

**федеральное государственное автономное образовательное учреждение
высшего образования
ПЕРВЫЙ МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ
УНИВЕРСИТЕТ имени И.М. СЕЧЕНОВА
Министерства здравоохранения Российской Федерации
(Сеченовский Университет)**

**Институт фармации им. А.П. Нелюбина
Кафедра биотехнологии**

Гориченко Иван Вадимович
студент 5 курса 09-01 группы

**Исследование генных сетей, ассоциированных с семейной
гиперхолестеринемией, с помощью онлайн-инструментов биоинформатики**

ДИПЛОМНАЯ РАБОТА
по направлению подготовки (специальности)
06.05.01 Биоинженерия и биоинформатика

**МОСКВА
2023**

Работа выполнена в ФГАОУ ВО Первый МГМУ имени И.М. Сеченова Минздрава России (Сеченовский Университет) на кафедре биотехнологии и кафедре информационных и интернет-технологий ИЦМ.

зав. кафедрой биотехнологии,
доктор биологических наук,
профессор

Луценко С.В.

Научный руководитель дипломной работы:

профессор РАН,
доктор биологических наук

Орлов Ю.Л.

С дипломной работой можно ознакомиться на кафедре биотехнологии Института фармации им. А.П. Нелюбина Сеченовского университета по адресу: пр-т Вернадского, д. 96, корп. 1.

Оглавление

ВВЕДЕНИЕ.....	5
Актуальность темы.....	5
Цель и задачи исследования.....	6
1. ОБЗОР ЛИТЕРАТУРЫ.....	8
1.1. Семейная гиперхолестеринемия.....	8
1.1.1. Основные факты.....	8
1.1.2. Обмен холестерина.....	11
1.2 Клиническая диагностика.....	17
1.3. Терапия.....	19
1.4. Биоинформационные методы исследования заболеваний.....	21
2. МАТЕРИАЛЫ И МЕТОДЫ.....	23
2.1. Материалы.....	23
2.1.1. База данных GeneCards.....	23
2.1.2. База данных KEGG.....	24
2.1.3. База данных Ensembl.....	24
2.2. Методы.....	25
2.2.1. Ресурс STRING-DB.....	25
2.2.2. Ресурс DAVID.....	26
2.2.3. Ресурс PANTHER.....	27
3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ.....	28
3.1. Получение списка генов, связанных с наследственной предрасположенностью к семейной гиперхолестеринемии.....	28
3.2. Расчет категорий генных онтологий.....	30
3.3. Реконструкция генных сетей для генов семейной гиперхолестеринемии.....	46
3.4. Детальный анализ генов.....	52

3.4.1 Ген LDLR.....	52
3.4.2 Ген APOA1.....	56
3.4.3 Ген PCSK9.....	57
ВЫВОДЫ.....	60
СПИСОК ЛИТЕРАТУРЫ.....	62
СПИСОК ТЕРМИНОВ.....	66
СПИСОК СОКРАЩЕНИЙ.....	66
ПРИЛОЖЕНИЯ.....	68
Приложение А.....	68
Приложение Б.....	74
Сертификат участника школы по биоинформатике.....	74

ВВЕДЕНИЕ

Актуальность темы

Генетические заболевания, такие как семейная гиперхолестеринемия, являются серьезной медицинской проблемой, требующей постоянного внимания и исследований.

Генетические заболевания – это большая проблема для медицины. В настоящее время существует множество генетических заболеваний, которые причиняют боль и страдания многим людям во всем мире. Одним из таких заболеваний является семейная гиперхолестеринемия, вероятность развития которого составляет 1 случай на 250 человек [1, 19]. Эта болезнь представляет угрозу для жизни и здоровья пациентов, вызывая сердечно-сосудистые заболевания и повышенный риск преждевременной смерти. Поэтому изучение генных сетей, связанных с этим заболеванием, является необходимостью для науки и медицины. Благодаря использованию биоинформатических инструментов и баз данных возможно получить более глубокое понимание механизмов заболевания и его проявлений у пациентов. Такое исследование сможет привести к более точной диагностике гиперхолестеринемии и более эффективным методам терапии. Выявление генов, ответственных за семейную гиперхолестеринемия, открывает новые перспективы для исследования и лечения многих других генетических заболеваний. Более глубокое изучение генных сетей может показать нам новые пути для развития медицинских технологий и создания новых лекарственных препаратов. Жизнь и здоровье людей должны быть главным приоритетом для науки и медицины. Поэтому исследование генных сетей, связанных с семейной гиперхолестеринемией, является неотъемлемой частью работы в этой области. Дальнейшие исследования в этом направлении могут дать нам надежду на более эффективные методы диагностики и лечения многих генетических заболеваний, что, в свою очередь, поможет улучшить качество жизни многих людей в мире.

С помощью биоинформатических инструментов возможно изучение генных сетей, ассоциированных с этим заболеванием, что может привести к усовершенствованию методов диагностики и терапии.

С использованием биоинформатических инструментов имеется возможность проведения детального исследования генных сетей, связанных с семейной гиперхолестеринемией. Это открывает дорогу к более точной диагностике, разработке эффективных методов терапии и позволяет разобраться в сложных молекулярных взаимоотношениях и регуляторных механизмах в организме. Кроме этого, картирование таких генных сетей не только улучшает понимание механизмов возникновения и развития заболевания, но также может способствовать выявлению связанных с ним генетических мутаций и полиморфизмов, что является важным шагом в борьбе с болезнью. Благодаря возможностям биоинформатики, широкому спектру доступных методов и баз данных, исследователи и врачи могут получить больше информации о генах, связанных с семейной гиперхолестеринемией, выявить новые гены и мишени для лечения и улучшения диагностики и предотвращения развития заболевания в будущем.

Цель и задачи исследования

Ввиду наследственного развития болезни медикаментозная терапия (наиболее популярная при СГХС) недостаточна для устранения заболевания.

В связи с этим актуальной задачей является определение генетических причин заболевания и генетического риска, а также рассмотрение причины осложнений с использованием инструментов биоинформатики.

Целями данной работы являются анализ современного состояния по исследованию семейной гиперхолестеринемии по международным литературным источникам, определение списка генов, ассоциированных с развитием заболевания по генетическим базам данных, компьютерная реконструкция и

визуализация генных сетей для анализа семейной гиперхолестеринемии с целью поиска перспективных генов-мишеней для диагностики и терапии.

В соответствии с заданной целью были сформированы следующие задачи:

- 1 Изучение литературных данных, связанных с семейной гиперхолестеринемией;
- 2 Построение списка генов, ассоциированных с семейной гиперхолестеринемией по интернет-доступным базам данных;
- 3 Составление таблиц представленности категорий генных онтологий сетей по полученному списку генов семейной гиперхолестеринемии на основе инструментов биоинформатики;
- 4 Компьютерная реконструкция генной сети генов семейной гиперхолестеринемии, ее визуализация, статистический анализ и сопоставление с генными сетями сопутствующих заболеваний;
- 5 Анализ полученных результатов, оценка связи полученных результатов с клиническими и литературными данными, обсуждение перспективных генов как мишеней для терапевтического воздействия;

1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Семейная гиперхолестеринемия

1.1.1. Основные факты

Семейная гиперхолестеринемия – это аутосомно-доминантно-наследственное генетическое заболевание, которое приводит к повышению уровня холестерина в крови. Семейная гиперхолестеринемия может проявляться в виде резко повышенного уровня общего холестерина и холестерина липопротеинов низкой плотности (ЛПНП) или в виде преждевременной ишемической болезни сердца (ИБС) [1].

Семейная гиперхолестеринемия, также известная как семейная гиперлиппротеинемия типа 2 или гиперлипидемия Фредриксона класса 2а, является аутосомно-доминантно-наследственным генетическим заболеванием, которое приводит к повышению уровня холестерина в крови. Как правило, пациент наследует только один из дефектных генов, являясь гетерозиготным. В редких случаях пациент наследует аномальный ген от обоих родителей, становясь гомозиготным; гомозиготность при семейной гиперхолестеринемии приводит к чрезвычайно высокому уровню холестерина в крови. Международный фонд семейной гиперхолестеринемии и Национальный институт здоровья и совершенствования медицинской помощи (бывший Национальный институт здоровья и клинического совершенствования) опубликовали рекомендации по выявлению, оценке и лечению семейной гиперхолестеринемии, чтобы помочь врачам, ответственным за лечение пациентов с семейной гиперхолестеринемией [2].

Семейная гиперхолестеринемия может проявляться в виде резко повышенного уровня общего холестерина и холестерина липопротеинов низкой плотности (ЛПНП) и липопротеинов очень низкой плотности (ЛПОНП) или в виде преждевременной ишемической болезни сердца (ИБС). По оценкам, 5%

инфарктов миокарда (ИМ) у пациентов моложе 60 лет и 20% ИМ у пациентов моложе 45 лет связаны с СГХС. У мужчин с семейной гиперхолестеринемией вероятность развития ишемической болезни сердца к 50 годам составляет 50%, а у женщин с семейной гиперхолестеринемией вероятность развития ишемической болезни сердца к 60 годам составляет 30%. Гомозиготная форма семейной гиперхолестеринемии, хотя и встречается редко, является особенно разрушительной. У пациентов с гомозиготной формой семейной гиперхолестеринемии атеросклероз развивается в детстве, а ишемическая болезнь сердца может проявиться в возрасте до 20 лет [4].

Согласно клиническим рекомендациям Минздрава РФ ("Клинические рекомендации "Семейная гиперхолестеринемия" (утв. Минздравом России)) распространенность гетерозиготной семейной гиперхолестеринемии в мире составляет 1 на 250 человек. По данным эпидемиологического исследования, проведенного в двух регионах Российской Федерации, распространенность семейной гиперхолестеринемии составляет 1 на 108 человек. Распространенность гомозиготной семейной гиперхолестеринемии - значительно меньше (1 на 300 тыс. - 1 млн. человек). СГХС является причиной развития ИМ до 45 лет в 20% случаев [4]. Пациенты с гомозиготной СГХС имеют повышенный риск развития преждевременной смерти. У мужчин, больных гетерозиготной СГХС, в случае отсутствия лечения ИБС развивается к 30 годам у 5,4%, к 50 годам - 51,4%, к 60 годам - 85,4%, а у женщин к 60 годам - у 53,3%. Согласно докладу ВОЗ, 50% мужчин с гетерозиготной СГХС умирают в возрасте до 60 лет из-за ИБС. В России продолжительность жизни у мужчин с гетерозиготной СГХС - 53 года, у женщин 62 года. У нелеченых пациентов с гомозиготной СГХС атеросклероз развивается в возрасте до 20 лет, и продолжительность жизни составляет не более 30 лет [4].

Гетерозиготы встречаются с частотой 1:500 человек, при этом у некоторых народов Африки – 1:100. Количество рецепторов ЛПНП на поверхности клеток у гетерозигот снижено вдвое, а концентрация ХС в плазме, соответственно, вдвое

повышается. У таких больных к 35-40-летнему возрасту концентрация ХС в крови достигает 400-500 мг/дл (при норме 200+-50 мг/дл) [31].

Вероятность развития гомозиготной СГХС составляет 1 случай на 1 000 000 человек. С раннего возраста у таких пациентов уровень ХС в крови увеличен в 5–6 раз. В таких случаях в организме макрофаги захватывают ЛПНП, и откладываются в коже и сухожилиях, образуя ксантомы. ХС откладывается на стенках артерий, образуя атеросклеротические бляшки. Высока вероятность гибели ребенка в 5-6 лет [31].

В связи с этим рутинный скрининг холестерина рекомендуется проводить в разном возрасте в зависимости от риска. Обнаружение семейной гиперхолестеринемии должно привести к каскадному тестированию для выявления других членов семьи с этим заболеванием.

Пациенты с выявленной семейной гиперхолестеринемией должны быть проконсультированы относительно терапевтического изменения образа жизни и должны принимать высокие дозы препаратов, снижающих синтез ХС печенью, статинов, для достижения 50% снижения уровня ЛПНП. Пациентам с тяжелой формой семейной гиперхолестеринемии может быть полезна дополнительная терапия другими препаратами, снижающими уровень холестерина, и аферез ЛПНП. Дети с гомозиготной семейной гиперхолестеринемией и чрезвычайно высоким уровнем ЛПНП могут быть кандидатами на ортотропную трансплантацию печени.

На рисунке ниже представлено соотношение сердечно-сосудистых заболеваний, встречающихся у пациентов с семейной гиперхолестеринемией.

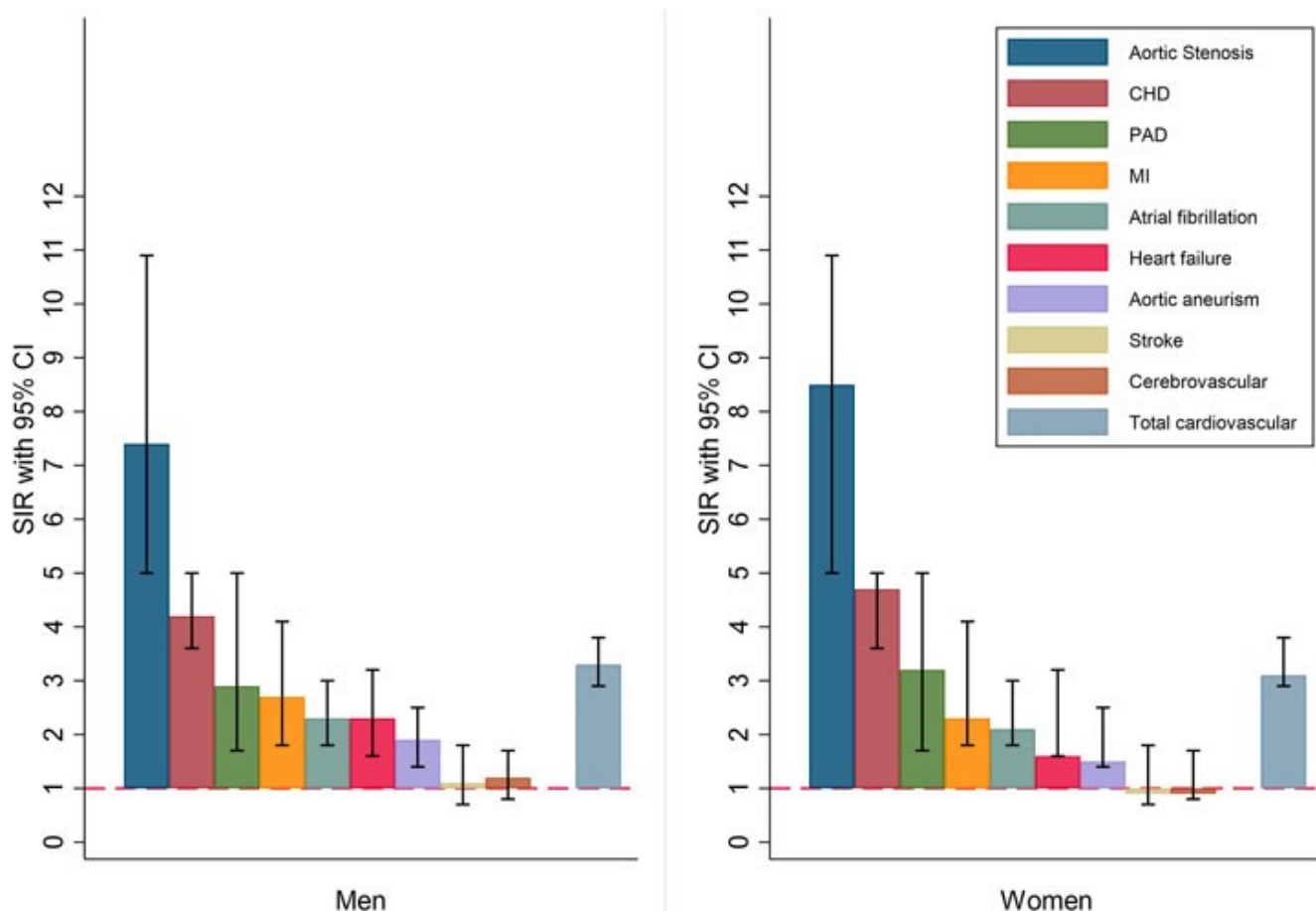


Рисунок 1. Заболевания, связанные с СГХС [7]

1.1.2. Обмен холестерина

Для определения генов, чьи мутации приводят к СГХС, и их анализа необходимо учитывать роль холестерина в органах и тканях организма. Для этого важно акцентировать внимание на реакциях, протекающих в клетках и межклеточном пространстве в обмене холестерина.

Холестерол – основной стероид организма человека. Его функции:

- 1 входит в состав большинства клеточных мембран.
- 2 компонент монослоя липидов на поверхности липопroteинов.
- 3 Является предшественником желчных кислот, стероидных гормонов, витамина D₃.

В полости 12-перстной кишки молекулы экзогенного ХС и его эфиров входят в состав больших липидных капель. Ассимиляция поступающего с пищей холестерина состоит из следующих этапов:

- 1 Эмульгирование жиров солями желчных кислот.
- 2 Гидролиз эфиров холестерина при участии фермента панкреатического сока – холестеролэстеразы.
- 3 Далее ХС и ЖК включаются в состав смешанных мицелл и проникают в энтероциты.
- 4 Синтез эфиров холестерина в энтероцитах.
- 5 ЭХС и небольшое количество свободного ХС включаются в состав незрелых хиломикронов (ХМн).
- 6 Через лимфу ХМн поступают в кровь, где происходит их созревание.
- 7 Зрелые ХМ подвергаются действию липопротеинлипазы (ЛП-липазы), в результате чего образуются остаточные хиломикроны (ХМост).
- 8 ХМост. взаимодействуют с ЛПНП-рецепторами печени и поглощаются печенью с помощью рецептор-опосредованного эндоцитоза. (Лигандами ЛПНП-рецепторов служат белки апо-В и апо-Е).
- 9 Компоненты ХМост. расщепляются лизосомальными ферментами. Освобождающийся ХС включается в общий фонд ХС в организме.

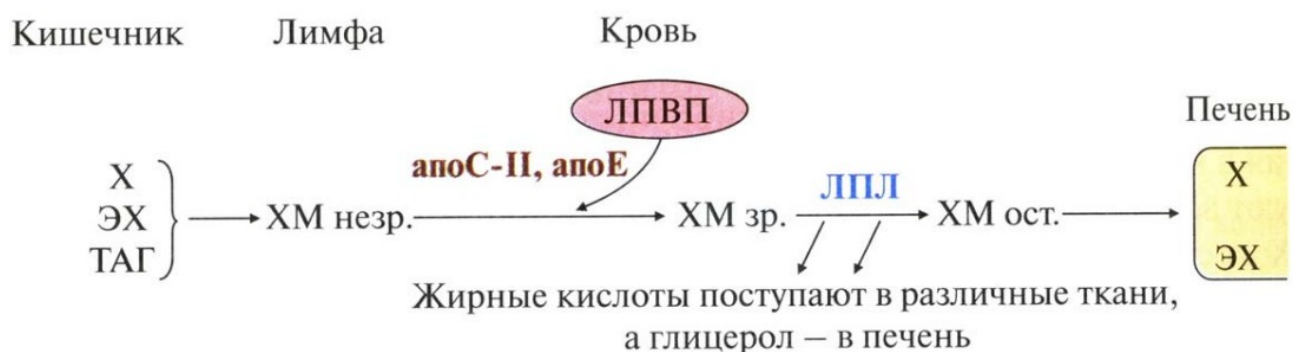


Рисунок 2. Анаболический путь ХС (Х — холестерин, ЭХ — эфиры холестерина, ТАГ — триацилглицерол, ЛПЛ — липопротеинлипаза, ХМ — хиломикрон)

Основной синтез холестерина происходит в печени (80%). ХС синтезируется из мевалоната, который в свою очередь синтезируется из ацетилкофермента-А. Синтез ХС протекает также в клетках кишечника и коре надпочечников. Синтез ХС происходит в абсорбтивный период, в цитозоле и эндоплазматическом ретикулуле клеток печени одновременно с синтезом жирных кислот и триацилглицеринов [32].

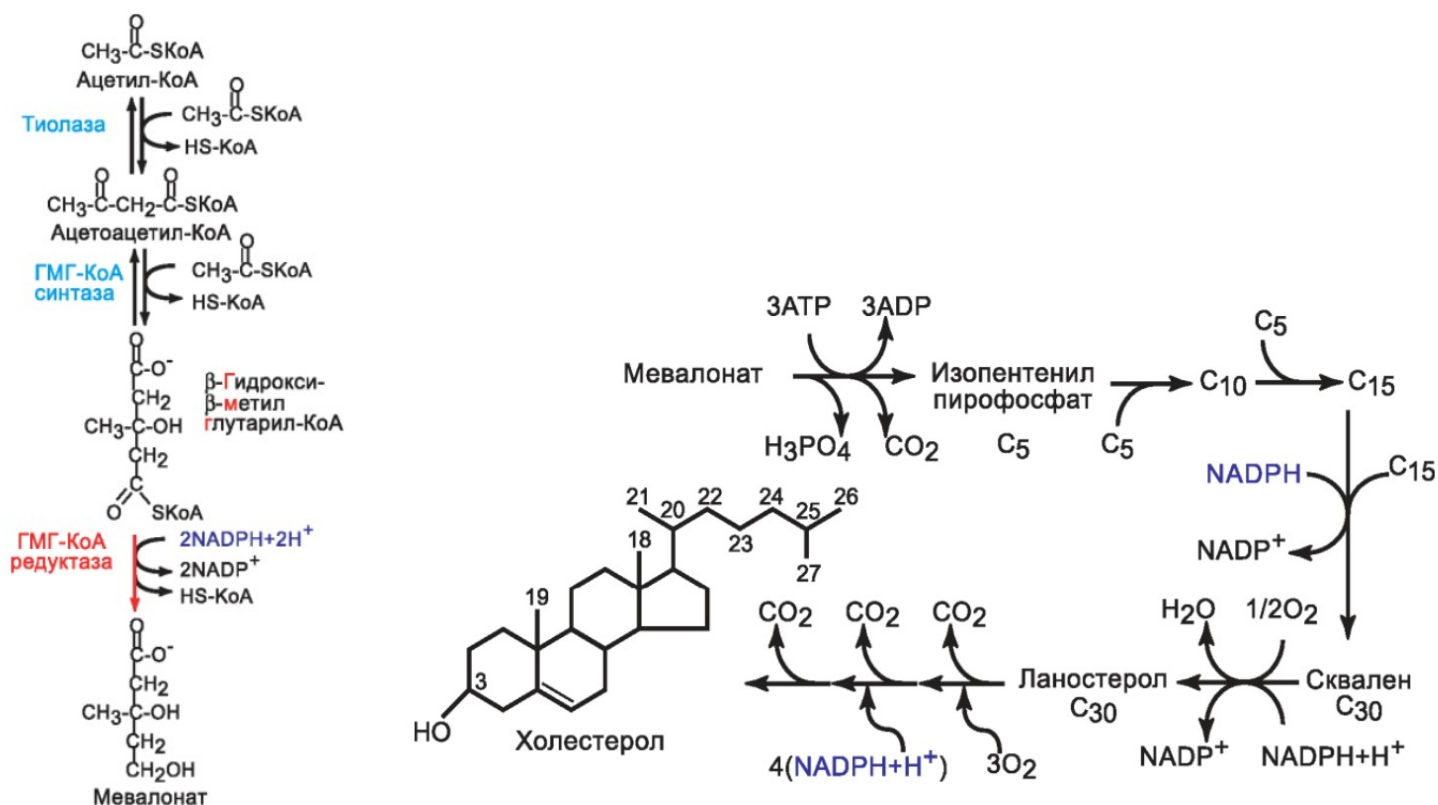


Рисунок 3. Синтез холестерина (C₁₀ — геранилпирофосфат, C₁₅ — фарнезилпирофосфат) [32]

Реакции входящие в синтез ХС имеют высокую роль, так как влияние на ферменты, катализирующие данные реакции, может привести к изменению уровня ХС во всем организме, что может быть использовано в терапии СГХС.

Метаболизм липопротеинов включает в себя синтез, транспортировку и поглощение липидов, таких как холестерин и триглицериды, в различные ткани и клетки организма. Этот процесс включает в себя несколько различных типов липопротеинов. Холестерин и триглицериды синтезируются в печени. Холестерин также может быть получен из пищи. Печень синтезирует холестерин из ацетата,

который образуется из цикла лимонной кислоты, и формирует комплекс с аполипопротеинами. Триглицериды синтезируются из углеводов, белков и жиров.

После синтеза холестерина и триглицериды собираются в липопротеины, которые состоят из ядра липидов, включая эфиры холестерина и триглицериды, окруженного оболочкой из белков, фосфолипидов и неэстерифицированного холестерина. Липопротеины различаются по размеру, составу и функциям. Различные типы липопротеинов включают хиломикроны, липопротеины очень низкой плотности (ЛПОНП), липопротеины промежуточной плотности (ЛППП), липопротеины низкой плотности (ЛПНП) и липопротеины высокой плотности (ЛПВП). Большая часть ХС и ЭХС удаляется из печени в ЛПОНП. ЛПОНП содержат аполипопротеин-B100.

Липопротеины транспортируются в крови. Хиломикроны переносят диетические липиды в печень и периферические ткани. ЛПОНП переносят эндогенные липиды из печени в периферические ткани, включая жировую ткань и мышцы. ЛПОНП в крови созревают, подвергаются действию липопротеин-липазы (ЛП-липаза) и превращаются сначала в ЛППП, а затем в ЛПНП. ЛПНП – основная транспортная форма холестерина.

ЛПНП состоят из частично израсходованных ЛПВП и переносят липиды из печени в остальные органы. ЛПВП забирают избыток холестерина из периферических тканей и возвращают его в печень для выведения.

Когда липопротеины достигают тканей-мишеней, они поглощаются клетками через рецепторы на поверхности клеток. ЛПНП поглощается рецептором ЛПНП, а ЛППП – либо рецептором ЛПНП, либо рецептором ЛППП. ЛПНП-рецепторы взаимодействуют своим N-концевым гидрофобным доменом с белками апо-В (В48 и В100) и апо-Е на поверхности липопротеинов (в основном: ЛПНП). Хиломикроны и ЛПОНП поглощаются тканями, в которые они доставляют липиды.

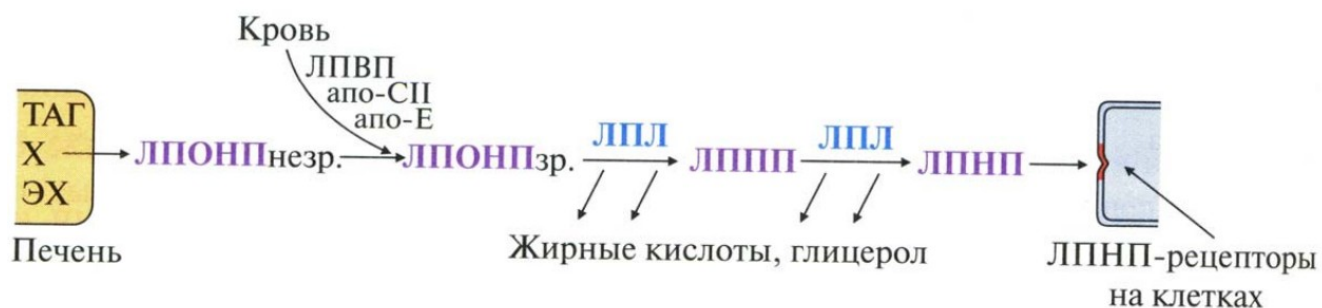


Рисунок 4. Цикл липопротеинов [32]

ЛПВП синтезируется в печени и кишечнике. На печень приходится большая часть производства ЛПВП. Вновь синтезированные частицы ЛПВП имеют небольшие размеры и содержат мало холестерина. Они взаимодействуют с периферическими тканями, включая жировую ткань и макрофаги, и получают свободный холестерин и фосфолипиды через АТФ-связывающие кассетные транспортеры. Это увеличивает размер и содержание холестерина в частицах ЛПВП. Созревая, частицы ЛПВП становятся более крупными, сферическими и богатыми липидами. Этот процесс называется созреванием.

Частицы ЛПВП способны удалять избыток холестерина из периферических тканей, включая макрофаги артериальной стенки. Этот процесс называется обратным транспортом холестерина. После доставки холестерина в печень из системной циркуляции он выводится в желчь или используется для биосинтеза стероидных гормонов.

Частицы ЛПВП обладают плейотропными эффектами, выходящими за рамки липидного обмена. К ним относятся противовоспалительные, антитромботические и антиоксидантные эффекты.

выводится из организма с продуктами жизнедеятельности, в основном, в виде желчных кислот.

1.2 Клиническая диагностика

Многие люди с семейной гиперхолестеринемией могут не иметь явных симптомов до проявления ишемической болезни сердца, инсульта или заболеваний периферических артерий. СГХС может вызвать появление характерных жировых отложений вокруг локтей, коленей, вдоль сухожилий и вокруг роговицы глаз. Эти жировые отложения называются ксантомами.



Рисунок 7. Ксантомы

Отложения холестерина на веках, называемые ксантелазмами, также распространены.



Рисунок 8. Ксантелазмы

Основными симптомами и признаками семейной гиперхолестеринемии являются [42]:

- Высокий уровень общего холестерина и холестерина ЛПНП.
уровень ХС в норме: 200 ± 50 мг/дл (5,2-1,2 ммоль/л)
уровень ХС при СГХС: 400-500 мг/дл (гетерозигота 35-40 лет)
- Сильный семейный анамнез высокого уровня общего холестерина и холестерина ЛПНП и/или раннего сердечного приступа.
- Повышенные и устойчивые к терапии уровни ЛПНП у одного или обоих родителей.
- Ксантомы (восковые отложения холестерина в коже или сухожилиях).
- Ксантелазма (отложения холестерина в веках).
- Дуга роговицы (отложение холестерина в уголке глаза).
- Стенокардия

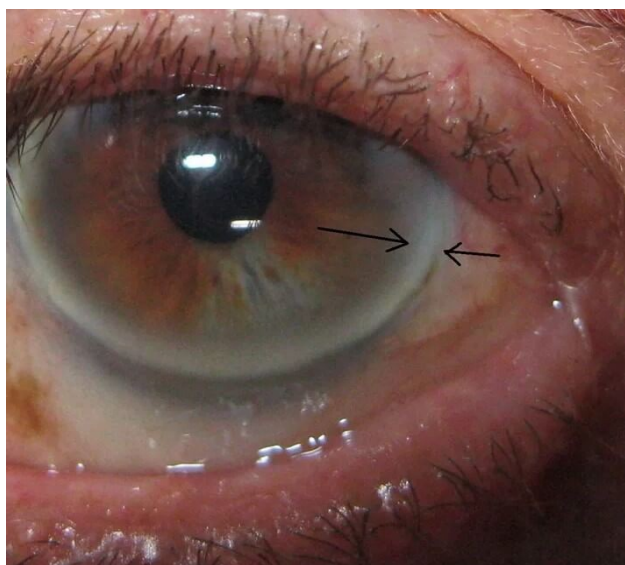


Рисунок 9. Дуга роговицы



Рисунок 10. Уплотнение сухожилий

При гомозиготной СГХС ксантомы развиваются в младенческом возрасте. Сердечные приступы и смерть могут наступить до 30 лет [4].

1.3. Терапия

Для снижения риска сердечно-сосудистых событий рекомендуется раннее выявление и лечение СГХС. Целью лечения является снижение уровня ЛПНП на 50% от исходного уровня. Это было основано на результатах двух исследований, которые показали снижение интимальной медиальной толщины сонной артерии у пациентов, достигших снижения уровня ЛПНП на 50% на терапии статинами. Изменение образа жизни, липидоснижающая терапия, аферез ЛПНП и в редких случаях трансплантация печени являются краеугольными камнями снижения уровня ЛПНП у пациентов с СГХС.

Важно, чтобы врач знал о различных вариантах лечения СГХС, поскольку для достижения целевого снижения уровня ЛПНП потребуется комбинированная терапия. По данным исследователей, до 77% пациентов с диагнозом СГХС, принимающих лекарства от холестерина, не достигают целевого уровня ЛПНП в течение двух лет после постановки диагноза. Наиболее распространенной причиной неоптимального контроля уровня холестерина у пациентов (32%) было "согласие" лечащего врача с неоптимальным уровнем ЛПНП. Другие причины неоптимального контроля включают неправильное дозирование лекарств и поздний возраст постановки диагноза.

Модификация образа жизни является основой лечения СГХС. Первичная профилактика, включающая консультирование по вопросам курения, избегания диабета и регулярной физической активности, является неотъемлемым компонентом лечения. Также рекомендуется направление к диетологам и диетологам соответствующей квалификации. Поощряется ежедневная физическая активность, а для пациентов с избыточным весом следует подчеркнуть необходимость достижения оптимального веса. На каждые 10 кг потерянного веса приходится 8 мг/дл снижения ЛПНП. Хотя модификация образа жизни является первым шагом в лечении пациентов с СГХС, одной этой стратегии будет недостаточно для достижения цели по снижению уровня ЛПНП. Модификация

образа жизни максимум снизит концентрацию ЛПНП на 10-15%, и большинству пациентов потребуется дальнейшая фармакологическая терапия.

В фармакологической терапии СГХС используют несколько групп лекарственных препаратов. К первой группе относятся статины. Статины ингибируют гидроксиметилглутарилкоэнзим-А-редуктазу (ГМГ-КоА-редуктаза), ключевой фермент в синтезе холестерина в печени, в конечном итоге снижая выработку ЛПНП (Рисунок 3. Синтез холестерина).

Статины также приводят к повышению регуляции рецепторов ЛПНП, однако этот эффект ограничен из-за усиленной компенсаторной деградации таких рецепторов. У пациентов с СГХС эффективность статинов снижается, поскольку механизм их действия связан с повышением уровня рецепторов ЛПНП в печени. Это касается также эзетимиба и секвестрантов желчных кислот.

Помимо статинов часто используется эзетимиб (Ezetimibe). Эзетимиб – это селективный ингибитор абсорбции холестерина, который блокирует всасывание пищевого холестерина и доставку кишечного холестерина в печень, что приводит к понижению уровня ХС в организме. Статины и эзетимиб, принимаемые вместе, усиливают снижение уровня ЛПНП. Документально подтверждено, что снижение уровня ЛПНП составляет 10–40 процентов [28, 29, 30].

К другой группе препаратов от СГХС относятся секвестранты желчных кислот. Они связывают холестерин в кишечнике, тем самым уменьшая абсорбцию ХС в кишечнике, усиливая его выведение и снижая энтерогепатический оборот желчных кислот и холестерина. Уменьшение количества желчных кислот приводит к компенсаторному увеличению конверсии холестерина в желчные кислоты, повышая экспрессию печеночных рецепторов ЛПНП и, следовательно, увеличивая печеночное удаление частиц ЛПНП из циркуляции. При СГХС секвестранты желчных кислот считаются препаратами второй линии терапии после лечения статинами и эзетимибом.

Несколько лет назад в практике начали использовать ниацин, мипомерсен (ингибитор синтеза аполипопротеина-В), ломитапид (ингибитор микросомального

белка-переносчика) и ингибиторы PCSK9 (Пропротеиновая конвертаза субтилизин-кексинового типа 9).

1.4. Биоинформационные методы исследования заболеваний

Биоинформационные методы широко используются для анализа генов предрасположенности к развитию заболеваний. К ним относят базы данных медицинской генетики и онлайн инструменты биоинформатики: инструменты анализа последовательностей макромолекул, инструменты для прогнозирования и оценки структуры молекул, инструменты аннотации генов, банки данных по белкам, нуклеиновым кислотам.

Международным научным сообществом собраны и продолжают дополняться открытые базы данных по общей аннотации генома и транскриптома человека, например: Human Genome Resources at NCBI (<https://www.ncbi.nlm.nih.gov/genome/guide/human>), Ensembl (<http://ensembl.org>), справочная база данных белков человека HPRD (<http://hprd.org>) и PDB (<https://www.rcsb.org>), биохимических реакций KEGG (<http://www.genome.jp/kegg>). Базу данных генов и фенотипов заболеваний OMIM (Online Mendelian Inheritance in Man, <https://omim.org/>) используется для получения базового списка генов человека, ассоциированных с заболеванием [4].

Для подробной визуализации генных зависимостей будут реконструированы модели генных сетей для генов, ассоциированных с СГХС.

Генные сети - это набор взаимосвязанных генов и белков, которые работают вместе для регулирования различных биологических процессов в клетке или ткани. Эти сети дают полное представление о взаимодействии между генами и их продуктами, а также о том, как они работают вместе для выполнения определенных функций в физиологических или патологических процессах.

Генные сети состоят из множества генов и белков, которые связаны через сложные регуляторные пути, включающие различные типы молекулярных взаимодействий, такие как транскрипционная регуляция, посттранскрипционная

регуляция и белок-белковые взаимодействия. Транскрипционная регуляция - это контроль экспрессии генов путем изменения скорости транскрипции гена в РНК. Это может быть достигнуто путем связывания факторов транскрипции или других регуляторных белков с определенными участками ДНК для активации или репрессии транскрипции. Посттранскрипционная регуляция - это контроль экспрессии генов путем изменения обработки или стабильности молекул РНК. Сюда входят сплайсинг РНК, альтернативное полиаденилирование, редактирование РНК и деградация РНК. Белково-белковые взаимодействия означают связывание белков друг с другом для формирования комплексов, которые могут регулировать экспрессию генов или выполнять другие функции в клетке. Понимание генных сетей важно для понимания глубинных механизмов регуляции генов и выяснения молекулярной основы сложных заболеваний. В последние годы анализ генных сетей стал мощным инструментом для выявления связанных с заболеваниями генов и путей, предсказания мишеней для лекарств и определения потенциальных кандидатов в лекарственные препараты для разработки новых методов лечения. Для анализа генных сетей обычно используются математические алгоритмы и компьютерные модели. Эти инструменты объединяют различные биомедицинские данные, включая транскрипционные и протеомные данные, для обеспечения всестороннего обзора клеточных функций и выявления важных регуляторных путей и взаимодействий генов. Генные сети - это сложные системы взаимодействий между генами и белками, которые регулируют различные биологические процессы и играют важную роль в здоровье и болезни. Понимание этих сетей требует интеграции многочисленных биомедицинских данных и использования вычислительных методов[3,5].

2. МАТЕРИАЛЫ И МЕТОДЫ

2.1. Материалы

2.1.1. База данных GeneCards

С помощью базы данных GeneCards (<https://www.genecards.org>) получен список генов, играющих ключевую роль в развитии заболевания [12]. GeneCards — база данных, предоставляющая информацию о генах и их функциях у разных организмов. Этот ресурс содержит данные о более чем 73 000 генах, которые могут быть связаны с различными медицинскими состояниями. GeneCards также предлагает информацию о последовательностях генома и их экспрессии в разных тканях, что может быть полезным для идентификации потенциальных мишеней для новых лекарственных средств. Одной из главных функций GeneCards является графический интерфейс, который позволяет исследователям просматривать информацию о генах, включая их определение, функцию, экспрессию и ассоциированные с ними заболевания. Кроме того, на ресурсе имеется возможность искать любую информацию, используя разные подходы, такие как поиск по названию гена, поиск по медицинскому состоянию или поиска по SNP-маркерам (Single-nucleotide polymorphism). Этот инструмент также позволяет создавать интерактивные диаграммы, которые показывают связи между разными генами. Одно из главных преимуществ GeneCards заключается в том, что он предоставляет доступ к множеству веб-ресурсов, таких как PubMed, UniProt и Ensembl. Это позволяет исследователям быстро получать доступ к информации, которая усиливает их исследования и расширяет их знания о конкретных генах и функциях организма. GeneCards также предоставляет информацию о смежных генах, что может быть полезным для исследования механизмов их взаимодействия [9–11].

2.1.2. База данных KEGG

KEGG (Kyoto Encyclopedia of Genes and Genomes, <https://www.genome.jp/kegg>) – это база данных генетических путей и метаболических карт, которая широко используется в биоинформатике. KEGG собирает информацию о генных функциях, взаимодействиях между молекулами, а также о различных путях метаболизма в различных организмах. Эта база данных содержит пути метаболизма, пути сигнализации, пути болезней и связи между ними [39].

2.1.3. База данных Ensembl

Ensembl (<https://www.ensembl.org>) - база данных, созданная и поддерживаемая совместными усилиями Европейского биоинформатического института (EMBL-EBI) и Университета Кембриджа. Она содержит геномные данные более чем 70 организмов, включая человека, мышей, рыб, насекомых и других животных. Ensembl предоставляет исчерпывающую информацию о расположении генов, альтернативных сплайс-вариантов, РНК и белков. Кроме того, база данных позволяет сравнивать геномы различных организмов, выделять консервативные и изменчивые области генома, а также отслеживать эволюционные изменения в геномах. База данных Ensembl содержит информацию о генах из HGNC (<https://www.genenames.org>), что позволяет обеспечить единый системный подход к аннотации генов. Кроме того, Ensembl интегрирован с GeneCards, что позволяет получить информацию о функциях и связях генов с заболеваниями, а также с DrugBank (<https://go.drugbank.com>), чтобы получить информацию о лекарствах, которые могут влиять на гены. Ensembl также интегрирован с KEGG, что позволяет смотреть на гены, связанные с болезнью, с точки зрения метаболических путей [40].

2.2. Методы

2.2.1. Ресурс STRING-DB

Для реконструкции генной сети взаимодействий генов СГХС использовался ресурс STRING-DB (<https://string-db.org>). STRING-DB - биоинформатический инструмент, позволяет проводить анализ взаимодействий между белками и генными сетями. Он представляет собой базу данных, которая содержит множество информации о белках, их взаимодействиях, функциях и патологических свойствах. STRING обладает набором функций, которые упрощают исследование генных сетей, позволяет искать белки с помощью различных критериев, включая название, идентификаторы и их гены, а также проводить анализ взаимодействий между белками. STRING также предоставляет информацию о молекулярных путях, которые включают взаимодействовавшие белки, и может оценивать силу связи между ними на основе различных критериев, таких как функциональная связь, физическое взаимодействие, совместное включение в определенный путевой анализ, а также сходство между паттернами экспрессии генов. Инструмент даёт возможность проводить анализы на различных уровнях сложности: от поиска взаимодействий между отдельными белками до анализа комплексных генных сетей. Также у него есть функция анализа функциональной связи между различными белками в генной сети, что позволяет идентифицировать ключевые белки, которые играют важную роль в регуляции метаболизма и других жизненно важных функций. Кроме того, STRING позволяет интегрировать данные из различных источников, таких как базы данных о генах, экспериментальные данные и данные об экспрессии генов, что позволяет получить более полное представление о функциях и взаимодействиях белков. Использование STRING позволяет значительно

упростить анализ генных сетей и выявление ключевых факторов, которые участвуют в различных биологических процессах [41].

2.2.2. Ресурс DAVID

Для анализа категорий генных онтологий использовался ресурс DAVID (Database for Annotation, Visualization and Integrated Discovery) (<https://david.ncifcrf.gov/summary.jsp>). DAVID – это бесплатный онлайн-инструмент для анализа функций генов, который объединяет широкий спектр биологических источников для выведения биологической информации. DAVID предоставляет интерфейс для исследования с использованием функционального аннотирования генов, а также с помощью анализа путей, генных функций и показателей биологических процессов. Основными функциями DAVID являются генерация групп генов на основе кластеризации по сходству экспрессии генов, а также обнаружение функциональных и тематических групп генов на основе аннотации генов. DAVID также поддерживает анализ списков генов, создание графических отчетов для генных анализов, и анализ дифференциально экспрессируемых генов. Основное преимущество DAVID заключается в том, что он позволяет одновременно использовать несколько биологических баз данных и сравнивать их результаты анализа. Базы данных, используемые в DAVID, включают Gene Ontology (<http://geneontology.org>), KEGG (<https://www.genome.jp/kegg>), Panther (<http://www.pantherdb.org>), Reactome (<https://reactome.org>), COG (<https://www.ncbi.nlm.nih.gov/research/cog-project>), и InterPro (<https://www.ebi.ac.uk/interpro>). Таким образом, DAVID предоставляет более широкий набор функций, чем большинство других биоинформатических инструментов. DAVID является мощным инструментом для анализа функций генов, позволяющим выявлять сходства и различия между группами генов, а также предоставляющим полные функциональные аннотации для отдельных генов. Он также позволяет создавать интуитивно понятные графические отчеты для облегчения визуализации результатов анализа. Рекомендуется использовать

DAVID вместе с другими биоинформатическими инструментами, чтобы получить максимальную пользу от анализа генных данных [38].

2.2.3. Ресурс PANTHER

Аналогично с DAVID, в целях получения более точного результата, для анализа категорий генных онтологий также использовался ресурс PANTHER. Система классификации PANTHER (Protein ANalysis THrough Evolutionary Relationships, <http://www.pantherdb.org>) была разработана для классификации белков и их генов с целью облегчения высокопроизводительного анализа. Основой PANTHER является всеобъемлющая аннотированная библиотека филогенетических деревьев семейств генов. Все узлы дерева имеют постоянные идентификаторы, которые сохраняются между версиями PANTHER, обеспечивая стабильную основу для аннотаций свойств белка, таких как подсемейство и функция. Каждое филогенетическое дерево используется для аннотирования каждого белка, входящего в семейство, по его: семейству и классу, подсемейству, ортологам, паралогам, функциям (с использованием терминов GO, аннотированных на деревьях проектом GO Phylogenetic Annotation Project, https://wiki.geneontology.org/Phylogenetic_Annotation_Project), путям (курируемые PANTHER и Reactome). PANTHER предоставляет информацию о классификации белок-кодирующих генов из 143 геномов, включенных в деревья PANTHER [37].

3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

3.1. Получение списка генов, связанных с наследственной предрасположенностью к семейной гиперхолестеринемии.

При анализе генов, мутации в которых имеют влияние на развитие семейной гиперхолестеринемии, информационная база данных «GeneCards» выдала 1,620 генов. На рисунке №11 представлено распределение генов, у которых показатель релевантности более 35, полный список генов прикреплен в приложении А. Показатель релевантности отображает связь гена и интересующего нас заболевания, он высчитывается с использованием алгоритмов и моделей информационного поиска [13,14, 15].

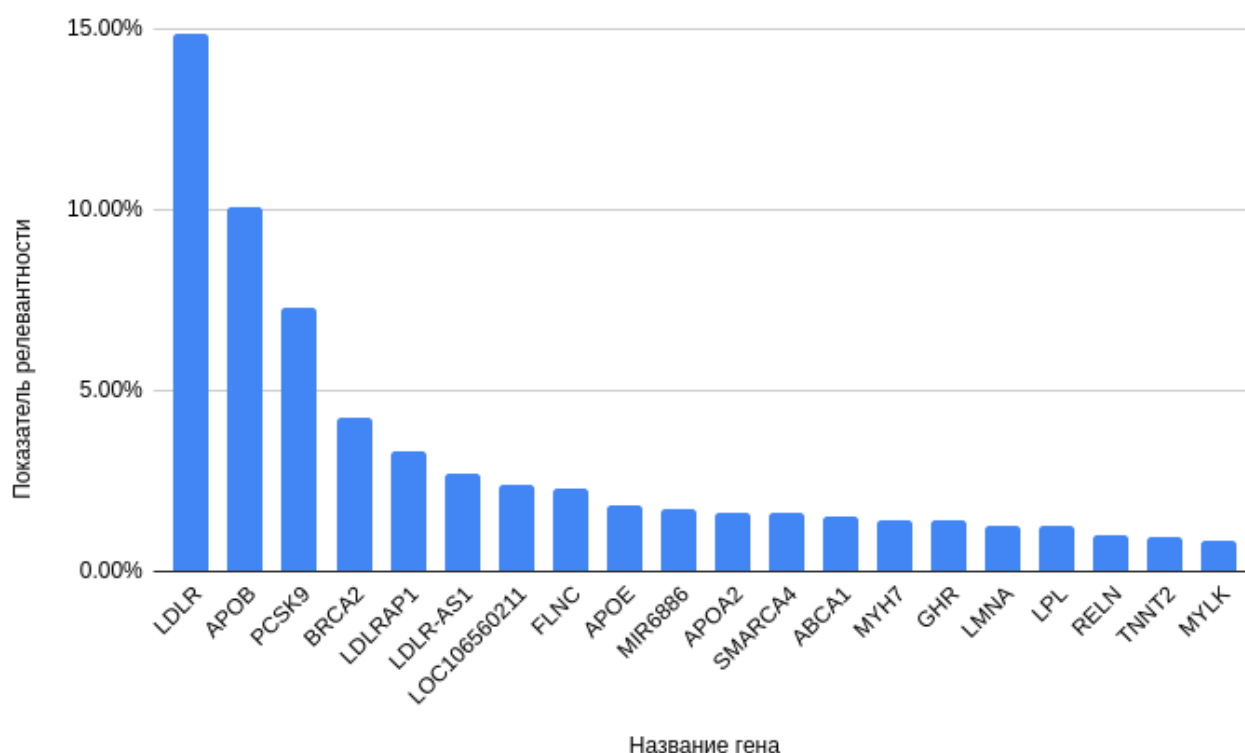


Рисунок 11. Распределение наиболее часто мутлирующих генов.

Ниже в таблице приведены наиболее часто мутлирующие при СГХС гены, их описание и показатели актуальности (условные единицы, показывающие степень связи заболевания с геном).

Таблица 1. Гены ассоциированные с семейной гиперхолестеринемией

Ген	Описание	Категория	Показатель релевантнос ти
LDLR	Low Density Lipoprotein Receptor	Protein Coding	235.00
APOB	Apolipoprotein B	Protein Coding	216.31
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9	Protein Coding	149.20
BRCA2	BRCA2 DNA Repair Associated	Protein Coding	109.99
LDLRAP1	Low Density Lipoprotein Receptor Adaptor Protein 1	Protein Coding	94.77
LDLR-AS1	LDLR Antisense RNA 1	RNA Gene	70.18
LOC106560211	APOB 5' Regulatory Region	Biological Region	56.96
FLNC	Filamin C	Protein Coding	53.78
APOE	Apolipoprotein E	Protein Coding	50.09
MIR6886	MicroRNA 6886	RNA Gene	50.03
APOA2	Apolipoprotein A2	Protein Coding	47.78
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4	Protein Coding	41.43
ABCA1	ATP Binding Cassette Subfamily A Member 1	Protein Coding	40.82
MYH7	Myosin Heavy Chain 7	Protein Coding	39.13
GHR	Growth Hormone Receptor	Protein Coding	39.09
LMNA	Lamin A/C	Protein Coding	39.03
LPL	Lipoprotein Lipase	Protein Coding	38.92
RELN	Reelin	Protein Coding	38.19
TNNT2	Troponin T2, Cardiac Type	Protein Coding	37.77
MYLK	Myosin Light Chain Kinase	Protein Coding	37.45

Интернет-ресурс OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) ассоциировал с семейной гиперхолестеринемией 1,784 ген, при этом явных отличий в наиболее ассоциированных генах не присутствовало.

3.2. Расчет категорий генных онтологий

Список 150 генов человека был загружен через интерфейс Gene Ontology для поиска значимых категорий генных онтологий для этой группы генов. В качестве референсного генома выбран геном человека. Из списка генов все идентификаторы были распознаны. Всего в референсном геноме по использовалось 20589 генов. Применена коррекция Бенжамини [8]. Было выполнено три аналогичных анализа для разных групп генных онтологий с ограничением значения с P-Value до E-06 для биологических процессов, E-04 для молекулярных функций и E-05 для клеточных компартментов для представления наиболее информативных результатов. Построены таблицы категорий генных онтологий (табл. 2-4).

Были выбраны только 3 основные генные онтологии (Gene Ontology):

- GO biological process - биологические процессы;
- GO cellular component - клеточные компартменты.
- GO molecular function - молекулярные функции;

Для анализа категорий генных онтологий использовались ресурсы DAVID. Список 1,620 генов человека (использовался список генов из «GeneCards») был загружен через интерфейс DAVID для поиска значимых категорий генных онтологий для этой группы генов. Было распознано 1584 идентификатора (Таблица №3.2.1). Применена коррекция Бенжамини [8]. Было выполнено четыре аналогичных анализа для разных групп генных онтологий с ограничением значения с P-Value до E-15 – для общей аннотации, E-12 – для биологических процессов, E-15 – для клеточных компартментов и E-6 – для молекулярных функций для представления наиболее информативных результатов.

Таблица 2. Категории генных онтологий для генов СГХС по DAVID

Группа	Категория	Кол-во генов	P - Value	P-Значение по Бенжамини
GOTERM_CC_DIRECT	extracellular space	283	4,70E-52	3,50E-49
GOTERM_BP_DIRECT	response to drug	115	1,50E-43	9,00E-40
GOTERM_BP_DIRECT	cholesterol homeostasis	49	1,80E-37	5,50E-34
GOTERM_CC_DIRECT	extracellular region	285	2,20E-37	8,30E-35
GOTERM_BP_DIRECT	cholesterol metabolic process	50	8,10E-37	1,70E-33
GOTERM_BP_DIRECT	response to hypoxia	78	4,90E-36	7,50E-33
GOTERM_MF_DIRECT	protein binding	975	9,40E-32	1,70E-28
GOTERM_BP_DIRECT	positive regulation of transcription from RNA polymerase II promoter	192	7,10E-27	8,70E-24
GOTERM_BP_DIRECT	cilium assembly	57	8,80E-27	9,00E-24
GOTERM_BP_DIRECT	inflammatory response	103	1,80E-25	1,60E-22
GOTERM_BP_DIRECT	positive regulation of cell proliferation	116	4,60E-25	3,30E-22
GOTERM_BP_DIRECT	aging	64	4,80E-25	3,30E-22
GOTERM_MF_DIRECT	enzyme binding	93	1,40E-24	1,30E-21
GOTERM_BP_DIRECT	cilium morphogenesis	57	2,50E-24	1,60E-21
GOTERM_CC_DIRECT	primary cilium	42	3,20E-24	6,00E-22
GOTERM_CC_DIRECT	extracellular exosome	378	3,40E-24	6,00E-22
GOTERM_CC_DIRECT	cytosol	428	4,00E-24	6,00E-22
GOTERM_BP_DIRECT	platelet degranulation	49	6,80E-24	3,80E-21
GOTERM_BP_DIRECT	negative regulation of apoptotic process	112	8,80E-24	4,50E-21
GOTERM_CC_DIRECT	cell surface	118	1,30E-22	1,70E-20
GOTERM_CC_DIRECT	cilium	56	3,20E-22	3,40E-20

Группа	Категория	Кол-во генов	P - Value	P-Значение по Бенжамини
GOTERM_CC_DIRECT	plasma membrane	495	2,00E-21	1,90E-19
GOTERM_CC_DIRECT	ciliary tip	29	4,60E-20	3,90E-18
GOTERM_CC_DIRECT	platelet alpha granule lumen	32	7,60E-20	5,80E-18
GOTERM_BP_DIRECT	response to lipopolysaccharide	57	9,80E-20	4,70E-17
GOTERM_BP_DIRECT	lipid metabolic process	55	3,00E-19	1,30E-16
GOTERM_BP_DIRECT	lipoprotein metabolic process	27	3,30E-19	1,30E-16
GOTERM_MF_DIRECT	protein homodimerization activity	138	9,40E-19	5,70E-16
GOTERM_BP_DIRECT	positive regulation of smooth muscle cell proliferation	33	1,30E-18	4,90E-16
GOTERM_CC_DIRECT	membrane raft	61	1,30E-18	9,30E-17
GOTERM_BP_DIRECT	triglyceride homeostasis	22	2,90E-18	1,10E-15
GOTERM_BP_DIRECT	response to nutrient	36	4,70E-18	1,60E-15
GOTERM_BP_DIRECT	response to ethanol	43	4,90E-18	1,60E-15
GOTERM_BP_DIRECT	positive regulation of ERK1 and ERK2 cascade	56	1,50E-17	4,50E-15
GOTERM_CC_DIRECT	membrane	292	1,60E-17	1,00E-15
GOTERM_BP_DIRECT	positive regulation of angiogenesis	44	3,80E-17	1,10E-14
GOTERM_BP_DIRECT	photoreceptor cell maintenance	24	5,60E-17	1,60E-14
GOTERM_CC_DIRECT	ciliary basal body	39	7,40E-17	4,30E-15
GOTERM_BP_DIRECT	leukocyte migration	45	8,00E-17	2,10E-14
GOTERM_BP_DIRECT	triglyceride metabolic process	24	1,50E-16	3,90E-14
GOTERM_CC_DIRECT	external side of plasma membrane	59	1,50E-16	8,20E-15
GOTERM_CC_DIRECT	high-density lipoprotein particle	19	1,70E-16	8,40E-15

Как видно из таблицы, наиболее значимыми категориями для генов СГХС являются категории онтологий внеклеточного пространства, гомеостаза и метаболических путей холестерина (cholesterol homeostasis, cholesterol metabolic process), а также белковое связывание (protein binding).

Анализируя полученные категории, видно, что большую роль в развитии заболевания играет нарушение процессов обмена липидами, в особенности их транспорт.

Таблица 3. Категории генных онтологий для генов СГХС (биологические процессы)

Категории генных онтологий для молекулярных функций	Количество генов	P - Value
triglyceride homeostasis	79	1.00E-12
response to toxic substance	68	1.23E-12
regulation of cytosolic calcium ion concentration	79	1.35E-12
regulation of protein kinase B signaling	55	1.43E-12
positive regulation of endocytosis	44	1.55E-12
positive regulation of catabolic process	102	1.64E-12
regulation of cell development	102	1.79E-12
phospholipid metabolic process	84	1.81E-12
response to light stimulus	78	1.84E-12
regulation of lipid storage	33	1.95E-12
multi-multicellular organism process	62	2.09E-12
camera-type eye development	78	2.13E-12
cell death	159	2.15E-12
positive regulation of lipid transport	40	2.16E-12
negative regulation of cellular protein metabolic process	157	2.27E-12
epithelial cell differentiation	106	2.52E-12
transmembrane receptor protein tyrosine kinase signaling pathway	90	2.73E-12

Категории генных онтологий для молекулярных функций	Количество генов	P - Value
regulation of developmental growth	29	2.75E-12
acylglycerol homeostasis	29	2.75E-12
positive regulation of cell-cell adhesion	73	2.88E-12
response to ketone	59	3.01E-12
regulation of cellular component biogenesis	155	3.91E-12
taxis	103	4.05E-12
regulation of response to wounding	53	4.09E-12
monocarboxylic acid biosynthetic process	54	4.14E-12
neuron development	139	4.44E-12
regulation of organelle organization	181	5.22E-12
angiogenesis	76	5.31E-12
hepaticobiliary system development	50	5.73E-12
chemotaxis	102	6.43E-12
regulation of biosynthetic process	463	6.57E-12
positive regulation of cytosolic calcium ion concentration	72	7.25E-12
regulation of ERK1 and ERK2 cascade	74	7.27E-12
reproductive structure development	90	7.30E-12
regulation of lipase activity	40	7.35E-12
regulation of epithelial cell migration	64	7.44E-12
regulation of leukocyte cell-cell adhesion	78	7.88E-12
positive regulation of ion transport	71	8.38E-12
programmed cell death	153	8.39E-12
muscle structure development	98	8.51E-12
regulation of cold-induced thermogenesis	50	9.03E-12

Для представления наиболее информативных результатов отобраны гены с значениями p-value до E-12. Из таблицы видно, что наиболее значимыми категориями для генов СГХС являются гены, кодирующие белки негативной регуляции биологических процессов, отвечающих за метаболизм липидов, ответные реакции на токсические вещества, внутриклеточные процессы (катаболизм, регуляция ионов кальция в цитозоле, эндоцитоз), а также за развитие, пролиферацию и дифференцировку клеток эпителия, лейкоцитов, нейронов.

Таблица 4. Категории генных онтологий для генов СГХС
(клеточные компартменты)

Категории генных онтологий для клеточных компартментов	Количество генов	P - Value
high-density lipoprotein particle	1273	2.18E-67
extracellular space	527	1.85E-50
membrane-bounded organelle	1300	2.41E-48
side of membrane	571	3.06E-47
extracellular region	609	7.66E-46
endomembrane system	635	1.59E-45
organelle	1343	1.11E-44
cellular anatomical entity	1584	1.08E-42
cellular_component	1587	4.80E-41
intracellular anatomical structure	1388	2.72E-40
cell projection	388	3.57E-39
plasma membrane bounded cell projection	373	7.61E-38
cytoplasmic vesicle	389	5.08E-36
intracellular vesicle	389	5.75E-36
intracellular organelle	1245	1.70E-29
extracellular organelle	330	3.07E-28

Категории генных онтологий для клеточных компартментов	Количество генов	P - Value
extracellular membrane-bounded organelle	330	3.07E-28
extracellular vesicle	329	6.95E-28
extracellular exosome	323	1.23E-26
plasma membrane region	227	1.46E-25
endoplasmic reticulum	309	3.55E-25
cell periphery	705	2.40E-23
intracellular membrane-bounded organelle	1145	4.07E-23
secretory vesicle	192	3.02E-22
organelle membrane	459	3.96E-22
secretory granule	169	3.29E-21
vesicle lumen	95	6.58E-21
cytoplasmic vesicle lumen	93	4.89E-20
cell surface	172	1.67E-19
secretory granule lumen	90	9.56E-19
plasma membrane	638	5.89E-18
cell junction	292	2.04E-17
membrane microdomain	90	2.42E-17
membrane	955	2.70E-16
protein-lipid complex	32	4.89E-15
lipoprotein particle	31	6.59E-15
plasma lipoprotein particle	31	6.59E-15
cytosol	582	7.97E-15

Порог по значимости категорий был взят на уровне E-15. Анализ полученных результатов онтологий генов, относящихся к разделу клеточных компартментов, показал, что самыми значимыми являются категории

относящиеся к мембранам органелл и клеток, что вполне логично учитывая метаболизм холестерина, так как его внутриорганизменный транспорт и запасание осуществляется только в фосфолипидных мембранах.

Далее, аналогично предыдущим, была построена таблица онтологий для генов, регулирующих молекулярные функции.

Таблица 5. Категории генных онтологий для генов СГХС
(молекулярные функции)

Категории генных онтологий для молекулярных функций	Количество генов	P - Value
lipoprotein particle binding	89	1.36E-12
protein homodimerization activity	125	1.55E-12
anion binding	304	1.56E-12
hormone activity	46	1.15E-11
transcription factor binding	112	1.44E-11
RNA polymerase II-specific DNA-binding transcription factor binding	73	1.61E-11
nuclear receptor activity	30	1.39E-10
ligand-activated transcription factor activity	30	1.39E-10
kinase binding	128	1.41E-10
oxidoreductase activity	124	3.31E-10
cation binding	470	3.54E-10
protein-lipid complex binding	23	5.98E-10
DNA-binding transcription factor binding	23	5.98E-10
olfactory receptor activity	1	6.86E-10
transition metal ion binding	164	7.31E-10
protein kinase binding	115	2.02E-09
G protein-coupled receptor binding	67	3.32E-09
heme binding	44	3.86E-09

Категории генных онтологий для молекулярных функций	Количество генов	P - Value
tetrapyrrole binding	45	7.93E-09
O-acyltransferase activity	27	1.68E-08
steroid binding	36	2.54E-08
glycosaminoglycan binding	56	2.69E-08
protease binding	41	4.58E-08
organic acid binding	40	1.04E-07
heparin binding	45	1.57E-07
organic cyclic compound binding	594	1.75E-07
sulfur compound binding	58	3.81E-07
peptide binding	65	4.08E-07
carboxylic acid binding	45	4.13E-07
nucleotide binding	256	4.72E-07
cytokine activity	53	4.77E-07
nucleoside phosphate binding	256	4.77E-07
metal ion binding	443	5.78E-07
acyltransferase activity	56	7.07E-07
hydrolase activity	273	7.87E-07
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	22	1.52E-06
amide binding	73	2.08E-06
hormone binding	30	2.21E-06
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	43	2.56E-06
acyltransferase activity, transferring groups other than amino-acyl groups	50	3.09E-06
monocarboxylic acid binding	27	3.66E-06

Категории генных онтологий для молекулярных функций	Количество генов	P - Value
apolipoprotein binding	15	5.92E-06
antioxidant activity	29	7.98E-06
integrin binding	38	9.71E-06
amyloid-beta binding	28	9.76E-06
lipid transporter activity	40	9.93E-06

Здесь также для краткости таблицы порог значимости был принят за E-06. Из таблицы видно, что самыми значимыми являются категории связывания макромолекул: липопротеинов, ферментов, белков, циклических биомолекул. Часто встречается ферментативная, гормональная, цитокиновая активность. Как и в онтологиях клеточных компонентов можно наблюдать категории клеточной адгезии, что возможно обусловлено ненормально большим содержанием липидов, в основном холестерина, в крови и в тканях.

Таким образом, с помощью онлайн-ресурса DAVID было подтверждено, что большинство генов СГХС относятся к категориям генных онтологий мембран, транспорта липидов, адгезии клеток и внутриклеточных нарушений метаболизма липидов.

Дополнительно, для визуализации генных онтологий был использован онлайн - инструмент GOST (<http://biit.cs.ut.ee/gprofiler/gost>). Тот же список генов был загружен через интерфейс инструмента. На рисунке 12 представлен график поточечных значений категорий генных онтологий генов СГХС, рассчитанный с помощью GOST. Применена коррекция Бенжамини [8].

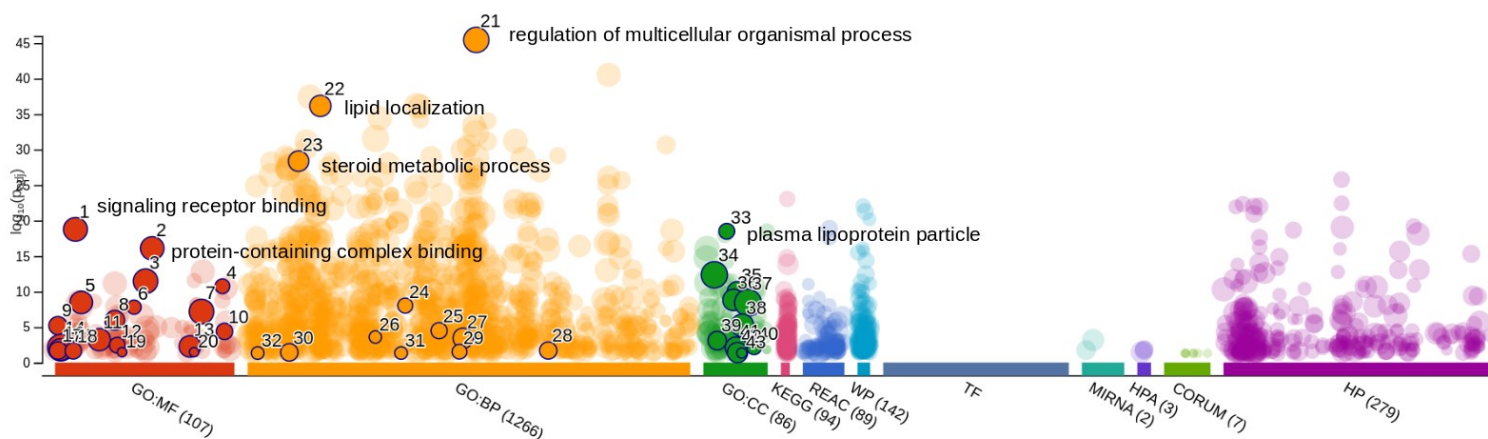


Рисунок 12. График статистической значимости найденных категорий генных онтологий в логарифмической шкале.

Ниже приведены таблицы с наиболее значимыми генными онтологиями молекулярных функций, клеточных компартментов и биологических процессов для генов, ассоциированных с семейной гиперхолестеринемией.

Таблица 6. Список категорий генных (GO biological process) онтологий по GOST для семейной гиперхолестеринемии.

Категория генных онтологий	p adj (значимость)
response to organic substance	1.66E-132
response to chemical	1.84E-129
response to oxygen-containing compound	5.37E-128
cellular response to chemical stimulus	3.30E-125
multicellular organismal process	4.69E-123
response to stimulus	4.17E-122
response to endogenous stimulus	4.97E-110
homeostatic process	3.13E-105
cellular response to stimulus	3.17E-102
lipid metabolic process	1.18E-101

Категория генных онтологий	p adj (значимость)
system development	5.14E-101
multicellular organism development	5.79E-98
cellular response to organic substance	7.39E-96
regulation of multicellular organismal process	1.59E-95
developmental process	4.27E-95
response to external stimulus	1.26E-92
animal organ development	1.35E-92
anatomical structure development	5.19E-92
localization	3.94E-89
cellular response to oxygen-containing compound	8.70E-89
small molecule metabolic process	4.61E-88
regulation of biological quality	9.56E-88
cellular response to endogenous stimulus	5.99E-82
response to organic cyclic compound	1.43E-81
transport	5.78E-81
response to hormone	8.06E-81
response to lipid	6.15E-80
response to stress	2.04E-79
establishment of localization	2.34E-78
regulation of localization	3.02E-78
response to nitrogen compound	3.95E-78
positive regulation of multicellular organismal process	6.46E-78
response to organonitrogen compound	7.49E-78
inflammatory response	9.42E-78
regulation of transport	1.88E-75
cell communication	5.25E-74

Категория генных онтологий	p adj (значимость)
chemical homeostasis	1.21E-73
regulation of cell population proliferation	1.74E-73
positive regulation of biological process	2.43E-72
circulatory system development	3.81E-71
anatomical structure morphogenesis	4.75E-71
signaling	5.17E-71
cell population proliferation	8.59E-71
cell surface receptor signaling pathway	8.79E-70
regulation of response to stimulus	9.81E-69
response to abiotic stimulus	9.12E-68
positive regulation of transport	2.35E-66
signal transduction	6.30E-66
organic hydroxy compound metabolic process	1.34E-65
circulatory system process	2.19E-65
cellular lipid metabolic process	3.42E-65

Таблица 7. Список категорий генных (GO cellular component) онтологий по GOST для семейной гиперхолестеринемии.

Категория генных онтологий	p adj (значимость)
cytoplasm	6.75E-82
extracellular space	9.91E-67
vesicle	8.00E-65
endomembrane system	8.69E-65
extracellular region	3.59E-64

Категория генных онтологий	p adj (значимость)
cytoplasmic vesicle	4.69E-51
intracellular vesicle	6.42E-51
cell projection	5.80E-47
plasma membrane bounded cell projection	6.76E-45
cell periphery	4.24E-40
extracellular vesicle	1.29E-37
extracellular organelle	1.42E-37
extracellular membrane-bounded organelle	1.42E-37
endoplasmic reticulum	2.42E-37
extracellular exosome	4.30E-36
plasma membrane region	1.43E-35
secretory vesicle	2.26E-34
membrane	8.36E-33
secretory granule	9.21E-33
organelle membrane	4.48E-32
plasma membrane	8.72E-32
cilium	6.62E-30
vesicle lumen	1.26E-29
cytoplasmic vesicle lumen	2.26E-28
cell surface	4.78E-28
lipoprotein particle	3.46E-27
plasma lipoprotein particle	3.46E-27
protein-lipid complex	5.75E-27
secretory granule lumen	1.61E-26
cell junction	1.93E-25
membrane raft	2.19E-25

Категория генных онтологий	p adj (значимость)
membrane microdomain	2.79E-25
neuron projection	2.06E-24
non-motile cilium	6.88E-23
cytosol	1.22E-22
bounding membrane of organelle	6.04E-22
high-density lipoprotein particle	5.14E-21
ciliary transition zone	2.49E-20
somatodendritic compartment	2.05E-19
membrane-bounded organelle	2.24E-19
ciliary tip	2.94E-19
9+0 non-motile cilium	1.30E-18
apical part of cell	1.55E-18
intraciliary transport particle	1.78E-18
platelet alpha granule	3.08E-18
photoreceptor cell cilium	7.00E-18
endoplasmic reticulum lumen	9.32E-18
side of membrane	1.30E-17
cell projection membrane	1.61E-16
platelet alpha granule lumen	2.56E-16
very-low-density lipoprotein particle	6.09E-16
triglyceride-rich plasma lipoprotein particle	6.09E-16

Таблица 8. Список категорий генных (GO molecular function) онтологий по GOST для семейной гиперхолестеринемии.

Категория генных онтологий	p adj (значимость)
protein binding	9.23E-52
signaling receptor binding	6.22E-49
identical protein binding	2.93E-36
protein-containing complex binding	7.54E-32
enzyme binding	7.84E-31
molecular function regulator activity	1.16E-27
signaling receptor activator activity	9.20E-27
receptor ligand activity	3.20E-26
lipid binding	4.12E-26
signaling receptor regulator activity	1.19E-25
catalytic activity	1.08E-23
molecular function activator activity	3.64E-23
small molecule binding	2.02E-22
ion binding	8.46E-21
protein dimerization activity	2.45E-20
protein homodimerization activity	1.50E-19
protein-lipid complex binding	3.83E-18
lipoprotein particle binding	3.83E-18
RNA polymerase II-specific DNA-binding transcription factor binding	1.37E-17
nuclear receptor activity	5.09E-17
ligand-activated transcription factor activity	5.09E-17
hormone activity	6.19E-17

Категория генных онтологий	p adj (значимость)
cytokine receptor binding	1.01E-16
carbohydrate derivative binding	1.52E-16
transcription factor binding	2.35E-16
anion binding	4.28E-16
DNA-binding transcription factor binding	4.47E-16

К самым значимым категориям для данного списка генов можно отнести следующие онтологии:

- 1 Для биологических процессов – внутриорганизменная ответная реакция на различного рода вещества и стимулы, процессы сигнализации и коммуникации клеток, транспорт липидов.
- 2 Для клеточных компартментов – внутри- и внеклеточное пространство, мембранные комплексы.
- 3 Для молекулярных функций – связывание макромолекул: белков, ферментов, липидов, сигнальных молекул.

На основе проведенного анализа, можно отметить связь развития семейной гиперхолестеринемии с нарушением транспорта липидов, возникающих при синтезе молекул переносчиков холестерина, непосредственно во время их транспортировки и связывании с клеточными рецепторами.

3.3. Реконструкция генных сетей для генов семейной гиперхолестеринемии.

Для реконструкции генной сети взаимодействий генов СГХС использовался ресурс STRING-DB (<https://string-db.org>).

При реконструировании генной сети всех 1620 генов, было распознано 1543 гена. Полученная сеть получилась слишком громоздкая и неподдающаяся адекватному анализу, из-за чего было принято решение, для реконструирования модели генной сети взять 150 генов, наиболее тесно связанных с семейной гиперхолестеринемией.

На рисунке №13 представлена генная сеть генов СГХС, реконструированная с помощью STRING-DB. Все гены были распознаны. Сеть плотно связанная между собой, хотя связи выставлялись только по параметрам белок-белковые взаимодействия и ко-экспрессии. На рисунке можно видеть, что одними из центральных генов являются гены LDLR, APOB, PCSK и APOE кодирующие рецептор липопротеинов низкой плотности и аполипопротеин В и Е и пропротеиновую конвертазу субтилизин-кексинового типа 9, гидролизующую рецепторы липопротеинов низкой плотности. На изображении гены отведены в стороны для наглядной визуализации количества связей с другими генами.

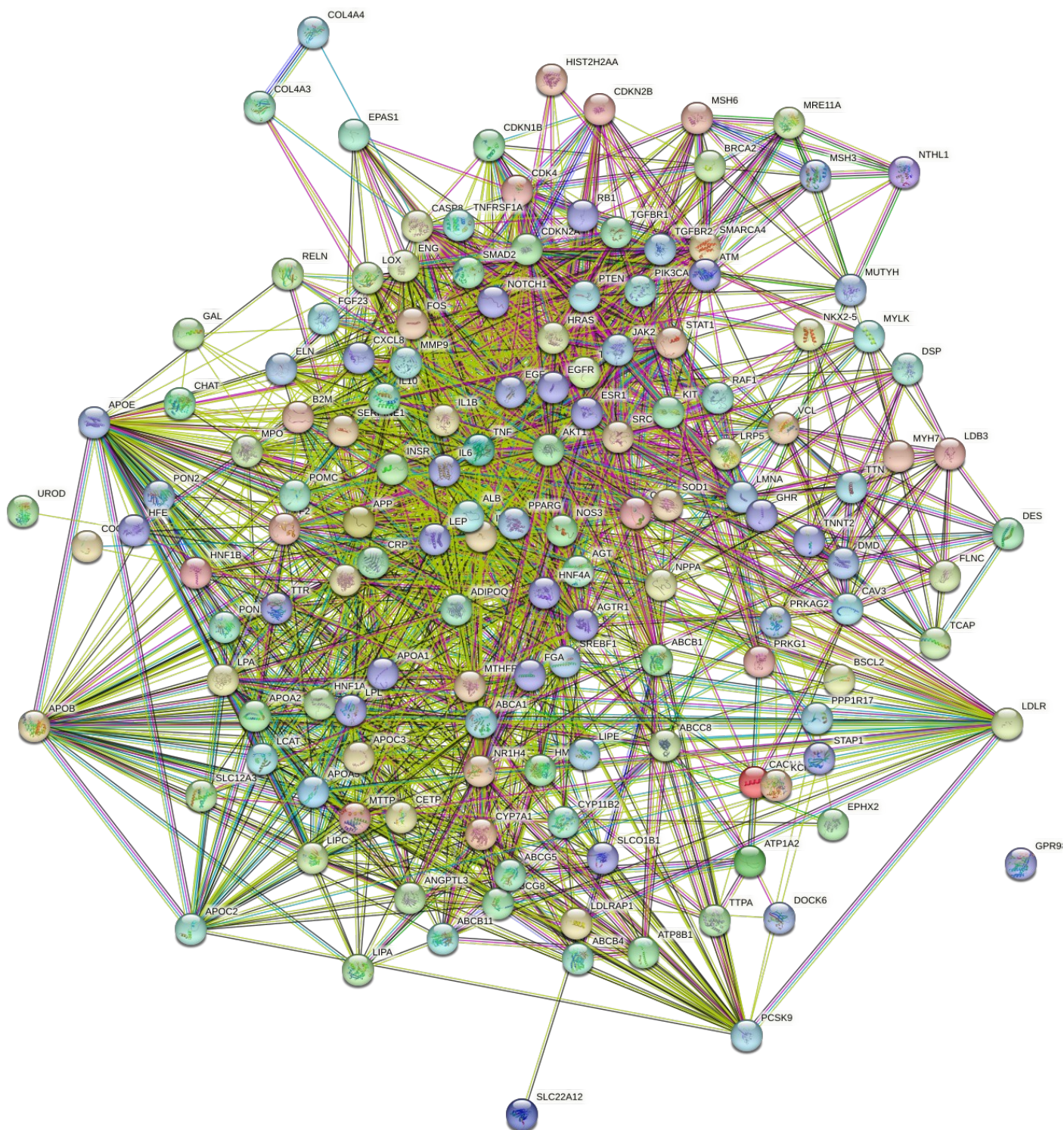


Рисунок 13. Генная сеть СГХС

Статистика сети:

- количество узлов – 150;
- количество ребер – 2023;
- средняя степень связности узла – 28.1;
- средний локальный коэффициент кластеризации – 0.614;

- ожидаемое количество ребер – 677;
- значимость плотности белок-белковых взаимодействий – $1.0E-16$.

Далее полученная сеть была перестроена для получения более подробной визуализации генных связей. Из сети были удалены одиночные ни с кем не связанные гены и не подтвержденные экспериментально связи межгенных взаимодействий. А также дублирующиеся связи между генами заменены одной.

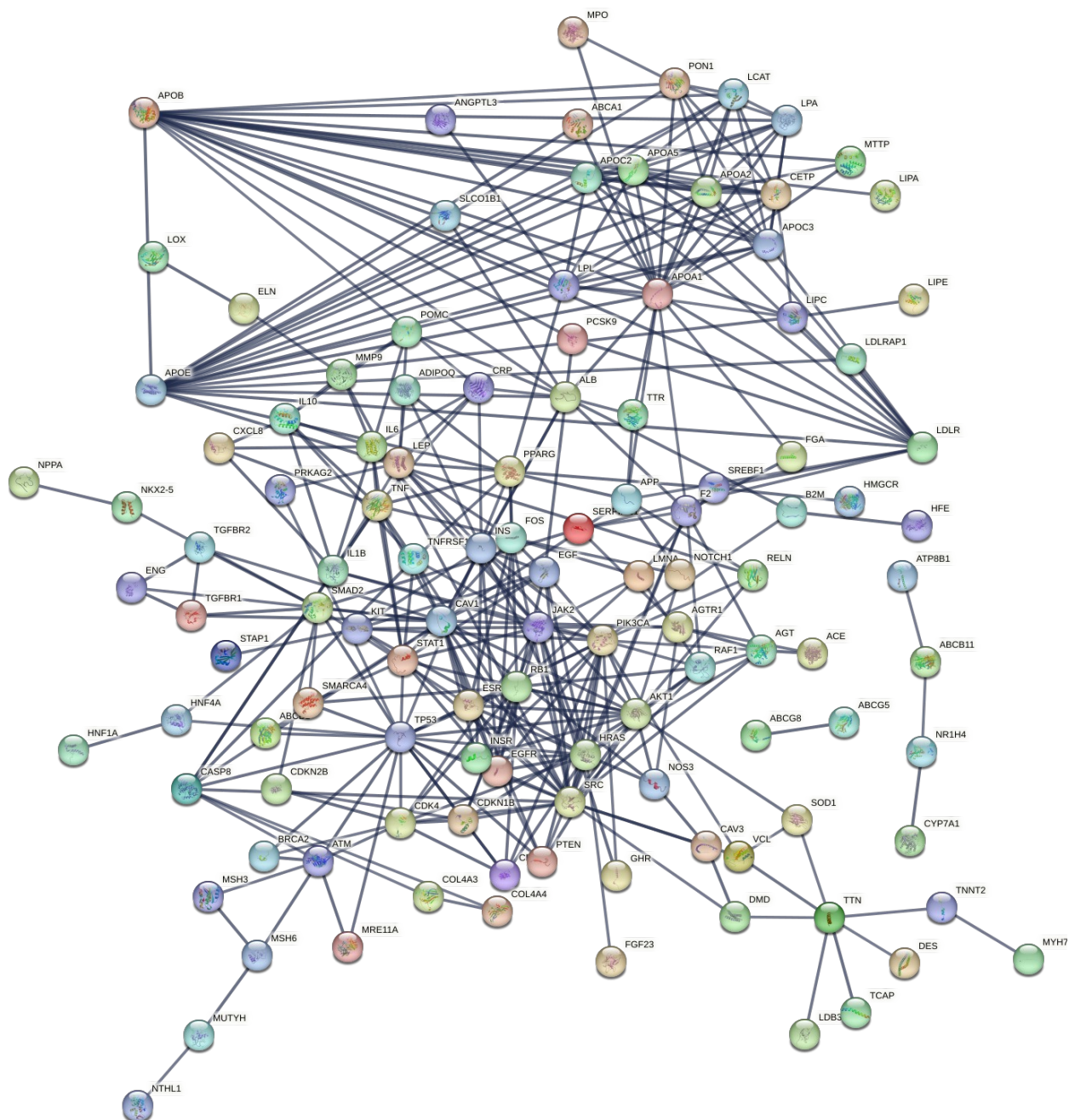


Рисунок 14. Итеративно перестроенная генная сеть СГХС

Статистика сети:

- количество узлов – 144;
 - количество ребер – 358;
 - средняя степень связности узла – 4.97;
 - средний локальный коэффициент кластеризации – 0.442;
 - ожидаемое количество ребер – 122;
- значимость плотности белок-белковых взаимодействий – $1.0E-16$.

В ходе работ была произведена кластеризация полученной генной сети. Сеть была разбита на 10 обособленных кластеров. Из всех кластеров выделялись два наиболее многочисленных по составу генов, данные кластеры представлены ниже на рисунке 15.

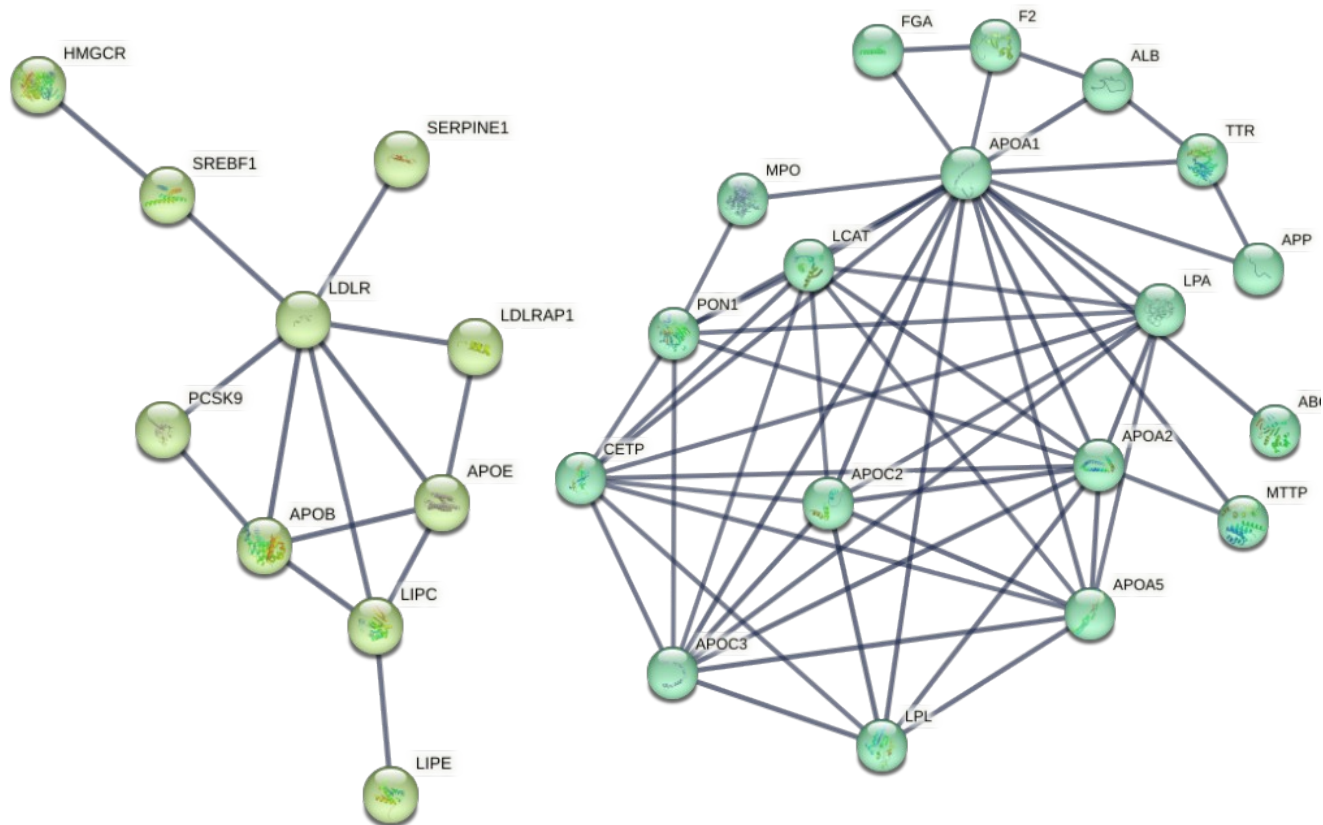


Рисунок 15. Кластеризация генной сети СГХС

Кластер с генами, выделенными желтым цветом, функционально связан с рецептор-опосредованным эндоцитозом холестерина в клетки и его регуляцией. Центральный ген кластера -- LDLR, ген, кодирующий рецептор липопротеинов низкой плотности, уменьшение его активности приводит к повышению концентрации ЛПНП и холестерина в крови. Другими двумя наиболее значимыми генами кластера являются гены APOB и PCSK9. Это обусловлено тем, что ген APOB является лигандом для рецептора LDLR, а PCSK9 кодирует фермент, который ингибирует LDLR путем протеолиза.

Следующий, зеленый, кластер отвечает за обратный транспорт холестерина в печень, а также за ремодуляцию липопротеинов низкой плотности. Наиболее

значимый ген кластера с наибольшим ранговым значением и количеством связей — ген APOA1. Ген кодирует аполипопротеин A1, основной белок липопротеинов высокой плотности. Белок является активатором лецитинхолестеринацилтрансферазы (ЛХАТ), участвующей в процессе созревания ЛПВП. В кластер также входят другие аполипопротеины: APOA2, APOC2, APOC3, LPA и APOA5.1

В результате анализа генных сетей было выяснено, что наиболее значимыми являются группы генов, ассоциированных с транспортировкой холестерина внутри организма человека. Среди всех генов наиболее предрасположенные к развитию СГХС оказались гены рецептора ЛПНП (LDLR, LDLRAP1) и поверхностных белков липопротеинов (APOA1, APOB, APOE). Также высока роль ферментов, регулирующих активность белков, упомянутых ранее, наиболее значимый из них — ген PCSK9.

3.4. Детальный анализ генов

Для геномной аннотации были взяты гены из генной сети, мутации которых наиболее предрасположены к развитию семейной гиперхолестеринемии.

3.4.1 Ген LDLR

Ген LDLR (Low-density lipoprotein receptor), кодирует рецептор липопротеинов низкой плотности (ЛПНП). Идентификатор белка был взят с ресурса “UniProt” (<https://www.uniprot.org>):

LDLR – P01130

Функция белка: связывает ЛПНП, основной липопротеин плазмы, несущий холестерин, и транспортирует его в клетки путем эндоцитоза, сначала лиганд с рецептором объединяются, а затем погружаются в клетку [34].

Структура белка представлена в основном альфа-спиралями. Ниже представлены изображения структуры белка LDLR взяты с интернет-ресурсов «PDBe-KB»

(Protein Data Bank in Europe Knowledge Base)(<https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P01130>) и «UniProt» (<https://www.uniprot.org>).

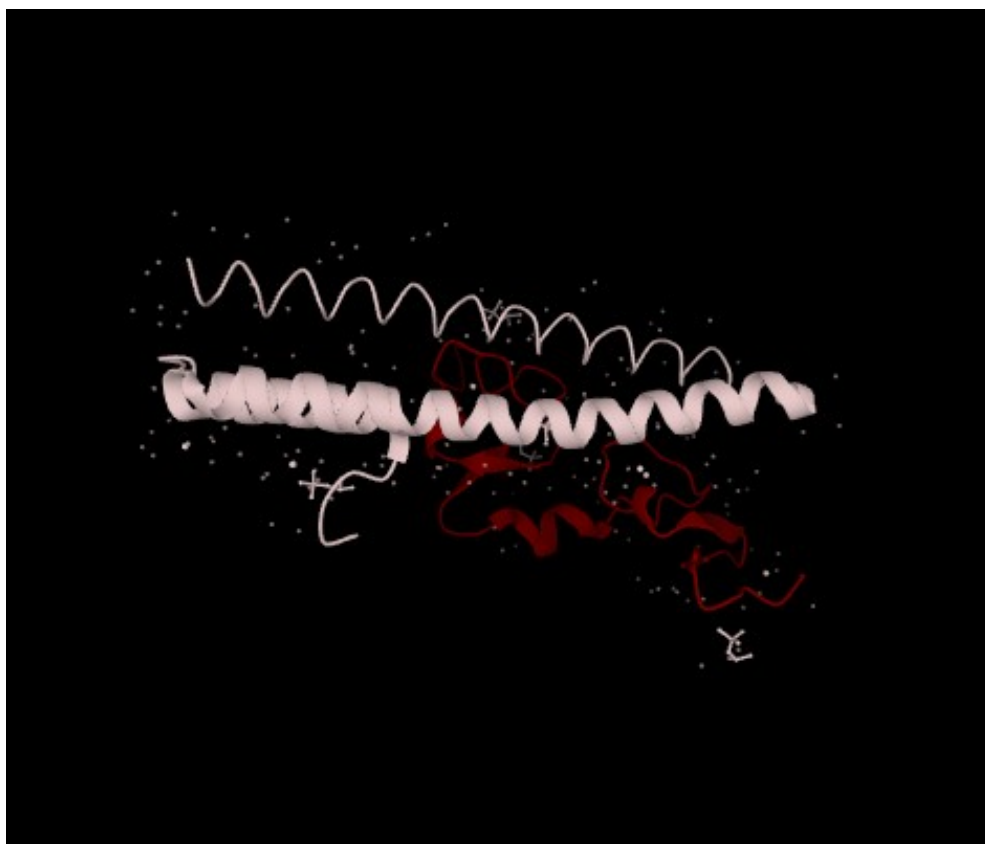


Рисунок 16. Структура белка P01130 (взято с «PDBe-KB»)

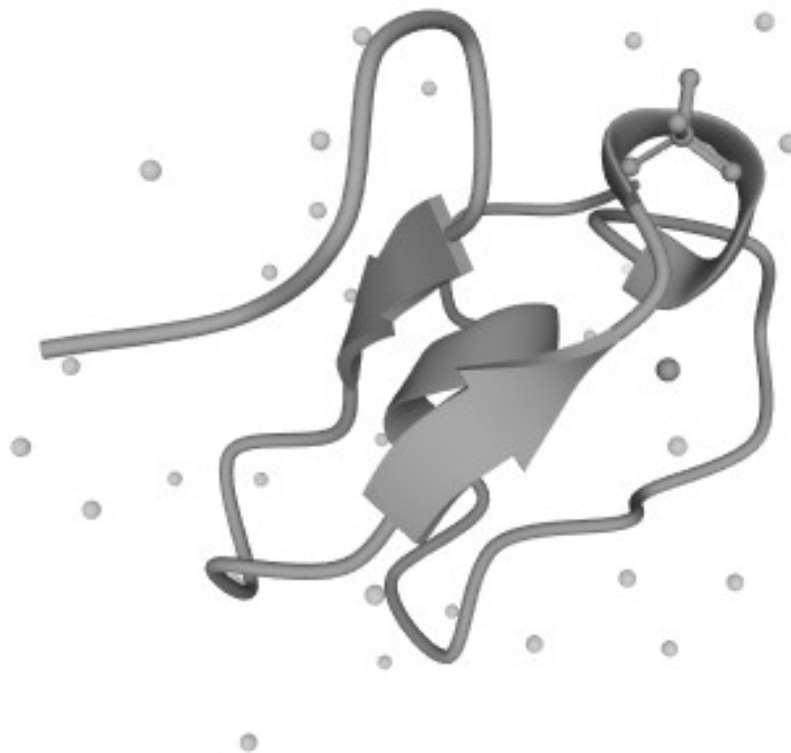


Рисунок 17. Схематическое изображение белка P01130 (взято с «UniProt»)

Ген LDLR расположен на коротком плече хромосомы 19 (19p13.1-13.3) и имеет длину около 45 кб, кодирует 18 экзонов и 17 интронов. LDLR - это белок из 839 аминокислот, который синтезируется в эндоплазматическом ретикулуме, где он сворачивается и частично гликозилируется.

доставляет функциональную копию человеческого гена рецептора липопротеинов низкой плотности (LDLR) в клетки печени. Этот препарат в настоящее время исследуется для лечения гомозиготной семейной гиперхолестеринемии.

3.4.2 Ген APOA1

Этот ген кодирует аполипопротеин A-I, который является основным белковым компонентом липопротеина высокой плотности (ЛПВП) в плазме крови. Кодированный белок способствует оттоку холестерина из тканей в печень для выведения, а также является кофактором для лецитинхолестеринацилтрансферазы (ЛХАТ) - фермента, ответственного за образование большинства холестероловых эфиров плазмы, а также созревание ЛПВП (Рисунок №5) [20]. Этот ген тесно связан с двумя другими генами аполипопротеинов на хромосоме 11. Белок также взаимодействует с белками NAXE [21] и CLU [22]. Дефекты этого гена связаны с дефицитом ЛПВП, включая болезнь Танжера, и с системным невропатическим амилоидозом [23, 24].

В базе данных «UniProt» белок имеет идентификационный номер P02647.

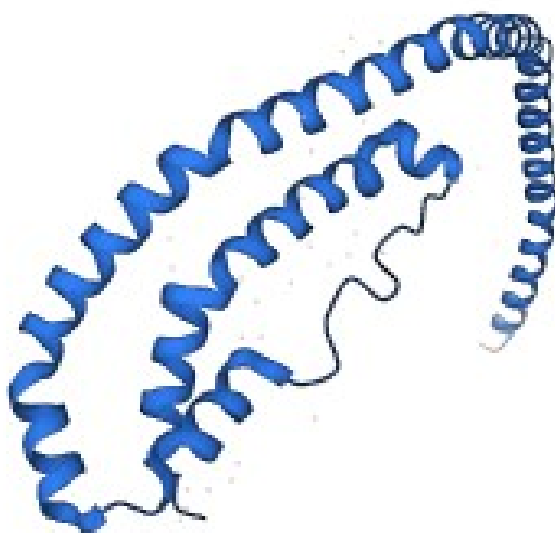


Рисунок 19. Структура белка P02647 (взято с «PDBe-KB»)

3.4.3 Ген PCSK9

Протеинконвертаза субтилизин/кексин типа 9 (PCSK9) - это фермент, кодируемый геном PCSK9 у человека на первой хромосоме. Это фермент семейства протеинконвертаз - белков, активирующих другие белки. Как и многие другие белки, PCSK9 при первом синтезе неактивен, поскольку участок пептидной цепи блокирует его активность; пропротеинконвертазы удаляют этот участок, чтобы активировать фермент. Ген PCSK9 также содержит один из 27 локусов, связанных с повышенным риском развития ишемической болезни сердца [25, 26].

PCSK9 повсеместно экспрессируется во многих тканях и типах клеток. PCSK9 связывается с рецептором ЛПНП во внеклеточной жидкости и разрушает его. Рецептор LDLR на мембранах печени и других клеток связывает и инициирует проникновение ЛПНП из внеклеточной жидкости в клетки и направляет комплекс в лизосомы для разрушения. Если блокировать PCSK9, комплекс ЛПНП- LDLR разделяется в процессе транспортировки, при этом ЛПНП переваривается в лизосоме, а LDLR возвращается на поверхность клетки и таким образом может удалить дополнительные ЛПНП-частицы из внеклеточной жидкости. Поэтому блокирование PCSK9 может снизить концентрацию ЛПНП-частиц в крови.

PCSK9 имеет важное медицинское значение. Препараты, блокирующие PCSK9, могут снижать концентрацию частиц ЛПНП. Первые два ингибитора PCSK9, алирокумаб и эволокумаб, были одобрены «Управлением по контролю за продуктами и лекарствами США» в 2015 году для снижения концентрации частиц ЛПНП, когда статины и другие препараты недостаточно эффективны или плохо переносятся.

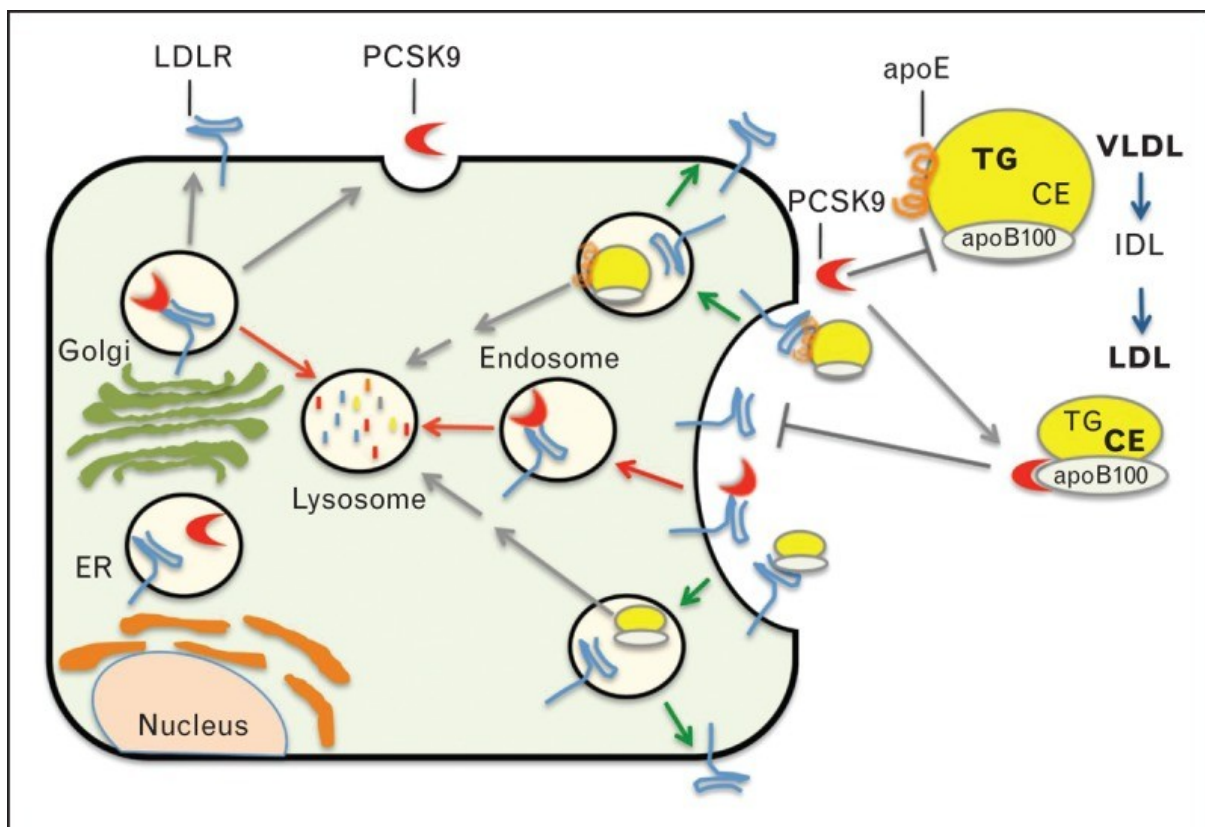


Рисунок 20. Активирование рецептора ЛПНП с помощью PCSK9 [27]

В базе данных «UniProt» белок имеет идентификационный номер Q8NBP7. Ниже представлена структура фермента, кодируемого геном PCSK9. Изображения взяты с сайтов баз данных «UniProt» и «PDB».

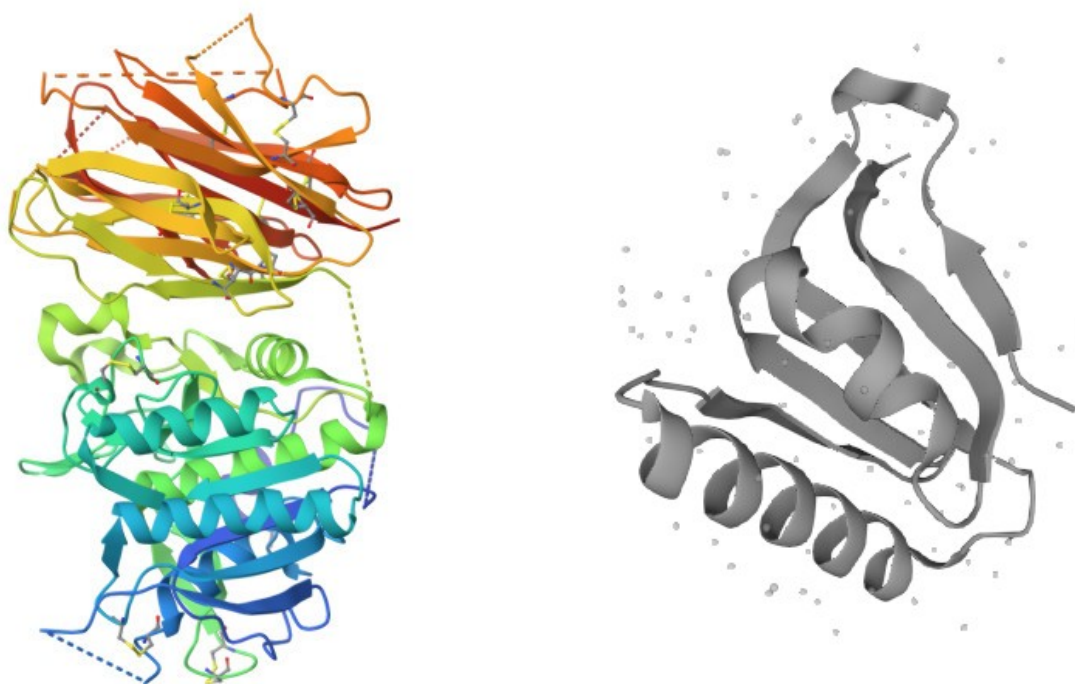


Рисунок 21. Структура белка Q8NBP7 (взято с «PDBe-KB» и «UniProt»)

С помощью базы данны «DrugBank» были найдены два лекарственных средства, взаимодействующих с PCSK9. Оба средства являются моноклональными антителами, ингибирующими PCSK9.

Алирокумаб - биофармацевтический препарат, получивший одобрение FDA в июле 2015 года в качестве препарата второй линии для лечения повышенного уровня холестерина у взрослых, у которых уровень ЛПНП-холестерина не контролируется комбинацией диеты и лечения статинами. Это человеческое моноклональное антитело, входящее в семейство ингибиторов PCSK9, которые представляют собой новый класс антихолестериновых терапевтических средств. Ингибирование PCSK9 способствует лучшему выведению ЛПНП из крови [29]. Второй препарат — эволокумаб, моноклональное антитело, разработанное для лечения гиперлипидемии. Это подкожная инъекция, одобренная для лиц, принимающих максимальную терапию статинами и нуждающихся в дополнительном снижении уровня холестерина ЛПНП. Он одобрен для лечения гомозиготной и гетерозиготной семейной холестеринемии в качестве дополнения к другим методам терапии первой линии [30].

ВЫВОДЫ

Семейная гиперхолестеринемия является опасным заболеванием, особенно в случае гомозиготности носителя. У людей с гомозиготной семейной гиперхолестеринемией ксантомы развиваются под кожей на локтях, коленях и ягодицах, а также в сухожилиях в возрасте 4-5 лет [33]. У таких пациентов в большинстве случаев развиваются сердечные приступы до 30 лет, зачастую со смертельным исходом [4]. В случае гетерозиготности, данное заболевание, в большинстве случаев протекает бессимптомно до первых случаев сердечной недостаточности, которые в среднем проявляются в возрасте 40 лет.

На данный момент легкое течение болезни поддается лекарственному лечению. Для снижения симптомов пациентам прописывают статины [28], которые структурно похожи на ГМГ – КоА, из-за чего конкурентно блокируют ГМГ-КоА-редуктазу в мевалонатном пути, предотвращая синтез холестерина в организме. В последнее время набирают популярность препараты на основе моноклональных тел для блокировки ингибиторов рецепторов липопротеинов низкой плотности: алирокумаб и эволокумаб [29, 30]. В тяжелых случаях СГХС используют плазмаферез. Но статины и плазмаферез представляют из себя симптоматическое лечение, которое не способно устранить причину СГХС. Исследование структуры генной сети показывало высокую связность генов и их продуктов. Анализ литературы (PubMed) показал продолжающийся рост публикаций по данной теме - всего 10,727 публикаций по настоящее время.

- 1 Изучена имеющаяся актуальная научная литература по заболеванию семейная гиперхолестеринемия. Несмотря на то, что исследуемое заболевание имеет относительно большую распространенность и в некоторых случаях тяжелое течение болезни, СГХС считается неизлечимым заболеванием, от которого имеется только симптомальное лечение, неспособное устранить полностью причину заболевания.

- 2 Построен список генов, ассоциированных с семейной гиперхолестеринемией по интернет-доступным базам данных.
- 3 Составлены таблицы представленности категорий генных онтологий сетей по полученному списку генов семейной гиперхолестеринемии с использованием инструментов биоинформатики. Среди наиболее значимых категорий генных онтологий оказались категории, связанные с нарушением транспорта липидов, а также категории взаимодействий клеточных рецепторов.
- 4 Проведена компьютерная реконструкция генной сети генов семейной гиперхолестеринемии, ее визуализация, статистический анализ и сопоставление с генными сетями сопутствующих заболеваний.
- 5 Проведен анализ выявленных центральных генов, ассоциированных с семейной гиперхолестеринемией: LDLR, APOB и PCSK9 – которые могут использоваться для таргетной терапии, а также в качестве маркерных генов для диагностики заболевания.

СПИСОК ЛИТЕРАТУРЫ

1. Varghese M. Familial hypercholesterolemia: A review // *Annals of Pediatric Cardiology*. 2014. № 2 (7). С. 107.
2. Назарова О. В., Назаров А. В. Практические аспекты использования игровых технологий в процессе организации долговременной самостоятельной работы студентов высшей школы // *Общество: социология, психология, педагогика*. 2019. № 12. С. 177–182.
3. Bragina E. Yu. [и др.]. Insights into pathophysiology of dystrophy through the analysis of gene networks: an example of bronchial asthma and tuberculosis // *Immunogenetics*. 2014. № 7–8 (66). С. 457–465.
4. Ежов М. В. ., Бажан С. С., Ершова А. И. ., Мешков А. Н., Соколов А. А., Кухарчук В. В., Гуревич В. С., Воевода М. И., Сергиенко И. В., Шахтшнейдер Е. В., Покровский С. Н., Коновалов Г. А., Леонтьева И. В., Константинов В. О., Щербакова М. Ю., Захарова И. Н., Балахонова Т. В., Филиппов А. Е., Ахмеджанов Н. М., Александрова О. Ю., Липовецкий Б. М. Клинические рекомендации по семейной гиперхолестеринемии // *Атеросклероз и Дислипидемии*. 2019. Т. № 1 (34). СС. 5–43.
5. Reconstruction of bacterial gene networks, metabolic pathways and transcription regulation Колчанов Н.А. French-Russian workshop on annotation of bacterial genomes. Франция, Тулуза 2006, октябрь
6. Amberger J. S., Bocchini C. A., Scott A. F. et al. OMIM. org: leveraging knowledge across phenotype–gene relationships // *Nucleic acids research*. – 2019. – Т. 47. – №. D1. – С. D1038-D1043.
7. Hovland A. [и др.]. The risk of various types of cardiovascular diseases in mutation positive familial hypercholesterolemia; a review // *Frontiers in Genetics*. 2022. (13). С. 1072108.
8. Benjamini, Y. and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57:289-300.

9. Электронный ресурс GeneCards URL:
<https://www.genecards.org/cgi-bin/carddisp.pl?gene=LDLR&keywords=LDLR>
10. Электронный ресурс GeneCards URL:
<https://www.genecards.org/cgi-bin/carddisp.pl?gene=APOB&keywords=apob>
11. Электронный ресурс GeneCards URL:
<https://www.genecards.org/cgi-bin/carddisp.pl?gene=PCSK9&keywords=pcsk9>
12. Safran M. [и др.]. GeneCards Version 3: the human gene integrator // Database. 2010. № 0 (2010). С. баq020–баq020.
13. Электронный ресурс GeneCards URL:
<https://www.genecards.org/Guide/Search#relevance>
14. A. H. Lashkari, F. Mahdavi and V. Ghomi, "A Boolean Model in Information Retrieval for Search Engines" 2009 International Conference on Information Management and Engineering, Kuala Lumpur, Malaysia, 2009, pp. 385-389.
15. Электронный ресурс ElasticSearch URL:
<https://www.elastic.co/guide/en/elasticsearch/guide/current/scoring-theory.html>
16. Nordestgaard B. G. [и др.]. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society // European Heart Journal. 2013. № 45 (34). С. 3478–3490a.
17. М.В. Ежов, И.В. Сергиенко, Т.А. Рожкова [и др.]. Диагностика и лечение семейной гиперхолестеринемии (российские рекомендации) // Вестник современной клинической медицины. — 2017. — Т. 10, вып. 2. — С.72—79.
18. Austin M. A. Genetic Causes of Monogenic Heterozygous Familial Hypercholesterolemia: A HuGE Prevalence Review // American Journal of Epidemiology. 2004. № 5 (160). С. 407–420.

19. Nordestgaard B. G. [и др.]. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society // *European Heart Journal*. 2013. № 45 (34). С. 3478–3490a.
20. Akerlöf E. [и др.]. Identification of apolipoprotein A1 and immunoglobulin as components of a serum complex that mediates activation of human sperm motility // *Biochemistry*. 1991. № 37 (30). С. 8986–8990.
21. Ehnholm C. [и др.]. The apolipoprotein A-I binding protein of placenta and the SP-40,40 protein of human blood are different proteins which both bind to apolipoprotein A-I // *Biochimica Et Biophysica Acta*. 1991. № 3 (1086). С. 255–260.
22. Ritter M. [и др.]. Cloning and characterization of a novel apolipoprotein A-I binding protein, AI-BP, secreted by cells of the kidney proximal tubules in response to HDL or ApoA-I // *Genomics*. 2002. № 5 (79). С. 693–702.
23. Hamidi Asl K. [и др.]. A novel apolipoprotein A-1 variant, Arg173Pro, associated with cardiac and cutaneous amyloidosis // *Biochemical and Biophysical Research Communications*. 1999. № 2 (257). С. 584–588.
24. Lachmann H. J. [и др.]. Misdiagnosis of hereditary amyloidosis as AL (primary) amyloidosis // *The New England Journal of Medicine*. 2002. № 23 (346). С. 1786–1791.
25. Nassoury N. [и др.]. The cellular trafficking of the secretory proprotein convertase PCSK9 and its dependence on the LDLR // *Traffic (Copenhagen, Denmark)*. 2007. № 6 (8). С. 718–732.
26. Poirier S. [и др.]. The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2 // *The Journal of Biological Chemistry*. 2008. № 4 (283). С. 2363–2372.
27. Lagace T. A. PCSK9 and LDLR degradation: regulatory mechanisms in circulation and in cells // *Current Opinion in Lipidology*. 2014. № 5 (25). С. 387–393.

28. Raal F. J., Hovingh G. K., Catapano A. L. Familial hypercholesterolemia treatments: Guidelines and new therapies // *Atherosclerosis*. 2018. (277). С. 483–492.
29. Fundación Hipercolesterolemia Familiar clinicaltrials.gov. Clinical Trial to Evaluate the Effect of Alirocumab on the Volume, Architecture and Composition of Atherosclerotic Plaque in Patients With Familial Hypercholesterolemia. 2022.
30. Amgen clinicaltrials.gov. A Multicenter, Open-label, Single-arm, Study to Evaluate Safety and Tolerability of Repatha in Patients With Homozygous Familial Hypercholesterolemia (HoFH) in India. 2020.
31. Singh S., Bittner V. Familial hypercholesterolemia--epidemiology, diagnosis, and screening // *Current Atherosclerosis Reports*. 2015. № 2 (17). С. 482.
32. Северин С. Биологическая химия / С. Северин, 2-е изд., ГЭОТАР-МЕД, 2011.
33. About Familial Hypercholesterolemia // Genome.gov [Электронный ресурс]. URL: <https://www.genome.gov/Genetic-Disorders/Familial-Hypercholesterolemia> (дата обращения: 10.06.2023).
34. Davis C. G. [и др.]. Deletion of clustered O-linked carbohydrates does not impair function of low density lipoprotein receptor in transfected fibroblasts // *The Journal of Biological Chemistry*. 1986. № 6 (261). С. 2828–2838.
35. Yamamoto T. [и др.]. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA // *Cell*. 1984. № 1 (39). С. 27–38.
36. Ebhardt M. [и др.]. Mutation analysis in 46 German families with familial hypercholesterolemia: identification of 8 new mutations. Mutations in brief no. 226. Online // *Human Mutation*. 1999. № 3 (13). С. 257.
37. Электронный ресурс PANTHER URL: <http://www.pantherdb.org/about.jsp>
38. Электронный ресурс DAVID URL: <https://david.ncifcrf.gov/content.jsp?file=fact.html>
39. Электронный ресурс KEGG URL: <https://www.genome.jp/kegg/kegg1a.html>

- 40.Электронный ресурс Ensembl URL:
<https://www.ensembl.org/index.html>
- 41.Электронный ресурс STRING-DB URL: <https://string-db.org/cgi/help?sessionId=bsB0mXM80t9s>
- 42.Civeira F. Guidelines for the diagnosis and management of heterozygous familial hypercholesterolemia // Atherosclerosis. 2004. № 1 (173). С. 55–68.

СПИСОК ТЕРМИНОВ

каскадное тестирование – это процесс информирования членов семьи об обнаруженном в семье генетическом заболевании, после чего члены семьи проходят тестирование на это заболевание.

скэвенджер-рецептор – это семейство рецепторов клеточной поверхности, которые разнообразны по своей структуре и биологической функции и делятся на различные классы. СР могут связываться с целым рядом лигандов и усиливать элиминацию измененных собственных или несобственных мишеней.

Функциональные механизмы, которые приводят к очищению от вредных веществ, включают фагоцитоз, эндоцитоз, адгезию и сигнализацию.

СПИСОК СОКРАЩЕНИЙ

ФН — семейная гиперхолестеринемия (Familial hypercholesterolemia)

ЛПНП - липопротеин низкой плотности

ИБС - ишемическая болезнь сердца

ИМ - инфаркт миокарда

СГХС – семейная гиперхолестеринемия

ХС — холестерол

ЭХС – эфир холестерина

ЖК – жирные кислоты

ХМ – хиломикрон

ТАГ – триацилглицерол

ХМост – остаточные хиломикроны

ХМн – незрелые хиломикроны

ЛХАТ – лецитин-холестерол-ацилтрансфераза

ЛП – липопротеин

ЛПВП – липопротеин высокой плотности

ЛПНП – липопротеин низкой плотности

ЛПОНП – липопротеин оень низкой плотности

БПЭХ – белка, переносящего ЭХС

ЛП-липаза – липопротеинлипаза

PCSK9 – Пропотеиновая конвертаза субтилизин-кексинового типа 9

SNP – Single-nucleotide polymorphism

АРОВ – аipoprotein B

LDLR – lowdensity lipoprotein receptor

РНК – рибонуклеиновая кислота

ДНК – дезоксирибонуклеиновая кислота

ГМГ-КоА – 3-гидрокси-3-метилглутарил-кофермент А

GO – gene ontology

SIR – standardized incidence ratio

ПРИЛОЖЕНИЯ

Приложение А

Список генов, ассоциированных с семейной гиперхолестеринемией

Название гена	Описание	Категория	Uniprot ID	Релевантность
LDLR	Low Density Lipoprotein Receptor	Protein Coding	P01130	209.8613586
APOB	Apolipoprotein B	Protein Coding	P04114	141.6577606
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9	Protein Coding	Q8NBP7	102.9853287
LDLR-AS1	LDLR Antisense RNA 1	RNA Gene		59.98997116
LDLRAP1	Low Density Lipoprotein Receptor Adaptor Protein 1	Protein Coding	Q5SW96	46.98540497
MIR6886	MicroRNA 6886	RNA Gene		38.24206924
LOC106560211	APOB 5' Regulatory Region	Functional Element		33.51564407
APOE	Apolipoprotein E	Protein Coding	P02649	32.15110016
EPHX2	Epoxide Hydrolase 2	Protein Coding	P34913	25.88716316
GHR	Growth Hormone Receptor	Protein Coding	P10912	24.58562279
APOA2	Apolipoprotein A2	Protein Coding	P02652	23.18382835
STAP1	Signal Transducing Adaptor Family Member 1	Protein Coding	Q9ULZ2	23.10738182
ABCG5	ATP Binding Cassette Subfamily G Member 5	Protein Coding	Q9H222	21.51917458
ABCG8	ATP Binding Cassette Subfamily G Member 8	Protein Coding	Q9H221	20.19938469
DOCK6	Dedicator Of Cytokinesis 6	Protein Coding	Q96HP0	19.60373878
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4	Protein Coding	P51532	17.89723778
LPL	Lipoprotein Lipase	Protein Coding	P06858	17.4499073
ABCA1	ATP Binding Cassette Subfamily A Member 1	Protein Coding	O95477	13.96794319
LOC126862855	BRD4-Independent Group 4 Enhancer GRCh37_chr19:11202176-11203375	Functional Element		13.31741619
LPA	Lipoprotein(A)	Protein Coding	P08519	11.90921307

HMGCR	3-Hydroxy-3-Methylglutaryl-CoA Reductase	Protein Coding	P04035	11.79646683
LOC126862856	P300/CBP Strongly-Dependent Group 1 Enhancer GRCh37_chr19:11205098- 11206297	Functional Element		11.23807049
CYP27A1	Cytochrome P450 Family 27 Subfamily A Member 1	Protein Coding	Q02318	10.97994614
CETP	Cholesteryl Ester Transfer Protein	Protein Coding	P11597	10.91513538
COG2	Component Of Oligomeric Golgi Complex 2	Protein Coding	Q14746	10.25444412
PON1	Paraoxonase 1	Protein Coding	P27169	9.603778839
APOA1	Apolipoprotein A1	Protein Coding	P02647	9.314496994
NGID-106632268	APOB 3' Scaffold/Matrix Attachment Region	Functional Element		9.220497131
LIPC	Lipase C, Hepatic Type	Protein Coding	P11150	9.086737633
PPP1R17	Protein Phosphatase 1 Regulatory Subunit 17	Protein Coding	O96001	8.785656929
PON2	Paraoxonase 2	Protein Coding	Q15165	8.376523018
APOC3	Apolipoprotein C3	Protein Coding	P02656	8.00844574
APOA4	Apolipoprotein A4	Protein Coding	P06727	8.00844574
LCAT	Lecithin-Cholesterol Acyltransferase	Protein Coding	P04180	7.830331326
SCARB1	Scavenger Receptor Class B Member 1	Protein Coding	Q8WTV0	7.507217407
NOS3	Nitric Oxide Synthase 3	Protein Coding	P29474	7.39097023
DYNC2LI1	Dynein Cytoplasmic 2 Light Intermediate Chain 1	Protein Coding	Q8TCX1	7.374056816
MTTP	Microsomal Triglyceride Transfer Protein	Protein Coding	P55157	7.26172924
LRP1	LDL Receptor Related Protein 1	Protein Coding	Q07954	7.155556202
APOA5	Apolipoprotein A5	Protein Coding	Q6Q788	7.053258896
LIPA	Lipase A, Lysosomal Acid Type	Protein Coding	P38571	6.950642586
INS	Insulin	Protein Coding	P01308	6.77882719
CYP7A1	Cytochrome P450 Family 7 Subfamily A Member 1	Protein Coding	P22680	6.734242916
ANGPTL3	Angiopoietin Like 3	Protein Coding	Q9Y5C1	6.69925642
MIR7-3HG	MIR7-3 Host Gene	RNA Gene	Q8N6C7	6.670752525
FABP2	Fatty Acid Binding Protein 2	Protein Coding	P12104	6.485346794
SELE	Selectin E	Protein Coding	P16581	6.451208591
MSH6	MutS Homolog 6	Protein Coding	P52701	6.447939396

SERPINE1	Serpin Family E Member 1	Protein Coding	P05121	6.317782879
LMF1	Lipase Maturation Factor 1	Protein Coding	Q96S06	6.25308609
SOAT1	Sterol O-Acyltransferase 1	Protein Coding	P35610	6.009261131
SREBF2	Sterol Regulatory Element Binding Transcription Factor 2	Protein Coding	Q12772	5.972795486
ABCB1	ATP Binding Cassette Subfamily B Member 1	Protein Coding	P08183	5.953897476
FLNB	Filamin B	Protein Coding	O75369	5.953897476
EGF	Epidermal Growth Factor	Protein Coding	P01133	5.888225555
SCNN1A	Sodium Channel Epithelial 1 Subunit Alpha	Protein Coding	P37088	5.888225555
ADRB2	Adrenoceptor Beta 2	Protein Coding	P07550	5.888225555
FGB	Fibrinogen Beta Chain	Protein Coding	P02675	5.888225555
MYLIP	Myosin Regulatory Light Chain Interacting Protein	Protein Coding	Q8WY64	5.810139656
CBS	Cystathionine Beta-Synthase	Protein Coding	P35520	5.655964851
PLA2G7	Phospholipase A2 Group VII	Protein Coding	Q13093	5.607188225
EGFR	Epidermal Growth Factor Receptor	Protein Coding	P00533	5.486152172
APOA1-AS	APOA1 Antisense RNA	RNA Gene		5.486152172
SREBF1	Sterol Regulatory Element Binding Transcription Factor 1	Protein Coding	P36956	5.400567055
APOC2	Apolipoprotein C2	Protein Coding	P02655	5.381000042
ABCC6	ATP Binding Cassette Subfamily C Member 6	Protein Coding	O95255	5.193945885
NPC1L1	NPC1 Like Intracellular Cholesterol Transporter 1	Protein Coding	Q9UHC9	5.046584129
CLTC	Clathrin Heavy Chain	Protein Coding	Q00610	4.547242641
DAB2	DAB Adaptor Protein 2	Protein Coding	P98082	4.547242641
ADPRS	ADP-Ribosylserine Hydrolase	Protein Coding	Q9NX46	4.547242641
CLASRP	CLK4 Associating Serine/Arginine Rich Protein	Protein Coding	Q8N2M8	4.547242641
SORT1	Sortilin 1	Protein Coding	Q99523	4.515461445
OTC	Ornithine Transcarbamylase	Protein Coding	P00480	4.223255157
SPIRE2	Spire Type Actin Nucleation Factor 2	Protein Coding	Q8WWL2	4.223255157
MTHFR	Methylenetetrahydrofolate Reductase	Protein Coding	P42898	3.935173035
LMNA	Lamin A/C	Protein Coding	P02545	3.6772964
MT-TP	Mitochondrially Encoded TRNA-Pro (CCN)	RNA Gene		3.444824696

ALOX5AP	Arachidonate 5-Lipoxygenase Activating Protein	Protein Coding	P20292	3.329940081
ACE	Angiotensin I Converting Enzyme	Protein Coding	P12821	3.246582985
AGT	Angiotensinogen	Protein Coding	P01019	3.177936554
ADD1	Adducin 1	Protein Coding	P35611	3.106548309
ADIPOQ	Adiponectin, C1Q And Collagen Domain Containing	Protein Coding	Q15848	2.770146847
AGTR1	Angiotensin II Receptor Type 1	Protein Coding	P30556	2.632549047
CD93	CD93 Molecule	Protein Coding	Q9NPY3	2.632549047
HNRNPUL1	Heterogeneous Nuclear Ribonucleoprotein U Like 1	Protein Coding	Q9BUJ2	2.632549047
OR13G1	Olfactory Receptor Family 13 Subfamily G Member 1	Protein Coding	Q8NGZ3	2.632549047
F2	Coagulation Factor II, Thrombin	Protein Coding	P00734	2.447497368
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	Protein Coding	Q8N726	2.354623318
NR3C1	Nuclear Receptor Subfamily 3 Group C Member 1	Protein Coding	P04150	2.354623318
CDKN2B	Cyclin Dependent Kinase Inhibitor 2B	Protein Coding	P42772	2.354623318
ESR1	Estrogen Receptor 1	Protein Coding	P03372	2.187401533
SLCO1B1	Solute Carrier Organic Anion Transporter Family Member 1B1	Protein Coding	Q9Y6L6	2.187401533
IL6	Interleukin 6	Protein Coding	P05231	2.14330411
F7	Coagulation Factor VII	Protein Coding	P08709	2.094527483
CRP	C-Reactive Protein	Protein Coding	P02741	2.049357891
ITGB3	Integrin Subunit Beta 3	Protein Coding	P05106	2.039163589
NPPA	Natriuretic Peptide A	Protein Coding	P01160	2.039163589
SORL1	Sortilin Related Receptor 1	Protein Coding	Q92673	1.992626905
GPT	Glutamic--Pyruvic Transaminase	Protein Coding	P24298	1.98368597
C3	Complement C3	Protein Coding	P01024	1.947694302
NR1H4	Nuclear Receptor Subfamily 1 Group H Member 4	Protein Coding	Q96RI1	1.943452954
VCAM1	Vascular Cell Adhesion Molecule 1	Protein Coding	P19320	1.919431686
HNF4A	Hepatocyte Nuclear Factor 4 Alpha	Protein Coding	P41235	1.851513386
NR1H2	Nuclear Receptor Subfamily 1 Group H Member 2	Protein Coding	P55055	1.851513386
SP1	Sp1 Transcription Factor	Protein Coding	P08047	1.851513386
PPARA	Peroxisome Proliferator Activated Receptor Alpha	Protein Coding	Q07869	1.799456358

CRABP2	Cellular Retinoic Acid Binding Protein 2	Protein Coding	P29373	1.799456358
LEPR	Leptin Receptor	Protein Coding	P48357	1.733784437
GPIHBP1	Glycosylphosphatidylinositol Anchored High Density Lipoprotein Binding Protein 1	Protein Coding	Q8IV16	1.733784437
MPO	Myeloperoxidase	Protein Coding	P05164	1.66497004
TNF	Tumor Necrosis Factor	Protein Coding	P01375	1.66497004
CFH	Complement Factor H	Protein Coding	P08603	1.66497004
XDH	Xanthine Dehydrogenase	Protein Coding	P47989	1.66497004
CYP11B2	Cytochrome P450 Family 11 Subfamily B Member 2	Protein Coding	P19099	1.66497004
XK	X-Linked Kx Blood Group Antigen, Kell And VPS13A Binding Protein	Protein Coding	P51811	1.66497004
MIR130A	MicroRNA 130a	RNA Gene		1.66497004
PLTP	Phospholipid Transfer Protein	Protein Coding	P55058	1.648199201
ADRB3	Adrenoceptor Beta 3	Protein Coding	P13945	1.648199201
CYP27B1	Cytochrome P450 Family 27 Subfamily B Member 1	Protein Coding	O15528	1.644892335
NR1H3	Nuclear Receptor Subfamily 1 Group H Member 3	Protein Coding	Q13133	1.644892335
CRYAA	Crystallin Alpha A	Protein Coding	P02489	1.644892335
CYP7B1	Cytochrome P450 Family 7 Subfamily B Member 1	Protein Coding	O75881	1.644892335
NR1I2	Nuclear Receptor Subfamily 1 Group I Member 2	Protein Coding	O75469	1.644892335
FDX1	Ferredoxin 1	Protein Coding	P10109	1.644892335
LPXN	Leupaxin	Protein Coding	O60711	1.579384804
EMSLR	E2F1 MRNA Stabilizing LncRNA	RNA Gene		1.579384804
LIPE	Lipase E, Hormone Sensitive Type	Protein Coding	Q05469	1.44157815
USF1	Upstream Transcription Factor 1	Protein Coding	P22415	1.44157815
RXRG	Retinoid X Receptor Gamma	Protein Coding	P48443	1.44157815
TXNIP	Thioredoxin Interacting Protein	Protein Coding	Q9H3M7	1.44157815
USF2	Upstream Transcription Factor 2, C-Fos Interacting	Protein Coding	Q15853	1.44157815
FADS3	Fatty Acid Desaturase 3	Protein Coding	Q9Y5Q0	1.44157815
HYPLIP2	Hyperlipidemia, Combined, 2	Genetic Locus		1.44157815
SNORD118	Small Nucleolar RNA, C/D Box 118	RNA Gene		1.432709336

TMX2-CTNND1	TMX2-CTNND1 Readthrough (NMD Candidate)	RNA Gene		1.432709336
VLDLR	Very Low Density Lipoprotein Receptor	Protein Coding	P98155	1.401305914
CXCL8	C-X-C Motif Chemokine Ligand 8	Protein Coding	P10145	1.383932829
PSRC1	Proline And Serine Rich Coiled-Coil 1	Protein Coding	Q6PGN9	1.383932829
SNX17	Sorting Nexin 17	Protein Coding	Q15036	1.377227902
TRIB1	Tribbles Pseudokinase 1	Protein Coding	Q96RU8	1.328568816
CD40LG	CD40 Ligand	Protein Coding	P29965	1.262896895
IL1RN	Interleukin 1 Receptor Antagonist	Protein Coding	P18510	1.262896895
ANXA5	Annexin A5	Protein Coding	P08758	1.262896895
CCR3	C-C Motif Chemokine Receptor 3	Protein Coding	P51677	1.262896895
CCR4	C-C Motif Chemokine Receptor 4	Protein Coding	P51679	1.262896895
FGF23	Fibroblast Growth Factor 23	Protein Coding	Q9GZV9	1.262896895
CXCR1	C-X-C Motif Chemokine Receptor 1	Protein Coding	P25024	1.262896895
TNNT1	Troponin T1, Slow Skeletal Type	Protein Coding	P13805	1.262896895
MIR505	MicroRNA 505	RNA Gene		1.262896895

Приложение Б

Сертификат участника школы по биоинформатике



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКИЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
МЕДИЦИНСКИЙ УНИВЕРСИТЕТ

имени Н.И. Пирогова

МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

СЕРТИФИКАТ

выдан

*Гориченко
Ивану Вадимовичу*

за участие в школе по биоинформатике
"Анализ данных ChIP-seq"

Продолжительность — 16 академических часов

Время проведения: 23,24 апреля 2021 года

Ректор

С.А. Лукьянов

Москва 2021