскому сборнику. СПб. 1885. С.23.
- Попова С.Н., Соломинский П.А., Галушкин С.Н. и др., Полиморфизм глутатион-S трансфераз *GSTM1* и *GSTT1* в ряде популяций России. ГЕНЕТИКА, 2002, т.38, № 2, с.281-284.
- Хуснутдинова Э.К., Лимборская С.А., Этногеномика. В кн.: Геномика-медицине. Научное издание. / Под ред. Академика РАМН Иванова В.И. и академика РАН Киселева Л.Л. М., ИКЦ «Академкнига», 2005. с. 312-349.
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56-65.
- Adler V., Yin Z., Fuchs S., et al. Regulation of JNK signaling by *GSTp*. EMBO J., 1999. № 18. P.1321–1334.

- Allocati N., Favaloro B., Masulli M., et al. Proteusmirabilis glutathione S-transferase B1-1 is involved in protective mechanisms against oxidative and chemical stress. *Biochem*, 2003. № 373. P.305–11.
- Amicarelli F., Falone S., Cattani F., et al. Amphibian transition to the oxidant terrestrial environment affects the expression of glutathione S-transferases isoenzymatic pattern. *Biochim. Biophys. Acta*, 2004. № 1691. P.181–92.
- An J., Blackwell T. SKN-1 links *C. elegans* mesendodermal specification to a conserved oxidative stress response. *Genes Dev.*, 2003. №17. P.1882–93.
- Ardlie K., Kruglyak L., Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet.*, 2002. V.3. №4. P.299-309.
- Armstrong R. Structure, Catalytic Mechanism, and Evolution of the Glutathione Transferases. *Chem. Res. Toxicol.*, 1997. № 10. P. 2-18.
- Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, Bustamante CD. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* 2009; 19(5):795-803.
- Axarli I., Ridgen D., Labrou N. Characterization of the ligandin site of maize glutathione S-transferase I. *Biochem. J.*, 2004. № 382. P. 885–893.
- Balanovsky O., Rootsi S., Pshenichnov A. Two sources of the Russian patrilineal heritage in their Euro-Asian context, *Am. J. Hum. Genet.*, 2008. vol. 82. № 1. P.236–250.
- Barrett J., Fry B., Maller J. Haploview: analysis and visualization of LD and haplotype maps, *Applications note*, 2005. Vol. 21 №2. P.263–265.
- Bersaglieri T., Sabeti P., Patterson N. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet.*, 2004. V.74. №6. P.1111-20.
- Bhattacharjee P., Paul S., Banerjee M. Functional compensation of glutathione S-transferase *M1 (GSTM1)* null by another *GSTs* superfamily member, *GSTM2*. *Sci Rep.* 2013. №3. P.2704

- Board P. Biochemical genetics of glutathione-S-transferase in man. *American Journal of Human Genetics*, 1981. №33. P.36–43.
- Board P., Coggan M. Johnston P. Genetic heterogeneity of the human glutathione transferases: a complex of gene families. *Pharmacol Ther.*, 1990. №48. P.357–369.
- Bolt H., Thier R. Relevance of the deletion polymorphisms of the glutathione S-transferases *GSTT1* and *GSTM1* in pharmacology and toxicology. *Curr Drug Metab.*, 2006. № 7. P.613-28.
- Buchanan JA, Scherer SW. Contemplating effects of genomic structural variation. *Genet Med*. 2008; 10 (9):639-47.
- Cardon L., Abecasis G. Using haplotype blocks to map human complex trait loci. *Trends Genet*. 2003. №19. P. 135-40.
- Chakravarti A. Population genetics – making sense out of sequence. *Nature genetics supplements*, 1999. V.21. P.56-60.
- Cho S., Lee Y., Park H., et al. Glutathione S-transferase mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. *J. Biol. Chem.*, 2001. №276. P. 12749–12755.
- Clark A. The role of haplotypes in candidate gene studies. *Genet. Epidemiol.*, 2004. V.27. №4. P.321-33.
- Coggan M., Whitbread L., Whittington A. Structure and organization of the human theta-class glutathione S-transferase and D-dopachrome tautomerase gene complex. *Biochemical Journal*, 1998. №334. P.617–623.
- Coles B., Kadlubar F. Human alpha class glutathione S-transferases: genetic polymorphism, expression, and susceptibility to disease. *Methods in Enzymology*, 2005. №401. P.9–42
- Conrad DF, Hurles ME. The population genetics of structural variation. *Nat Genet*. 2007; 39 (7 Suppl):S30-6
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K,

- Valsesia A, Walter K, Wei J; Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010; 464(7289):704-12.
- Crawford D., Nickerson D. Definition and clinical importance of haplotypes. *Annu Rev Med.*, 2005. V.56. P.303-20.
  - da Fonseca R., Johnson W., O'Brien S. Molecular evolution and the role of oxidative stress in the expansion and functional diversification of cytosolic glutathione transferases. *BMC Evolutionary Biology*, 2010. №10. P.281.
  - Daly A. Pharmacogenetics of the major polymorphic metabolizing enzymes. *Fundam. Clin. Pharmacol.*, 2003 V. 17. №1. P. 27–41.
  - Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet*. 2001; 29(2):229-32
  - Dawson E., Abecasis G., Bumpstead S. A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, 2002. V.418. №6897. P.544-8.
  - Desikan R., A-H-Mackerness S., Hancock J., et al. Regulation of the Arabidopsis transcriptome by oxidative stress. *Plant Physiol.*, 2001. №127. P. 159–72.
  - Dhillon V., Shahid M., Husain S. Associations of *MTHFR DNMT3b* 4977 bp deletion in mtDNA and *GSTM1* deletion, and aberrant CpG island hypermethylation of *GSTM1* in non-obstructive infertility in Indian men. *Mol Hum Reprod*, 2007. №13. P.213-22.
  - Di Pietro G., Magno L., Rios-Santos F. Glutathione S-transferases: an overview in cancer research. *Expert Opin. Drug Metab. Toxicol.*, 2010. № 6. P. 153-70.
  - Dixon D., Cole D., Edwards R. Characterization of a Zeta class glutathione transferase from *Arabidopsis thaliana* with a putative role in tyrosine catabolism. *Arch. Biochem. Biophys.*, 2000. № 384. P.407–12.
  - Economopoulos K., Choussein S., Vlahos N. *GSTM1* polymorphism, *GSTT1* polymorphism, and cervical cancer risk: a meta-analysis. *Int J Gynecol Cancer*, 2010. №20. P.1576-80.

- Economopoulos K., Sergentanis T. *GSTM1, GSTT1, GSTP1, GSTA1* and colorectal cancer risk: a comprehensive meta-analysis. *Eur J Cancer*, 2010. №46. P.1617-31.
- Elding H., Lau W., Swallow D. Dissecting the Genetics of Complex Inheritance: Linkage Disequilibrium Mapping Provides Insight into Crohn Disease. *Am J Hum Genet.*, 2011. №89. P.798–805
- Eslahchi C., Katanforoush A., Pezeshk H. Haplotype block partitioning and tagSNP selection under the perfect phylogeny model. *Iranian Journal of biotechnology*, 2011. Vol. 9, №4.
- Evans W., McLeod H. Pharmacogenomics – drug disposition, drug targets, and side effects. *N Engl J Med.*, 2003. №348. P.538-49.
- Feng X., Dong C., Shi J. Lack of association of glutathione S-transferase *GSTM3* gene polymorphism with the susceptibility of lung cancer. *Asian Pac J Cancer Prev.*, 2012. №13. P.4465-8.
- Flagg E., Coates R., Eley J., et al. Dietary glutathione intake in humans and the relationship between intake and plasma total glutathione level. *Nutr. Cancer*, 1994. № 21. P.33–46.
- Ford J., Li Y., O'Sullivan M. Glutathione S-transferase M1 polymorphism and lung cancer risk in African-Americans. *Carcinogenesis*, 2000. №21. P.1971-5.
- Frova C. Glutathione transferases in the genomics era: New insights and perspectives. *Biomolecular Engineering*, 2006. № 23. p. 149–169.
- Gabriel SB1, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science*, 2002. 296(5576):2225-2229.
- Garte S., Gaspari L., Alexandrie A. Metabolic Gene Polymorphism Frequencies in Control Populations, *Cancer. Epidemiol. Biomarkers. Prev.*, 2001, vol.10, №12, P.1239–1248.

- Gibson J., Tapper W., Ennis S. Exome-based linkage disequilibrium maps of individual genes: functional clustering and relationship to disease. *Human Genetics*, 2013. V.132. I.2, P.233-243.
- Gibson J., Tapper W., Zhang W. Hum Genomics. Cosmopolitan linkage disequilibrium maps. *Hum Genomics*, 2005. №2. P.20-7.
- Ginsberg G., Molenski S., Hattis D., et al. Genetic polymorphism in glutathione transferases (GST): population distribution of *GSTM1*, *GSTT1*, and *GSTP1* conjugating activity. *Journal of Toxicology and Environmental Health*, 2009. № 12. P.389–439.
- Girault I., Lidereau R., Bieche I. Trimodal *GSTT1* and *GSTM1* genotyping assay by real-time PCR. *International Journal of Biological Markers*, 2005. №20. P.81–86.
- Glantz S. *Primer of BIOSTATISTICS*. 4 Edition. M. Praktica, 1999. P.105.
- Gray I., Campbell D., Spurr N. Single nucleotide polymorphisms as tools in human genetics. *Human molecular genetics*, 2000. V.9. №.16. P.2403-2408.
- Grundy J., Storey K. Antioxidant defences and lipid peroxidation damage in estivating toads, *Scaphiopus couchii*. *J. Comp. Physiol. B*, 1998. № 168. P.132–42.
- Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics*. 2008; 1(1):4.
- Guengerich FP. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol*. 2001; 14(6):611-50.
- Guy C., Hoogendoorn B., Smith S. Promoter polymorphisms in glutathione-S-transferase genes affects transcription. *Pharmacogenetics*, 2004. №14. P.45-51.
- Habdous M., Siest G., Herbeth B. Glutathione S-transferases genetic polymorphisms and human diseases: overview of epidemiological studies. *Ann Biol Clin.*, 2004. №62. P.15-24.
- Hayes J., Flanagan J., Jowsey I. Glutathione Transferases. *Annu. Rev. Pharmacol. Toxicol*. 2005. p. 88.

- Hayes J., Judah D., McLellan L. et al. Ethoxyquin-induced resistance to aflatoxin B1 in the rat is associated with the expression of a novel alpha-class glutathione S-transferase subunit, Yc2, which possesses high catalytic activity for aflatoxin B1-8,9-epoxide. *Biochem. J.*, 1991. № 279. P.385-398.
- Hayes J., McLellan L. Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. *Free Radic. Res.*, 1999. № 31. P.273–300.
- Hayes J., Strange R. Potential contribution of the glutathione S-transferase supergene family to resistance to oxidative stress. *Free Radical Research.*, 1995. №22. P.193–207.
- Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, Myers S. A genetic atlas of human admixture history. *Science*. 2014; 343(6172):747-51.
- Hinds D., Stuve L., Nilsen G. Whole-genome patterns of common DNA variation in three human populations. *Science*, 2005. V.307. №5712. P.1072-9.
- Iantomasi T., Favilli F., Marraccini P., et al. Glutathione transport system in human small intestine epithelial cells. *Biochim. Biophys.*, 1997. № 1330. P.274–83.
- Ilio D., Aceto C., Bucciarelli A., et al. Glutathione transferase isoenzymes from human prostate. *Biochem. J.*, 1990. №271. P. 481-485.
- International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449 (7164):851-61.
- Ishiia T., Matsusea T., Teramotoa S., et al. Glutathione S-transferase *P1 (GSTP1)* polymorphism in patients with chronic obstructive pulmonary disease. *Thorax*, 1999. № 54. P.693-696.
- Iyanagi T. Molecular mechanism of phase I and phase II drug-metabolizing enzymes: implications for detoxification. *Int Rev Cytol*. 2007; 260:35-112.
- Jancovaa P., Anzenbacherb P., Anzenbacherovaa E. Phase II drug metabolizing enzymes. *Fac Univ Palacky Olomouc Czech Repub*, 2010. № 154. 6. p. 103–11.

- Jefferies H., Bot J., Coster J., et al. The role of glutathione in intestinal dysfunction. *J Invest Surg.*, 2003. №16. P.315-23.
- Jefferies H., Coster J., Khalil A. Glutathione. *ANZ Journal of Surgery*, 2003 № 73(7). p.517-22.
- Jorde L., Bamshad M., Rogers A. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays*, 1998. V.20. №2. P.126-136.
- Josephy D. Genetic Variations in Human Glutathione Transferase Enzymes: Significance Human Genomics Proteomics. *Pharmacology and Toxicology*, 2010. DOI: 10.4061/2010/876940.
- Katju V, Bergthorsson U. Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front Genet.* 2013; 4:273.
- Ke X., Hunt S., Tapper W. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet.*, 2004. V.13. №6. P.577-88.
- Keen J., Jakoby W. Glutathione transferases. Catalysis of nucleophilic reactions of glutathione. *J. Biol. Chem.*, 1978. № 253. P.5654–57.
- Kenneth D. Manevich Y., Grek C., et al. The role of glutathione S-transferase P in signaling pathways and S-glutathionylation in cancer. *Free Radical Biology & Medicine*, 2011.
- Khojasteh-Bakht S., Nelson S., Atkins W. Glutathione S-transferase catalyzes the isomerization of (R)-2-hydroxymenthofuran to mintlactones. *Arch. Biochem. Biophys.*, 1999. № 370. P.59–65.
- Khoury M., Little J. Human genome epidemiologic reviews: the beginning of something HuGE. *American Journal of Epidemiology*, 2000. №151. P.2–3
- Khrunin A., Khokhrin D., Limborska S. Glutathione-S-Transferase Gene Polymorphism in Russian Populations of European Part of Russia, *Russian Journal of Genetics*, 2008. Vol. 44, №10, P.1241–1245.
- Khrunin A., Mihailov E., Nikopensius T. Analysis of Allele and Haplotype Diversity Across 25 Genomic Regions in Three Eastern European Populations., *Human Heredity*, 2009, №68, P.35–44.

- Khrunin, A., Bebyakova, N., Ivanov, V., et al., Polymorphism of Y-Chromosomal Microsatellites in Russian Populations from the Northern and Southern Russia as Exemplified by the Populations of Kursk and Arkhangelsk Oblast, *Russ. J. Genet.*, 2005, vol. 41, № 8, P.922–927.
- Kidd K., Kidd J. A nuclear perspective on human evolution. In: AJ Boyce, CGN Mascie-Taylor (eds.) *Molecular biology and human diversity*. Cambridge University Press, Cambridge, 1996. P.242–264.
- Kimura-Kataoka K., Ueki M., Takeshita H. Identification of the functional alleles of the nonsynonymous single-nucleotide polymorphisms potentially implicated in systemic lupus erythematosus in the human deoxyribonuclease I gene. *DNA Cell Biol.*, 2014. №33. P. 492-502.
- Kobayashi M., Itoh K., Suzuki T., et al. Identification of the interactive interface and hylogenic conservation of the Nrf2-Keap1 system. *Genes Cells*, 2002. № 7. P.807–20.
- Laborde E. Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death. Ed. by Finazzi Agró. *Cell Death and Differentiation*, 2010. № 17. P. 1373–1380.
- Lan T., Yang Z., Yang X., et al. Extensive functional diversification of the populus glutathione S-transferase supergene family. *Plant Cell*, 2009. № 12. p. 3749–3766.
- Josephy P., Mannervik B. Screening and characterization of variant Theta-class glutathione transferases catalyzing the activation of ethylene dibromide to a mutagen. *Molecular Toxicology*, № 47(9). P. 657-65.
- Leiers B., Kampkotter A., Grevelding C., et al. A stress-responsive glutathione S-transferase confers resistance to oxidative stress in *Caenorhabditis elegans*. *Free Radic. Biol. Med.*, 2003. №34. P.1405– 1527
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008; 319(5866):1100-4.

- Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human *ADME* genes. *Hum Mol Genet.* 2011; 20(3):528-40.
- Lindblom T., Dodd A. Xenobiotic detoxification in the nematode *Caenorhabditis elegans*. *Journal of Experimental Zoology*, 2006. № 9. p. 720–730.
- Liu K., Muse S. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 2005. №21. P.2128-9.
- Lo H., Ali-Osman F. Genomic cloning of *hGSTP1C* and allelic human pi class glutathione S- transferase gene variant and functional characterization of its retinoic acid response elements. *J Biol Chem*, 1997. № 272. P.32743–9.
- Lohmueller K., Albrechtsen A., Li Y. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet.*, 2011. №7. e1002326. doi: 10.1371/journal.pgen. 1002326.
- Luca D, Ringquist S, Klei L, Lee AB, Gieger C, et al. (2008) On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am J Hum Genet.* 2008; 82(2):453–63.
- Lumjuan N., Stevenson B., Prapanthadara L., et al. The *Aedes aegypti* glutathione transferase family. *Insect Biochemistry and Molecular Biology*, 2007. № 10. p. 1026–1035.
- Major J., Yu K., Weinstein S. Genetic variants reflecting higher vitamin e status in men are associated with reduced risk of prostate cancer. *J Nutr.*, 2014. №144. P.729-33.
- Manichaikul A., Rich S., Perry H. A functionally significant polymorphism in *ID3* is associated with human coronary pathology. *PLoS One*, 2014. №9. doi: 10.1371/journal.pone.0090222.
- Mannervik B. The isoenzymes of glutathione transferase. *Advan. Enzymol. Relat. Areas Mol. Biol.*, 1985. № 57. P. 357-417.
- Mannervik B., Awasthit Y., Boardt Ph., et al. Nomenclature for human glutathione Transferases. *Biochem. J.*, 1992. № 282. P. 305-308.

- Mathew C. The isolation of high molecular weight eukaryotic DNA. In: *Methods in Molecular Biology* / Humana Press (ed. Walker JM), 1984. V.2. P.31 - 34.
- McEvoy BP, Powell JE, Goddard ME, Visscher PM. Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res.* 2011;21(6):821-9.
- McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM; International HapMap Consortium. Common deletion polymorphisms in the human genome. *Nat Genet.* 2006;38(1):86-92.
- McGonigle B., Keeler S., Lau, S.-M., et al. A genomics approach to the comprehensive analysis of the glutathione-S transferase gene family in soybean and maize. *Plant Physiol.*, 2000. №124. P.1105–1120.
- McIlwain C., Townsend D., Tew K. Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene*, 2006. V. 25. №11. P. 1639–1648.
- Minelli C., Granell R., Newson R. Glutathione-S-transferase genes and asthma phenotypes. A Human Genome Epidemiology (HuGE) systematic review and meta-analysis including unpublished data. *International Journal of Epidemiology*, 2010. №39. P.539–562.
- Mitchell A., Morin D., Lame M. Purification, mass spectrometric characterization, and covalent modification of murine glutathione S-transferases. *Chemical Research in Toxicology*, 199. № 8. p. 1054–1062
- Mannervik B., Board P. G., Hayes J., et al. Nomenclature for mammalian soluble glutathione transferases. *Methods in Enzymology*, 2005. № 401. p. 1–8.
- Morel F., Rauch C., Coles B. The human glutathione transferase alpha locus: genomic organization of the gene cluster and functional characterization of the genetic polymorphism in the *hGSTA1* promoter. *Pharmacogenetics*, 2002. №12. P.277–286.

- Moyer A., Salavaggione O., Hebring S. Glutathione S-transferase *GSTT1* and *GSTM1*: gene sequence variation and functional genomics. *Clin Cancer Res.*, 2007. №13. P.7207-16.
- Mueller J., Lohmussaar E., Magi R. Linkage disequilibrium patterns and tag SNP transferability among European populations. *Am J Hum Genet.*, 2005. V.76. №3. P.387-98.
- Mueller L., Godman C., Silady R. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid binding protein. *Plant Physiol.*, 2000. № 123. P. 1561–1570.
- Nebert D., Vasiliou V. Analysis of the glutathione S-transferase (*GSTs*) gene family. *Hum Genomics*, 2004. № 1(6). P.460-4.
- Nelis M., Esko T., Magi R. Genetic Structure of Europeans: A View from the North–East, *PLoS ONE*, 2009.e5472.doi:10.1371/journal.pone.0005472.
- Nothnagel M., Rohde K. The Effect of Single-Nucleotide Polymorphism Marker Selection on Patterns of Haplotype Blocks and Haplotype Frequency Estimates. *Am. J. Hum. Genet.*, 2005. №77. P.988–998.
- Oakley A. Glutathione transferases: a structural perspective. *Drug Metabolism Reviews*, 2011. № 43(2). P.138–151.
- Ohashi J., Naka I., Patarapotikul J. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am J Hum Genet.*, 2004. V.74. №6. P.1198-208.
- Pajaud J., Kumar S., Rauch C. Regulation of Signal Transduction by Glutathione Transferases. *International Journal of Hepatology*, 2012. doi:10.1155/2012/137676.
- Pandya U., Srivastava S., Singhal S., et al. Activity of allelic variants of Pi class human glutathione S-transferase toward chlorambucil. *Biochem Biophys Res Commun.*, 2000. № 278. P. 258-62.
- Parl F. A need for true *GSTM1* and *GSTT1* genotyping. *Cancer Epidemiology Biomarkers and Prevention*, 2009, №18. P.2793.

- Pearson W., Vorachek W., Xu S. Identification of class-mu glutathione transferase genes *GSTM1-GSTM5* on human chromosome 1p13. *Am J Hum Genet.*, 1993. №53. P.220-33.
- Pemble S., Schroeder K., Spencer S. Human glutathione S-transferase theta (*GSTT1*): cDNA cloning and the characterization of a genetic polymorphism. *Biochem.*, 1994. № 300. P.271–276.
- Phillips M., Lawrence R., Sachidanandam R. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet.*, 2003. V.33. №3. P.382-7.
- Piacentini S., Polimanti R., Porreca F. *GSTT1* and *GSTM1* gene polymorphisms in European and African populations. *Mol Biol Rep.*, 2011 №38. P.1225-30.
- Polimanti R., Piacentini S., Fuciarelli M. HapMap-based study of human soluble glutathione S-transferase enzymes: the role of natural selection in shaping the single nucleotide polymorphism diversity of xenobiotic-metabolizing genes. *Pharmacogenetics and Genomics*, 2011, Vol. 21 №10. DOI: 10.1097/FPC.0b013e328349da4d.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38(8):904-9.
- R x C: a Windows (TM) program for the analysis of contingency tables via the Metropolis algorithm <http://www.marksgeneticsoftware.net/rxc.htm>.
- Ramsay E., Dilda P. Glutathione S-conjugates as prodrugs to target drug-resistant tumors. *Front Pharmacol.*, 2014. №11. doi: 10.3389/fphar.2014.00181.
- Ranson H., Collin F., Hemingway J. The role of alternative mRNA splicing in generating heterogeneity within the *Anopheles gambiae* class I glutathione S-transferase family. *Proc. Natl. Acad. Sci.*, 1998. № 95. P. 14284–14289.
- Rebbeck T. Molecular Epidemiology of the Human Glutathione S-Transferase Genotypes *GSTM1* and *GSTT1* in Cancer susceptibility. *Cancer Epidemiology, Biomarkers & Prevention*, 1997. Vol.6, P.33-743.

- Resler A., Makar K., Heath L. Genetic variation in prostaglandin synthesis and related pathways, NSAID use and colorectal cancer risk in the Colon Cancer Family Registry. *Carcinogenesis*, 2014. №35. P.2121-6.
- Roodi N., Dupont W., Moore J. Association of homozygous wild-type glutathione S-transferase *GSTM1* genotype with increased breast cancer risk. *Cancer Research*, 2004. №64. P.1233–1236.
- Sabeti P., Reich D., Higgins J. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 2002. V.419. 6909. P.832-7.
- Sabeti PC, Varilly P, Fry B, Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007; 449(7164):913-8.
- Sata F., Yamada H., Kondo T. Glutathione S-transferase *GSTM1* and *GSTT1* polymorphisms and the risk of recurrent pregnancy loss. *Mol Hum Reprod*, 2003. №9. P.165-9.
- Schaibley V., Zawistowski M., Wegmann D. The influence of genomic context on mutation patterns in the human genome inferred from rare variants. *Genome Research*, 2013. №23. P.1974-1984.
- Semino O., Passarino G., Oefner P. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*, 2000, Vol.290, N5494, P.1155-1159.
- Sheehan D., Meade G., Foley V. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.*, 2001. № 360. P.1–16.
- Sheridan J., Mack D., Amre D. A non-synonymous coding variant (I616F) in the *tlr5* gene is potentially associated with Crohn's disease and influences responses to bacterial flagellin. *PLoS One.*, 2013. №8. e61326. doi: 10.1371/journal.pone.0061326.
- Shi X., Zhou S, Wang Z. *CYP1A1* and *GSTM1* polymorphisms and lung cancer risk in Chinese populations: a meta-analysis. *Lung Cancer*, 2008. №59. P.155-63.

- Shifman S., Kuypers J., Kokoris M. Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet.*, 2003. V.12. 7. P.771-6.
- Sies H., Packer L. *Glutathione Transferases and Gamma-Glutamyl Transpeptidases*. Academic Press, 2005.
- Singh S., Zimniak L., Zimniak P. The human *hGSTA5* gene encodes an enzymatically active protein. *Biochimica et biophysica acta*, 2010. № 1. p. 16–22.
- Smith A., DeRidder B., Guo W.-J., et al. Proteomic analysis of Arabidopsis glutathione S-transferases from benoxacor- and copper-treated seedlings. *J. Biol. Chem.* 2004. № 279. P.26098–26104.
- Smolenski S., Hattis D., Kathryn Z. Genetic Polymorphism in Glutathione Transferases (GST): Population Distribution of *GSTM1*, *GSTT1*, and *GSTP1* Conjugating Activity. *J Toxicol Environ Health B Crit Rev*, 2009. №12. P.389-439.
- Soranzo N., Gorla S., Mizzi L., et al. Organisation and structural evolution of the rice glutathione S-transferase gene family. *Mol. Gen. Genomics*, 2004. №271. P. 511–521.
- Steffens M., Lamina C., Illig T. SNP-Based Analysis of Genetic Substructure in the German Population. *Hum Hered*, 2006. №62. P.20–29.
- Tamer L., Calikoğlu M., Ates N. Glutathione-S-transferase gene polymorphisms (*GSTT1*, *GSTM1*, *GSTP1*) as increased risk factors for asthma. *Respirology*, 2004. №9. P.493-8.
- Tan K-L., Chelvanayagam G., Parker M. Mutagenesis of the active site of the human Theta-class glutathione transferase *GSTT2-2*: catalysis with different substrates involves different residues. *Biochemical Journal*, 1996. №319. P.315–321.
- Tan X., Wang Y., Han Y. Genetic variation in the *GSTM3* promoter confer risk and prognosis of renal cell carcinoma by reducing gene expression. *Br J Cancer*, 2013. №109. P.3105-15.

- Tasa G., Juronen E., Viikmaa M. Distribution of Glutathione S-Transferase *GSTT1* Phenotypes in the Estonian Population, *Gene. Geogr.*, 1996, vol.10, №3, P.181–189.
- Tatewaki N., Maekawa K., Katori N. Genetic Variations and Haplotype Structures of the Glutathione S-transferase Genes, *GSTT1* and *GSTM1*, in a Japanese Patient Population. *Drug Metab. Pharmacokinet.*, 2009. № 24. P. 118–126.
- The International HapMap Consortium, 2003. <http://hapmap.ncbi.nlm.nih.gov>
- Timofeeva M., Jäger B., Rosenberger A. A multiplex real-time PCR method for detection of *GSTM1* and *GSTT1* copy numbers. *Clinical Biochemistry*, 2009. №42. P.500–509.
- Tsuchida S., Sato K. Glutathione transferases and cancer. *Crit Rev Biochem Mol Biol.*, 1992. №27. P.337-84.
- van Ommen B., Bogaards J., Peters W. Quantification of human hepatic glutathione s-transferases. *Biochem. J.*, 1990. № 269. P. 609-613.
- Veal E., Toone W., Jones N., et al. Distinct roles for glutathione S-transferases in the oxidative stress response in *Schizosaccharomyces pombe*. *J. Biol. Chem.*, 2002. № 277. P.5523–31.
- Vogel F. ABO blood groups and diseases. *Amer. J. Hum. Genet.*, 1970. V.22. P. 464.
- Wagner U., Edwards R., Dixon D. Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family. *Plant Mol. Biol.*, 2002. № 49. P.515–532.
- Wall J., Pritchard J. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet.*, 2003. V.4. №8. P.587-97.
- Wang M., Bohmann D., Jasper H. JNK signaling confers tolerance to oxidative stress and extends lifespan in *Drosophila*. *Develop. Cell*, 2003. № 5. P.811–16.
- Wang N., Akey J., Zhang K. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am J Hum Genet.* 2002. V.71. №5. P.1227-34.

- Ward R., Frazier B., Dew K. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Nat. Acad. Sci. US*, 1991, Vol. 88, p.8720-8724.
- Watson M., Stewart R., Smith G. Human glutathione S-transferase *GSTP1* polymorphisms: relationship to lung tissue enzyme activity and population frequency distribution. *Carcinogenesis*, 1998. №19. P.275-80.
- Weir BS, Cockerham CC: Estimating F-statistics for the analysis of population structure. *Evolution* 1984; 38: 1358–1370.
- Whitbread A., Masoumi A., Tetlow N. Characterization of the omega class of glutathione transferases. *Methods in Enzymology*, 2005. №401. P. 78–99.
- Whitbread A., Tetlow N., Eyre H. Characterization of the human Omega class glutathione transferase genes and associated polymorphisms. *Pharmacogenetics*, 2003. №13. P.131–144.
- William H., Michael J., Jakoby W. Glutathione S-Transferases: the first enzymatic step in mercapturic acid formation. *J. Biol. Chem*, 1974. № 249. P.7130-7139.
- Xu S., Wang Y., Roe B. Characterization of the Human Class Mu Glutathione S-Transferase Gene Cluster and the *GSTM1* Deletion *J Biol Chem.*, 1998. №273. P.3517-27.
- Zhang X., Huang M., Wu X. *GSTM1* copy number and promoter haplotype as predictors for risk of recurrence and/or second primary tumor in patients with head and neck cancer. *Pharmgenomics Pers Med.*, 2013. №6. P.9-17.
- Zhang X., Lin J., Wu X. Association between *GSTM1* copy number, promoter variants and susceptibility to urinary bladder cancer. *Int J Mol Epidemiol. Genet.*, 2012. №3. P.228-236.
- Zhao Y., Deng X., Song G. The *GSTM1* null genotype increased risk of gastric cancer: a meta-analysis based on 46 studies. *PLoS One.*, 2013. №8. e81403. doi: 10.1371/journal.pone.0081403.
- Zhao Y., Marotta M., Eichler E. Linkage disequilibrium between two high-frequency deletion polymorphisms: implications for association studies involving

the glutathione-S transferase (*GSTs*) genes. PLoS Genetics. 2009. №5 (5, article e1000472).

- Zimniak P., Nanduri B., Piękała S., Naturally occurring human glutathione S-transferase *GSTP1-1* isoforms with isoleucine and valine in position 104 differ in enzymic properties. Eur J. Biochem., 1994. № 224. P. 893-9.