

*На правах рукописи*

Шарапов Содбо Жамбалович

**ПОЛНОГЕНОМНОЕ ИССЛЕДОВАНИЕ  
АССОЦИАЦИЙ УРОВНЕЙ N-ГЛИКОЗИЛИРОВАНИЯ  
БЕЛКОВ ПЛАЗМЫ КРОВИ ЧЕЛОВЕКА**

1.5.7. - Генетика

**АВТОРЕФЕРАТ**

Диссертации на соискание учёной степени  
кандидата биологических наук

Новосибирск  
2022

Работа выполнена в лаборатории рекомбинационного и сегрегационного анализа ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

**Научный руководитель:** **Аульченко Юрий Сергеевич** д.б.н., г.н.с. лаборатории рекомбинационного и сегрегационного анализа ФГБНУ «ФИЦ Институт цитологии и генетики СО РАН», г. Новосибирск

**Официальные оппоненты:** **Лебедев Игорь Николаевич** д.б.н., профессор РАН, заместитель директора по научной работе, руководитель лаборатории НИИ медицинской генетики Федерального государственного бюджетного научного учреждения "Томский национальный исследовательский медицинский центр Российской академии наук", г. Томск.

**Кулемин Николай Александрович** к.б.н., научный сотрудник лаборатории молекулярной генетики человека, «Федеральный научно-клинический центр физико-химической медицины Федерального Медико-биологического Агентства», г. Москва

**Ведущее учреждение:** ФГБУН «Институт общей генетики им. Н.И. Вавилова РАН», г. Москва

Защита диссертации состоится «\_\_»\_\_\_\_\_ 2022г. на утреннем заседании диссертационного совета 24.1.239.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН», в конференц-зале Института по адресу: пр. академика Лаврентьева 10, г. Новосибирск, 630090, тел.: (383) 363-49-06 (1321); факс: (383) 333-12-78; e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте института [www.bionet.nsc.ru](http://www.bionet.nsc.ru).

Автореферат разослан «\_\_»\_\_\_\_\_ 2022г.

Учёный секретарь  
диссертационного совета,  
доктор биологических наук

Хлебодарова Т.М.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** Гликозилирование - присоединение углеводного остатка (гликана) - является одной из самых распространенных пост- и ко-трансляционных изменений белков. Гликозилирование влияет на биохимические свойства белков и на их биологические функции. Прогресс в области методов анализа гликозилирования белков позволил к началу 2010-х годов проводить исследования популяционной изменчивости N-гликома белков плазмы крови - набора N-гликанов, связанных с белками плазмы крови, и их концентраций - и ее ассоциации с распространенными заболеваниями человека. Эти исследования показали, что уровни различных N-гликанов белков плазмы крови ассоциированы с рядом мультифакторных заболеваний человека.

С биохимической точки зрения, процесс биосинтеза N-гликанов хорошо изучен, однако понимание механизмов общей и тканеспецифической регуляции этих биохимических реакций *in vivo* ограничено. Методы генетики позволяют изучать сложные биологические системы *in vivo*. В частности, определение генов и регуляторных областей генома, вариация в которых приводит к изменениям в N-гликоме, может пролить свет на механизмы регуляции N-гликозилирования. Самым широко используемым методом картирования локусов мультифакторных признаков и заболеваний человека является полногеномное исследование ассоциаций (ПГИА). К настоящему времени было проведено два ПГИА уровней N-гликанов белков плазмы крови человека (Huffman и др., 2011; Lauc и др., 2010), в которых измерение уровней N-гликанов плазмы крови производилось с помощью технологии высокопроизводительной жидкостной хроматографии (ВЭЖХ). В результате была найдена ассоциация шести локусов генома, в том числе локуса, содержащего ген фактора транскрипции гепатоцитов *HNF1A*, роль которого в контроле N-гликозилирования белков ранее не была известна. В последующем функциональном исследовании (Lauc и др., 2010) было показано, что продукт гена *HNF1A* регулирует экспрессию генов фукозилтрансфераз (*FUT3*, *FUT5*, *FUT6*, *FUT8*, *FUT10*, *FUT11*) в клетках линии HepG2, полученной из гепатоцитов. Таким образом, исследование генетического контроля гликозилирования методом ПГИА позволяет формировать гипотезы о регуляции данного процесса.

В течении семи лет, прошедших после публикации последнего ПГИА N-гликома белков плазмы крови в 2011 году, появились новые технологии измерения гликанов. Среди современных технологий анализа уровней гликозилирования белков плазмы крови человека, наибольшее распространение получил метод сверхвысокоэффективной жидкостной хроматографии (СВЭЖХ). Число гликанов, измеряемых методом СВЭЖХ, и точность их измерения выше, чем у метода ВЭЖХ. Генетический анализ расширенного спектра гликанов, измеренных с большей точностью, позволит увеличить мощность анализа генетических ассоциаций, и, как следствие, более точно охарактеризовать регуляцию процессов гликозилирования найденными локусами.

Однако проведение ПГИА гликома плазмы крови, измеренного технологией СВЭЖХ, ранее было невозможным из-за того, что в зависимости от исследования, число измеряемых признаков варьировало от 36 до 42. Данные различия обусловлены изменениями в протоколах анализа СВЭЖХ, как на этапах проведения

хроматографического анализа, так и на этапе интеграции - определения границ хроматографических пиков. Проведение полногеномного исследования ассоциаций на материале нескольких выборок возможно только при условии того, что в каждой из выборок анализ ассоциаций проводится для единого (гармонизированного) набора признаков. Таким образом, для проведения ПГИА гликома плазмы крови человека с последующим подтверждением результатов на независимых выборках, требуется разработка и применение метода гармонизации гликомных профилей СВЭЖХ, полученных в анализируемых выборках.

Принимая во внимание ограниченное число работ, посвященных исследованию генетического контроля гликозилирования, в том числе из-за недостаточного методологического обеспечения, исследование генетического контроля гликозилирования белков плазмы крови человека с использованием современных методов измерения гликома является актуальной проблемой современной генетики человека.

**Цель исследования** - поиск генов, участвующих в контроле N-гликозилирования белков плазмы крови человека.

**Задачи исследования:**

1. Провести полногеномное исследование ассоциаций генетических маркеров с уровнями N-гликанов белков плазмы крови человека на материале выборки TwinsUK.
2. Разработать и валидировать метод гармонизации данных об уровнях N-гликанов белков плазмы крови человека, необходимый для сопоставления результатов, полученных в различных выборках, и применить его для гармонизации данных независимых выборок.
3. Подтвердить результаты анализа ассоциаций на материале независимых выборок.
4. Приоритизировать в найденных локусах гены-кандидаты, наиболее вероятно влияющие на уровни N-гликанов белков плазмы крови человека.

**Научная новизна работы.** В данной работе впервые проведено полногеномное исследование ассоциации уровней N-гликанов белков плазмы крови человека, измеренных технологией сверхвысокоэффективной жидкостной хроматографии (СВЭЖХ). Для проведения исследования на материале нескольких выборок впервые был разработан метод гармонизации гликомных профилей, измеренных технологией СВЭЖХ. Впервые была показана и подтверждена ассоциация 9 локусов генома с уровнями N-гликанов белков плазмы крови человека. Общее число локусов с показанной и подтвержденной ассоциацией было увеличено с 6 до 15. Для 15 локусов были предложены гены – кандидаты, наиболее вероятно участвующие в регуляции процессов гликозилирования. Впервые показана возможная роль генов *RUNX3*, *IKZF1*, *SMARCB1*, *DERL3*, *CHCHD10*, *IGH*, *TMEM121* и *HLA* в генетическом контроле уровней N-гликозилирования белков плазмы крови человека.

**Теоретическая и научно-практическая ценность.** Результаты данной работы расширяют представление о генетической регуляции N-гликозилирования белков плазмы крови человека. Полногеномное исследование ассоциации уровней гликозилирования белков плазмы крови человека, измеренных технологией СВЭЖХ, позволило установить роль пятнадцати локусов в регуляции процессов гликозилирования. Биоинформатический анализ найденных локусов расширил наше

представление о возможных генах и процессах, участвующих в регуляции N-гликозилирования белков. Роль одного из предложенных генов-кандидатов – транскрипционного фактора *IKZF1* в регуляции фукозилирования белков – была доказана в результате *in vitro* эксперимента, что подкрепляет научную состоятельность сформулированных гипотез о генах-кандидатах.

Получение и публикация в открытом доступе результатов анализа генетических ассоциаций для 113 N-гликомных признаков и более 8 миллионов ОНП (однонуклеотидных полиморфизмов) позволит всем заинтересованным исследователям использовать их для совместного анализа с интересующими их данными функциональной геномики и количественной генетики. В частности, полученные результаты ПГИА N-гликома плазмы крови будут востребованы в исследованиях по поиску биомаркеров и терапевтических мишеней гликом-ассоциированных заболеваний.

**Методология и методы диссертационного исследования.** Материалом для данной работы являлись результаты измерения уровней N-гликанов белков плазмы крови человека и генотипы испытуемых пяти выборок. Контроль качества данных СВЭЖХ образцов исследуемых выборок проводился согласно принятым в данной области стандартам и включал в себя этапы гармонизации, логарифмирования, нормализации, коррекции бэтч-эффектов, определения образцов – статистических выбросов и расчета производных признаков. Контроль качества генотипов испытуемых проводился согласно принятым и опубликованным стандартам (Anderson и др., 2010). Картирование геномных локусов, вариация в которых ассоциирована с изменением уровней N-гликанов белков плазмы крови человека проводилась с помощью метода полногеномного исследования ассоциации. Биоинформатический анализ полученных результатов проводился с использованием данных функциональной геномики – GTEx (GTEx Consortium и др., 2017), CEDAR (Momozawa и др., 2018), и с использованием методов приоритизации функциональных вариантов – VEP (McLaren и др., 2016), FATHMM-XF (Rogers и др., 2017), FATHMM-InDel (Ferlandino и др., 2017), и генов - SMR/HEIDI (Zhu и др., 2016) и DEPICT (Pers и др., 2015).

**Степень достоверности результатов.** Степень достоверности результатов подтверждается согласованностью результатов полногеномного анализа ассоциации, выполненного в данной работе, с опубликованными ранее результатами. Полученные результаты были подтверждены на материале нескольких независимых выборок, набранных в различных популяциях людей, что говорит о высокой степени достоверности полученных результатов и их обобщающей способности (генерализуемости).

#### **Положения, выносимые на защиту:**

1. Популяционная изменчивость уровней N-гликанов, связанных с белками плазмы крови человека, контролируется как минимум 15 локусами генома, 9 из которых определены впервые.
2. Генами-кандидатами, вовлеченными в процесс N-гликозилирования, являются гены регуляторов транскрипции (*IKZF1*, *SMARCB1* и *RUNX3*), деградации гликопротеинов (*DERL3*), тяжелой цепи иммуноглобулинов (*IGH*) и гены с неизвестной функцией (*TMEM121* и *CHCHD10*).

**Структура и объём работы.** Работа состоит из введения, обзора литературы, материалов и методов, результатов, обсуждения, заключения, выводов, списка литературы (167 источников) и приложений. Объём работы составляет 153 страницы. Работа включает 10 таблиц, 17 рисунков и 4 таблицы в приложении.

**Личный вклад автора.** Цели и задачи были сформулированы автором совместно с научным руководителем. Основные результаты, изложенные в диссертации, получены и проинтерпретированы автором лично. Материалы для исследования - первичные данные измерения уровней гликозилирования белков плазмы крови человека и генотипы испытуемых - были предоставлены коллегами в рамках научного сотрудничества.

**Апробация результатов.** Результаты работы были представлены лично автором на 9 международных научных конференциях в виде 7 устных и 2 стендовых докладов: «European Human Genetics Conference 2021 (ESHG 2021)», Вена, Австрия, 2021; «10th Moscow Conference on Computational Molecular Biology», Москва, Россия, 2021; International Conference on Quantitative Genetics (ICQG6), Брисбен, Австралия, 2020; «VI съезд биохимиков России», Дагомыс, Россия, 2019; BGRS/SB-2018,2020, Новосибирск, Россия, 2018,2020; «ISABS Conference on Forensic and Anthropologic Genetics», Сплит, Хорватия, 2019; «2nd GlycoCom», Дубровник, Хорватия, 2018.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

### Глава 1. Обзор литературы

В разделе 1.1 дано краткое описание области и предмета изучения гликомики – раздела гликобиологии, описана роль гликозилирования в функционировании белков и важность изучения гликома для решения задач диагностики, предсказания рисков, профилактики и лечения заболеваний человека. В разделе 1.2 описано строение и разнообразие гликанов (см. Рис. 1). В разделе 1.3 описаны ключевые этапы биосинтеза N-гликанов. Методы анализа N-гликома представлены в разделе 1.4. В разделе 1.5 приведены результаты изучения наследуемости уровней N-гликанов белков плазмы крови. В разделе 1.6 дано описание метода полногеномного исследования ассоциации (ПГИА), приведена общая схема проведения ПГИА на примере количественного признака. В разделе 1.7 описаны результаты ранее опубликованных полногеномных исследований ассоциаций уровней N-гликанов белков плазмы крови человека. В разделе 1.8 приведено краткое заключение, обосновывающее цели и задачи диссертации.

### Глава 2. Материалы и методы

Данная работа была выполнена на основе материала пяти выборок – TwinsUK, EPIC-Potsdam, PainOmics, SOCCS и SABRE. Приведена схема исследования. Демографические, геномные данные и данные измерений уровней N-гликанов белков плазмы крови были предоставлены нашими коллегами в рамках совместного проекта по исследованию генетического контроля гликозилирования белков плазмы крови человека. Описание исследуемых выборок, методы генотипирования и измерения уровней N-гликанов методом СВЭЖХ описаны в разделе 2.2.

В разделе 2.3.1 описано проведение контроля качества данных об уровнях N-гликанов белков плазмы крови образцов исследуемых выборок. Методики

проведения ПГИА уровней N-гликанов белков плазмы крови человека на выборке TwinsUK и определения локусов, достоверно ассоциированных с изучаемыми признаками, приведены в разделах 2.3.2 и 2.3.3. В разделах 2.3.4 и 2.3.5 описаны методы подтверждения результатов анализа генетических ассоциаций на независимых выборках. В разделах 2.3.6 – 2.3.9 описаны количественно-генетические и биоинформатические методы приоритизации каузативных генов (расположенных в найденных локусах), вероятнее всего обуславливающих найденные ассоциации. В разделе 2.3.6 описан поиск вероятных каузативных замен в найденных локусах. В разделе 2.3.7 описан анализ биологических путей и экспрессии генов найденных локусов. В разделе 2.3.8 описан метод SMR/HEIDI позволяющий определить гены, чей уровень экспрессии в определенной ткани может опосредовать найденные ассоциации между локусом и признаков. В разделе 2.3.9 описан поиск возможных плеiotропных эффектов найденных локусов на признаки и заболевания человека.

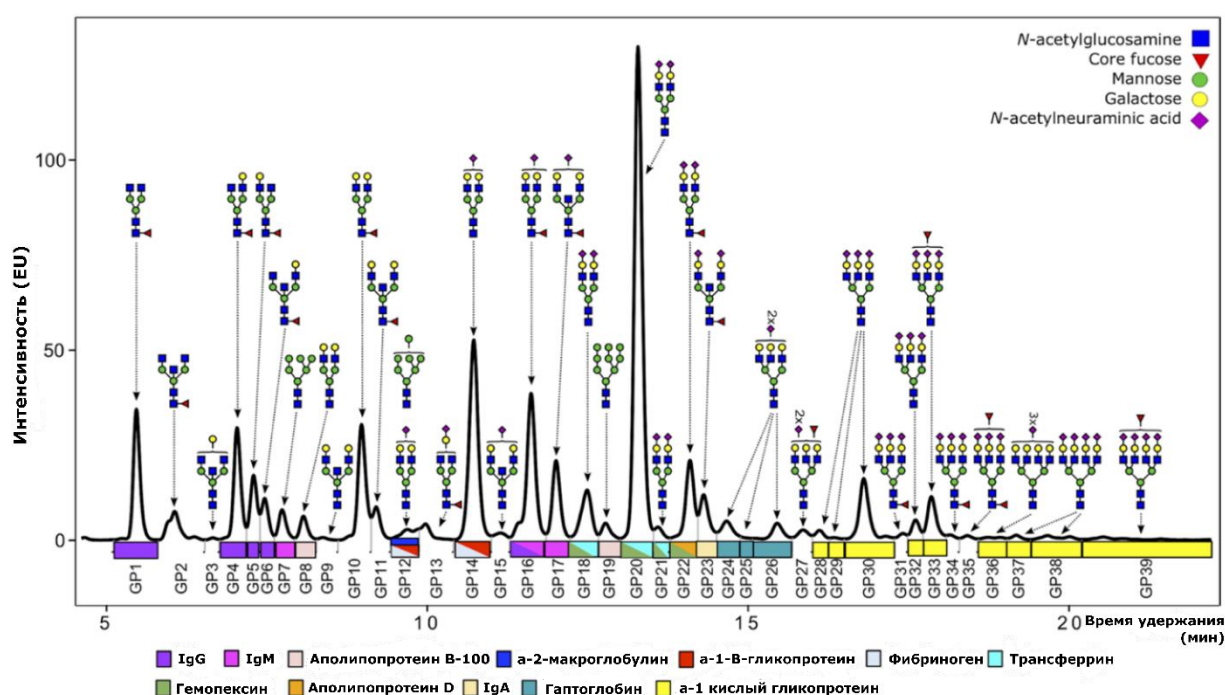


Рис. 1. Пример СВЭЖХ хроматограммы, полученной для образца из выборки TwinsUK. По оси X отложено время удерживания, по оси Y – интенсивность сигнала флуоресценции. Каждому образцу приписаны структуры гликанов, входящих в состав пика согласно аннотации 2020 года (Zaytseva и др., 2020). На графике представлены гликопротеины, с которыми связаны представленные N-гликаны.

## Глава 3. Результаты

### 3.1 ПГИА и определение локусов, ассоциированных с уровнями N-гликанов белков плазмы крови человека

В данной работе впервые проведен ПГИА уровней N-гликанов белков плазмы крови, измеренных технологией СВЭЖХ. Результаты описаны в разделе 3.1. Полногеномное исследование ассоциаций с целью выявления новых локусов, ассоциированных с уровнями N-гликанов белков плазмы крови человека, было проведено на выборке TwinsUK. Геномные и гликомные данные 2,763 образцов успешно прошли контроль качества. На основе сходства биохимической структуры N-гликанов, входящих в состав 36 пиков на хроматограмме, были рассчитаны 77

производных гликомных признаков. Перед проведением ПГИА все признаки были предварительно скорректированы на эффекты пола и возраста и приведены к нормальному распределению методом квантильной нормализации. ПГИА был проведен для 113 N-гликомных признаков и 8,557,543 ОНП, измеренных у 2,763 участников исследования TwinsUK. Использовалась смешанная линейная модель ассоциации, учитывающая возможную генетическую структурированность выборки.

В общей сложности, 908 ОНП показали значимую ( $P\text{-value} < 1.7 \cdot 10^{-9}$ ) ассоциацию как минимум с одним из 68 гликаных признаков (всего обнаружено 5,090 значимых ассоциаций ОНП-признак). Данные ОНП были расположены в 14 локусах (см. Рис. 2 и Табл. 1). Сравнительный анализ полученных результатов с опубликованными ранее показал, что среди найденных локусов присутствуют четыре из шести локуса, выявленных в предыдущих исследованиях. Мы впервые показали ассоциацию десяти локусов генома с уровнями N-гликанов белков плазмы крови. Таким образом, общее число локусов, с когда-либо показанной ассоциацией с уровнями N-гликанов белков плазмы крови человека, было увеличено с 6 до 16. Полученные результаты опубликованы в работе (Sharapov и др., 2019).

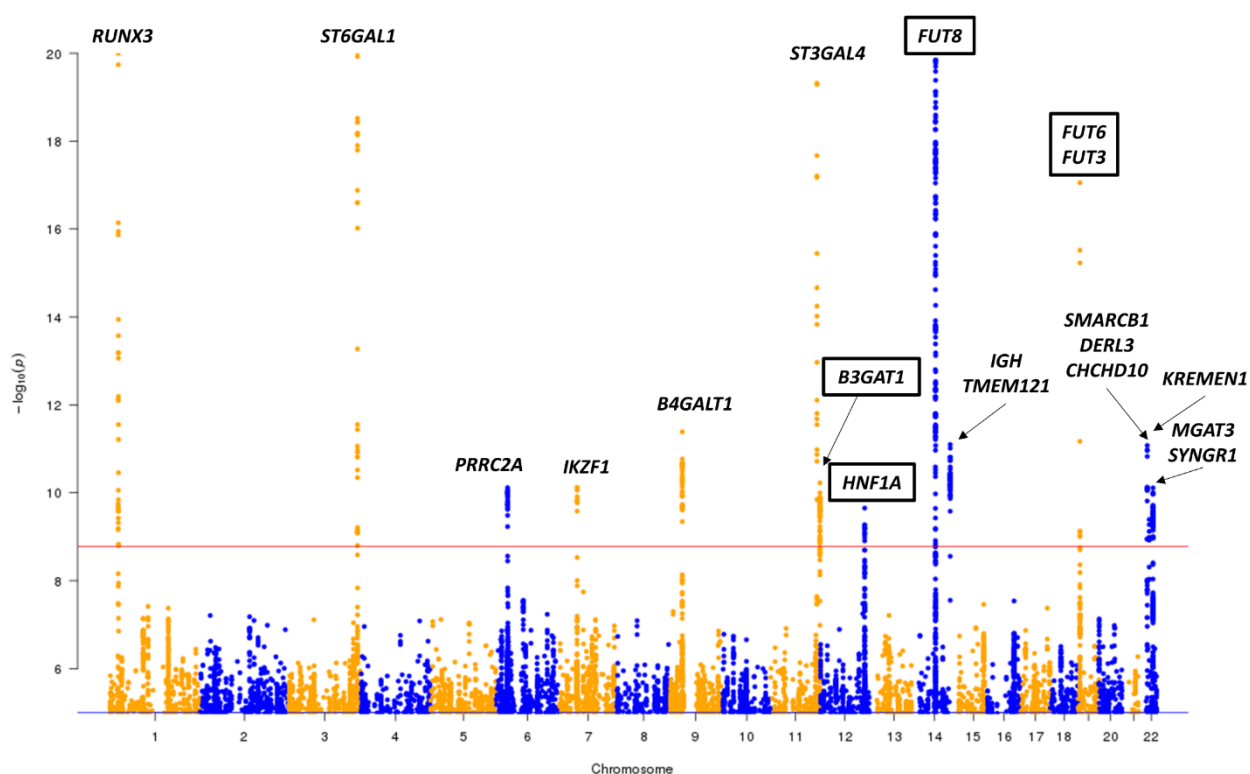


Рис. 2. Результаты ПГИА гликома плазмы крови человека. Показано расположение локусов, ассоциированных с гликомом плазмы крови человека. По оси X - геномная координата ОНП (номера обозначены номера хромосом). По оси Y – отрицательный десятичный логарифм P-value ассоциации:  $-\log_{10}(P\text{-value})$ . Для каждого ОНП выбрано наименьшее P-value среди 113 признаков. Красная линия соответствует уровню значимости ( $1.7 \cdot 10^{-9}$ ). На рисунке показаны ОНП с  $P\text{-value} < 1 \cdot 10^{-5}$ . Точки с  $-\log_{10}(P\text{-value}) > 20$  показаны на уровне  $-\log_{10}(P\text{-value}) = 20$ . Локусы, прошедшие порог полногеномной значимости отмечены названиями генов, приоритизированных в результате биоинформатического анализа (см. Раздел 3.4). Рамкой выделены локусы, найденные в предыдущих исследованиях.



Табл. 1. Результаты ПГИА N-гликома плазмы крови человека. В верхней части таблицы приведены результаты ПГИА десяти новых локусов; в нижней части таблицы приведены результаты ПГИА для четырех локусов, ассоциация которых с гликомом была показана ранее. Позиция – позиция ОНП (хромосома:п.н.) в геномных координатах (сборки генома GRCh37); ген – предложенный кандидатный ген в локусе; Эфф/Реф – эффекторный/референтный аллель; EAF – частота эффекторного аллеля; BETA (SE) – оценка эффекта эффекторного аллеля на признак и его стандартная ошибка; P – P-value ассоциации ОНП с признаком; P<sub>GC</sub> – P-value ассоциации после коррекции геномным контролем; признак – признак с минимальным P-value ассоциации для данного ОНП

| Локус              |              |                                    | TwinsUK     |      |               |          |                 |                           |                    |
|--------------------|--------------|------------------------------------|-------------|------|---------------|----------|-----------------|---------------------------|--------------------|
| ОНП                | Позиция      | Ген                                | Эфф/<br>Реф | EAF  | BETA (SE)     | P        | P <sub>GC</sub> | Признак                   | Число<br>признаков |
| <b>rs186127900</b> | 1:25318225   | <i>RUNX3</i>                       | G/T         | 0.99 | -1.26 (0.119) | 1.35E-24 | 4.04E-24        | FBG1n/G1n (PGP82)         | 26                 |
| <b>rs59111563</b>  | 3:186722848  | <i>ST6GAL1</i>                     | D/I         | 0.74 | 0.34 (0.031)  | 9.51E-27 | 1.09E-26        | FG1S1/(FG1+FG1S1) (PGP41) | 3                  |
| <b>rs3115663</b>   | 6:31601843   | <i>HLA</i>                         | T/C         | 0.80 | 0.26 (0.040)  | 6.31E-11 | 7.65E-11        | M9 (PGP18)                | 1                  |
| <b>rs6421315</b>   | 7:50355207   | <i>IKZF1</i>                       | G/C         | 0.59 | 0.19 (0.029)  | 7.50E-11 | 7.57E-11        | A2[6]BG1n (PGP60)         | 2                  |
| <b>rs13297246</b>  | 9:33128617   | <i>B4GALT1</i>                     | G/A         | 0.83 | -0.26 (0.038) | 3.46E-12 | 4.11E-12        | FA2G2n (PGP67)            | 2                  |
| <b>rs3967200</b>   | 11:126232385 | <i>ST3GAL4</i>                     | C/T         | 0.88 | -0.49 (0.043) | 8.64E-28 | 1.51E-27        | A2G2S[3,6+3]2 (PGP17)     | 7                  |
| <b>rs35590487</b>  | 14:105989599 | <i>IGH; TMEM121</i>                | C/T         | 0.77 | -0.24 (0.034) | 5.90E-12 | 7.98E-12        | FA2[3]G1n (PGP62)         | 2                  |
| <b>rs9624334</b>   | 22:24166256  | <i>SMARCB1; DERL3;<br/>CHCHD10</i> | G/C         | 0.85 | 0.28 (0.040)  | 4.79E-12 | 8.38E-12        | FA2[6]BG1n (PGP63)        | 2                  |
| <b>rs140053014</b> | 22:29550678  | <i>KREMEN1</i>                     | I/D         | 0.98 | -0.67 (0.106) | 4.01E-10 | 4.05E-10        | G3S2/G3S3 (PGP109)        | 1                  |
| <b>rs909674</b>    | 22:39859169  | <i>MGAT3</i>                       | C/A         | 0.27 | 0.22 (0.033)  | 6.31E-11 | 7.72E-11        | FBS2/FS2 (PGP56)          | 3                  |
| <b>rs1866767</b>   | 11:134274763 | <i>B3GAT1</i>                      | C/T         | 0.87 | 0.28 (0.043)  | 5.15E-11 | 5.95E-11        | A4G4S[3,3,3,3]4 (PGP33)   | 3                  |
| <b>rs1169303</b>   | 12:121436376 | <i>HNF1A</i>                       | A/C         | 0.51 | 0.19 (0.029)  | 2.07E-10 | 2.23E-10        | A4G4S[3,3,3]3 (PGP30)     | 2                  |
| <b>rs7147636</b>   | 14:66011184  | <i>FUT8</i>                        | T/C         | 0.33 | -0.39 (0.030) | 4.71E-37 | 6.63E-37        | FA2G2S[3+6,6+3]2 (PGP20)  | 17                 |
| <b>rs7255720</b>   | 19:5828064   | <i>FUT3; FUT6</i>                  | G/C         | 0.96 | 1.14 (0.068)  | 2.13E-55 | 2.53E-55        | G4S3/G4S4 (PGP110)        | 18                 |

### 3.2 Разработка и валидация метода гармонизации гликомных профилей

Для подтверждения генетических ассоциаций, найденных при анализе выборки TwinsUK, использовались данные независимых выборок EPIC, PainOmics, SOCCS и SABRE. Данные N-гликозилирования белков плазмы крови образцов этих выборок содержали от 36 (SABRE) до 42 (SOCCS) пиков на хроматограмме. В зависимости от эксперимента некоторые из пиков оказались неразделимыми на хроматограмме (в результате чего они объединяются в один пик), при этом порядок пиков и структуры N-гликанов, входящих в эти пики, сохраняются (см. Рис. 4). Сравнение результатов анализа генетических ассоциаций нескольких выборок подразумевает, что в каждой из выборок анализ ассоциаций проводился для одного и того же (гармонизированного) набора признаков. Гармонизацию гликомных профилей можно проводить методом изменения границ интегрирования пиков. В этом случае, хроматограммы каждого из образцов каждой из выборок требуют повторной ручной обработки, в ходе которой будут установлены единые для всех выборок границы пиков. Однако, ручная обработка хроматограмм является трудоемким процессом. Так как количественным признаком, отражающим уровень N-гликана, является площадь под соответствующим пиком, а в результате объединения двух пиков в один происходит объединение их площадей, то простым способом гармонизации пиков между выборками является суммирование площадей соответствующих пиков (см. Рис. 3).

Для реализации разработанного подхода, была создана таблица соответствия гликомных пиков, полученных для исследуемых выборок на основе данных о биохимической структуре гликанов, входящих в состав гликомных пиков. Был составлен список из 36 гармонизированных гликомных пиков и схема гармонизации для каждой из выборок. Разработанный протокол гармонизации был апробирован на данных 35 случайно выбранных образцов из исследуемых выборок. Для этих образцов была проведена гармонизация пиков GP24 и GP25 в пик RGP22 двумя методами – ручным объединением пиков на хроматограмме и методом суммирования площадей. Коэффициент корреляция Пирсона между значениями пика RGP22, полученного двумя методами был более 0.9999, при этом разница между значением RGP22, полученным методом суммирования площадей и методом ручной интеграции, наблюдалась лишь в восьмом значащем знаке. Ввиду простоты, низких временных затрат, и высокой точности в качестве метода гармонизации был выбран разработанный в данной работе метод суммирования площадей.

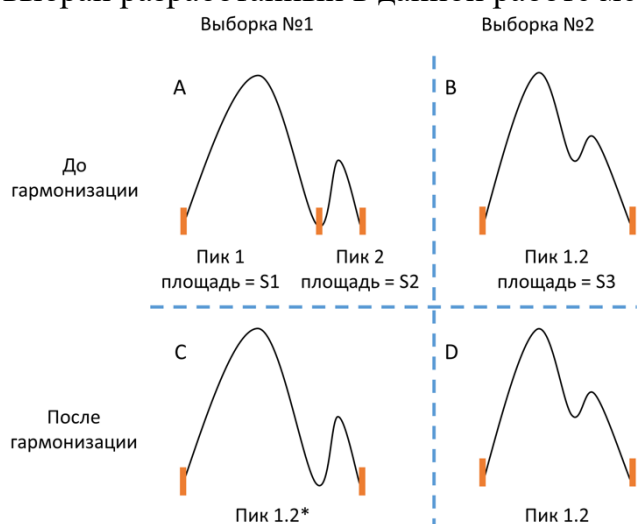


Рис. 3. Схема гармонизации гликомных пиков. А и В – участки исходных хроматограмм (до гармонизации) образцов из двух выборок. А – в выборке №1 пики 1 и 2 разделены. В – в выборке №2 пики 1 и 2 не разделены и образуют пик 1.2. С и D – участки хроматограмм, после проведения гармонизации. С – площадь пика 1.2\* равна сумме площадей двух исходных пиков 1 и 2. D – пик 1.2 соответствует гармонизированному пику 1.2\*.

### 3.3 Подтверждение результатов ПГИА на независимых выборках

Подтверждение генетических ассоциаций 16 локусов проводилось на материале выборок EPIC-Potsdam, PainOmics, SOCCS и SABRE. Эти выборки не анализировались в предыдущих исследованиях, и на момент проводимого исследования они представляли собой самую большую коллекцию образцов геномных и N-гликомных данных. Уровни N-гликанов образцов исследуемых выборок были измерены методом СВЭЖХ. Гликомные данные образцов были гармонизированы разработанным методом суммирования площадей. В общей сложности N-гликомные и геномные данные 4,802 образцов четырех выборок прошли контроль качества.

Для каждого из 16 локусов был выбран ОНП для проведения анализа ассоциаций. Для локусов *FUT3/FUT5/FUT6*, *FUT8*, *HNFI1A*, *B3GAT1*, *MGAT5* и *SLC9A9*, чья ассоциация была показана в предыдущих ПГИА, были выбраны ОНП, показавшие наименьшее P-value в исследовании (Huffman и др., 2011). Для 10 новых локусов были выбраны ОНП, показавшие наименьшее P-value ассоциации на выборке TwinsUK (см. Табл. 1). Анализ генетических ассоциаций 16 локусов проводился с использованием того же протокола, что и для поискового ПГИА, выполненного на выборке TwinsUK. Объединение результатов, полученных на четырех выборках, проводилось методом мета-анализа с фиксированными эффектами. В данном анализе использовалась обновленная (по сравнению с данными выборки TwinsUK) панель N-гликомных признаков. Для получения обновленного списка признаков, ассоциированных с 16 локусами, тестировалась ассоциация локусов со всеми 117 N-гликомными признаками из обновленной панели. Ассоциация локуса считалась подтвержденной, если P-value ассоциации ОНП с хотя бы одним из 117 признаков было меньше использованного порога. Использовался уровень значимости  $P\text{-value} < 0.05/(16 * 117) = 2.67 * 10^{-5}$ , где 16 – число локусов, а 117 – число N-гликомных признаков.

Объем выборки обеспечивал 95% статистическую мощность подтверждения ассоциаций для 15 локусов. Для локуса, расположенного на 22 хромосоме, 29.5 м.п.н. (rs140053014, интронный вариант гена *KREMEN1*) статистическая мощность подтверждения равнялась 9%. Это связано с тем, что данный ОНП прошел контроль качества только в выборке SOCCS (N = 472).

В результате была подтверждена ассоциация 15 из 16 локусов (см. Табл. 2). Высокий процент подтвержденных ассоциаций локусов (94%) свидетельствует о надежности результатов ПГИА, протоколов и стандартов, принятых сообществом, изучающим генетическую регуляцию гликозилирования белков.

Ассоциация локуса *KREMEN1* не была подтверждена. Это может объясняться недостаточной статистической мощностью (9%) подтверждения сигнала ассоциации. В этом случае, повторное подтверждение ассоциации локуса *KREMEN1* на выборке размером более 2,480 образцов позволит подтвердить ассоциацию данного локуса с 95% мощностью. На данный момент, на основании полученных результатов, нельзя сделать однозначный вывод об отсутствии или наличии роли локуса *KREMEN1* в контроле N-гликозилирования белков плазмы крови человека.

Табл. 2. Подтверждение ассоциаций 16 локусов. В верхней части таблицы приведены результаты для шести локусов, ассоциация которых с гликомом была показана ранее (Huffman и др., 2011); в нижней части таблицы приведены результаты для десяти локусов, ассоциация которых показана впервые в данном исследовании. ОНП – ОНП, показавшие наименьшее P-value в исследовании (Huffman и др., 2011) (верхняя часть таблицы), в данном исследовании (нижняя часть таблицы); Позиция – позиция ОНП (хромосома:п.н.) в геномных координатах (сборки генома GRCh37); ген – предложенный кандидатный ген в локусе; Эфф/Реф – эффекторный/референтный аллель; EAF – частота эффекторного аллеля; признак – признак с минимальным P-value ассоциации для данного ОНП; BETA (SE) – оценка эффекта эффекторного аллеля на признак и его стандартная ошибка; P – P-value ассоциации ОНП с признаком

| ОНП                     | Позиция      | Ген                         | Эфф/Реф | EAF  | Признак                 | BETA (SE)           | P                      | Признак           | BETA (SE)    | P                       | EAF  |
|-------------------------|--------------|-----------------------------|---------|------|-------------------------|---------------------|------------------------|-------------------|--------------|-------------------------|------|
| Huffman et al. 2011     |              |                             |         |      |                         | Данное исследование |                        |                   |              |                         |      |
| rs1257220               | 2:135015347  | MGAT5                       | A/G     | 0.26 | Tetra-antennary glycans | 0.19 (0.03)         | $1.80 \times 10^{-10}$ | G4total, A4total  | 0.22 (0.02)  | $6.11 \times 10^{-20}$  | 0.26 |
| rs4839604               | 3:142960273  | SLC9A9                      | C/T     | 0.77 | Tetrasialylated glycans | -0.22 (0.03)        | $3.50 \times 10^{-13}$ | FBS2/FS2          | -0.20 (0.03) | $3.87 \times 10^{-11}$  | 0.80 |
| rs7928758               | 11:134265967 | B3GAT1                      | T/G     | 0.88 | A4F2G4                  | 0.23 (0.04)         | $1.66 \times 10^{-08}$ | A4G4S3            | 0.36 (0.03)  | $6.43 \times 10^{-27}$  | 0.85 |
| rs735396                | 12:121438844 | HNF1A                       | T/C     | 0.61 | A2F1G2                  | 0.18 (0.03)         | $7.81 \times 10^{-12}$ | G3Fa/G3total      | 0.21 (0.02)  | $4.91 \times 10^{-20}$  | 0.65 |
| rs11621121              | 14:65822493  | FUT8                        | C/T     | 0.43 | A2                      | 0.27 (0.03)         | $1.69 \times 10^{-23}$ | FG3/G3total       | -0.31 (0.02) | $8.94 \times 10^{-45}$  | 0.42 |
| rs3760776               | 19:5839746   | FUT3;<br>FUT6               | G/A     | 0.87 | A3F1G3                  | 0.44 (0.04)         | $3.18 \times 10^{-29}$ | G3Fa/G3total      | 0.48 (0.05)  | $3.85 \times 10^{-23}$  | 0.91 |
| Sharapov et al. 2019    |              |                             |         |      |                         | Данное исследование |                        |                   |              |                         |      |
| rs186127900             | 1:25318225   | RUNX3                       | G/T     | 0.99 | FBG1n/G1n               | -1.26 (0.12)        | $4.04 \times 10^{-24}$ | FA2G2S2           | 1.24 (0.19)  | $1.16 \times 10^{-10}$  | 0.99 |
| rs59111563 <sup>g</sup> | 3:186722848  | ST6GAL1                     | Del/Ins | 0.74 | FG1S1/(FG1+FG1S1)       | 0.34 (0.03)         | $1.09 \times 10^{-26}$ | FG1S1/(FG1+FG1S1) | 0.49 (0.02)  | $8.60 \times 10^{-97}$  | 0.74 |
| rs3115663               | 6:31601843   | HLA                         | T/C     | 0.80 | M9                      | 0.26 (0.04)         | $7.65 \times 10^{-11}$ | M9                | 0.15 (0.03)  | $1.63 \times 10^{-07}$  | 0.82 |
| rs6421315               | 7:50355207   | IKZF1                       | G/C     | 0.59 | A2[6]BG1n               | 0.19 (0.03)         | $7.57 \times 10^{-11}$ | A2[6]BG1n         | 0.23 (0.02)  | $1.19 \times 10^{-27}$  | 0.63 |
| rs13297246              | 9:33128617   | B4GALT1                     | G/A     | 0.83 | FA2G2n                  | -0.26 (0.04)        | $4.11 \times 10^{-12}$ | FA2G2n            | -0.31 (0.03) | $1.28 \times 10^{-24}$  | 0.83 |
| rs3967200               | 11:126232385 | ST3GAL4                     | C/T     | 0.88 | A2G2S[3,6+3]2           | -0.49 (0.04)        | $1.51 \times 10^{-27}$ | G4S3/G4S4         | 0.63 (0.03)  | $1.20 \times 10^{-106}$ | 0.86 |
| rs35590487              | 14:105989599 | IGH<br>TMEM121              | C/T     | 0.77 | FA2[3]G1n               | -0.24 (0.03)        | $7.98 \times 10^{-12}$ | FA2[3]G1n         | -0.20 (0.03) | $1.38 \times 10^{-09}$  | 0.75 |
| rs9624334               | 22:24166256  | SMARCB1<br>DERL3<br>CHCHD10 | G/C     | 0.85 | FA2[6]BG1n              | 0.28 (0.04)         | $8.38 \times 10^{-12}$ | FA2[6]BG1n        | 0.31 (0.03)  | $7.15 \times 10^{-26}$  | 0.83 |
| rs140053014             | 22:29550678  | KREMEN1                     | Ins/Del | 0.98 | G3S2/G3S3               | -0.67 (0.11)        | $4.05 \times 10^{-10}$ | FA2[3]G1n         | -0.68 (0.23) | 0.0027                  | 0.98 |
| rs909674                | 22:39859169  | MGAT3                       | C/A     | 0.27 | FBS2/FS2                | 0.22 (0.03)         | $7.72 \times 10^{-11}$ | FBn               | 0.22 (0.02)  | $1.88 \times 10^{-20}$  | 0.30 |

### 3.4 Биоинформатический анализ результатов ПГИА

Знание локуса является отправной точкой в исследовании биологических механизмов, лежащих в основе влияния генетической изменчивости на формирование исследуемого признака. Поскольку в локусе могут находиться тысячи сцепленных генетических вариантов и последовательности десятков генов, число возможных механистических гипотез, обуславливающих найденные ассоциации, может быть очень велико. В данном разделе диссертации описаны результаты биоинформатического и количественно-генетического исследования 15 локусов генома с подтвержденной ассоциацией, выполненного для отбора наиболее вероятных генов-кандидатов и формирования гипотез, объясняющих найденные ассоциации. В разделе 3.4.1 описаны результаты поиска каузативных ОНП в найденных локусах и генов, на которые влияют данные ОНП. В разделе 3.4.2 описана приоритизация генов методом DEPICT, основанном на идентификации биологических путей, влияющих на гликозилирование, и анализе тканеспецифичной экспрессии генов. В разделе 3.4.3 приведены результаты анализа SMR/HEIDI для определения генов, чей уровень экспрессии может опосредовать ассоциацию найденных локусов с уровнями N-гликанов белков плазмы крови. Основными секретами N-гликопротеинов в кровотоке являются клетки, секретирующие антитела, и печень. Поэтому в анализе использовались доступные нам данные об уровнях экспрессии генов в клетках и тканях, наиболее близких к ним - CD19+ В-лимфоцитах, крови и печени. В разделе 3.4.4 описаны возможные плеiotропные эффекты найденных локусов на признаки и заболевания человека. В разделе 3.4.5 приведено описание всех предложенных генов-кандидатов и гипотезы об их роли в регуляции N-гликозилирования белков плазмы крови человека.

В результате обобщения полученных результатов мы приоритизировали 18 генов в 15 локусах. В восьми локусах из пятнадцати были приоритизированы гены *MGAT5*, *MGAT3*, *FUT3/FUT6*, *FUT8*, *ST6GAL1*, *ST3GAL4*, *B4GALT1* и *B3GAT1*, кодирующие ферменты - гликозилтрансферазы, участвующие в биосинтезе N-гликанов. За исключением локуса *B3GAT1*, спектр N-гликомных признаков, ассоциированных с данными локусами, соответствовал известной ферментативной активности данных ферментов.

В локусе на 3-ей хромосоме, 142 м.п.н. был приоритизирован ген *SLC9A9*. Этот ген кодирует Na<sup>+</sup>/H насос, который предположительно регулирует уровень pH в аппарате Гольджи (АГ). Гликозилирование белков происходит в АГ, и, согласно имеющимся данным, этот процесс чувствителен к изменению pH (Kellokumpu, 2019). Процессы, происходящие в АГ, влияют на формирование гетеродимерных комплексов, осуществляющих гликозилирование (Hassinen и др., 2011). В работе (Rivinoja, Hassinen, Kokkonen, Kauppinen, & Kellokumpu, 2009) было показано, что увеличение pH в АГ может нарушить терминальное N-гликозилирование (включая сиалилирование) из-за неверной локализации гликозилтрансфераз. В согласии с этой гипотезой, в работе (Huffman и др., 2011) локус *SLC9A9* показал ассоциацию с уровнем тетра-сиалилированных N-гликанов, а в данной работе – с уровнем сиалилированных N-гликанов.

В локусе на 12-ой хромосоме, 121 м.п.н., был приоритизирован ген *HNF1A*. Подробное функциональное исследование данного локуса, проведенное в работе (Laus и др., 2010), показало, что ген *HNF1A*, кодирующий фактор транскрипции

гепатоцитов, регулирует экспрессию большинства генов, кодирующих фукозилтрансферазы *FUT3*, *FUT5*, *FUT6*, *FUT8*, *FUT10* и *FUT11* в клеточной линии HepG2, полученной из клеток печени.

В локусе на 22-ой хромосоме 39 м.п.н. были приоритизированы три гена – *SMARCB1*, *DERL3* и *CHCHD10*. Ген *SMARCB1* кодирует белок, являющийся частью комплекса hSWI/SNF – ремоделера хроматина в иммунных клетках. Метод DEPICT в данном локусе приоритизировал ген *DERL3*, кодирующий фермент, участвующий в деградации люминальных гликопротеинов с некорректной третичной структурой, происходящей в эндоплазматическом ретикулуме (Oda и др., 2006). Также в кодирующей последовательности гена *DERL3* расположен патогенный вариант rs3177243, определенный нами как потенциально каузативный в данном локусе. Ген *CHCHD10* кодирует митохондриальный белок, встречающийся в фибриллах крист митохондрий. Функция этого гена на данный момент не ясна. В нашем исследовании было показано, что ассоциация данного локуса с гликомом плазмы крови может быть опосредована плеiotропным эффектом локуса на уровень экспрессии гена *CHCHD10*. В 2017 году была опубликована работа, показавшая роль фрагментации митохондрий и числа контактов эндо-плазматического ретикулума с митохондриями в представленности сialiрированных гликанов на поверхности глиобластных клеток, что в свою очередь влияло на узнавание лимфоцитами клеток глиобластомы (Martinvalet, 2018).

Локус на хромосоме 14 в 105 м.п.н. содержит кластер генов *IGH*, которые кодируют тяжелые цепи иммуноглобулинов. Иммуноглобулин G (IgG) является наиболее представленным N-гликопротеином плазмы крови человека (Clerc и др., 2016). Однако анализ SMR/HEIDI показал, что ассоциация данного локуса с гликомом плазмы может опосредоваться экспрессией гена *TMEM121* в крови. Более того, в этом локусе была приоритизирована потенциальная каузативная замена rs10569304 – делеция, расположенная в экзоне гена *TMEM121*. Данный ген кодирует трансмембранный белок 121, и экспрессируется в тканях сердечной мышцы, поджелудочной железы, печени, скелетных мышц и селезенки.

Для локуса на 7 хромосоме в 50 м.п.н. был приоритизирован ген *IKZF1*. Ранее в работе (Lauc и др., 2013) была показана ассоциация данного локуса с N-гликозилированием IgG. Авторы предположили, что ген *IKZF1* является кандидатным для данного локуса. Ген *IKZF1* кодирует ДНК-связывающий белок Ikaros - регулятор транскрипции, связанный с ремоделированием хроматина. *IKZF1* рассматривается как важный регулятор дифференциации лимфоцитов (Marke, van Leeuwen, & Scheijen, 2018; Sellars, Reina-San-Martin, Kastner, & Chan, 2009). Интересно, что локус *IKZF1* показал ассоциацию с уровнями N-гликанов белков плазмы крови (и IgG в частности) с фукозилированным остовом, с которыми был ассоциирован локус *FUT8* (см. Рис. 4). Нами была выдвинута гипотеза о роли *IKZF1* в регуляции фукозилирования остова N-гликанов IgG путем регуляции экспрессии гена *FUT8*. В исследовании (Klarić и др., 2020), выполненном в соавторстве с соискателем, было показано, что нокаун гена *IKZF1* в IgG секретирующих клетках МАТАТ6 приводит к более чем трехкратному увеличению экспрессии *FUT8* и увеличению уровня фукозилирования остова N-гликана, связанного с уровнем секретируемого IgG. Таким образом, выдвинутая нами гипотеза нашла свое подтверждение.

В локусе на 1-ой хромосоме, 25 м.п.н., был приоритизирован ген *RUNX3*. Данный ген кодирует Runt-домен-содержащий белок - фактор транскрипции, который также, как и ген *IKZF1* (Sellars и др., 2009) играет важную роль в созревании и дифференцировке В-лимфоцитов.

Приоритизация генов в локусе *HLA* (локус главного комплекса гистосовместимости человека) на 6-ой хромосоме 25-32 м.п.н. была проведена, однако ее результаты имеют высокий шанс быть ложноположительными. Данный локус имеет самую высокую плотность генов в геноме человека; он имеет высочайшую степень полиморфности на нуклеотидном уровне; аллели данного локуса находятся в высоком неравновесии по сцеплению на протяжении всего локуса длиной в 8 м.п.н. Все это не позволяет проводить тонкое картирование и приоритизацию генов в данном локусе общепринятыми методами.

Таким образом, в результате проведённой работы мы предложили новые гены-кандидаты, вовлеченные в процесс N-гликозилирования - гены регуляторов транскрипции (*IKZF1*, *SMARCB1* и *RUNX3*), деградации гликопротеинов (*DERL3*), тяжелой цепи иммуноглобулинов (*IGH*) и гены с неизвестной функцией (*TMEM121* и *CHCHD10*).

### 3.5 Генная сеть регуляции N-гликозилирования

В данной работе был проведен анализ генетических ассоциаций с использованием самой крупной коллекции образцов с измеренными геномами и N-гликомами плазмы крови. Была показана достоверная ассоциация 15 локусов со 116 из 117 признаков. Суммарно 214 пар локус-признак показали достоверную ассоциацию (см. раздел Подтверждение результатов ПГИА на независимых выборках). Эти результаты были использованы для реконструкции генной сети регуляции уровней N-гликанов белков плазмы крови (см. Рис. 4). Данная сеть визуализирует ассоциацию между найденными локусами и уровнями N-гликанов белков плазмы крови.

Для построения сети была проведена классификация гликомных признаков. Признаки были классифицированы на четыре группы согласно ткани, секретирующей N-гликопротеины в плазму крови. Первая группа – признаки, отражающие уровни N-гликанов иммуноглобулинов (A, G, D, E, M), секретируемых плазмócитами. Вторая группа - признаки, отражающие уровни N-гликанов белков (трансферрин, гаптоглобин, и т.д.), секретируемых в основном гепатоцитами – клетками печени. Третья группа - признаки, отражающие уровни N-гликанов, которые могут быть связаны с белками, секретируемыми как плазмócитами, так и гепатоцитами. Четвертая группа - признаки, классификация которых не была произведена.

Локусы и ассоциированные признаки в данной сети могут быть визуально разделены на две частично перекрывающиеся подсети. Первая подсеть образована локусами *ST3GAL4*, *B3GAT1*, *MGAT5*, *FUT3/FUT6*, *HNF1A* и *SLC9A9*. Эта подсеть ассоциирована с уровнями N-гликанов, связанных с N-гликопротеинами, секретируемыми в кровотоке клетками печени – гепатоцитами. Большинство из этих признаков отражают уровни N-гликанов с тремя или четырьмя антеннами, не встречающимися в составе иммуноглобулинов. Полученные результаты позволяют



выдвинуть гипотезу о том, что данная подсеть генов регулирует процессы N-гликозилирования в гепатоцитах.

Вторая подсеть образована локусами *FUT8*, *IKZF1*, *MGAT3*, *RUNX3*, *SMARCB1/DERL3/CHCHD10*, *B4GALT1*, *ST6GAL1* и *IGH/TMEM121*. Эти локусы ассоциированы с уровнями N-гликанов, связанных с иммуноглобулинами, секретируемыми в кровотоке плазмочитами. Основываясь на том, что IgG является наиболее представленным гликозилированным белком плазмы крови, можно выдвинуть гипотезу о том, что гены-кандидаты из данной сети регулируют процессы N-гликозилирования в плазмочитах.

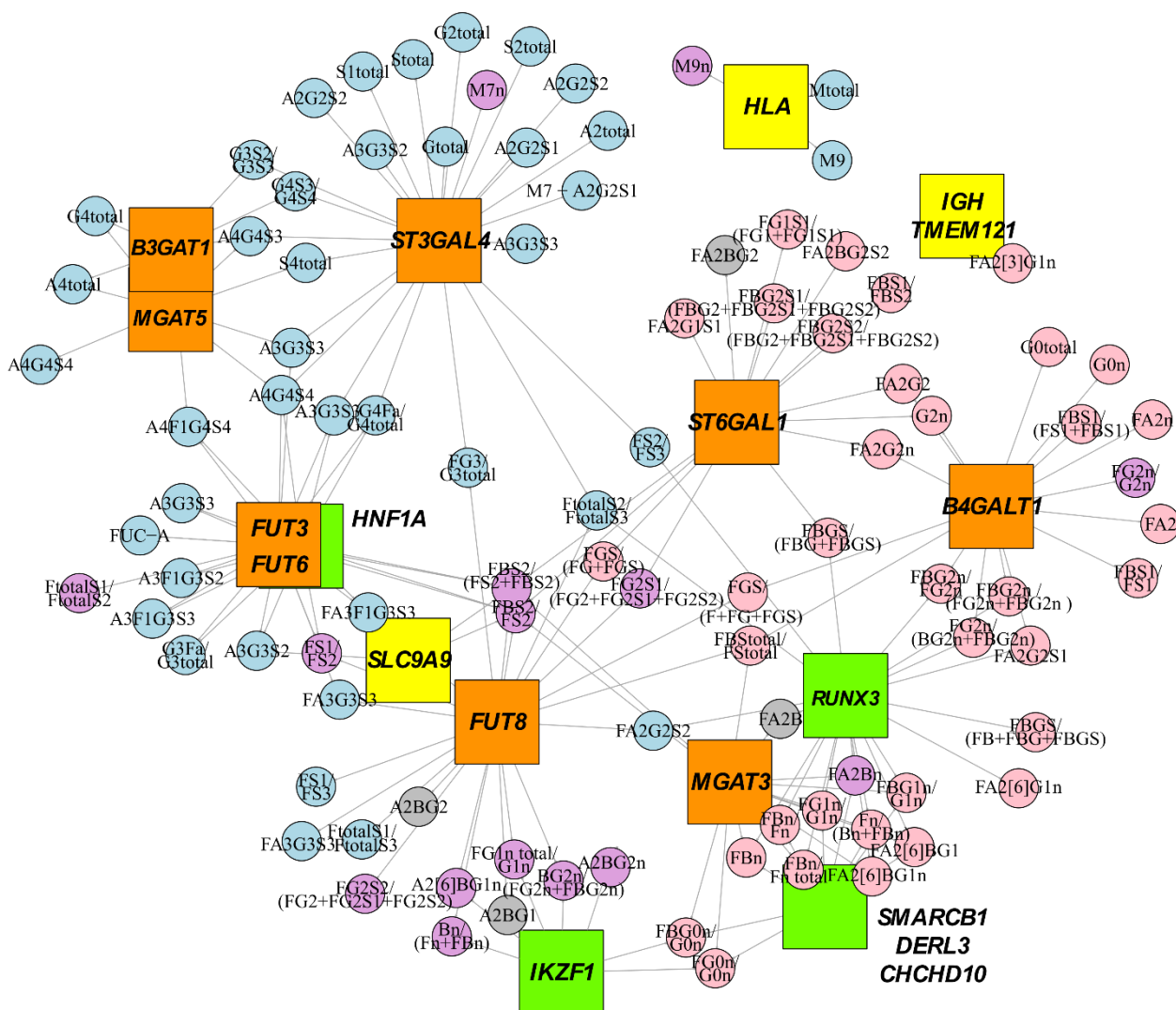


Рис. 4. Генная сеть регуляции уровней N-гликанов белков плазмы крови. Квадратами обозначены локусы, найденные в данном исследовании. Для каждого локуса приведены названия приоритизированных генов. Оранжевым цветом выделены локусы, для которых были приоритизированы гены гликозилтрансфераз. Зеленым цветом окрашены локусы, для которых были приоритизированы гены факторов транскрипции. Кругами обозначены уровни N-гликанов белков плазмы крови. Синим цветом окрашены уровни N-гликанов, связанных с гликопротеинами, секретируемыми гепатоцитами. Розовым цветом окрашены уровни N-гликанов, связанных с гликопротеинами (а именно иммуноглобулинами), секретируемыми клетками лимфоцитарного ряда. Пурпурным цветом окрашены уровни N-гликанов, которые могут быть связаны с гликопротеинами, секретируемыми как гепатоцитами, так и клетками лимфоцитарного ряда. Серым цветом окрашены уровни N-гликанов, для которых классификация не была проведена.



## ЗАКЛЮЧЕНИЕ

В диссертационной работе описаны результаты первого в мире полногеномного исследования ассоциаций уровней N-гликанов белков плазмы крови человека, измеренных методом СВЭЖХ. В результате данной работы число локусов, ассоциированных с гликомом плазмы крови, было увеличено с 6 до 16. Ассоциация 15 из 16 локусов была подтверждена в независимых выборках. Для найденных локусов было проведено биоинформатическое исследование с целью приоритизации генов-кандидатов и установления возможных механизмов влияния данных генов на регуляцию N-гликома плазмы крови. В результате было предложено 18 генов-кандидатов. Реконструкция генной сети регуляции N-гликозилирования позволила выдвинуть гипотезы о тканеспецифичной регуляции N-гликозилирования предложенными генами-кандидатами.

Результаты, описанные в диссертационной работе, подтверждают идею о том, что генетический контроль N-гликозилирования белков плазмы крови человека представляет собой сложный процесс, который находится под контролем генов, принадлежащих к разным путям и экспрессирующихся в разных тканях. Полученные результаты позволяют сформулировать функциональные гипотезы о механизме влияния найденных локусов на N-гликозилирование белков плазмы крови и будут востребованы в области молекулярно-генетических исследований N-гликома и изучения роли N-гликома в патогенезе социально и экономически важных заболеваний человека.

## ВЫВОДЫ

1. В результате проведения полногеномного исследования генетических ассоциаций уровней N-гликанов белков плазмы крови человека и последующего подтверждения найденных ассоциаций, показано участие 15 локусов генома в контроле исследованных признаков. В данных локусах приоритизированы 18 генов-кандидатов, чьи продукты наиболее вероятно вовлечены в регуляцию N-гликозилирования белков плазмы крови человека.
2. Для девяти локусов подтвержденная ассоциация с уровнями N-гликанов белков плазмы крови была показана впервые. Для этих локусов были предложены гены-кандидаты *IKZF1*, *RUNX3*, *SMARCB1*, *DERL3*, *CHCHD10*, *IGH* и *TMEM121*, чья роль в регуляции N-гликозилирования белков плазмы крови человека ранее не была известна.
3. Полученные результаты подтверждают вовлеченность генов метаболизма N-гликанов и их факторов транскрипции в регуляцию N-гликозилирования белков плазмы крови человека. Кроме того, результаты позволяют предположить участие в регуляции данного процесса генов деградации гликопротеинов и гомеостаза pH в аппарате Гольджи.
4. Реконструкция генной сети позволила предположить, что часть приоритизированных нами генов регулирует N-гликозилирование белков в печени, а часть - в клетках, секретирующих антитела.

## СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Sodbo Zh Sharapov**, Alexandra S Shadrina, Yakov A Tsepilov, Elizaveta E Elgaeva, Evgeny S Tiys, Sofya G Feoktistova, Olga O Zaytseva, Frano Vuckovic, Rafael Cuadrat, Susanne Jäger, Clemens Wittenbecher, Lennart C Karssen, Maria Timofeeva, Therese Tillin, Irena Trbojević-Akmačić, Tamara Štambuk, Najda Rudman, Jasminka Krištić, Jelena Šimunović, Ana Momčilović, Marija Vilaj, Julija Jurić, Anita Slana, Ivan Gudelj, Thomas Klarić, Livia Puljak, Andrea Skelin, Antonia Jeličić Kadić, Jan Van Zundert, Nishi Chaturvedi, Harry Campbell, Malcolm Dunlop, Susan M Farrington, Margaret Doherty, Concetta Dagostino, Christian Gieger, Massimo Allegri, Frances Williams, Matthias B Schulze, Gordan Lauc, Yurii S Aulchenko (2021). Replication of 15 loci involved in human plasma protein N-glycosylation in 4802 samples from four cohorts. *Glycobiology*, 31(2), 82–88. <https://doi.org/10.1093/glycob/cwaa053>; (IF=4.31, Wos, Scopus)
2. Lucija Klarić, Yakov A Tsepilov, Chloe M Stanton, Massimo Mangino, Timo Tõnis Sikka, Tõnu Esko, Eugene Pakhomov, Perttu Salo, Joris Deelen, Stuart J McGurnaghan, Toma Keser, Frano Vučković, Ivo Ugrina, Jasminka Krištić, Ivan Gudelj, Jerko Štambuk, Rosina Plomp, Maja Pučić-Baković, Tamara Pavić, Marija Vilaj, Irena Trbojević-Akmačić, Camilla Drake, Paula Dobrinić, Jelena Mlinarec, Barbara Jelušić, Anne Richmond, Maria Timofeeva, Alexander K Grishchenko, Julia Dmitrieva, Mairead L Bermingham, **Sodbo Zh Sharapov**, Susan M Farrington, Evropi Theodoratou, Hae-Won Uh, Marian Beekman, Eline P Slagboom, Edouard Louis, Michel Georges, Manfred Wuhler, Helen M Colhoun, Malcolm G Dunlop, Markus Perola, Krista Fischer, Ozren Polasek, Harry Campbell, Igor Rudan, James F Wilson, Vlatka Zoldoš, Veronique Vitart, Tim Spector, Yurii S Aulchenko, Gordan Lauc, Caroline Hayward. (2020). Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Science Advances*, 6(8), eaax0301. <https://doi.org/10.1126/sciadv.aax0301>, (IF=14.14, Wos, Scopus)
3. **Sodbo Zh Sharapov**, Yakov A Tsepilov, Lucija Klaric, Massimo Mangino, Gaurav Thareja, Alexandra S Shadrina, Mirna Simurina, Concetta Dagostino, Julia Dmitrieva, Marija Vilaj, Frano Vuckovic, Tamara Pavic, Jerko Stambuk, Irena Trbojevic-Akmacic, Jasminka Kristic, Jelena Simunovic, Ana Momcilovic, Harry Campbell, Margaret Doherty, Malcolm G Dunlop, Susan M Farrington, Maja Pucic-Bakovic, Christian Gieger, Massimo Allegri, Edouard Louis, Michel Georges, Karsten Suhre, Tim Spector, Frances MK Williams, Gordan Lauc, Yurii S Aulchenko (2019). Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Human Molecular Genetics*, 28(12), 2062–2077. <https://doi.org/10.1093/hmg/ddz054>, (IF=6.15, Wos, Scopus)
4. Xia Shen, Lucija Klarić, **Sodbo Sharapov**, Massimo Mangino, Zheng Ning, Di Wu, Irena Trbojević-Akmačić, Maja Pučić-Baković, Igor Rudan, Ozren Polašek, Caroline Hayward, Timothy D Spector, James F Wilson, Gordan Lauc, Yurii S Aulchenko (2017). Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nature Communications*, 8(1), 447. <https://doi.org/10.1038/s41467-017-00453-3>, (IF=14.92, Wos, Scopus)