

На правах рукописи

ЧЕКАЛИН ЕВГЕНИЙ ВИТАЛИЕВИЧ

**ВЛИЯНИЕ ЭПИГЕНЕТИЧЕСКИХ ФАКТОРОВ НА РАЗВИТИЕ
ИММУНОВОСПАЛИТЕЛЬНЫХ ЗАБОЛЕВАНИЙ КОЖИ**

03.02.07 – Генетика

03.01.09 – Математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2019

Работа выполнена в лаборатории функциональной геномики Федерального государственного бюджетного учреждения науки Института общей генетики им. Н.И. Вавилова Российской академии наук.

Научный руководитель: кандидат биологических наук, доцент
БРУСКИН Сергей Александрович

Научный консультант:

Официальные оппоненты:

- Доктор биологических наук, кандидат физико-математических наук, профессор **Миронов Андрей Александрович**, Федеральное государственное образовательное учреждение высшего профессионального образования "Московский государственный университет имени М.В. Ломоносова", факультет биоинженерии и биоинформатики, г. Москва

- Кандидат биологических наук **Поверенная Екатерина Владимировна**, старший научный сотрудник Федерального государственного бюджетного научного учреждения «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича», г. Москва

Ведущая организация: Федеральное государственное учреждение «Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук», г. Москва

Защита состоится «__» _____ 2019 г. в ____ часов на заседании диссертационного совета Д.002.214.01 в Федеральном государственном бюджетном учреждении науки Институте общей генетики им. Н.И. Вавилова Российской академии наук по адресу: 119991, Москва, ул. Губкина, 3.

С диссертацией и авторефератом можно ознакомиться в библиотеке и на сайте Института www.vigg.ru, тел. 8-499-135-14-31.

Автореферат разослан «__» _____ 2019 года.

Ученый секретарь диссертационного совета,
доктор биологических наук

Горячева И. И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы.

Псориаз - комплексное воспалительное заболевание кожи, в которое вовлечены гиперпролиферация кератиноцитов и их aberrантная дифференцировка, а также ангиогенез дермы [Gudjonsson et.al. 2010]. Псориаз – это сложная патология, относящаяся к иммуновоспалительным заболеваниям кожи, такие, для которой велик вклад наследуемых факторов. В патогенез псориаза вовлечено большое количество взаимодействующих генов и белков, при этом считается, что в патогенез заболевания большой вклад вносят нарушения регуляции транскрипции. Были предложены модели терапий, которые нацелены на ингибирование экспрессии транскрипционных факторов STAT3 и NFκB [Andres, 2013; Goldminz, 2013]. С появлением большого количества экспериментов по иммунопреципитации хроматина, появилась возможность предсказать активность транскрипционных факторов на основе сетей белок-белковых (PPI), ген-генных (GGI) и белок-ДНК взаимодействий, что дает возможность надеяться на прояснение механизмов возникновения псориаза, включая вклад регуляторных взаимодействий.

Последнее время активно обсуждается вопрос роли эпигенетических факторов в патогенезе псориаза, включая изменения метилирования ДНК [Zhang et.al., 2011; Zhang et.al., 2013]. Было показано, что в окрестностях генов p16INK4, P14ARF, LFA-1, ассоциированных с псориазом [Zarrabeitia *et al.*, 1989; Schon *et al.*, 2000; Rocha-Pereira *et al.*, 2004] картина метилирования ДНК при развитии псориаза изменяется. Также было показано, что мишенями метилирования являются сайты связывания транскрипционных регуляторов ДНК [Hu *et al.*, 2013]. Таким образом, детальное изучение роли метилирования ДНК и активности транскрипционных регуляторов в патогенезе псориаза может стать ключом для понимания молекулярных механизмов развития заболевания, а в перспективе для разработки новых терапий. Поэтому актуальным вопросом является изучение влияния метилирования ДНК на регуляцию экспрессии генов, ассоциированных с псориазом, в частности, путем изменения сайтов связывания транскрипционных факторов.

Цель исследования:

Основной целью исследования стала характеристика роли транскрипционных факторов и метилирования ДНК в патогенезе псориаза с помощью транскриптомных, эпигеномных и биоинформатических подходов.

Исходя из поставленной цели, были сформулированы следующие задачи исследования:

1. Провести анализ дифференциальной экспрессии генов и анализ дифференциального метилирования ДНК в образцах здоровой кожи и кожи, поражённой псориазом;
2. Выявить корреляцию между уровнями метилирования ДНК и экспрессии генов;
3. С помощью маркерных графов выявить список транскрипционных факторов, регулирующих дифференциально экспрессирующиеся гены в поражённой псориазом коже;
4. Оценить изменение профилей метилирования ДНК в областях связывания транскрипционных факторов;
5. Выявить список генов, изменение экспрессии которых при псориазе может объясняться изменениями в уровнях метилирования сайтов посадки их транскрипционных регуляторов.

Научная новизна и значимость работы.

Впервые проведён мета-анализ всех находящихся в открытом доступе RNA-Seq экспериментов по исследованию сравнительной экспрессии генов в образцах здоровой кожи, и кожи поражённой псориазом. Анализ позволил выявить список из 919 дифференциально экспрессирующихся генов (ДЭГ). Впервые проведён мета-анализ всех экспериментов профилирования метилирования ДНК в образцах здоровой кожи и кожи, поражённой псориазом, и выявлен список генов с дифференциально метилированными участками (ДМГ). Впервые продемонстрирована ассоциация дифференциального метилирования ДНК и дифференциальной экспрессии генов кластера S100A в образцах здоровой кожи и кожи, поражённой псориазом. Предсказаны ключевые транскрипционные факторы, участвующие в патологии псориаза, а также выдвинуто предположение о связи метилирования ДНК и вклада активности транскрипционных факторов в регуляцию экспрессии конкретных генов в патогенезе псориаза.

Впервые показана роль эпигенетических изменений в конкретной ткани при изменении транскрипционной активности, связанной с патологическими изменениями, что имеет большое значение для разработки модели заболевания.

Идентифицированные ДЭГ и ДМГ могут быть рассмотрены как мишени для разработки новых подходов к созданию высокоэффективных терапевтических препаратов для лечения псориаза, а также при оценке ответа больных на тот или иной тип лечения.

Апробация результатов работы

Результаты проведенных исследований были представлены на международных и российских конференциях.

Публикации

Автором опубликовано 6 печатных работ, в том числе 4 статьи по теме диссертационной работы в журналах, входящих в перечень рецензируемых научных журналов и изданий, рекомендованных ВАК Минобрнауки для опубликования основных научных результатов диссертации.

Структура и объем диссертации

Диссертационная работа состоит из введения, обзора литературы, материалов и методов исследования, результатов и обсуждения, а также выводов и списка цитируемой литературы. Работа изложена на 153 страницах машинописного текста, включает 23 таблицы и 15 рисунков, а также приложение, содержащее 5 таблиц. Список цитируемых литературных источников включает 142 наименования.

СОДЕРЖАНИЕ РАБОТЫ

Материалы и методы исследования

Для оценки роли эпигенетических факторов в развитии псориаза, нам было необходимо оценить, насколько картина экспрессии генов отличается в пораженной и здоровой коже, как меняется картина метилирования ДНК, экспрессия генов, кодирующих транскрипционные факторы, а также оценить возможную роль метилирования ДНК в качестве модификатора связывания транскрипционных факторов с ДНК (Рисунок 1). Поэтому, на первом этапе работы мы проанализировали все находящиеся в открытом доступе данные, полученные из RNA-seq экспериментов по псориазу, выявили списки генов (далее ДЭГ), для которых характерна дифференциальная экспрессия между образцами поражённой псориазом и здоровой кожи, а также оценили обогащение ДЭГ генами с охарактеризованными функциями. На втором этапе мы оценивали уровень метилирования ДНК в поражённой и здоровой коже, а также оценивали обогащение функциональных каскадов генами с дифференциально метилированными локусами (далее ДМЛ). После того как мы оценили уровень дифференциальной экспрессии и дифференциального метилирования ДНК, различными способами проводился поиск закономерностей между уровнем экспрессии генов и уровнем метилирования ДНК в соответствующих функциональных регионах. На следующем этапе работы мы выявляли транскрипционные факторы, которые участвуют в регуляции гиперпролиферации клеток кожи. Для этого мы собрали базу транскрипционных факторов (далее ТФ) и их мишеней, пользуясь архивами экспериментов по иммунопреципитации хроматина ChIPBase 2.0. Из этих данных был построен оргграф всех возможных ТФ и их мишеней и определен список ТФ, которые были обогащены ДЭГ, идентифицированными в нашей работе, чтобы выявить список наиболее важных ТФ для псориаза. Для того чтобы выявить ТФ, которые непосредственно регулируют работу гена-мишени, мы использовали алгоритм сборки графа с максимальной информацией. На последнем этапе работы мы оценивали пересечение сайтов связывания ТФ с сайтами дифференциального метилирования ДНК или с сайтами, в которых была обнаружена корреляция между экспрессией и метилированием ДНК.

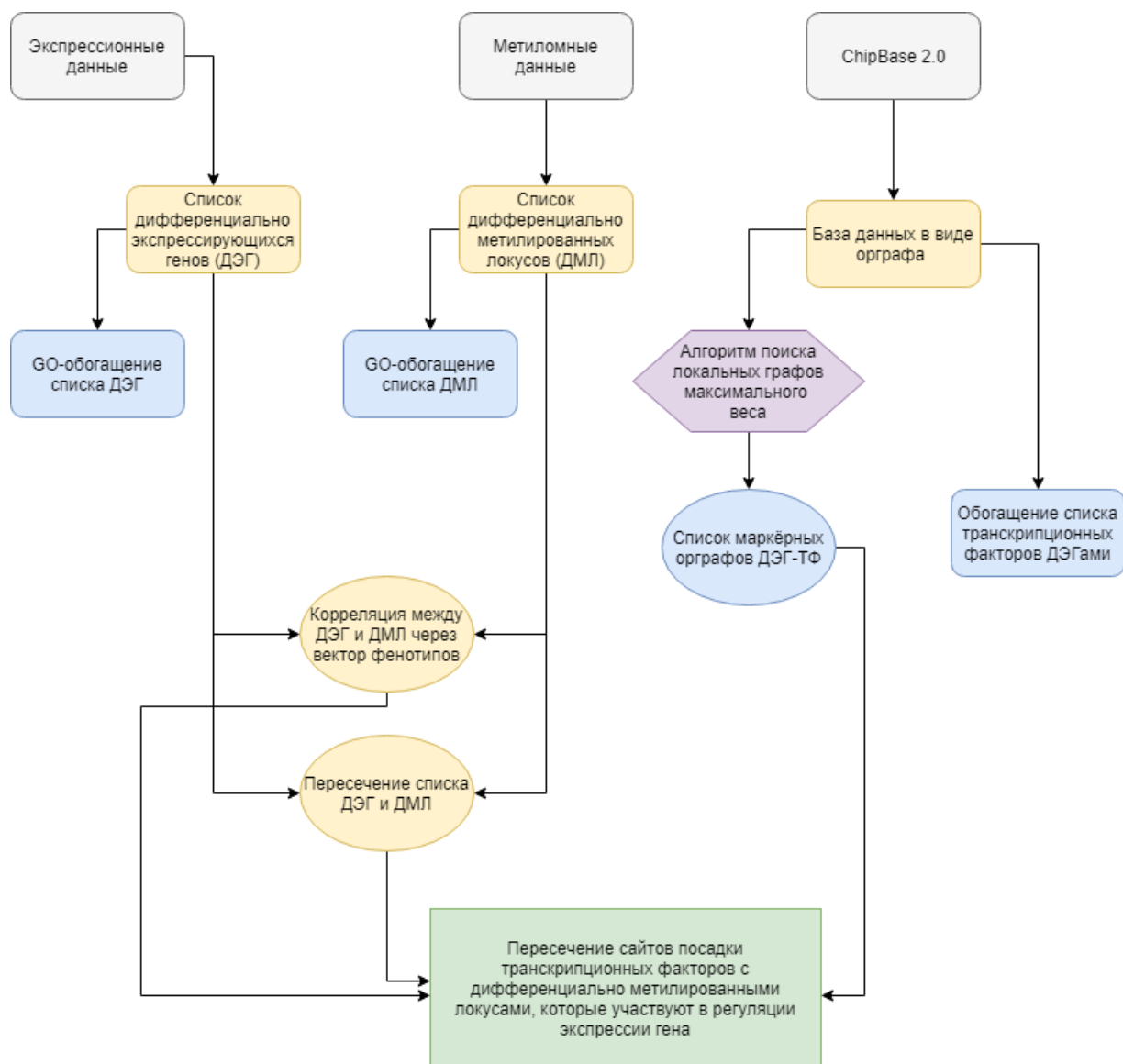


Рисунок 1. Принципиальная схема исследования. Серые блоки - входные данные, жёлтые блоки, жёлтые эллипсы и синие эллипсы - промежуточные результаты, зелёный блок - конечный результат исследования, на которое оно было направлено. В блоках в скобках курсивом указаны главы, в которых интерпретируются результаты.

Из открытых источников были собраны базы данных экспериментов по экспрессии генов в образцах кожи, взятых у больных псориазом и здоровых индивидуумов. Всего в собранную нами базу вошли данные по экспрессии генов 227 образцов поражённой псориазом кожи и 139 образцов здоровой кожи (GSE54456, GSE84204, GSE78023, GSE89725, GSE92472, GSE74697, GSE54456, GSE50598, GSE67785, GSE66511, SE63980, GSE47944, GSE41745). Образцы в этих исследованиях были обработаны с помощью утилит *trimmomatic*, *STAR*, а также библиотек *R combat* и *edgeR* для того, чтобы получить матрицу нормализованных значений FPKM, далее приступили к её обработке. Из полученной матрицы FPKM мы смогли оценить среднюю экспрессию генов по группам (рассчитав показатель *Log Fold-Change*), а также оценить достоверность

дифференциальной экспрессии с помощью t-теста Стьюдента и поправки FDR Бенджамини-Хохберга. За достоверную дифференциальную экспрессию гена мы принимали изменения экспрессии более чем в 1.5 раза ($FC \geq \pm 1,5$) и уровне значимости p-value с поправкой $FDR < 0.01$.

Из наших данных и открытых источников была собрана база данных экспериментов по анализу метилирования ДНК в коже больных псориазом и здоровых индивидуумов на чипах Illumina Methylation BeadChip 450k (наши данные и эксперименты GSE63315, GSE73894, GSE115797; количество образцов здоровой кожи=83, поражённой псориазом=189). Для того чтобы получить beta-values, отражающие уровень метилирования ДНК, полученных с помощью Illumina Infinium HumanMethylation450 BeadChip, мы использовали пакет программной среды R IMA. Полученные b-value подвергались квантильной нормализации. После квантильной нормализации данные были использованы для дифференциального анализа уровня метилирования генов. Нормализованные значения b-value использовали для получения списка дифференциально метилированных локусов. Для этого было вычислено среднее значение метилирование локуса внутри каждой группы, после чего получали разницу среднего метилирования ДНК между группами и считали достоверность разницы между векторами групп метилирования с помощью t-test. Локус считали достоверно дифференциально метилированным, если разница beta-значений была по модулю больше 0.1, а p-value t-теста с поправкой FDR был меньше 0.01. Так как пробы на чипе Illumina Methylation BeadChip 450k аннотированы только на геномы hg17 и hg18, а результаты иммунопреципитации хроматина в ChipBase 2.0 аннотированы с hg19, координаты проб модифицировали для соответствия hg19-сборке с помощью пакета R 3.4.4 rtracklayer v1.38.3 и GenomciRanges 1.30.3.

Для того чтобы оценить корреляцию между метилированием ДНК и уровнем генной экспрессии использовался принцип транзитивности корреляций. Несмотря на то, что в настоящее время отсутствуют данные, одновременно характеризующие метилирование регуляторных районов генов (например, промоторов) и активность транскрипции этих промоторов, до определенной степени такую корреляцию можно оценить, используя фенотипы (здоровая или больная кожа) в качестве обуславливающей переменной. Если коэффициенты экспрессии и фенотипа R_{EP} и корреляции метилирования ДНК и фенотипа R_{MP} велики по своему абсолютному значению, то это накладывает ограничения на корреляции экспрессии и метилирования ДНК R_{EM} . А именно, если ожидается, что имеет место антикорреляция $R_{EM} < 0$, а $R_{EP} > 0$ и $R_{MP} < 0$ то:

$$R_{EM} < R_{EP}R_{MP} - \sqrt{1 - R_{EP}^2 - R_{MP}^2 + R_{EP}^2 R_{MP}^2} < 0$$

В случае положительной корреляции $R_{EM} > 0$, а $R_{EP} < 0$ и $R_{MP} > 0$ то:

$$R_{EM} > R_{EP}R_{MP} + \sqrt{1 - R_{EP}^2 - R_{MP}^2 + R_{EP}^2R_{MP}^2} > 0$$

Для изучения связи ДНК-белкового взаимодействия и экспрессии гена использовался метод агрегации базы данных сайтов посадки транскрипционных регуляторов мы выбрали базу данных ChipBase 2.0 [Zhou, 2017]. После агрегации всех списков ТФ → ген-мишень, мы получили базу данных сайтов посадки транскрипционных факторов в диапазоне [-1kb:1kb] нуклеотид от точки старта транскрипции (рисунок 2).

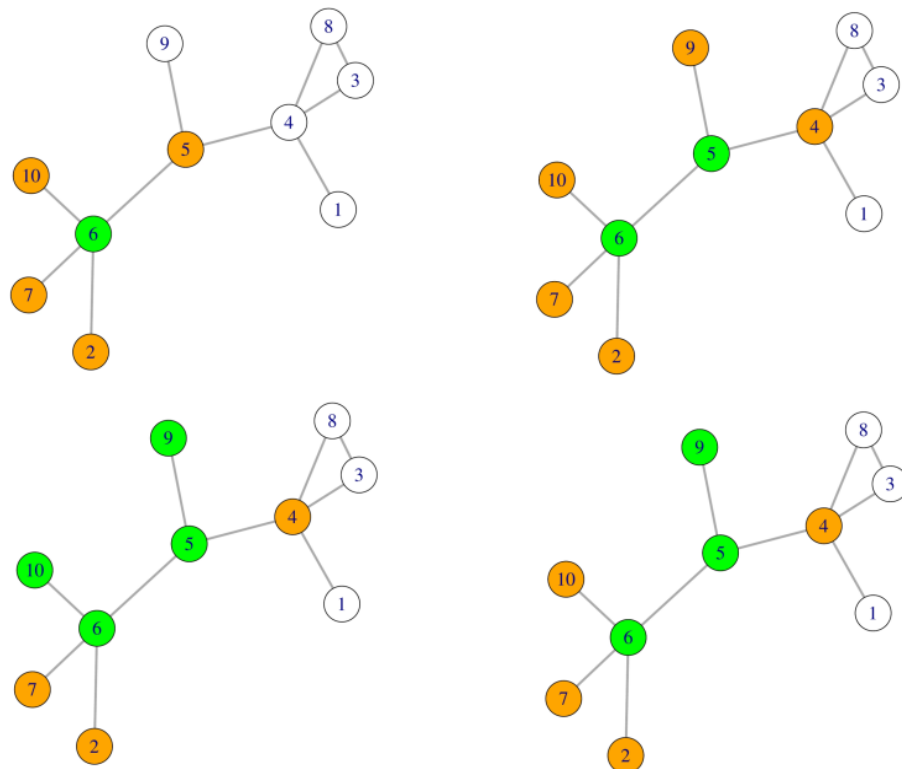


Рисунок 2. Жадная сборка подграфа активных ТФ. На первом этапе собираются соседи зерна подграфа (узел 6). В данном случае это узлы 2, 5, 7, 19. Далее выявляется пара с наибольшей MI между усреднённой экспрессией и вектором фенотипов (5). Узел 5 добавляется в подграф. Далее для подграфа 6-5 выбираются соседние узлы (2, 4, 7, 10) и, аналогично, вычисляется пара узел-граф с максимальной MI (узел 9). Присоединение узлов к подграфу происходит до тех пор, пока не будет собран подграф с максимальной MI между усреднённым уровнем экспрессии и вектором фенотипов.

С помощью пакета igraph v1.0.1 мы получили направленный граф $G = \{N, E\}$, состоящий из наборов вершин $\{N\}$ и рёбер $\{E\}$. Вершины графа $\{N\}$ содержат гены и атрибут type, где атрибут type- фактор из двух уровней: target, если вершина является белок-кодирующим геном-мишенью для ТФ, либо ТФ, если вершина является транскрипционным фактором. Ребро $\{E\} = \{N1, N2\}$

существует и ориентировано $\{N1 \rightarrow N2\}$, если из пары узлов $\{N1, N2\}$ $N1$ является транскрипционным фактором, а $N2$ - мишенью для этого транскрипционного фактора согласно базе данных ChIPBase v2.0.

Для идентификации графов активных транскрипционных факторов с помощью локального графа максимальной взаимной информации мы использовали жадную сборку маркерного графа на основе максимизации индекса взаимной информации между усреднённым вектором экспрессии графа и вектором фенотипов, где 0 - это поражённая кожа, а 1 - здоровая кожа (рис 1.)

Результаты исследования.

Оценка полногеномных профилей экспрессии при псориазе.

Всего в собранную нами базу вошли данные по экспрессии генов 227 образцов поражённой псориазом кожи, и 139 образцов здоровой кожи, что на данный момент является наиболее масштабным мета-анализом экспрессии генов между поражённой псориазом и здоровой кожей. Сырые прочтения в формате .SRA были преобразованы в .fastq формат, после чего были произведены процедуры тримминга и выравнивания триммированных прочтений на геном hg38 с помощью программы STAR 2.6.0a. Для получения матрицы FPKM последующий анализ производили в R с помощью пакета DESeq2. Из полученной матрицы FPKM мы смогли оценить среднюю экспрессию генов по группам (рассчитав показатель Log Fold-Change), а также оценить достоверность дифференциальной экспрессии с помощью t-теста Стьюдента и применения поправки FDR Бенджамини-Хохберга. За достоверную дифференциальную экспрессию гена мы принимали изменение экспрессии более чем в 2 раза ($\text{Log Fold-Change} \geq \pm 1,5$) и уровне значимости p-value с поправкой $\text{FDR} < 0.01$. Мы провели анализ дифференциальной экспрессии белок-кодирующих генов между поражённой и здоровой кожей. Всего было проанализировано 57393 транскриптов, из которых 919 были достоверными ДЭГ между поражённой псориазом и здоровой кожей. Причём экспрессия 389 генов была понижена в поражённой псориазом коже, а экспрессия остальных 530 была достоверно повышена в поражённой коже. Так как в процессе развития псориаза наряду с гиперпролиферацией кератиноцитов происходит привлечение иммунных клеток и инициация воспаления, логично было бы предположить, что среди дифференциально экспрессирующихся генов присутствуют гены, которые участвуют в иммунном ответе [Caruntu, 2014]. Сравнение генов со списками, полученными предыдущими исследователями показало, что для 75% ДЭГ в предыдущих работах уже была выявлена дифференциальная экспрессия.

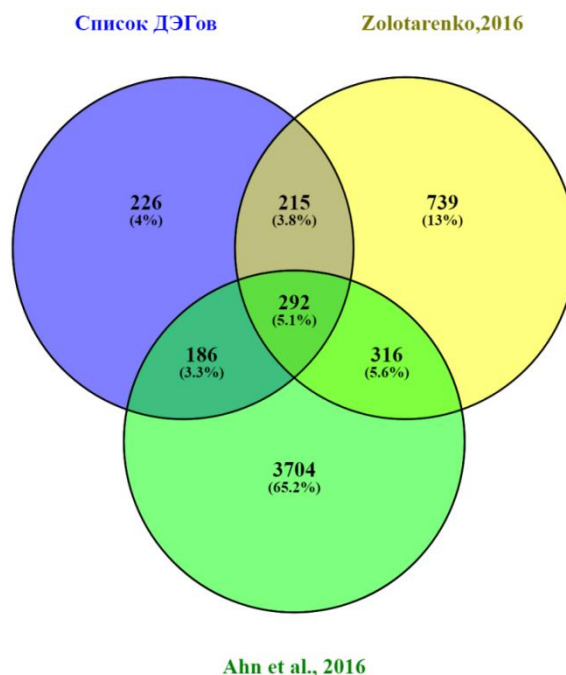


Рисунок 3. Пересечение списка ДЭГ полученных в этом исследовании, а также в исследованиях, которые были опубликованы ранее в работах 2016 года, Zolotarengo et al. и Ahn et al.

При этом, сравнение списка ДЭГ, полученных нами с, например, Zolotarengo et al, 2016 и Ahn et al., 2016 выявил т.н. «коровый» список из 292 транскриптов, которые присутствуют во всех трёх исследованиях (рисунок 3). Интересно, что в этот список попало большое количество генов, которые участвуют как в иммунном ответе, так и в дифференцировке клеток кожи и ее ороговении.

Таблица 1. Топ-10 ДЭГ, которые были обнаружены во всех трёх исследованиях.

Имя гена	logFC	logCPM	PValue	FDR
<i>S100A7A</i>	9.840	9.985	1E-158	4E-155
<i>DEFB4A</i>	9.244	6.101	3E-149	7E-146
<i>SPRR2A</i>	8.776	10.020	3E-210	5E-206
<i>SPRR2F</i>	8.740	8.218	9E-134	1E-130
<i>IL36A</i>	7.854	4.326	3E-151	7E-148
<i>SPRR2C</i>	7.357	4.374	2E-156	6E-153
<i>TMPRSS11D</i>	7.077	5.193	9E-199	9E-195
<i>S100A12</i>	6.943	4.538	1E-205	1E-201
<i>PI3</i>	6.691	10.659	9E-84	3E-81
<i>IL19</i>	6.266	2.921	3E-77	9E-75

Например, в топ-10 генов, экспрессия которых повышена в поражённой псориазом коже, попали гены антибактериального комплекса S100A7A, S100A12, IL19и IL36A (Таблица 1). Помимо списка генов, которые уже были ранее идентифицированы как ДЭГ, в ходе данной работы мы выявили гены, которые ранее не были ассоциированы с псориазом. Для ряда генов данного списка можно предположить роли, которые они играют в развитии изучаемого заболевания. Например, в список генов, для которых впервые показаны изменения экспрессии при псориазе, попал ген TMPRSS11A, который ассоциирован с клеточной пролиферацией, а также ген иммуноглобулина IGHG4. MFSD2B- это ген, который участвует в экспорте фактора S1P на поверхности эритроцитов, который, в свою очередь, привлекает в очаг поражения иммунные клетки и может являться мишенью для разработки терапий.

Оценка уровня метилирования ДНК в псориазе с помощью мета-анализа

Для полногеномного анализа метилирования ДНК в поражённой псориазом и здоровой кожи, мы, аналогично анализу экспрессии генов, провели агрегацию данных из статей, размещённых в базе данных GeoDataSets в NCBI и наших собственных результатов анализа дифференциального метилирования ДНК с помощью чипов анализа метилирования Illumina Methylation BeadChip 450k. В ходе такого анализа мы собрали базу данных по метилированию ДНК для 127 образцов поражённой псориазом кожи и 83 образцов здоровой кожи. В ходе данного анализа мы выявили 899 дифференциально метилированных локусов ДНК между поражённой псориазом и здоровой кожей, причём уровень метилирования ДНК 485 локусов был значимо понижен в поражённой коже, а остальные 414 локуса имели повышенный уровень метилирования ДНК в поражённой коже больных псориазом. Уровень метилирования ДНК в локусах, ассоциированных с белками S100A, был понижен в поражённой псориазом кожей. Так как было показано, что гены группы S100A дифференциально экспрессируются между поражённой и здоровой кожей, а все 3 гена, для которых был обнаружен достоверный уровень дифференциального метилирования ДНК, входили в этот список, логично предположить, что и в псориазе метилирование ДНК в значительной степени регулирует уровень активности этих генов (таблица 2).

Таблица 2. Дифференциально метилированные локусы ДНК между поражённой и здоровой кожей, которые ассоциированы с белками группы S100A.

Название пробы	Имя гена	Локус пробы	Δ значений поражённой и здоровой кожи	Бета-и	p-value поправкой Бонферрони
cg12067024	S100A7A	TSS1500	-0.116		1.73E-10
cg06355720	S100A9	3'UTR	-0.120		4.17E-15
cg01431057	S100A8	Body	-0.100		4.22E-11
cg20335425	S100A8	5'UTR	-0.127		7.93E-16
cg16139316	S100A9	5'UTR	-0.157		1.78E-15
cg03165378	S100A9	TSS1500	-0.109		1.38E-14
cg17496887	S100A7A	TSS1500	-0.124		7.95E-11
cg20169988	S100A9	Body	-0.133		8.75E-15
cg03514239	S100A9	TSS1500	-0.101		1.26E-13

Последующая интеграция обнаруженных нами результатов в результаты, полученные с помощью этих методов позволят нам идентифицировать новые локусы дифференциального метилирования ДНК и оценить роль этого механизма в регуляции заболевания [Chandra et al., 2018] (рисунок 4).

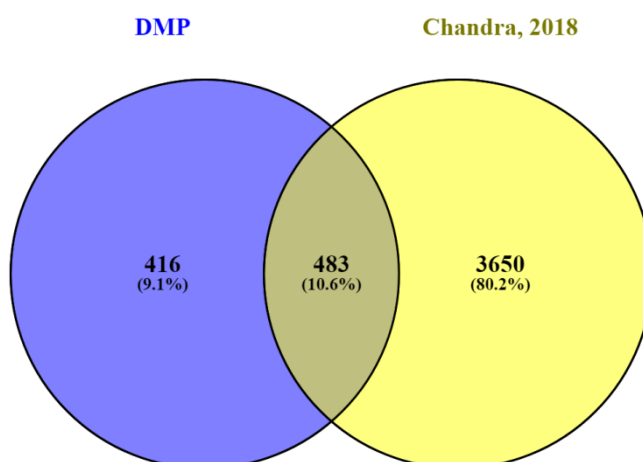


Рисунок 4. Диаграмма Венна для выявления пересечения списка ДМЛ, которые мы обнаружили в данной работе с ДМЛ, которые выявили Chandra с коллегами [Chandra et al., 2018].

Поиск локусов ДНК с достоверной корреляцией между уровнем экспрессии и уровнем метилирования ДНК

Для того чтобы корреляция между экспрессией и метилированием была достоверной для одного гена и локуса метилирования ДНК в этом гене, должно выполняться неравенство $R_{EP}^2 + R_{MP}^2 > 1$, где R_{EP} - корреляция между вектором

экспрессии этого гена и вектором фенотипов образцов, составляющих вектор экспрессии, R_{EM} - корреляция между вектором бета-значений метилирования ДНК и векторов фенотипов. (рисунок 5). Всего мы выявили 149 локусов с транзитивной корреляцией. Интересно, что в этот список попало большое количество локусов, ассоциированных с белками комплекса S100A. Это лишний раз подтверждает гипотезу о влиянии уровня метилирования ДНК на регуляцию экспрессии этих белков. Помимо белков комплекса S100A, корреляция была определена для генов RNYND1, PI3, ZBTB4 и ZC3H12A.

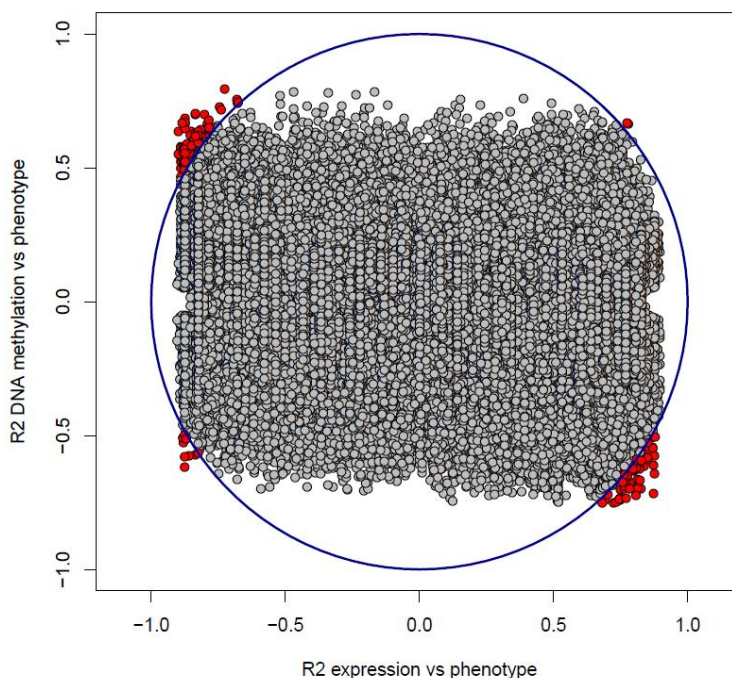


Рисунок 5. Гены и локусы на чипе Illumina Methylation BeadChip 450k с определённой экспрессией между уровнем экспрессии и метилирования ДНК. Красным отмечены пробы, для которых выполняется условие определённости корреляции. Для локусов, точки которых попали в левый-нижний угол, определена позитивная корреляция между экспрессией и метилированием ДНК, а для локусов, которые попали за пределы круга в левом-верхнем и правом-нижнем углу определена негативная корреляция между экспрессией генов и метилированием ДНК.

В регуляции гена RNYND1 ранее не была выявлена компонента метилирования ДНК, хотя в предыдущих работах этот ген был продемонстрирован как мишень для терапий, которые базируются на дифференциальном изменении процесса окислительно-восстановительных реакций в коже, поражённой псориазом [Gu, 2015]. Также выдвигалось предположение об альтернативном составе GC в регуляторных областях гена RNYND1 [Swindell, 2014], что подтверждается его корреляцией с вектором фенотипов (таблица 3).

Таблица 3. Топ-20 локусов с определённой корреляцией между уровнем метилирования ДНК локуса и уровнем экспрессии гена, который ассоциирован с этим локусом.

Ген	Probe Id	R^2_{ep}	R^2_{ep} p-value	R^2_{mp}	R^2_{mp} p-value	$R^2_{ep}+R^2_{mp}$
PHYHD1	cg14153069	0.872	3.90E-55	-0.612	2.24E-15	1.135
ANKRD33B	cg05524458	0.815	2.52E-43	-0.688	6.12E-21	1.138
MFHAS1	cg01274122	-0.875	7.94E-56	-0.616	1.11E-15	1.146
PLXNA2	cg19721764	0.813	6.27E-43	-0.698	1.02E-21	1.148
CYP2S1	cg19430423	-0.725	2.57E-30	0.793	6.57E-31	1.156
MCC	cg02551980	0.825	3.21E-45	-0.700	7.09E-22	1.172
SLC29A1	cg22330763	0.807	6.93E-42	-0.723	6.09E-24	1.174
S100A9	cg03514239	-0.874	1.01E-55	0.644	1.58E-17	1.180
ZC3H12A	cg07813495	-0.835	5.03E-47	0.701	6.03E-22	1.188
S100A9	cg20169988	-0.874	1.01E-55	0.664	5.08E-19	1.206
S100A9	cg03165378	-0.874	1.01E-55	0.665	4.05E-19	1.207
PI3	cg02733351	-0.899	2.27E-62	0.633	8.47E-17	1.210
S100A9	cg06355720	-0.874	1.01E-55	0.671	1.45E-19	1.215
S100A8	cg20335425	-0.876	3.27E-56	0.675	7.75E-20	1.224
SPRR2A	cg18766755	-0.884	2.40E-58	0.667	3.09E-19	1.228
S100A9	cg16139316	-0.874	1.01E-55	0.686	9.42E-21	1.235
PHYHD1	cg13613439	0.872	3.90E-55	-0.718	1.82E-23	1.276

Идентификация подграфов транскрипционных регуляторов, участвующих в регуляции экспрессии своих генов-мишеней

В ходе данного этапа работы мы искали регуляторы транскрипции конкретных генов с помощью разработанного нами алгоритма поиска графа с максимальной взаимной информацией между усреднённым значением генной экспрессии между всеми элементами графа и вектором фенотипов. Дело в том, что очень часто транскрипционные регуляторы не являются дифференциально экспрессирующимися генами. Для того чтобы выявить их, используют различные подходы. Наиболее консервативным подходом является поиск транскрипционных регуляторов, которые обогащены дифференциально экспрессирующимися генами по сравнению с базой данных, которая содержит информацию об этих ТФ и их мишенях. Тем не менее, такой анализ не даёт сведений о регуляторной активности транскрипционных регуляторов в вопросе контроля экспрессии конкретных генов-мишеней. Для того чтобы выявить генные сети, которые содержат транскрипционный регулятор и ген-мишень, можно применить модификацию алгоритма поиска маркеров, который впервые был предложен ещё в 2007 году [[Chuang, 2007](#)]. Если мы будем рассматривать пару транскрипционный фактор-мишень, то, при дифференциальной активности ТФ в поражённой и здоровой коже, взаимная информация усреднённого значения экспрессии и вектора фенотипов такой пары должна быть выше, чем взаимная информация вектора экспрессии гена-мишени и вектора фенотипов, так как псориаз - заболевание многофакторное и экспрессия одного гена может варьировать внутри выборки. А с помощью дополнительной энтропии, которая происходит из вектора экспрессии транскрипционного фактора, мы сможем «уточнить» разницу между поражённой и здоровой кожей. Более того, если в регуляции транскрипционного фактора участвует ещё один регулятор экспрессии, то такая структура, по идее, должна ещё лучше объяснять разницу между поражённой и здоровой кожей. Для того чтобы выявить такие сети-регуляторы, было предложено применить подход, основанный на поиске локального орграфа вокруг ДЭГ с максимальной взаимной информацией между усреднённым вектором экспрессии этого орграфа и вектором фенотипов. На вход мы подавали список из 930 дифференциально экспрессирующихся генов, которые использовали в качестве зёрен агрегации подграфов. На первом этапе работы мы проверяли вхождение всего списка дифференциально экспрессирующихся генов во множество генов, содержащихся в списке вершин орграфа, полученного нами после агрегации базы данных ChipBase 2.0, таких генов оказалось 741. После этого мы итерационно жадно собирали граф с максимальной взаимной информацией согласно алгоритму, приведённому в

главе «материалы и методы». На выходе мы получили таблицу, которая содержит информацию о зерне агрегации подграфа, вершинах подграфа, количестве вершин в графе, а также взаимной информации ДЭГ, подграфа и пермутационных тестов. Мы исключали подграфы, FDR хотя бы одного теста которых был больше, чем 0.01. Всего таким способом мы выявили 503 подграфа с достоверным значением p-value с поправкой FDR.

После агрегации подграфов, мы оценили количество ТФ, которые попали в разные графы. Наибольшую частоту встречаемости имел ТФ *NFKB1*, который в предыдущих работах неоднократно ассоциировался с регуляцией развития псориаза [Butt, 2005; Smyth, 2006]. Помимо него, также большую частоту (больше, чем в 20% графов) имели транскрипционные факторы *KLF15*, *ENF*, *STAT1*. Для транскрипционных регуляторов *STAT1* и *STAT2* уже показывали ассоциацию с псориазом в предыдущих работах неоднократно, это транскрипционный фактор, который участвует в положительной регуляции генов по сигналам интерферона 1-3 типа [Hald, 2013]. Для *ENF*, который является эпителий-специфичным транскрипционным фактором и участвует в регуляции клеточной дифференциации и онкогенезе, была показана ассоциация с развитием псориаза только единожды и полученные нами результаты подтверждают эту гипотезу [Swindell, 2015].

Таблица 4. Топ-4 транскрипционных факторов, которые участвуют в регуляции дифференциально экспрессирующихся генов в поражённой и здоровой коже.

TFs	Количество графов, содержащих ТФ	Частота присутствия в графе
<i>NFKB1</i>	195	0.388
<i>ENF</i>	134	0.266
<i>KLF15</i>	110	0.219
<i>STAT1</i>	107	0.213

Транскрипционные регуляторы, которые были найдены в большом количестве графов, можно, видимо, считать основными регуляторами заболевания, так как именно они присутствуют в большом количестве орграфов, которые хорошо объясняют разницу между поражённой псориазом и здоровой кожей (таблица 4).

Оценка уровня метилирования ДНК в сайтах связывания транскрипционных регуляторов

Долгое время метилирование ДНК рассматривалось как важный тип эпигенетического сайленсинга генов [Holliday et al., 1975]. Данная модификация играет важную роль во многих клеточных процессах, таких как транскрипция, геномный импринтинг и инактивация X-хромосом [Robertson, 2005]. В клетках

млекопитающих метилирование катализируется ДНК метилтрансферазами (DNMTs), для которых донором метильной группы служит S-аденозил-метионин (SAM). Перенос метильной группы при метилировании ДНК происходит строго на 5'-конце цитозина в CpG сайте. У человека обнаружено 3 активных метилтрансферазы: DNMT1, DNMT3a и DNMT3b [Jones et al., 2002]. DNMT1 обеспечивает стабильный уровень метилирования полуметилированных сайтов ДНК, в то время как DNMT3a и DNMT3b метилируют целевые участки de novo [Goll et al., 2006].

Мы оценивали пересечение пиков иммунопреципитации хроматина для различных ТФ с такими локусами, у которых сумма квадратов корреляции и между экспрессией гена этого локуса и фенотипа и корреляции между метилированием этого локуса и фенотипа была больше 1. Наличие таких локусов говорит об обратной связи между уровнем метилирования ДНК и уровнем экспрессии, т. е. о снижении уровня экспрессии гена через метилирование ДНК. В ходе такого анализа мы нашли 42 локуса, которые удовлетворяли требованиям для 4 уникальных транскрипционных факторов (STAT1, STAT2, RCOR1 и JUNB). Далее для этих локусов мы оценивали пересечение с сайтами связывания транскрипционного фактора с ДНК. Всего мы нашли пересечение для двух транскрипционных регуляторов (STAT1 и STAT3) в трёх локусах для трёх генов. Первым геном, в регуляции которого участвуют, по всей видимости, одновременно STAT1 и STAT3, оказался ген с выраженной антимикробной функцией P13. Для данного гена была показана ассоциация с псориазом в предыдущих исследованиях [Swindell, 2015], но механизм регуляции экспрессии гена через дифференциальное метилирование ДНК в сайтах связывания его транскрипционных регуляторов продемонстрирован впервые. Помимо гена P13, аналогичная закономерность была показана для генов S100A8 и S100A9, также для транскрипционного регулятора STAT1, что подтверждает выводы, сделанные в предыдущих главах: помимо того, что в регуляции как минимум этих двух белков класса S100A участвует метилирование ДНК, на локусы с дифференциальным метилированием ДНК также приходятся сайты связывания для различных транскрипционных факторов (таблица 5).

Таблица 5. Список генов и их локусов с определённой корреляцией между экспрессией гена и уровнем метилирования ДНК, которые попали в предполагаемый сайт посадки ТФ, который участвует в регуляции гена в столбце «Ген»

Название локуса	Ген	ТФ	R ² экспр.	FDR p-value	R ² метил.	FDR p-value	Квадрат общей корреляции
cg04880990	<i>KRT6A</i>	<i>STAT3</i>	-0.832	1.90E-46	0.597	1.74E-14	1.05
cg26348348	<i>LYPD1</i>	<i>STAT1</i>	-0.8	7.54E-41	0.699	8.62E-22	1.13
cg00085448	<i>OAS2</i>	<i>RCOR1</i>	-0.852	1.48E-50	0.59	4.75E-14	1.08
cg00175901	<i>OAS2</i>	<i>RCOR1</i>	-0.852	1.48E-50	0.526	1.11E-10	1.003
cg12560128	<i>OAS2</i>	<i>RCOR1</i>	-0.852	1.48E-50	0.635	6.72E-17	1.13
cg16399664	<i>OAS2</i>	<i>RCOR1</i>	-0.852	1.48E-50	0.535	4.21E-11	1.01
cg00085448	<i>OAS2</i>	<i>STAT1</i>	-0.852	1.48E-50	0.59	4.75E-14	1.08
cg00175901	<i>OAS2</i>	<i>STAT1</i>	-0.852	1.48E-50	0.526	1.11E-10	1.003
cg02733351	<i>PI3</i>	<i>STAT1</i>	-0.899	2.27E-62	0.633	8.47E-17	1.21
cg19931348	<i>PI3</i>	<i>STAT1</i>	-0.899	2.27E-62	0.551	6.48E-12	1.11
cg02733351	<i>PI3</i>	<i>STAT3</i>	-0.899	2.27E-62	0.633	8.47E-17	1.21
cg19931348	<i>PI3</i>	<i>STAT3</i>	-0.899	2.27E-62	0.551	6.48E-12	1.11
cg20335425	<i>S100A8</i>	<i>STAT1</i>	-0.876	3.27E-56	0.675	7.75E-20	1.22
cg02711163	<i>S100A9</i>	<i>STAT1</i>	-0.874	1.01E-55	0.496	2.43E-09	1.01
cg03165378	<i>S100A9</i>	<i>STAT1</i>	-0.874	1.01E-55	0.665	4.05E-19	1.208
cg14826683	<i>SPRR2D</i>	<i>JUNB</i>	-0.865	1.26E-53	0.564	1.43E-12	1.07
cg02772121	<i>TRIM15</i>	<i>RCOR1</i>	-0.815	2.40E-43	0.64	2.80E-17	1.08
cg05701418	<i>TRIM15</i>	<i>RCOR1</i>	-0.815	2.40E-43	0.637	4.67E-17	1.07

ЗАКЛЮЧЕНИЕ

Развитие постгеномных технологий привело к появлению большого количества методов высокоэффективного и высокопроизводительного полногеномного анализа, среди которых можно отметить анализ метилирования CpG на микрочипах и иммунопреципитацию хроматина. Метилирование - одна из основных эпигенетических модификаций ДНК, а совокупность всех таких модификаций в геноме человека образует метилом. Для многих болезней исследователи обнаружили эпигенетическую компоненту, которая, как правило, играет большую роль в их патогенезе. К таким заболеваниям относятся и иммуновоспалительные заболевания кожи, например, псориаз. В данной работе был проведен анализ всех доступных на сегодняшний момент данных экспериментов по дифференциальной экспрессии генов и метилированию ДНК в здоровой коже и в коже, пораженной псориазом. В результате удалось выделить список значимых генов, который хорошо согласуется с результатами других исследований, и содержит множество генов, отвечающих за формирование иммунного ответа и пролиферацию клеток кожи, что согласуется с основной гипотезой развития заболевания.

Предложенный в работе метод анализа транзитивности векторов фенотипов позволил впервые выявить корреляцию между экспрессией генов, ассоциированных с возникновением псориаза, и метилированием регуляторных участков ДНК. Среди генов, экспрессия которых значимо изменяется в гиперпролиферирующих клетках кожи, в результате одновременного изменения экспрессии регулирующих их транскрипционных факторов и метилирования регуляторных участков ДНК, присутствуют известные гены, участвующие в формировании реакции воспаления и иммунного ответа, такие как *S100A8* и *S100A9*, *OAS2*, *PI3* и *AIM2*. С помощью сопоставления генных сетей с анализом ДНК-белковых взаимодействий удалось выявить транскрипционные факторы, которые потенциально участвуют в регуляции процессов, приводящих к возникновению гиперпролиферации клеток кожи. Среди таких ТФ были найдены *NFKB1*, *EHF*, *KLF15*, *STAT1*, *GATA6*, *STAT3* и *RCOR1*, для которых уже неоднократно была показана роль в регуляции воспаления и пролиферации клеток кожи. При этом оказалось, что степень метилирования сайтов связывания таких транскрипционных факторов как *STAT1*, *STAT2*, *GATA3*, *GATA6*, *ESR1* и *RCOR1* часто изменяется в образцах кожи, взятой у больных псориазом. Суммируя результаты, полученные в ходе этого исследования, можно сделать вывод, что определены гены и транскрипционные факторы, которые играют важную роль в патогенезе псориаза, причем как для регуляторных участков выявленных генов, так и мишеней транскрипционных факторов характерно

изменение степени метилирования в пораженной коже. Эти результаты могут быть применены как при разработке новых терапий псориаза, так и в качестве маркерных признаков для оценки моделей псориатической кожи.

ВЫВОДЫ

1. Проведенный мета-анализ доступных опубликованных данных по RNA-seq образцов кожи больных псориазом позволил выявить 930 генов, экспрессия которых достоверно изменена в поражённой псориазом коже. Функции большинства найденных генов связаны с регуляцией иммунного ответа или пролиферацией клеток кожи.

2. С помощью сравнительного мета-анализа метилирования ДНК в поражённой и здоровой коже выявлены 899 дифференциально метилированных локусов, большинство из которых расположено в районах генов, функции которых связаны с апоптозом и активацией иммунных клеток, что согласуется с иммунновоспалительной природой заболевания;

3. С помощью предложенного нами метода транзитивной корреляции между экспрессией генов, метилирования ДНК и вектора фенотипов обнаружено, что в псориазе для генов, принадлежащих к семейству *S100A*, а также генов *AIM2*, *PI3* существует значимая корреляция между экспрессией генов и метилированием ДНК;

4. Использование маркерных графов позволило выявить транскрипционные факторы *NFKB1*, *EHF*, *KLF15*, *STAT1*, *STAT3*, *GATA6* в качестве важных регуляторов патогенеза заболевания, которые участвуют в контроле наибольшего количества дифференциально экспрессирующихся генов;

5. Для дифференциально экспрессирующихся генов семейства *S100A*, а также для генов *OAS2*, *PI3*, *PLBD1*, *CSCL2*, *PARP14* и *EMILIN3* показано изменение профилей метилирования ДНК в областях связывания транскрипционных факторов, которые регулируют экспрессию этих генов;

6. Показано изменение уровня метилирования ДНК в помоторах генов-мишеней транскрипционных факторов *STAT1*, *STAT2*, *GATA3*, *GATA6*, *ESR1* и *RCOR1*.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи

- Соболев, В. В., Денисова, Е. В., Соболева, А. Г., Корсунская, И. М., Чекалин, Е. В., Золотаренко, А. Д., & Брускин, С. А. (2013). Метилирование генов в CD3⁺-клетках при псориазе. *Российский журнал кожных и венерических болезней*, (1), (pp. 40-43)
- Zolotareno, A., Chekalin, E., Mesentsev, A., Kiseleva, L., Griбанова, E., Mehta, R., Tatarinova T., Piruzian E., Bruskin, S. (2016). Integrated computational approach to the analysis of RNA-seq data reveals new transcriptional regulators of psoriasis. *Experimental & molecular medicine*, 48(11), (pp. 268-279).
- Zolotareno, A., Chekalin, E., Piruzian, E., & Bruskin, S. (2018). FRA1 mediates the activation of keratinocytes: Implications for the development of psoriatic plaques. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1864(12), (pp. 3726-3734).

Глава в монографии

- Zolotareno, A., Chekalin, E., Mehta, R., Baranova, A., Tatarinova, T. V., & Bruskin, S. (2017). Identification of transcriptional regulators of psoriasis from RNA-Seq experiments. In *Biological Networks and Pathway Analysis*, Humana Press, New York, NY, (pp. 355-370).

Тезисы конференций

1. Золотаренко А. Д., Соболева А. Г., Чекалин Е. В., Шеленков А. А. , Брускин С. А., Мезенцев А. В.. Создание новой гуманизированной экспериментальной модели псориаза предназначенной для генной терапии. Тезисы Докл. на II Международной научной конференции "Генетика и Биотехнология в XXI веке: проблемы, достижения, перспективы", Минск, Беларусь, 2015, с. 224
2. Mezentsev A., Zolotareno A., Chekalin E., Shalenkov A., Soboleva A., Bruskin S.; The potential benefits of silencing the transcription factor Fos11 for controlling psoriasis. ESDR Annual Meeting, Amsterdam, *Journal of Investigative Dermatology*, 2015, Vol. 135, p. S58-S59.
3. Преловская А.Н., Сорокина К.С., Чекалин Е.В., Золотаренко А.Д., Брускин С.А. CARD14 регулирует воспаление и пролиферацию клеток при псориазе. Материалы седьмой Международной школы молодых учёных по молекулярной генетике «Геномика и биология живых систем», Москва, Звенигород, 2016, с. 52.

4. Сорокина К.С., Чекалин Е.В., Преловская А.Н., Золотаренко А.Д., Брускин С.А. Поиск новых белковых партнёров CARD14, участвующих в регуляции сигнальных каскадов воспаления. Материалы седьмой Международной школы молодых учёных по молекулярной генетике «Геномика и биология живых систем», Москва, Звенигород, 2016, с. 58.
5. Zolotarenko, A., Prelovskaya, A., Chekalin, E. V., Piruzian, E. S., & Bruskin, S. 361 FRA1 as the regulator of psoriasis-associated hyperproliferation and EMT transition of keratinocytes. *Journal of Investigative Dermatology*, 2016, Vol. 136(9), p. S222.
6. Zolotarenko, A., Chekalin, E. V., Prelovskaya, A., & Bruskin, Housekeeping gene selection and validation for qPCR analysis of psoriasis biopsies. *Journal of Investigative Dermatology*, 2017, Vol. 137(10), p. S269.
7. Chekalin, E., Zolotarenko, A., Piruzyan, E., & Bruskin, S. LB1545 Graph-based classification reveals probable subtypes of psoriasis vulgaris. *Journal of Investigative Dermatology*, 2018, Vol. 138(9), p. B13.
8. Zolotarenko, A., Gurskaya, A., Khan, A., Chekalin, E., Piruzyan, E., & Bruskin, S. LB1477 IQGAP3 could be a promising target link for breaking the vicious circle of psoriasis. *Journal of Investigative Dermatology*, 2018, Vol. 138(9), p. B2.
9. Khan, A., Zolotarenko, A., Piruzian, E., Chekalin, E., & Bruskin, S. Versatile reporter for multiplex analysis of the transcription factor activity in mammalian cells. In *FEBS OPEN BIO*, 2018, Vol. 8, p. 444.
10. Chekalin, E., Zolotarenko, A., Korsunskaya, I., Eleonora, P., & Bruskin, S. Graph-based pipeline for the classification and the marker genes identification of multifactorial skin diseases. In *FEBS OPEN BIO*, 2018, Vol. 8, p. 450.