

Федеральное государственное автономное образовательное
учреждение высшего образования
«Южный федеральный университет»

На правах рукописи

Романов Дмитрий Евгеньевич

**Анализ связи между структурно-функциональной
организацией генома в окрестностях регулирующих
рост генов и морфо-физиологическими
характеристиками млекопитающих**

03.02.07 – Генетика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

Научный руководитель

д. б. н., проф.

Шкурат Татьяна Павловна

Ростов-на-Дону – 2020

Оглавление

Список сокращений и условных обозначений	6
Введение	7
Глава 1. Обзор литературы	13
1.1. Генетический контроль регуляции соматического роста млекопитающих	13
1.1.1. Системные факторы контроля соматического роста млекопитающих	15
1.1.2. Локальные факторы контроля соматического роста млекопитающих	17
1.1.3. Генетическая программа регуляции роста млекопитающих	18
1.1.4. Роль теломеры в контроле соматического роста млекопитающих	19
1.1.5. Геномное расстояние как фактор, определяющий морфофизиологические характеристики	20
1.2. <i>Cis</i> -регуляторные элементы генома	22
1.2.1. <i>Cis</i> -элементы: механизмы действия	22
1.2.2. Методы идентификации <i>cis</i> -элементов	23
1.2.3. Биоинформатический поиск <i>cis</i> -элементов	24
1.2.4. Выявление консервативных элементов генома с помощью построения и анализа точечных матриц гомологий	26
Глава 2. Материалы и методы	28
2.1. Выбор видов млекопитающих	28
2.2. Гены, участвующие в регуляции роста млекопитающих	28
2.2.1. Набор скриптов <i>efunctions</i> для взаимодействия с базами данных NCBI	31

2.3.	Выявление консервативных элементов в окрестностях регулирующих рост генов млекопитающих	33
2.3.1.	Программа для построения точечных матриц гомологий <i>dotolog</i>	34
2.4.	Поиск известных элементов генома в окрестностях регулирующих рост генов млекопитающих	36
2.4.1.	Программный конвейер для поиска гомологичных последовательностей <i>mblast</i>	38
2.5.	Корреляционный анализ	40
2.6.	Распределение по геному человека гомологов консервативных элементов	42
2.6.1.	Анализ сверхпредставленности категорий генов в окрестностях гомологов консервативных элементов	43
2.6.2.	Взаимное расположение гомологов консервативных элементов и CpG сайтов, входящих в состав эпигенетических часов Хорвата	44
Глава 3.	Результаты	45
3.1.	Консервативные элементы генома в окрестностях регулирующих рост генов млекопитающих	45
3.2.	Корреляция между распределением консервативных элементов в окрестностях регулирующих рост генов и морфо-физиологическими характеристиками млекопитающих	47
3.2.1.	Статистический анализ на группе млекопитающих, за исключением приматов, кроме человека	52
3.3.	Известные элементы генома, перекрывавшиеся с консервативными элементами в окрестностях регулирующих рост генов	52
3.4.	Представленность в геноме человека гомологов консервативных элементов	59

3.4.1.	Анализ сверхпредставленности категорий Gene Ontology среди генов в окрестностях гомологов консервативных элементов	65
3.4.2.	Взаимное расположение гомологов консервативных элементов и CpG сайтов, входящих в состав эпигенетических часов Хорвата	72
3.5.	Корреляция между положением на хромосоме регулирующих рост генов и морфо-физиологическими характеристиками млекопитающих	75
3.5.1.	Статистический анализ на всей группе млекопитающих, за исключением приматов, кроме человека	77
Глава 4.	Обсуждение	84
4.1.	Эволюционное модулирование экспрессии регулирующих рост генов у разных видов млекопитающих	84
4.2.	Уменьшение с возрастом экспрессии регулирующих рост генов млекопитающих	86
4.3.	Регуляторные механизмы взаимодействия на расстоянии	87
4.4.	Связь между морфо-физиологическими характеристиками млекопитающих и расстоянием между консервативными элементами в окрестностях регулирующих рост генов	89
4.4.1.	Случай отрицательной корреляции	89
4.4.2.	Случай положительной корреляции	91
4.5.	Связь между морфо-физиологическими характеристиками млекопитающих и положением на хромосоме регулирующих рост генов	94
Заключение	97	
Выводы	99	

Список литературы	101
Список иллюстративного материала	116
Список таблиц	119

Список сокращений и условных обозначений

КЭ — консервативный элемент генома

ИЭГ — известный элемент генома

н.п. — нуклеотидная пара

SNP — single nucleotide polimorfism, однонуклеотидный полиморфизм

TPE-OLD — telomere position effect over long distances, эффект теломерного замолкания на длинных расстояниях

Введение

Актуальность проблемы. Проблема регуляции роста млекопитающих остается одной из самых давних загадок в биологии [1, 2, 3, 4].

Скорость соматического роста млекопитающих высока в ранние этапы развития организма, но с возрастом постепенно снижается, тем самым задавая конечный размер тела взрослого животного. Снижение скорости роста млекопитающих, заключающееся в уменьшении уровня пролиферации клеток в разных тканях, контролируется в основном локальными механизмами и наиболее тесно связано с уменьшением экспрессии следующих 10 генов, относящихся к семейству транскрипционных факторов и белков сигнальных путей: *Ezh2*, *Gpc3*, *Mdk*, *Mest*, *Mycn*, *Peg3*, *Plagl1*, *Smo*, *Igf2* и *E2f3* [4]. На сегодняшний день остаются не известны механизмы, контролирующие скоординированное уменьшение экспрессии этих генов, и факторы, которые лежат в основе эволюционного модулирования соответствующей генетической программы [3, 4].

В системную регуляцию роста млекопитающих вовлечены гены соматотропной оси, из которых основными являются следующие 7 генов: *Gh1*, *Ghrh*, *Ghrl*, *Igf1*, *Sst*, *Igfbp3* и *Igfbp1* [5]. Было показано ранее, что некоторые морфо-физиологические характеристики млекопитающих скоррелированы с размером некодирующих областей этих генов [6]. Тем не менее, механизмы, модулирующие системную регуляцию роста млекопитающих, остаются открытым вопросом [3, 5].

Регуляция экспрессии гена зависит от присутствия в его окрестности различных *cis-регуляторных элементов*. Известно, что многие *cis-регуляторные* элементы генов являются одновременно и *консервативными элементами* генома [7]. Следует отметить, что геномное расстояние между промотором и некоторыми консервативными элементами в окрестности генов *Mycn* и *Plagl1* значимо скоррелировано с массой млекопитающих, а в окрестности гена *Ezh2* — с продолжительностью жизни [8].

С другой стороны, на регуляцию экспрессии гена может оказывать влияние

геномное положение гена, в частности, положение гена на хромосоме по отношению к теломерам. Показано, что недавно открытый *теломерный эффект положения на длинных расстояниях* (telomere position effect over long distances, TPE-OLD) может регулировать гены в нескольких миллионах н.п. от теломеры [9, 10]. Данный эффект был продемонстрирован для генов *Isg15*, *Dsp*, *C1s*, *Tert*, *Notch1* и *Sorbs2* и может быть задействован в контроле регулирующих рост генов млекопитающих. Важно отметить, что расстояние от гена до теломеры для генов соматотропной оси *Ghrh* и *Sst* и для генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD, значимо скоррелировано с продолжительностью жизни и периодом полового созревания млекопитающих [11].

Цель и задачи работы. Целью работы является анализ связи между структурно-функциональной организацией генома в окрестностях регулирующих рост генов и следующими морфо-физиологическими характеристиками млекопитающих: масса и размер тела взрослого животного, период полового созревания и продолжительность жизни.

Задачи:

1. Разработать набор инструментов для автоматизации типовых задач по извлечению информации из баз данных NCBI и получить геномные последовательности окрестностей регулирующих рост генов у различных видов млекопитающих.
2. Разработать модификацию метода построения и анализа точечной матрицы гомологий, позволяющую проводить *множественное сравнение* геномных последовательностей и выявлять на основании сравнения *консервативные элементы генома* в окрестностях регулирующих рост генов млекопитающих.
3. Разработать программный конвейер для биоинформатического поиска в геномных последовательностях *известных элементов генома* и осуществить

поиск этих элементов в окрестностях регулирующих рост генов млекопитающих.

4. Изучить зависимость между морфо-физиологическими характеристиками млекопитающих и распределением консервативных элементов в окрестностях регулирующих рост генов.
5. Произвести полногеномный поиск *гомологов* выявленных консервативных элементов генома, получить гены, лежащие в окрестности этих гомологов, и выполнить *анализ сверхпредставленности категорий Geno Ontology* для этих генов.
6. Изучить зависимость между морфо-физиологическими характеристиками млекопитающих и положением в геноме регулирующими рост генов.

Научная новизна работы. Разработана оригинальная модификация компьютерного метода поиска консервативных элементов генома на основе построения и анализа точечной матрицы гомологий, позволяющая проводить множественное сравнение геномных последовательностей. С помощью этого метода выявлены консервативные участки генома в окрестностях регулирующих рост генов у различных видов млекопитающих.

Впервые показана связь между геномным расстоянием между консервативными элементами в окрестностях регулирующих рост генов и такими морфо-физиологическими характеристиками млекопитающих, как масса и длина тела взрослого животного и продолжительность жизни.

Показано, что геномное расстояние между регуляторными элементами генома может выступать фактором, эволюционно модулирующим экспрессию генов регуляции роста и в конечном итоге определяющим фенотипические различия между видами млекопитающих.

Предложены модели регуляции этих генов, объясняющие фенотипические различия между видами млекопитающих.

Впервые выявлена взаимосвязь между морфо-физиологическими характеристиками млекопитающих и положением на хромосоме регулирующих рост генов и указаны возможные механизмы регуляции некоторых из этих генов, объясняющие фенотипические различия между видами млекопитающих.

Методика, примененная в исследовании, может быть использована для поиска новых соотношений генотипа и фенотипа.

Наличие значимой корреляции между фенотипом и расстоянием между консервативными элементами генома может выступать дополнительным подтверждением функциональной значимости для предсказанных регуляторных сайтов.

Практическая значимость исследования. Работа является шагом на пути к пониманию, как связаны структурно-функциональная организация генома и его фенотипические проявления, в частности, масса и размер тела млекопитающих, период полового созревания и продолжительность жизни.

В работе показано, что геномное расстояние между некоторыми консервативными элементами генома в окрестностях регулирующих рост генов может выступать одним из основных факторов, определяющих указанные морфо-физиологические характеристики.

Исследован вопрос, как положение этих генов на хромосомах, в частности, расстояние до ближайшей теломеры может также влиять на эти морфо-физиологические характеристики у разных млекопитающих.

Предложены молекулярные механизмы регуляции экспрессии этих генов, причем предполагается, что геномное расстояние является определяющим фактором эволюционного модулирования экспрессии, и построены модели регуляции этих генов в онтогенезе.

Результаты, которые получены в данной работе, могут быть востребованы в исследованиях генетических нарушений, вызывающих задержку роста или преждевременное старение. Знание механизмов роста и развития организма может найти применение в регенеративной медицине.

С практической точки зрения возможность регуляции роста и размеров

животных может быть востребована в сельском хозяйстве с целью повышения морфо-физиологических показателей основных сельскохозяйственных животных.

Предлагаемый в работе метод исследования может быть применен в качестве подхода к решению похожих фундаментальных проблем связи фенотипа с генотипом.

Положения, выносимые на защиту:

1. У млекопитающих выявлена значимая корреляция между морфо-физиологическими характеристиками и геномным расстоянием между некоторыми консервативными элементами в окрестностях регулирующих рост генов *Mycn*, *Plagl1* и *Ezh2*.
2. Среди генов, находящихся в окрестностях гомологов в геноме человека этих консервативных элементов, сверхпредставлены гены, связанные с регуляцией роста.
3. У млекопитающих выявлена значимая связь между периодом полового созревания и геномным расстоянием от гена до ближайшей теломеры для двух генов соматотропной оси *Ghrh* и *Sst* и для двух генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD.

Апробация работы. Основные результаты диссертации были представлены на 12-ой Международной конференции «Bioinformatics of Genome Regulation and Structure/Systems Biology» (Новосибирск, 2020), конференции Ростовского общества генетиков и селекционеров (Ростов-на-Дону, 2020 и 2017), 25-ой Международной конференции студентов, аспирантов и молодых ученых «Ломоносов-2018» (Москва, 2018), 7-ой конференции «Генетика — фундаментальная основа инноваций в медицине и селекции» (Ростов-на-Дону, 2017), симпозиуме EMBL «The Non-Coding Genome» (Германия, Хейдельберг, 2015), 6-ой Международной конференции «Актуальные проблемы биологии, нанотехнологий и медицины» (Ростов-на-Дону, 2015), 7-ой Международной Школе молодых ученых «Системная

биология и биоинформатика» SBV-2015 (Новосибирск, 2015), 5-ой Международной конференции «Актуальные проблемы биологии, нанотехнологий и медицины» (Ростов-на-Дону, 2013).

Публикации. По материалам диссертационного исследования опубликовано 18 печатных работ: 3 статьи Scopus, 2 свидетельства о регистрации программ для ЭВМ, 13 тезисов в сборниках трудов конференций.

Личный вклад автора в проведение исследования. Все представленные в диссертации результаты получены лично автором. По мере выполнения работы был реализован ряд методов анализа последовательностей и программных средств для автоматизации обработки данных.

Структура и объем диссертации. Диссертация состоит из введения, обзора литературы, описания материалов и методов исследования, изложения результатов и их обсуждения, заключения, выводов, списка сокращений и списка цитируемой литературы (145 источников). Работа представлена на 123 страницах и содержит 11 рисунков и 36 таблиц.

Глава 1

Обзор литературы

1.1. Генетический контроль регуляции соматического роста млекопитающих

Контроль регуляции роста органов и организма остается центральным вопросом биологии. Масса тела взрослых млекопитающих может принимать значения в широком диапазоне, начиная от 1.5 г у карликовой многозубки (*Suncus etruscus*) и заканчивая 150 т у синих китов (*Balaenoptera musculus*), т.е. различается более чем на 8 порядков. При этом развитие начинается из одной небольшой клетки, и все млекопитающие обладают сходным планом строения тела и набором органов. Уже эти факты позволяют заключить, что рост организма млекопитающего является тонко контролируемым процессом [12].

Было предложено несколько потенциальных молекулярных механизмов контроля роста млекопитающих.

Соматический рост может быть вызван как увеличением уровня пролиферации клеток (гиперплазия), так и увеличением размеров самих клеток (гипертрофия). В делящихся клетках оба этих фактора взаимосвязаны, что обеспечивает постоянство среднего размера клеток на фоне увеличения общего числа клеток [13].

Показано, что уменьшение скорости соматического роста вызвано уменьшением уровня пролиферации клеток и последующим замедлением скорости роста самих клеток. К примеру, у крыс в период с момента рождения до достижения одномесячного возраста общее количество ДНК во всех клетках, напрямую связанное с количеством клеток, увеличивается семикратно, в то время как количество белка в каждой клетке, что отражает размер клетки, увеличивается трехкратно [14].

С другой стороны, тело человека, например, содержит около 10^{13} клеток, а 25-граммовая мышь — только 3×10^9 клеток, и 3000-кратная разница в весе между этими видами объясняется 3000-кратной разницей именно в количестве клеток [3]. Принимая во внимание, что размеры самих клеток практически одинаковы у всех млекопитающих, огромнейшая разница в размерах тела взрослых особей обусловлена в первую очередь разницей в количестве клеток, нежели в их размере. Таким образом, замедление скорости роста млекопитающих связано с уменьшением уровня пролиферации клеток.

Причинами уменьшения уровня пролиферации клеток могут выступать как увеличение времени клеточного цикла, так и уменьшение доли растущих клеток. У мышей показано, что при развитии почек и печени время клеточного цикла растущих клеток практически не меняется, в то время как доля растущих клеток значительно падает [15]. Схожие результаты были получены на крысах [16, 17]. Таким образом, уменьшение уровня пролиферации клеток обусловлено уменьшением доли делящихся клеток.

Замедление скорости соматического роста может быть вызвано изменением соотношения между стволовыми, дифференцирующимися и дифференцированными клетками. В частности, оно может быть связано с уменьшением доли пролиферирующих стволовых клеток или же с уменьшением числа этих клеток. Также может падать уровень пролиферации дифференцирующихся из стволовых клеток. Изучение у кролика хрящевой пластинки роста, в которой представлены все три указанных выше типа клеток, показало, что замедление роста ассоциировано с исчерпанием пула первичных клеток [18].

Наблюдения за ростом различных органов показывают, что рост скоординирован по времени, по реакциям на внешние или внутренние условия и эволюционно [19]. Показано, что замедление роста происходит одновременно во многих органах, но, быть может, с разными темпами [20]. К примеру, замедление роста центральной нервной системы происходит гораздо раньше, чем у большинства других органов [21]. Скоординированное замедление роста органов наблюдается

также при недостатке гормона роста GH1, гипотиреозе или неполноценном питании, причем пропорции тела сохраняются. При нормализации условий наблюдается явление «наверстывания» роста [22].

Нокаут некоторых регулирующих рост генов также может приводить к изменению размеров тела у мышей, однако различные органы в разной степени реагируют на это воздействие. В частности, при удалении гена *Ghr* рецептора гормона роста у мыши вес большинства органов пропорционально уменьшается, за исключением почек и селезенки, которые уменьшаются в большей степени, и мозга, который, наоборот, уменьшается в меньшей степени [23]. Схожий эффект вызывают недостаток IGF1 или тиреоидного гормона у мышей и плацентарная недостаточность у человека, когда уменьшение роста мозга происходит в меньшей степени, нежели всего тела, что ведет к возрастанию массы мозга по отношению к массе тела [24, 25, 26]. Наконец, известно, что гомологичные органы у разных видов млекопитающих пропорционально изменяются согласно размерам тела животного.

1.1.1. Системные факторы контроля соматического роста млекопитающих

Скоординированное уменьшение скорости роста может быть обусловлено системными факторами. В частности, гормоны играют ключевую роль в развитии насекомых, когда по достижении определенного размера уровень ювенильного гормона падает одновременно с увеличением уровня экдизона, что вызывает остановку роста и начало метаморфоза [27].

У млекопитающих системная регуляция роста осуществляется семейством генов соматотропной оси, основными представителями которого являются гены *Gh1*, *Ghrh*, *Ghrl*, *Igf1*, *Sst*, *Igfbp3* и *Igfbp1* [3, 5].

В частности, гормон роста GH1 может оказывать существенное влияние на массу тела взрослого животного. Мыши, лишённые рецептора гормона роста, оказываются на 60% меньше по массе, нежели нормальные мыши. Недостаток IGF1

ведет еще к большему уменьшению массы тела, и масса таких мышеч составляет 30% от нормы [23]. И наоборот, сверхэкспрессия этих гормонов ведет к увеличению размеров тела [28]. Аналогичные эффекты наблюдаются и у человека.

Гормон роста GH1 оказывает влияние в большей степени на постнатальное развитие, в то время как IGF1 влияет и на пре-, и на постнатальный рост [29]. В целом, уровень GH1 в зародышах человека, овцы и грызунов оказывается значительно выше уровня этого гормона у взрослой особи [30].

Интересно отметить, что масса тела различных пород собак ассоциирована с разными аллельными вариантами гена *Igf1* и уровнем этого гормона [31, 32].

Тем не менее, существует множество доказательств, демонстрирующих вторичность роли системных факторов в регуляции роста [3]. Во-первых, показано, что уровень GH1 не влияет на нормальный рост плода [23, 33]. Во-вторых, введение постоянных доз гормона GH1 людям с недостатком этого гормона также ведет к нормальному развитию. В-третьих, замедление роста происходит и при сверхэкспрессии гормонов GH1 или IGF1, хотя при этом наблюдается существенное увеличение размеров тела [28, 34]. В-четвертых, уменьшение уровня GH1 в начале жизни не сопровождается уменьшением уровней IGF1 или IGFBP3 [35, 36]. Более того, концентрация IGF1, через который и проявляется основное действие гормона роста GH1, продолжает увеличиваться на фоне замедления скорости роста [37]. Рост концентрации IGF1 сопровождается увеличением уровня IGFBP3, который стабилизирует свободный IGF1 [35, 36], что, однако, может уменьшать биодоступность IGF1. Тем не менее, концентрация свободного IGF1 повышается с возрастом [36].

Таким образом, оба гормона соматотропной оси GH1 и IGF1 хотя и осуществляют системную регуляцию роста, но, по-видимому, не оказывают существенного влияния на замедление роста. Существуют, однако, доказательства, что IGF1 может являться паракринным регулятором [38].

Питание также может модулировать рост. Известно, что недостаток питания приводит к задержке роста [20, 39]. Тем не менее, избыточное питание не останав-

ливают замедление роста. В частности, у крыс, свиней и кур избыточное питание хотя и приводит к увеличению массы тела, но в основном за счет накопления жира, а не белка [40, 41, 42].

Интересно отметить, что женский половой гормон — эстроген — также оказывает влияние на замедление роста, в частности, на развитие хрящевой пластинки роста [3]. Показано, что недостаток эстрогена приводит к удлинению этапа слияния эпифизов и таким образом продлевает рост костей [43].

1.1.2. Локальные факторы контроля соматического роста млекопитающих

Многочисленные эксперименты по трансплантации органов между животными разного возраста показывают, что орган из молодой особи продолжает расти с той же скоростью, будучи пересаженным взрослой особи [3, 5]. Таким образом, программа регуляции роста органа заложена в самом органе и в меньшей степени подвержена влиянию внешних факторов. Следует отметить, что это не исключает участия системных факторов в замедлении роста, однако показывает главенствующую роль локальных механизмов регуляции роста.

Локальные механизмы регуляции роста могут в свою очередь быть как автономными, внутриклеточными, так и паракринными, основанными на взаимодействии клеток друг с другом, а контроль роста может осуществляться некоторой генетической программой [5].

В частности, показано снижение *in vivo* пролиферативной активности клеток зоны покоя хондроцитов в хрящевой пластинке роста [18], однако те же клетки, будучи помещенными в культуру, демонстрируют независимость способности к пролиферации от возраста донора [44].

В другом случае удаление половины печени у мышей вызывает регенерацию ткани до изначального объема [12, 3]. Следует отметить, что пересадка печени большего размера может приводить к уменьшению массы органа [45], однако щитовидная железа, почки, кишки или хрящи не изменяют своего размера при пере-

садке [3].

1.1.3. Генетическая программа регуляции роста млекопитающих

Существует большое число данных экспериментов, когда нокаут генов приводил к задержке роста или гигантизму, что свидетельствует о роли этих генов или ассоциированных сигнальных путей в регуляции роста. В частности, показана роль онкогенного сигнального пути *c-Myc* [46] и каскада киназы *Hippo* в регуляции роста [47]. Тем не менее, не известно, влияет ли модуляция этих регуляторных систем на замедление роста.

Было выдвинуто предположение о существовании единой генетической программы регуляции роста [1, 2, 3, 4, 48, 49]. Впервые существование такой программы было продемонстрировано с помощью полногеномного анализа экспрессии генов у растущих мышей, крыс и овец [1, 2, 4, 48]. Наличие такой единой программы регуляции объясняет скоординированное замедление роста органов при сохранении пропорций тела.

Было показано постепенное уменьшение с возрастом экспрессии множества генов. Сюда входят гены белков факторов роста *Igf2* и *Mdk* и генов белков транскрипционных факторов *Mycn*, *Plagl1*, *Ezh2*, *Mest*, *Smo*, *E2f3*, *Peg3* и *Gpc3* [1, 2, 4, 48]. Эксперименты по нокауту этих генов действительно подтверждают участие этих генов в регуляции роста.

На настоящий момент остаются не известны молекулярные механизмы, контролирующие скоординированное уменьшение экспрессии этих генов. Предполагается, что в основе этого явления могут лежать эпигенетические механизмы.

Особо следует подчеркнуть, что остаются также невыясненными факторы, которые могли бы лежать в основе эволюционного модулирования соответствующей генетической программы и таким образом объясняющие существенное различие в массе тела между различными видами млекопитающих.

Важно упомянуть о связи микроРНК и замолкании регулирующих рост ге-

нов. Поскольку одна микроРНК может иметь множество мишеней, было выдвинуто предположение, что уменьшение с возрастом экспрессии указанных выше генов может быть вызвано увеличением с возрастом экспрессии некоторой общей микроРНК. Эксперименты с использованием ДНК-микрочипов показали, что 4 вида микроРНК, 3 из которых принадлежат семейству *MIR29*, увеличивали свою экспрессию с возрастом во многих органах [50, 51].

Биоинформатический анализ показал, что предсказанные мишени *MIR29* сверхпредставлены в генах, уменьшающих свою экспрессию с возрастом во многих тканях. Для генов *Igf1*, *Mest* и *Igf2bp1* было экспериментально показано, что они действительно являются мишенями этих микроРНК [50].

Предполагалось, что ген *MIR29* негативно регулирует рост органов и что увеличение экспрессии гена *MIR29* во время ранней жизни может помочь уменьшить экспрессию регулирующих рост генов, что в итоге приведет к постепенному замедлению роста с возрастом. Также предполагалось, что нокаут гена *MIR29* приведет к увеличенному размеру тела и скорости роста. Тем не менее, нокаутные по гену *MIR29* мыши не показали сверхроста и вместо этого показали уменьшение роста и умерли в течение 4 недель. Проверка этих мышей показала, что наблюдались серьезные дефекты в дифференциации гладкой мускулатуры легких, что приводило к проблемам с дыханием и ранней гибели. Однако, неясно, может ли ген *MIR29* служить основным негативным регулятором постнатального роста, несмотря на то, что он играет существенную роль в развитии легких [50, 51].

1.1.4. Роль теломеры в контроле соматического роста млекопитающих

Развитие организма может быть основано на некотором молекулярном механизме, вычисляющим количество клеточных делений [3]. В частности, уменьшение на раннем этапе развития числа прогениторных клеток селезенки не компенсируется в дальнейшем, что приводит к развитию органа меньшего размера [52]. Предполагается, что пролиферация прогениторных клеток ограничена автоном-

ными внутриклеточными механизмами, и каждая такая клетка может развиваться лишь в фиксированное количество ткани. Напротив, уменьшение на раннем этапе развития числа прогениторных клеток печени не ведет к существенному изменению конечного размера органа, что говорит о другом типе регуляции роста [52]. Рост хрящевой пластинки, по-видимому, также регулируется механизмами, основанными на подсчете количества клеточных делений [18].

Одним из наиболее изученных молекулярных механизмов, подсчитывающих количество делений, является *эффект укорочения теломеры* [53]. Показано, что эффект укорочения теломеры играет важную роль в процессах клеточного старения, антираковой защите [54] и, возможно, старении всего организма [55, 56], однако маловероятно, чтобы этот эффект имел центральное значение в контроле роста на начальных этапах жизни [3]. В частности, мутации в гене теломеразы вызывают преждевременное старение у человека и мыши на фоне нормального роста на начальных этапах развития [57].

Важно отметить, что недавно был открыт *эффект теломерного замолкания на длинных расстояниях* (telomere position effect over long distances, TPE-OLD), состоящий в физическом сближении теломеры и гена, что ведет к репрессии гена [9, 10]. По мере укорочения теломеры происходит разделение этих локусов и таким образом ген получает возможность экспрессироваться. Этот эффект был показан для генов *Isg15*, *Dsp*, *C1s* [9, 58], *Tert* [59], *Notch1* [60, 61] и *Sorbs2* [58, 62]. Вероятно, такой механизм мог бы контролировать экспрессию некоторого гена-репрессора регулирующих рост генов.

1.1.5. Геномное расстояние как фактор, определяющий морфо-физиологические характеристики

Сравнения последовательностей генов показывают, что *внутривидовое* варьирование морфо-физиологических характеристик обуславливается наличием геномных вариаций в окрестности генов, нежели накоплением точечных мутаций

в генах [31, 63, 64].

В другой работе продемонстрировано, что фенотипическое разнообразие внутри видов современных одомашненных животных скорее продиктовано высокой скоростью эволюции последовательности гена путем изменения размера микросателлитной ДНК, нежели скоростью накопления мутаций в том же гене [65].

Важно отметить, что длина вариации может качественно и количественно влиять на экспрессию генов [63, 66].

С другой стороны, существует множество подтверждений, что даже небольшие генетические изменения могут приводить к значительным фенотипическим различиям как внутри, так и между видами [67]. Например, показано, что *полиморфизм длины повторяющихся фрагментов* может оказывать значимое влияние на экспрессию генов и в результате выступать в качестве качественного и количественного фактора, вызывающего фенотипическую вариацию признаков [68]. Сравнительный анализ между породами собак числа повторов в генах, отвечающих за развитие скелета и черепа, показал наличие тесной положительной корреляции между размером черепа и отношением числа полиглутаминов к числу полиаланинов в домене повторов внутри гена *Runx2* [63, 69, 70].

Тем не менее, *внутривидовая* вариация числа повторов и ее связь с фенотипом не может объяснять *межвидовое* варьирование признака, т.к. тогда наличие определенного числа повторов было бы эволюционно закреплено в последовательности каждого из видов и легко было бы выявлено при сравнении последовательностей ортологичных генов.

Особо следует подчеркнуть, что наличие вариации длины повторов автоматически приводит к возникновению вариации геномного расстояния между возможными *cis*-регуляторными участками гена, что приводит к целесообразности проверки гипотезы о связи геномного расстояния между *cis*-регуляторными элементами генома в окрестности гена и морфо-физиологическими характеристиками животных.

1.2. *Cis*-регуляторные элементы генома

Геном человека представлен 20000–25000 генов, практически столько же генов содержится в геноме мыши, и это количество лишь немного превосходит число генов у простой нематоды *Caenorhabditis elegans*. Общепринятым считается, что эволюционное наращивание сложности организмов происходит за счет усложнения регуляции генов [71].

Несмотря на то, что последовательности многих генов были проаннотированы, определение *cis*-регуляторных элементов остается одной из основных задач геномики, главным образом, из-за сложности и трудоемкости идентификации этих элементов [72].

Белок-кодирующая ДНК составляет всего 1–2% генома человека; почти половину ДНК представляют повторяющиеся элементы [73]. Показано, что оставшаяся часть генома представлена, по-видимому, регуляторными последовательностями, необходимыми для тонкого контроля экспрессии генов. Отсутствие данных о геномном положении этих последовательностей означает, что еще меньше информации известно об ассоциации мутаций в таких локусах с наследственными заболеваниями. Однако известно, что многие элементы могут играть роль в предрасположенности к различным заболеваниям [72].

Кроме того, остается открытым вопрос, какие факторы отвечают за модулирование работы генов, т.е. тонкую градуальную подстройку экспрессии гена.

Таким образом, идентификация и аннотирование некодирующих *cis*-регуляторных элементов является одной из основных задач современной геномики.

1.2.1. *Cis*-элементы: механизмы действия

Для связывания РНК-полимеразы с промотором гена требуется участие транскрипционных факторов, которые в свою очередь должны связаться с *cis*-регуляторными элементами. *Cis*-регуляторные элементы могут быть располо-

жены на расстоянии тысяч н.п. от промотора гена в некодирующих последовательностях как перед геном, так и внутри интронов и после гена.

Три основные класса дистальных *cis*-элементов представляют *энхансеры*, *сайленсеры* и *инсуляторы*.

Энхансеры составляют большинство *cis*-элементов. Модель действия энхансера основана на сближении промотора и удаленного регуляторного сайта, содержащего сайты связывания транскрипционных факторов [74].

Механизм действия сайленсеров также может быть основан на сближении промотора и сайта, несущего, однако, сайты связывания репрессорных белков. Помимо этого, связывание белка с сайленсером может мешать связыванию факторов с близлежащим энхансером [75].

Наконец, инсуляторы защищают гены от воздействия как энхансеров, так и сайленсеров. С одной стороны, инсуляторные белки могут на основе конкуренции препятствовать связыванию других регуляторных белков с сайтом. С другой стороны, инсуляторные белки могут ограничивать распространение как активного, так и конденсированного состояния хроматина вдоль молекулы ДНК [76].

1.2.2. Методы идентификации *cis*-элементов

Одним из экспериментальных методов определения *cis*-элементов является делеция участков ДНК и наблюдение за изменением транскрипции гена. Определение сайтов чувствительности к ДНКазе I и ДНК-футпринтинг позволяют идентифицировать последовательности, потенциально связывающие транскрипционные белки [77].

Секвенирование геномных последовательностей различных видов млекопитающих открыло новые способы поиска *cis*-элементов с помощью методов компаративной геномики. Идентификация функциональных элементов путем сравнения последовательностей различных видов животных основана на предположении, что эти элементы изменяются в гораздо меньшей степени под влиянием отбора, в то

время как соседние нефункциональные участки ДНК свободно мутируют. Таким образом, консервативные элементы генома с большей вероятностью могут иметь некоторую функциональную роль [72].

1.2.3. Биоинформатический поиск *cis*-элементов

Сравнение геномов часто используется как способ идентификации регуляторных элементов. Первые исследования, основанные на сравнении геномов *Drosophila*, показали эффективность такого подхода [78]. Схожие исследования геномов комаров позволили идентифицировать ряд регуляторных мотивов [79].

Появление последовательностей геномов большого числа организмов стимулировало дальнейшую разработку биоинформатических методов выравнивания и визуального анализа последовательностей. Ряд исследований позволил успешно идентифицировать регуляторные элементы в геномах млекопитающих на основе сравнения геномов мыши и человека [80, 81]. Тем не менее, ввиду относительно небольшой дивергенции между видами млекопитающих и низкой скорости накопления нейтральных мутаций некоторые консервативные элементы могут в действительности не иметь функционального значения [82].

Были предложены критерии отличия истинных функциональных консервативных элементов от ложных. В частности, требование 70% совпадения нуклеотидов в выравнивании длиной не менее 100 н.п. позволило успешно выявить ряд регуляторных элементов при сравнении последовательностей мыши и человека [80, 83]. Применение такого критерия к сравнению полных геномов мыши и человека выявило более 300000 консервативных элементов (что составляет около 1% генома человека), расположенных в некодирующих последовательностях с малой транскрипционной активностью [84, 85].

Было разработано несколько альтернативных методов идентификации некодирующих консервативных элементов на основе межвидовых выравниваний. В одном из них вместо прямой оценки сходства было использовано вычисление ско-

рости накопления нейтральных мутаций и времени дивергенции последовательностей друг от друга, полученной на основе построения филогенетических деревьев. Другой подход был основан на множественном выравнивании последовательностей многих видов млекопитающих [86]. Существуют вычислительные методы выявления некодирующих консервативных элементов, в частности, на основе статистических моделей эволюции геномных последовательностей [87, 88].

Тем не менее, остается невыясненным, какая доля консервативных элементов на самом деле играет некоторую регуляторную или, по меньшей мере, функциональную роль. В частности, было показано, что делеция двух протяженных некодирующих участков длиной 1 млн н.п. на 3 и 19 хромосомах мыши и содержащих более 1000 выявленных на основании сравнения геномов мыши и человека консервативных элементов не повлияла на жизнеспособность особей и не имела каких-либо других явных фенотипических проявлений. С помощью количественной ПЦР было установлено, что делеции не привели к изменению уровня экспрессии ближайших к удаленному региону генов. Анализ на трансгенных мышах энхансерной активности 15 наиболее консервативных в делетированных регионах элементов показал, что лишь одна последовательность приводила к увеличению экспрессии репортерного гена [89]. Хотя эффект делеции этих фрагментов мог не проявиться в условиях проведенных экспериментов, следует учитывать возможность, что консерватизм последовательности не всегда может указывать на наличие функциональной роли.

Стоит отметить, что экспериментальная проверка предсказанных потенциальных регуляторных элементов все еще остается очень дорогой и сложной процедурой. Таким образом, ввиду того, что функции многих консервативных элементов остаются невыясненными, а также ввиду высокой трудоемкости и временных затрат на экспериментальную проверку отдельных элементов особую важность имеет разработка подходов, которые позволили бы выделить ряд консервативных элементов с наиболее высоким регуляторным потенциалом. В частности, поиск *ультра-консервативных элементов* длиной более 200 н.п. на основе

сравнения геномов мыши и человека выявил почти 500 последовательностей, из которых более половины приходилось на некодирующий геном. По сравнению с другими видами консервативных последовательностей млекопитающих, эти элементы располагались в окрестностях генов, вовлеченных в регуляцию процессов развития и перекрывались с рядом известных энхансеров, что предполагает, что такие последовательности с высокой долей вероятности являются *одновременно* и *cis*-регуляторными элементами генома [90]. Сравнение геномов сильно отдаленных животных таких, как рыбы и млекопитающие, также позволило выявить и подтвердить функциональную роль многих консервативных элементов [72].

Таким образом, сравнение ортологичных последовательностей с целью выявления *консервативных элементов* генома является одним из основных методов поиска и предсказания *cis*-регуляторных элементов генома.

1.2.4. Выявление консервативных элементов генома с помощью построения и анализа точечных матриц гомологий

Построение и анализ точечных матриц гомологий является одним старейших методов для определения гомологичных участков двух последовательностей [91]. Такое построение обладает большой наглядностью, так как гомологичные участки легко выявляются в виде диагональных линий.

Данный метод обладает высокой чувствительностью выявления гомологичных последовательностей, позволяет легко обнаруживать локальные перестройки последовательности такие, как инсерции и делеции, определять наличие повторов и учитывать взаимное расположение гомологичных участков. Кроме того, возможность отображения *аннотаций* и добавления пользовательских аннотаций позволила бы существенно облегчить визуальный анализ и дальнейшее исследование последовательности с помощью других биоинформатических инструментов.

Однако, требование вычислительных ресурсов и время построения дотплотов растет *квадратично* вместе с увеличением длины сравниваемых последова-

тельностью, что ограничивает применимость данного метода для больших последовательностей (порядка 1 млн. н.п.).

Помимо этого, к недостаткам данного метода следует отнести возможность одновременного сравнения только *двух* последовательностей. Одним из способов преодоления этого недостатка является построение *матрицы* из дотплотов, что позволяет проводить сравнение последовательностей «каждый с каждым» [92].

Для большого числа входных последовательностей данный вид анализа становится трудоемким. Кроме того, отсутствие возможности отображения аннотаций еще больше осложняет сравнение. Тем не менее, возможность *в интерактивном режиме* добавлять пользовательские аннотации позволила бы легко провести сравнение «один со многими» и, таким образом, осуществлять *множественное* сравнение последовательностей.

Глава 2

Материалы и методы

2.1. Выбор видов млекопитающих

Для исследования были отобраны 36 видов млекопитающих, обладающих различными морфо-физиологическими характеристиками (таблица 2.1). В качестве морфо-физиологических характеристик были рассмотрены масса и размер тела взрослого животного, возраст полового созревания и продолжительность жизни. Значения морфо-физиологических характеристик были получены из базы данных PanTHERIA [93].

Виды млекопитающих были выбраны как исходя из наличия в публичном доступе их геномных последовательностей, так и исходя из максимальности охвата диапазона варьирования морфо-физиологических характеристик. В частности, в исследование были включены как млекопитающие с наименьшей массой тела взрослого животного (*Sorex araneus*, обыкновенная бурозубка, масса 9.18 г), так и с наибольшей (*Physeter catodon*, кашалот, масса 14.5 т), а размах значений массы составил более 6 порядков.

В исследование вошли все виды млекопитающих, для которых была доступная полная версия генома.

2.2. Гены, участвующие в регуляции роста млекопитающих

В исследование были включены следующие 17 генов: 10 регулирующих рост генов, уменьшающих свою экспрессию с возрастом *Ezh2*, *Gpc3*, *Mdk*, *Mest*, *Mycn*, *Peg3*, *Plagl1*, *Smo*, *Igf2* и *E2f3* и 7 генов соматотропной оси *Gh1*, *Ghrh*, *Ghrl*, *Sst*, *Igf1*, *Igfbp3*, *Igfbp1*.

Информация по каждому гену у рассмотренных видов млекопитающих (в частности, уникальный идентификатор гена) была получена полнотекстовым по-

Таблица 2.1. Виды млекопитающих и некоторые их морфо-физиологические характеристики

Вид	Масса тела взрослого животного (г)	Длина тела взрослого животного (мм)	Половая зрелость (дни)	Продолжитель- ность жизни (месяцы)
<i>Mus musculus</i>	19.3	100	60	24
<i>Sorex araneus</i>	9.18	72	290	20
<i>Cricetulus griseus</i>	37	150	40	28
<i>Microtus ochrogaster</i>	42.5	121	34	18
<i>Tupaia chinensis</i>	180	180	90	120
<i>Rattus norvegicus</i>	282.89	200	50	30
<i>Callithrix jacchus</i>	290.21	210	430	150
<i>Pteropus alecto</i>	610.13	240	540	235
<i>Otolemur garnettii</i>	811.17	250	590	200
<i>Oryctolagus cuniculus</i>	1590.57	400	180	145
<i>Felis catus</i>	2884.8	540	280	168
<i>Macaca mulatta</i>	6455.19	550	1100	204
<i>Macaca fascicularis</i>	4569.32	520	1200	216
<i>Chlorocebus sabaues</i>	4300	560	1700	240
<i>Nomascus leucogenys</i>	7320	600	2800	336
<i>Canis lupus familiaris</i>	31756.51	900	600	220
<i>Papio anubis</i>	17728.56	700	3000	310
<i>Capra hircus</i>	47386.47	1000	450	190
<i>Pan paniscus</i>	35119.95	750	5000	490
<i>Acinonyx jubatus</i>	50577.92	1420	680	225
<i>Pan troglodytes</i>	45000	1500	3600	630
<i>Homo sapiens</i>	58540.63	1750	5300	840
<i>Ovis aries</i>	39097.89	1300	750	160
<i>Sus scrofa</i>	84471.54	1350	330	175
<i>Pongo abelii</i>	39696.12	1400	5500	360
<i>Lipotes vexillifer</i>	112138.32	1700	3100	300
<i>Gorilla gorilla</i>	112588.99	1600	3050	500
<i>Panthera tigris altaica</i>	161914.66	1830	1460	215
<i>Ursus maritimus</i>	371703.81	2000	1720	320
<i>Equus caballus</i>	403598.53	2050	760	330
<i>Tursiops truncatus</i>	281040.55	3000	3200	250
<i>Bos taurus</i>	618642.42	2500	700	180
<i>Loxodonta africana</i>	3824539.93	5000	4600	800
<i>Balaenoptera acutorostrata</i>	5587093.59	5500	2500	510
<i>Orcinus orca</i>	5628758.66	8600	5700	600
<i>Physeter catodon</i>	14540959.82	12000	3700	850

Таблица 2.2. Информация из базы данных NCBI Gene об исследуемых генах человека

Идентификатор	Ген	Описание	Хромосома	Положение в геноме
1871	E2F3	E2F transcription factor 3	6	6p22.3
3481	IGF2	insulin like growth factor 2	11	11p15.5
2146	EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit	7	7q36.1
2719	GPC3	glypican 3	X	Xq26.2
4232	MEST	mesoderm specific transcript	7	7q32.2
5178	PEG3	paternally expressed 3	19	19q13.43
5325	PLAGL1	PLAG1 like zinc finger 1	6	6q24.2
6608	SMO	smoothened, frizzled class receptor	7	7q32.1
4192	MDK	midkine	11	11p11.2
4613	MYCN	MYCN proto-oncogene, bHLH transcription factor	2	2p24.3
3484	IGFBP1	insulin like growth factor binding protein 1	7	7p12.3
2688	GH1	growth hormone 1	17	17q23.3
2691	GHRH	growth hormone releasing hormone	20	20q11.23
51738	GHRL	ghrelin and obestatin prepropeptide	3	3p25.3
3479	IGF1	insulin like growth factor 1	12	12q23.2
6750	SST	somatostatin	3	3q27.3
3486	IGFBP3	insulin like growth factor binding protein 3	7	7p12.3

иском в базе данных NCBI Gene по имени гена и последующей фильтрацией полученной таблицы по видам млекопитающих; информация об указанных генах человека приведена в таблице 2.2.

Материалом для исследования выступали *геномные последовательности окрестностей* указанных 17 генов у каждого из 36 рассмотренных видов млекопитающих. Под окрестностью гена понимался участок ДНК, содержащий последовательность самого гена и фланкирующие участки ДНК длиной 50000 н.п. до и после соответствующего гена.

Последовательности были получены в формате GenBank из базы данных NCBI Gene с помощью разработанного набора bash-скриптов *efunctions* на основе вызовов утилит из набора *E-Direct* для командной строки UNIX (раздел 2.2.1).

Таблица 2.3. Информация из базы данных NCBI Gene об исследуемых генах *человека*, регулируемых механизмом TPE-OLD

Идентификатор	Ген	Описание	Хромосома	Положение в геноме
7015	TERT	telomerase reverse transcriptase	5	5p15.33
8470	SORBS2	sorbin and SH3 domain containing 2	4	4q35.1
4851	NOTCH1	notch receptor 1	9	9q34.3
9636	ISG15	ISG15 ubiquitin like modifier	1	1p36.33
1832	DSP	desmoplakin	6	6p24.3
716	C1S	complement C1s	12	12p13.31

Всего была извлечена 561 последовательность.

При анализе связи морфо-физиологических характеристик с положением генов на хромосоме дополнительно были рассмотрены 6 генов *Isg15*, *Dsp*, *C1s*, *Tert*, *Notch1* и *Sorbs2*, регулируемых механизмом TPE-OLD (таблица 2.3).

2.2.1. Набор скриптов *efunctions* для взаимодействия с базами данных NCBI

Одной из решаемых в работе задач было извлечение в автоматическом режиме различной геномной информации из баз данных NCBI. Сюда входили задачи получения геномных последовательностей окрестностей генов, получения списка SNP (база данных NCBI dbSNP) и вариаций (база данных NCBI dbVar) в заданном геномном диапазоне. Для решения этих задач был создан набор bash-скриптов *efunctions*. Реализация была выполнена на основе вызовов утилит из набора *E-Direct* для командной строки UNIX [94].

Ниже приведено взаимодействие с набором скриптов *efunctions* на примере гена гормона роста *Ghl* человека.

Стандартное представление в базе данных NCBI Gene локуса гормона роста *Ghl* человека (сборка генома GRCh38.p7), находящегося в геномном интервале (63917193..63918852) хромосомы 17 (идентификатор NC_000017.11) на обратнo-комплементарной цепочке (2), имеет вид:

```
location="NC_000017.11_63917193_63918852_2"
```

Получение последовательности интервала в формате GenBank осуществляется функцией `fetch_seq`:

```
echo "$location" | fetch_seq "gb" > seq.gb
```

Получение всех SNP в формате docsum (XML-формат) в интервале осуществляется вызовом функции `fetch_snps_in_location`:

```
echo "$location" | fetch_snps_in_location "docsum" > snps.docsum
```

Аналогично, получение всех вариаций, *перекрывающихся* с интервалом, в формате docsum осуществляется вызовом функции: `fetch_vars`:

```
echo "$location" | fetch_vars "docsum" > vars.docsum
```

Для преобразования полученных файлов в формате docsum в файлы с аннотациями в формате GenBank были разработаны *конвертеры* `dbsnp_docsum2gbs` и `dbvar_docsum2gbs` соответственно (указание геномного интервала требуется для коррекции относительных координат):

```
java dbsnp_docsum2gbs "$location" snps.docsum snps.gbs
```

```
java dbvar_docsum2gbs "$location" vars.docsum vars.gbs
```

Наконец, слияние аннотаций из файлов `snps.gbs` и `vars.gbs` с файлом с геномной последовательности `seq.gb` осуществляется вызовом подпрограммы `mergegb` из набора *mblast* (раздел 2.4.1):

```
java mergegb seq.gb snps.gbs vars.gbs
```

При работе с ортологичными генами удобнее манипулировать не геномными интервалами, а уникальными *идентификаторами* генов.

В частности, ген гормона роста *Gh1* человека имеет идентификатор Gene ID 2688. Получение информации о гене в формате docsum осуществляется с помощью функции `fetch_gene_data`:

```
fetch_gene_data 2688 "docsum" > 2688.docsum
```

Из полученного XML-файла `2688.docsum` информацию о гене можно извлечь различными способами; для получения в стандартном виде геномного положения была реализована функция `extract_gene_location`:

```
cat 2688.docsum | extract_gene_location
NC_000017.11 63917193 63918852 2
```

Интервал *окрестности* гена можно получить с помощью вызова функции `make_flanking_location`, возвращающей расширенный интервал, включающий *фланкирующие* области (для примера взяты 50000 н.п. *перед* геном и 30000 н.п. *после* гена; цепочка учитывается):

```
echo "NC_000017.11_63917193_63918852_2" | make_flanking_location 50000 30000
NC_000017.11 63887193 63968852 2
```

Таким образом, разработанный набор скриптов *efunctions* позволяет автоматизировать стандартный спектр задач по извлечению геномных данных, что *существенно* сокращает время на получение и первичную обработку этих данных.

2.3. Выявление консервативных элементов в окрестностях регулирующих рост генов млекопитающих

Было проведено попарное сравнение последовательностей окрестности каждого из рассмотренных генов у исследованных видов млекопитающих с последовательностью окрестности соответствующего гена *человека*.

Сравнение проводилось с помощью построения и анализа *точечной матрицы гомологий*. Данный метод обладает высокой чувствительностью выявления гомологичных последовательностей и позволяет учитывать *контекст* последовательности, в частности, структуру соседних с гомологией участков последовательности.

Для работы с точечной матрицей гомологий была разработана программа *dotolog* (раздел 2.3.1), уникальной особенностью которой является возможность *множественного* сравнения последовательностей и добавления пользовательских аннотаций.

На основании визуального анализа гомологий в последовательностях окрестностей исследованных генов были выявлены *консервативные элементы* (КЭ) ге-

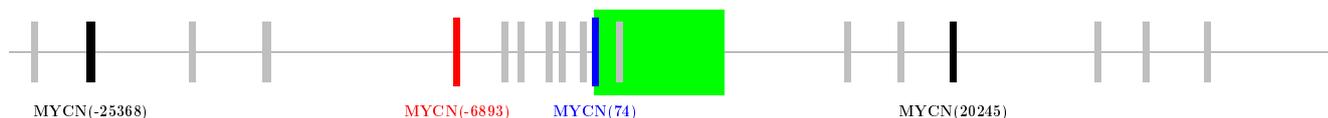


Рис. 2.1. Окрестность гена MYCN у человека и консервативные элементы внутри нее. Элемент MYCN(-6893) подсвечен красным, MYCN(74) — синим, ген *Musp* подсвечен зеленым, остальные консервативные элементы — серым

нома. Участок ДНК считался консервативным, если его последовательность сохранялась в окрестности соответствующего гена у каждого из рассмотренных видов млекопитающих.

Каждому КЭ был присвоен уникальный идентификатор, например, MYCN(-6893), где MYCN обозначало название гена, в окрестности которого находился данный КЭ, а -6893 — расстояние в н.п. от середины КЭ до начала соответствующего гена в геноме *человека* (рис. 2.1).

2.3.1. Программа для построения точечных матриц гомологий *dotolog*

Программа *dotolog* [95] предназначена для построения *точечной матрицы гомологий* (дотплота) для двух и более последовательностей. Главным нововведением является возможность построения *матрицы* из дотплотов, что позволяет проводить *множественное* выравнивание последовательностей. Отличительной особенностью программы также является поддержка отображения аннотаций к последовательностям и возможность *добавления* и *редактирования* пользовательских аннотаций, полученных по результатам визуального анализа.

Интерфейс программы реализует концепцию «*as Google Maps*», предоставляя удобство ориентирования по последовательностям. Главное окно программы дано на рисунке 2.2.

При отображении широко используется цветовое кодирование, в частности, *слова* (непрерывные последовательности совпадений) подсвечиваются зеленым цветом. Цвета стандартных геномных аннотаций даны в таблице 2.4. Пользова-



Рис. 2.2. Главное окно программы *dotolog*, демонстрирующее возможность построения ступенчатой матрицы из дотплотов

Таблица 2.4. Цвета подсветки различных элементов последовательности в программе *dotolog*

Цвет	Элемент
Темно-синий	Экзоны
Темно-зеленый	Интроны
Темно-красный	UTR
Темно-сиреневый	Другие элементы генома

тельские аннотации подсвечиваются цветом, сгенерированным по уникальному идентификатору элемента. Реализована возможность *фильтрации* отображаемых элементов.

Таким образом, предоставляемый программой *dotolog* функционал позволяет эффективно выявлять гомологичные *консервативные* элементы при анализе ортологичных последовательностей.

На программу выдано свидетельство о регистрации программы для ЭВМ №2016663454 «*Dotolog*: программа для автоматизации визуального анализа доплот-изображений нуклеотидных последовательностей ДНК» от 7 декабря 2016 г. (авторы: Романов Д. Е., Шкурат Т. П.).

2.4. Поиск известных элементов генома в окрестностях регулирующих рост генов млекопитающих

В окрестности каждого из исследуемых генов был произведен поиск последовательностей *известных элементов генома* (ИЭГ) с помощью разработанного программного конвейера *mblast*, реализующего поиск на основе алгоритма BLAST (раздел 2.4.1).

Последовательности ИЭГ были получены из баз данных, приведенных в таблице 2.5. Дополнительно были рассмотрены 697 полиморфизмов, для которых показана ассоциация с длиной тела человека [96].

В тех случаях, когда были доступны лишь данные о *положении* ИЭГ в ге-

Таблица 2.5. Базы данных известных элементов генома

База данных	Описание	Количество элементов	Сайт
EPD	The Eukaryotic Promoter Database	44599	http://epd.vital-it.ch/
JASPAR	The high-quality transcription factor binding profile database	1179986	http://jaspar.genereg.net/
ORegAnno	A community-driven resource for curated regulatory annotation	770220	http://www.oreganno.org/
LNCipedia	The long non-coding RNA database	79730	http://lncipedia.org/
miRBase	The microRNA database	40419	http://www.mirbase.org/
Pseudogene	A database of pseudogenes	78469	http://www.pseudogene.org/
dbSUPER	A database of super-enhancers	1807	http://bioinfo.au.tsinghua.edu.cn/dbsuper/
DENdb	A repository of predicted enhancers	766169	http://www.cbrc.kaust.edu.sa/dendb/
VISTA	A database of tissue-specific human enhancers	2315	https://enhancer.lbl.gov/
ENCODE	Transcription Factor ChIP-seq from ENCODE with Factorbook Motifs	4380271	http://genome.ucsc.edu/
CONDOR	A database of conserved non-coding elements	6800	http://condor.crick.ac.uk/
NCBI dbSNP	NCBI database for nucleotide variations	325658303	https://www.ncbi.nlm.nih.gov/SNP/
NCBI dbVar	NCBI database of large scale genomic variants	5226685	https://www.ncbi.nlm.nih.gov/dbvar/
NCBI BLAST	Alu repeats from NCBI BLAST databases	327	ftp://ftp.ncbi.nih.gov/blast/db/FASTA/

номе, геномные последовательности были получены с помощью применения набора программ *bedtools* [97] к соответствующей референсной последовательности генома животного. Референсные последовательности геномов животных были извлечены из баз данных NCBI и UCSC.

В качестве параметров поиска были выбраны следующие значения:

- степень покрытия 0.9,
- степень сходства 0.9 для последовательностей длиной менее 50 н.п.,
- степень сходства 0.85 для последовательностей длиной не менее 50 н.п.

Для каждого выявленного консервативного элемента был получен список перекрывававшихся с ним ИЭГ посредством реализации подпрограмм на языке Java с использованием функционала Java-библиотеки *BioJava* [98].

2.4.1. Программный конвейер для поиска гомологичных последовательностей *mblast*

Одной из решаемых в работе задачей был поиск заданных в формате FASTA известных нуклеотидных последовательностей в некоторой геномной последовательности, заданной в формате GenBank, и добавление результатов поиска в виде аннотаций в эту последовательность для удобства дальнейших расчетов. Программный конвейер *mblast* [99] был создан с целью автоматизации решения этой задачи. Разработанный инструмент позволяет производить в том числе и полногеномный поиск последовательностей [100].

Реализация была выполнена на основе использования функционала Java-библиотеки *BioJava* [98].

Семейство компьютерных программ BLAST является одним из основных инструментов для поиска гомологичных последовательностей [101]. В частности, программа `blastn` позволяет осуществлять быстрый поиск гомологичных нуклеотидных последовательностей.

Программа `blastn` требует на входе предварительно сгенерированную базу данных BLAST `seq.bdb` и искомые последовательности `query.fasta` и генерирует на выходе таблицу `blast.out` с результатами поиска:

```
blastn -db seq.bdb -query query.fasta -out blast.out
```

Генерация базы данных BLAST `seq.bdb` по исходной последовательности `seq.fasta` осуществляется вызовом:

```
makeblastdb -dbtype nucl -in seq.fasta -out seq.bdb
```

Таким образом, для решения поставленной задачи требуется:

1. конвертация исходной последовательности `seq.gb` в формате GenBank в последовательность `seq.fasta`;
2. генерация базы данных BLAST с помощью `makeblastdb`;
3. поиск гомологичных последовательностей с помощью `blastn`;

4. обработка результатов поиска и их добавление в виде аннотаций в исходную последовательность `seq.gb`.

Для подзадач 2 и 3 решения уже даны выше.

Для решения подзадачи 1 была реализована подпрограмма `seqconv`. Подпрограмма принимает на стандартный ввод конвертируемую последовательность и выдает результат на стандартный вывод:

```
cat seq.gb | java seqconv -f GENBANK -t FASTA_DNA > seq.fasta
```

Решение подзадачи 4 было разбито на 2 этапа: преобразование выходной таблицы `blast.out` в аннотации в формате GenBank `blast.gbs` и их добавление в исходный файл `seq.gb`.

Для преобразования выходной таблицы `blast.out` в аннотации формате GenBank `blast.gbs` была реализована подпрограмма `blast2gbs`:

```
java blast2gbs blast.out blast.gbs
```

Для осуществления слияния исходной последовательности `seq.gb` и полученных аннотаций `blast.gbs` была реализована подпрограмма `mergegb`:

```
java mergegb seq.gb blast.gbs
```

В результате, решение поставленной задачи выглядит следующим образом (на входе даны искомые последовательности `query.fasta` и геномная последовательность `seq.gb` в формате GenBank; на выходе — последовательность `seq.gb` с добавленными в виде аннотаций результатами поиска):

```
cat seq.gb | java seqconv -f GENBANK -t FASTA_DNA > seq.fasta
makeblastdb -dbtype nucl -in seq.fasta -out seq.bdb
blastn -db seq.bdb -query query.fasta -out blast.out
java blast2gbs blast.out blast.gbs
java mergegb seq.gb blast.gbs
```

Такая организация программного конвейера позволяет легко автоматизировать и *распараллелить* процедуру поиска сразу во многих последовательностях, что *существенно* сокращает затрачиваемое на расчеты время.

2.5. Корреляционный анализ

Был проведен корреляционный анализ зависимости между морфо-физиологическими характеристиками млекопитающих и расстоянием между консервативными элементами (КЭ) в окрестностях исследованных генов. В окрестности каждого гена были рассмотрены все возможные пары КЭ. Под расстоянием между КЭ понималось геномное расстояние в н.п. между серединами этих КЭ. Статистически значимыми считались зависимости, для которых коэффициент ранговой корреляции Спирмена между морфо-физиологической характеристикой и расстоянием между КЭ составлял не менее 0.7 при уровне значимости $p < 0.05$.

Для тех видов млекопитающих, для которых была доступна полная версия генома, был проведен корреляционный анализ зависимости между морфо-физиологическими характеристиками и расстоянием от начала гена до ближайшей теломеры как в абсолютных (н.п.), так и в относительных (доли длины хромосомы) единицах. Нулевая позиция на хромосоме соответствовала позиции первой теломеры; позиция с координатой, равной длине хромосомы, соответствовала позиции второй теломеры. Статистически значимыми считались зависимости, для которых коэффициент ранговой корреляции Спирмена между морфо-физиологической характеристикой и расстоянием от начала гена до ближайшей теломеры составлял не менее 0.7 при уровне значимости $p < 0.05$.

Для учета эффекта множественных сравнений результирующие р-значения были умножены на число сравнений в соответствии с процедурой коррекции Бонферрони.

Указанные статистические расчеты были также проведены на группе животных, из которой были исключены все приматы, кроме человека (таблица 2.6). Исключенные виды приматов приведены в таблице 2.7.

Корреляционный анализ проводился посредством реализации подпрограмм на языке Java с использованием функционала Java-библиотеки *Apache Commons Math*.

Таблица 2.6. Виды млекопитающих, за исключением приматов, кроме человека, и некоторые их морфо-физиологические характеристики

Вид	Масса тела взрослого животного (г)	Длина тела взрослого животного (мм)	Половая зрелость (дни)	Продолжитель- ность жизни (месяцы)
<i>Mus musculus</i>	19.3	100	60	24
<i>Sorex araneus</i>	9.18	72	290	20
<i>Cricetulus griseus</i>	37	150	40	28
<i>Microtus ochrogaster</i>	42.5	121	34	18
<i>Rattus norvegicus</i>	282.89	200	50	30
<i>Pteropus alecto</i>	610.13	240	540	235
<i>Oryctolagus cuniculus</i>	1590.57	400	180	145
<i>Felis catus</i>	2884.8	540	280	168
<i>Canis lupus familiaris</i>	31756.51	900	600	220
<i>Capra hircus</i>	47386.47	1000	450	190
<i>Acinonyx jubatus</i>	50577.92	1420	680	225
<i>Homo sapiens</i>	58540.63	1750	5300	840
<i>Ovis aries</i>	39097.89	1300	750	160
<i>Sus scrofa</i>	84471.54	1350	330	175
<i>Lipotes vexillifer</i>	112138.32	1700	3100	300
<i>Panthera tigris altaica</i>	161914.66	1830	1460	215
<i>Ursus maritimus</i>	371703.81	2000	1720	320
<i>Equus caballus</i>	403598.53	2050	760	330
<i>Tursiops truncatus</i>	281040.55	3000	3200	250
<i>Bos taurus</i>	618642.42	2500	700	180
<i>Loxodonta africana</i>	3824539.93	5000	4600	800
<i>Balaenoptera acutorostrata</i>	5587093.59	5500	2500	510
<i>Orcinus orca</i>	5628758.66	8600	5700	600
<i>Physeter catodon</i>	14540959.82	12000	3700	850

Таблица 2.7. Вошедшие в исследование виды приматов, за исключением человека, и некоторые их морфо-физиологические характеристики

Вид	Масса тела взрослого животного (г)	Длина тела взрослого животного (мм)	Половая зрелость (дни)	Продолжитель- ность жизни (месяцы)
<i>Tupaia chinensis</i>	180	180	90	120
<i>Callithrix jacchus</i>	290.21	210	430	150
<i>Otolemur garnettii</i>	811.17	250	590	200
<i>Macaca mulatta</i>	6455.19	550	1100	204
<i>Macaca fascicularis</i>	4569.32	520	1200	216
<i>Chlorocebus sabaeus</i>	4300	560	1700	240
<i>Nomascus leucogenys</i>	7320	600	2800	336
<i>Papio anubis</i>	17728.56	700	3000	310
<i>Pan paniscus</i>	35119.95	750	5000	490
<i>Pan troglodytes</i>	45000	1500	3600	630
<i>Pongo abelii</i>	39696.12	1400	5500	360
<i>Gorilla gorilla</i>	112588.99	1600	3050	500

2.6. Распределение по геному человека гомологов консервативных элементов

Для наиболее скоррелированных с морфо-физиологическими характеристиками консервативных элементов (КЭ) был произведен поиск гомологичных последовательностей в геноме человека с помощью сервиса *Ensembl BLAST* (алгоритм: *BLASTN*; чувствительность поиска: *Distant homologies*).

С помощью разработанного набора bash-скриптов *efunctions* (раздел 2.2.1) из базы данных NCBI Gene для каждого гомолога был получен список генов в окрестности ± 50000 н.п.

Таблица 2.8. Категории Gene Ontology сервиса *PANTHER*

Категория Gene Ontology	Описание
GO:0003674	Molecular function
GO:0008150	Biological process
GO:0005575	Cellular component
ANNOT_TYPE_ID_PANTHER_GO_SLIM_MF	PANTHER GO Slim Molecular Function
ANNOT_TYPE_ID_PANTHER_GO_SLIM_BP	PANTHER GO Slim Biological Process
ANNOT_TYPE_ID_PANTHER_GO_SLIM_CC	PANTHER GO Slim Cellular Location
ANNOT_TYPE_ID_PANTHER_PC	Protein class

2.6.1. Анализ сверхпредставленности категорий генов в окрестностях гомологов консервативных элементов

Для полученного списка генов с помощью сервиса *PANTHER* [102] был произведен анализ сверхпредставленности категорий генов (statistical overrepresentation test) по всем доступным категориям Gene Ontology (таблица 2.8).

Запросы были выполнены с помощью реализации подпрограмм на языке Java с использованием функционала Java-библиотеки *Apache HttpComponents*.

В качестве параметров анализа были выбраны следующие значения:

- test type: Fisher's exact,
- correction: Bonferroni,
- reference list: *Homo sapiens* (all genes in database).

Гены были сгруппированы по консервативному элементу или рассмотрены все вместе. В качестве уровня значимости при группировке по консервативному элементу были выбраны значения $p < 0.1$, при рассмотрении всех генов — $p < 0.05$.

Для наиболее представленных во всех категориях генов была получена подробная информация о принадлежности к подкатегориям Gene Ontology.

2.6.2. Взаимное расположение гомологов консервативных элементов и CpG сайтов, входящих в состав эпигенетических часов Хорвата

Ранее было показано, что линейная комбинация степени метилирования группы из 353 CpG сайтов в геноме человека является хорошим предиктором возраста (т.н. *эпигенетические часы Хорвата*) [103].

Геномные координаты этих 353 CpG сайтов даны для сборки генома *hg18*. С помощью сервиса *UCSC hgLiftOver* было произведено приведение геномных координат к сборке генома *hg38*.

С помощью набора программ *bedtools* [97] был получен список CpG сайтов, находящихся на расстоянии не более 65000 н.п. от положений гомологов консервативных элементов (вызов с ключом `bedtools window`). Для перекрывающихся с такими CpG сайтами генов была получена информация о принадлежности к категориям Gene Ontology, и приведены общие для нескольких генов категории.

Глава 3

Результаты

3.1. Консервативные элементы генома в окрестностях регулирующих рост генов млекопитающих

Были исследованы окрестности 17 регулирующих рост генов *Gh1*, *Ghrh*, *Ghrl*, *Igf1*, *Sst*, *Igfbp3*, *Igfbp1*, *E2f3*, *Igf2*, *Ezh2*, *Gpc3*, *Mest*, *Peg3*, *Plagl1*, *Smo*, *Mdk* и *Mycn* у 36 видов млекопитающих с разными морфо-физиологическими характеристиками (разделы 2.1 и 2.2).

На основании попарного сравнения последовательности окрестности каждого из этих генов с последовательностью окрестности того же гена человека было выявлено 505 консервативных элементов генома (таблица 3.1). Минимальная длина консервативных элементов составила 31 н.п., максимальная — 5334 н.п., средняя — 398 н.п.

Окрестность гена *Mdk* была высококонсервативна у исследованных видов млекопитающих и была исключена из дальнейшего исследования, т.к. заведомо не могла удовлетворять гипотезе исследования.

Ген *Igfbp1* у исследованных видов млекопитающих в большинстве случаев располагался в окрестности гена *Igfbp3*, в связи с чем окрестности обоих генов были рассмотрены одновременно.

Таблица 3.1. Количество и размер выявленных консервативных элементов в окрестностях регулирующих рост генов млекопитающих

Ген	Число консервативных элементов в окрестности	Средняя длина элемента (н.п.)	Минимальная длина элемента (н.п.)	Максимальная длина элемента (н.п.)
<i>E2f3</i>	47	368	100	3556
<i>Igf2</i>	36	313	62	1360
<i>Ezh2</i>	76	282	68	2072
<i>Gpc3</i>	34	630	105	4929
<i>Mest</i>	23	234	31	971
<i>Peg3</i>	9	365	89	912
<i>Plagl1</i>	21	505	57	2162
<i>Smo</i>	15	264	55	709
<i>Mycn</i>	25	327	98	1383
<i>Gh1</i>	39	272	129	513
<i>Ghrh</i>	37	291	70	816
<i>Ghrl</i>	5	512	316	912
<i>Igf1</i>	86	697	109	5334
<i>Sst</i>	41	523	118	1577
<i>Igfbp3/1</i>	11	380	101	783
Совокупно по всем генам	505	398	31	5334

3.2. Корреляция между распределением консервативных элементов в окрестностях регулирующих рост генов и морфо-физиологическими характеристиками млекопитающих

Был проведен корреляционный анализ зависимости между морфо-физиологическими характеристиками млекопитающих и расстоянием между консервативными элементами (КЭ) в окрестностях регулирующих рост генов (раздел 2.5). Всего было произведено 44336 сравнений. Для учета эффекта множественных сравнений была сделана поправка Бонферрони с коэффициентом 50000.

Значимая связь была обнаружена только между массой и длиной тела взрослого животного и расстоянием между некоторыми КЭ в окрестностях генов *Musp* и *Plagl1* и между продолжительностью жизни и расстоянием между некоторыми КЭ в окрестности гена *Ezh2*, причем для пар КЭ в окрестности генов *Musp* и *Ezh2* коэффициент корреляции был отрицательный, а для пар КЭ в окрестности гена *Plagl1* — положительный (таблица 3.2, рис. 3.1).

Ожидается, что корреляция как с массой, так и с размером тела взрослого животного была практически одинакова, т.к. обе эти морфо-физиологические характеристики сильно скоррелированы друг с другом.

Важно отметить, что наиболее скоррелированные пары КЭ включали КЭ, перекрывающийся с *промотором* соответствующего гена.

Корреляция между возрастом полового созревания и расстоянием между КЭ обнаружена не была.

Графическое представление в *полулогарифмической* шкале обнаруженных зависимостей для наиболее скоррелированных пар КЭ (таблица 3.3) дано на рис. 3.2. Примечательно, что на каждом графике точки выстраиваются вдоль прямой линии, что указывает на наличие *экспоненциальной* зависимости между морфо-физиологической характеристикой и расстоянием между КЭ.

Таблица 3.2. Корреляция между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием между некоторыми консервативными элементами в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Элементы, перекрывавшиеся с промотором гена, отмечены жирным

Коэфф. корр. Спирмена	P-значение	Скорректированное по Бонферрони р-значение $\times 50000$	1-ый элемент	2-ой элемент	Морфо-физиологическая характеристика
-0.81	1.45×10^{-8}	0.0007	MYCN(-6893)	MYCN(74)	Масса тела взрослого животного
-0.77	1.32×10^{-7}	0.007	MYCN(-6893)	MYCN(1273)	
-0.79	1.5×10^{-7}	0.008	MYCN(-6893)	MYCN(-1587)	
-0.78	1.58×10^{-7}	0.008	MYCN(-6893)	MYCN(3711)	
-0.75	6.51×10^{-7}	0.03	MYCN(-25368)	MYCN(-4493)	
-0.75	7.31×10^{-7}	0.04	MYCN(-25368)	MYCN(74)	
0.73	6.48×10^{-7}	0.03	PLAGL1(79389)	PLAGL1(112171)	
-0.8	4.71×10^{-8}	0.002	MYCN(-6893)	MYCN(74)	Длина тела взрослого животного
-0.77	2.12×10^{-7}	0.01	MYCN(-6893)	MYCN(1273)	
-0.77	3.08×10^{-7}	0.02	MYCN(-6893)	MYCN(3711)	
-0.77	3.79×10^{-7}	0.02	MYCN(-6893)	MYCN(-1587)	
-0.76	5.9×10^{-7}	0.03	MYCN(-25368)	MYCN(-4493)	
0.74	3.37×10^{-7}	0.02	PLAGL1(79389)	PLAGL1(112171)	
-0.79	4.07×10^{-7}	0.02	EZH2(-8314)	EZH2(2278)	Продолжительность жизни
-0.78	5.37×10^{-7}	0.03	EZH2(-8314)	EZH2(194)	
-0.78	5.99×10^{-7}	0.03	EZH2(-8314)	EZH2(1627)	

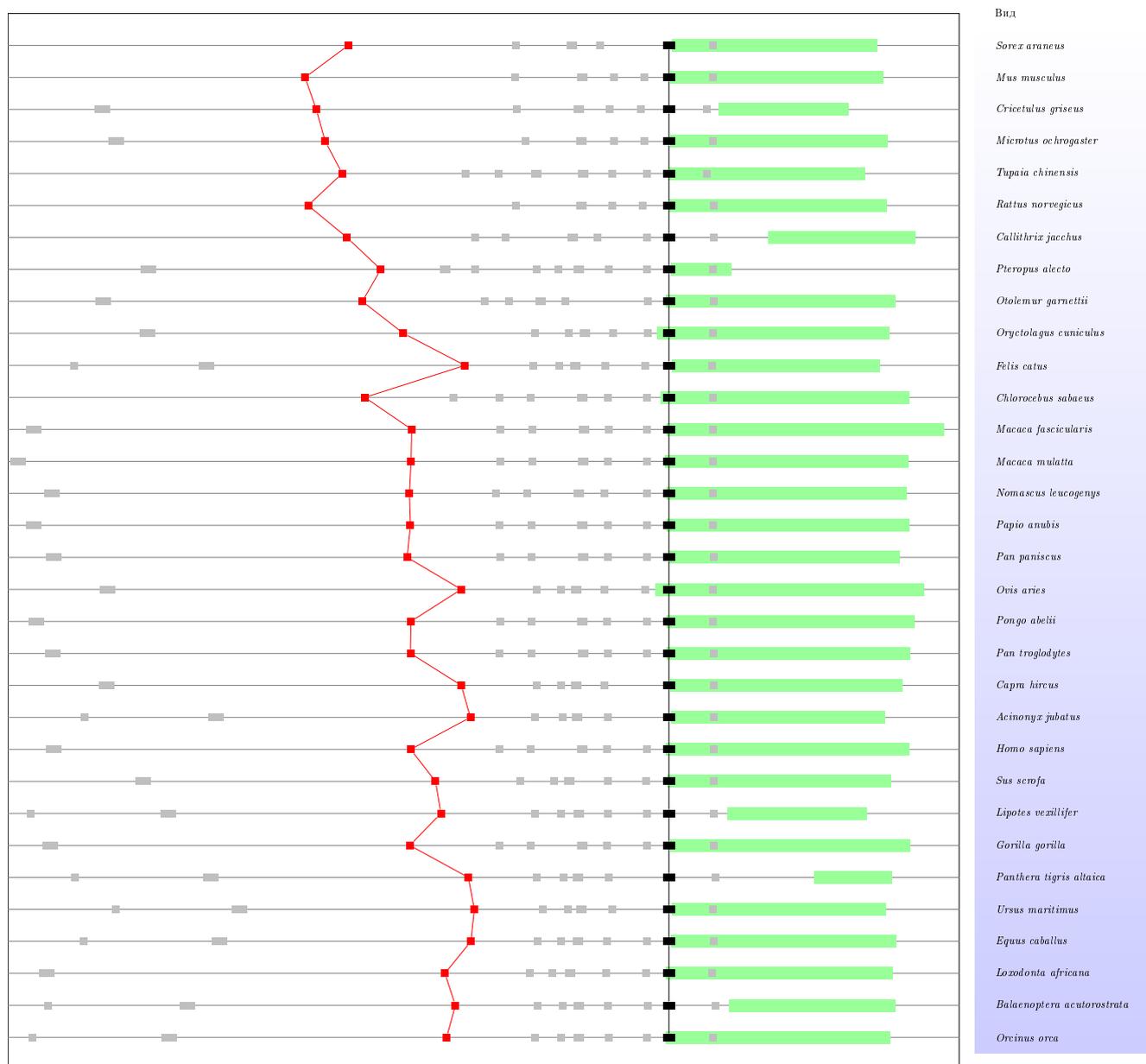


Рис. 3.1. Изменение расстояния между консервативными элементами MYCN(-6893) (подсвечен красным) и MYCN(74) (подсвечен черным) в окрестности гена *Musn* у различных видов млекопитающих при возрастании массы тела взрослого животного. Ген подсвечен зеленым, остальные консервативные элементы — серым. Последовательности окрестностей гена *Musn* центрированы по элементу MYCN(74). Млекопитающие упорядочены по массе тела взрослого животного

Таблица 3.3. Морфо-физиологические характеристики млекопитающих и геномное расстояние между наиболее скоррелированными парами консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Везде дано скорректированное по Бонферрони р-значение $\times 50000$. Элементы, перекрывающиеся с промотором гена, отмечены жирным

Вид	Масса тела взрослого животного (г)	Длина тела взрослого животного (мм)	Продолжитель- ность жизни (месяцы)	Расстояние между элементами пары (н.п.)		
				MYCN(-6893) – MYCN(74)	PLAGL1(79389) – PLAGL1(112171)	EZH2(-8314) – EZH2(194)
<i>Mus musculus</i>	19.3	100	24	9813	16700	
<i>Sorex araneus</i>	9.18	72	20	8653	20565	
<i>Cricetulus griseus</i>	37	150	28	9513	20016	
<i>Microtus ochrogaster</i>	42.5	121	18	9270	19221	
<i>Tupaia chinensis</i>	180	180	120	8813	34230	
<i>Rattus norvegicus</i>	282.89	200	30	9732	16954	
<i>Callithrix jacchus</i>	290.21	210	150	8686	35168	10860
<i>Pteropus alecto</i>	610.13	240	235	7772	29476	7823
<i>Otolemur garnettii</i>	811.17	250	200	8269	28891	
<i>Oryctolagus cuniculus</i>	1590.57	400	145	7160	30031	11345
<i>Felis catus</i>	2884.8	540	168	5501	36567	9551
<i>Macaca mulatta</i>	6455.19	550	204	6956	30704	9128
<i>Macaca fascicularis</i>	4569.32	520	216	6939	30693	9267
<i>Chlorocebus sabaeus</i>	4300	560	240	8202	31018	8559
<i>Nomascus leucogenys</i>	7320	600	336	6999	33339	8767
<i>Canis lupus familiaris</i>	31756.51	900	220		37555	9167
<i>Papio anubis</i>	17728.56	700	310	6972	30711	9329
<i>Capra hircus</i>	47386.47	1000	190	5600	34969	11560
<i>Pan paniscus</i>	35119.95	750	490	7061	32755	8524
<i>Acinonyx jubatus</i>	50577.92	1420	225	5356	36016	9348
<i>Pan troglodytes</i>	45000	1500	630	6971	32774	8519
<i>Homo sapiens</i>	58540.63	1750	840	6967	32782	8508
<i>Ovis aries</i>	39097.89	1300	160	5601	34013	11277
<i>Sus scrofa</i>	84471.54	1350	175	6295	33227	10880
<i>Pongo abelii</i>	39696.12	1400	360	6958		8555
<i>Lipotes vexillifer</i>	112138.32	1700	300	6150	35416	8493
<i>Gorilla gorilla</i>	112588.99	1600	500	6990	32822	9607
<i>Panthera tigris altaica</i>	161914.66	1830	215	5410	36125	9688
<i>Ursus maritimus</i>	371703.81	2000	320	5249	37651	9319
<i>Equus caballus</i>	403598.53	2050	330	5342	35752	9313
<i>Tursiops truncatus</i>	281040.55	3000	250		36951	9023
<i>Bos taurus</i>	618642.42	2500	180		33986	10410
<i>Loxodonta africana</i>	3824539.93	5000	800	6041	38866	6780
<i>Balaenoptera acutorostrata</i>	5587093.59	5500	510	5760	35595	8381
<i>Orcinus orca</i>	5628758.66	8600	600	5996	35185	8297
<i>Physeter catodon</i>	14540959.82	12000	850		36767	8040
Корреляция с массой тела взрослого животного				$r_s = -0.81,$ $p = 0.0007$	$r_s = 0.73,$ $p = 0.03$	
Корреляция с длиной тела взрослого животного				$r_s = -0.8,$ $p = 0.002$	$r_s = 0.74,$ $p = 0.02$	
Корреляция с продолжительностью жизни						$r_s = -0.78,$ $p = 0.03$

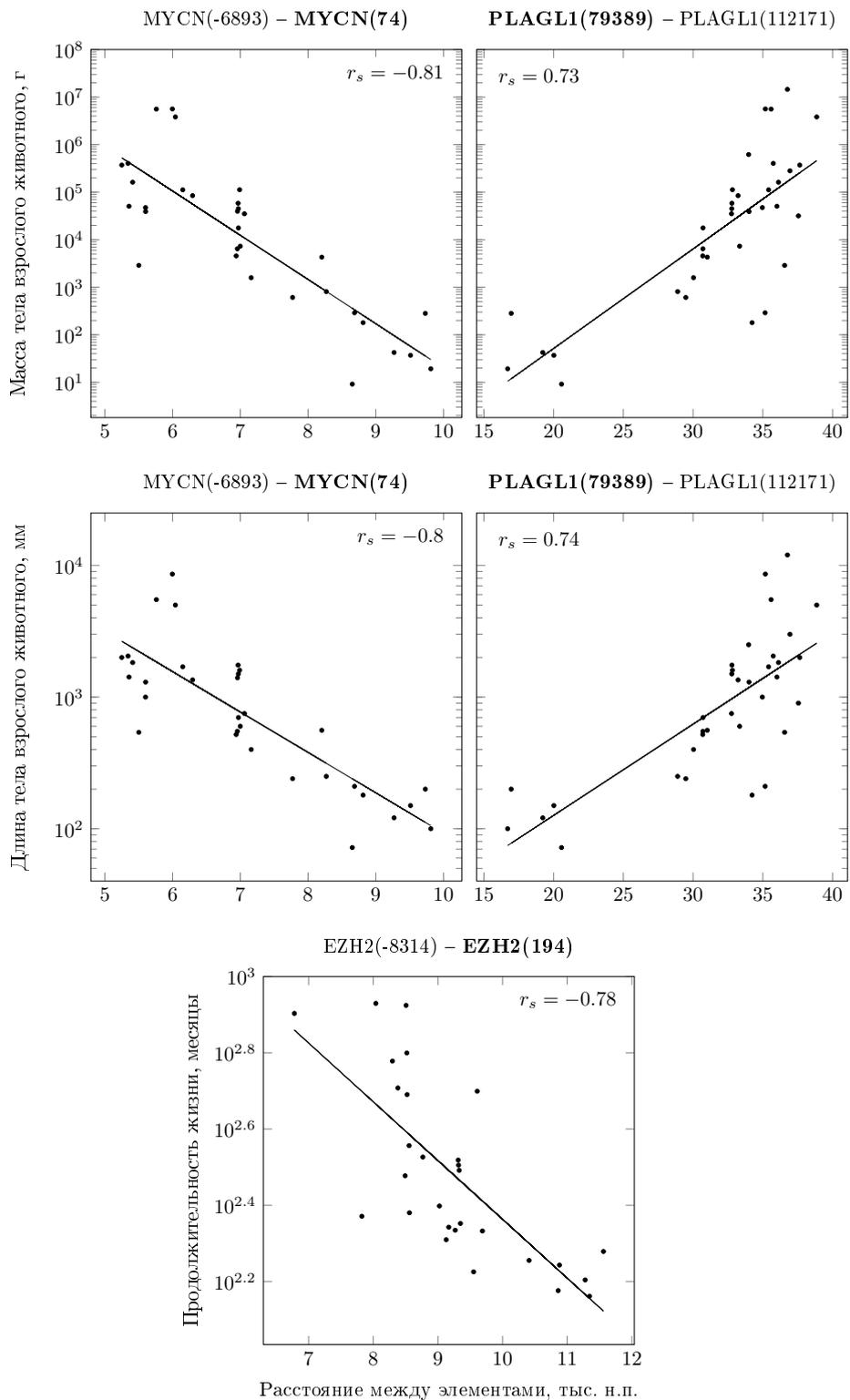


Рис. 3.2. Графическое представление в *полулогарифмической* шкале зависимости между морфофизиологическими характеристиками млекопитающих и геномным расстоянием между наиболее скоррелированными парами консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Элементы, перекрывавшиеся с промотором гена, отмечены жирным. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

Последовательности ДНК КЭ из наиболее скоррелированных пар КЭ приведены в таблице 3.4.

3.2.1. Статистический анализ на группе млекопитающих, за исключением приматов, кроме человека

Наличие в выборке животных большого количества видов приматов (раздел 2.5, таблица 2.7) может повлиять репрезентативность выборки и результаты. В связи с этим статистические исследования были также проведены на группе животных, из которой были исключены все приматы, кроме человека (раздел 2.5, таблица 2.6). Результаты корреляционного анализа приведены в таблице 3.5.

Графическое представление в *полулогарифмической* шкале обнаруженных зависимостей (таблица 3.6) дано на рис. 3.3.

Следует отметить, что пары консервативных элементов, демонстрировавшие максимальную корреляцию с морфо-физиологическими характеристиками на всей группе животных, также показывали максимальную корреляцию на выборке без приматов, что согласуется с вышеполученными результатами. Однако, в связи с уменьшившимся объемом выборки наблюдаемые р-значения оказываются значительно больше соответствующих р-значений на выборке со всеми животными.

3.3. Известные элементы генома, перекрывавшиеся с консервативными элементами в окрестностях регулирующих рост генов

Для каждого консервативного элемента (КЭ) был получен список перекрывавшихся с ним *известных элементов генома* (раздел 2.4; таблица 3.7).

Ни один КЭ не перекрывался ни с одним из следующих элементов генома: Alu-повторы, длинные некодирующие РНК, псевдогены, VISTA-энхансеры, пре-миРНК и зрелые миРНК. Кроме того, ни один КЭ не содержал внутри себя ни

Таблица 3.4. Последовательности ДНК консервативных элементов, входивших в пары наиболее скоррелированных с морфо-физиологическими характеристиками консервативных элементов.

Элементы, перекрывававшиеся с промотором гена, отмечены жирным

Элемент	Последовательность	Длина (н.п.)
MYCN(-6893)	TCCAGGACAGTCTCCAGCTGAACTTTGAGGAAAGAATTTGAATAAGCAGA AATGAAAGGAAAAGCCTTTCTAGGCCAGAAGTAGCAGCTTCTGGAAAT	98
MYCN(74)	GAAAAGCAAGTGGCTTTTGGCGCGAAAGCCTTGGCGCCTCCCCGTGATTTT TATGGAAATCAGGAGGGCGGGGTAAAGCCGCTTTCCTCTCCTTTCTCCCT CCCCCTTGTCTGCGCCACAGCCCCCTTCTCTCCCCGCCCGGGTGTGT CAGATTTTTCAGTTAATAATATCCCCGAGCTTCAAAGCGCAGGCTGTGA CAGTCATCTGTCTGGACGCGCTGGGTGGATGCGGGGGCTCCTGGGAACT GTGTGGAGCCGAGCAAGCGCTAGCCAGGCGCAAGCGCGCACAGACTGTA GCCATCC	307
PLAGL1(79389)	TTTAAATATGCATATGAACCTACAGTGATCTCAAATATATAATTAACAAA TGCATGTGCAAATATTATACAAAACCTAGTAATTTTATTTTGTGTGCAG ATTGTTCCCTGCCTGATTCATAAACATCGTG	130
PLAGL1(112171)	TCTTCTTCCCGTCTGATCCTCTCTCATCCCCAAAATAATCAACATGAGCA GAACGAAGCAGATGGCACAAGCCTTCTTTGGCATACTGGCAGATCTGTAT CACTCAGTATTTT TAGAATCAAGGGGACTGAAAATATATGTCCTGAGATGT TTTGGGGTAAGACTTACTATCTTCTAGGGTGACAGGTGTGAAATAGCA CCGTATTTAAGACTCAAATGCAATTTCTATGTCAAGTCACAGAACAGAAT GACTCGCATCAGGACCATTGCTCTGTAAGTCTGTTAGGATCGTTGG GTCTTCTTTCACTTAACACTATATTTATCAGATTCATTCACATGTGTGA TGTAGCTATAGATCTCTGCATCATGTAGGATATCCATTATGTAAT AACCACAATTTATTCATCTACCTTTGATGGGCATTTGGATAGTTCCAG TTTGCATTATGA	297
EZH2(-8314)	GTCTTCTTTCACTTAACACTATATTTATCAGATTCATTCACATGTGTGA TGTAGCTATAGATCTCTGCATCATGTAGGATATCCATTATGTAAT AACCACAATTTATTCATCTACCTTTGATGGGCATTTGGATAGTTCCAG TTTGCATTATGA	162
EZH2(194)	CGCGGGGAAACGAGCGCGCGGT TAAAACCGTTACCACCCCGAGTTTGG AACTGGTTCAAACCTGGCTTCCAGCACCCGCCCGCCCCCTCCCCGCCCG GGAACCTGCGGCGCCGGTTCCCGCCAAGAGCCGCGGCGCTTCGTCCCG CCCTTCGGCCGGTTCCCGCCACCTATCTCCCGCCTCCCGTCCGCGGCG GGCTCCGGGCCCCGCGATGTCTCCCGGTCCCGCGTGCCTGCACACCGC CTTCC TGAGAGGCGCGTGTGTTCAGCGAAAGAACAAGAGACGGCGGCG GCGCTTCCACACGGCCAGTGGCGTCCCTTACAGCGAACCCCGCCGCCG CGCGCGCGCACGCGCTGCCAGTGCCCGCCCGCCACGAGCCCTGAGCGCA CTCTGCGTGGGGCTGGCTCGGCGCCTCCGAGCCCGGCGGGCCCTGTGATT GGACGGGCGCCCGCCTCGCGTCCCGCCAATCGGGGCGGCGCTTGATTGGG CTGGGGGGGCAAATAAAAGCGATGGCGATTGGGCTGCCGCTTTGGCGC TCGGTCCGGTTCGCTCCGACACCCGGTGGGACTCAGAAGGCAGTGGAGCC CCGGCGGCGGCGGCGGCGCGGGGGCGACGCGCGGGAACAACGCGA GTCGGCGCGGGGACGAAGGTAACGCGCCGCTGCGGGCGGCCCGGCCGCG GGGGCTCCGGGAGTGCGAACCGGGCGGCG	729

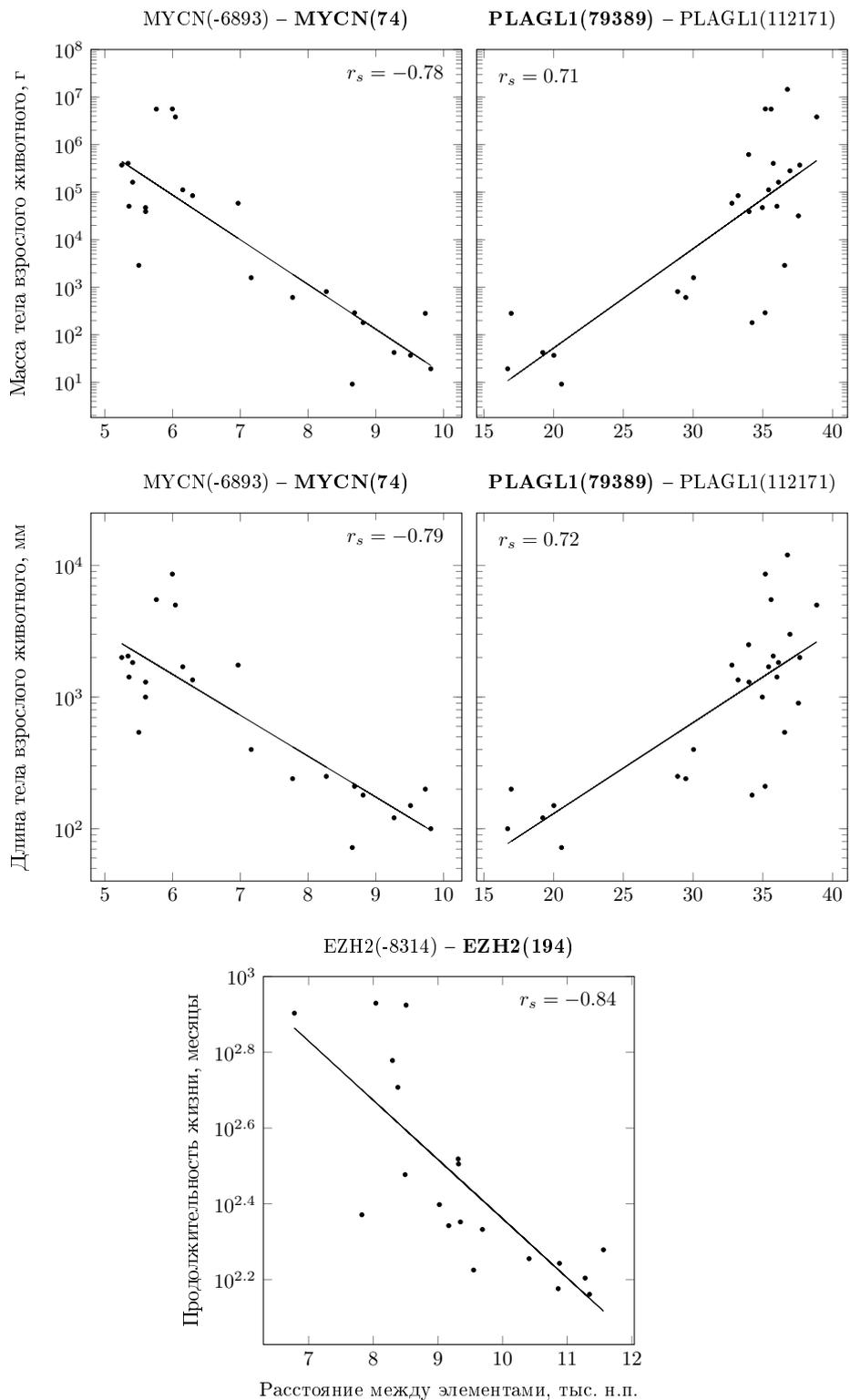


Рис. 3.3. Графическое представление в *полулогарифмической* шкале зависимости между морфофизиологическими характеристиками млекопитающих и геномным расстоянием между наиболее скоррелированными парами консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. **Все виды приматов, кроме человека, исключены.** Элементы, перекрывавшиеся с промотором гена, отмечены жирным. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

Таблица 3.5. Корреляция между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием между некоторыми консервативными элементами в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения. Элементы, перекрывавшиеся с промотором гена, отмечены жирным

Коэфф. корр. Спирмена	Р-значение	1-ый элемент	2-ой элемент	Морфо-физиологическая характеристика
-0.78	1.02×10^{-5}	MYCN(-6893)	MYCN(74)	Масса тела
0.71	3.92×10^{-5}	PLAGL1(79389)	PLAGL1(112171)	взрослого животного
-0.79	8.56×10^{-6}	MYCN(-6893)	MYCN(74)	Длина тела
0.72	2.73×10^{-5}	PLAGL1(79389)	PLAGL1(112171)	взрослого животного
-0.84	2.97×10^{-6}	EZH2(-8314)	EZH2(194)	Продолжительность жизни

одного из 697 полиморфизмов, ассоциированных с ростом [96].

Ожидается, что большинство КЭ, ассоциированных с началом гена, перекрывались с сайтами чувствительности к ДНКазе и содержали множество сайтов связывания различных транскрипционных факторов.

Важно отметить, что для каждого из генов *Mycn* и *Ezh2* пара наиболее скоррелированных КЭ содержала элемент, который перекрывался с промотором того же гена. Для гена *Plagl1* единственная значимо скоррелированная пара КЭ не содержала такого элемента, однако она содержала элемент, перекрывавшийся с началом этого гена у некоторых млекопитающих (*Tupaia chinensis*, *Pan paniscus*, *Equus caballus*) и предположительно несущий в себе промоторные последовательности.

Геномные элементы, перекрывавшиеся с выявленными консервативными элементами, могут быть задействованы в регуляции соответствующего гена. Были рассмотрены следующие пары элементов (перекрывающиеся геномные элементы даны в квадратных скобках; элементы, перекрывавшиеся с промотором гена, вы-

Таблица 3.6. Морфо-физиологические характеристики млекопитающих и геномное расстояние между наиболее скоррелированными парами консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения. Элементы, перекрывающиеся с промотором гена, отмечены жирным

Вид	Масса тела взрослого животного (г)	Длина тела взрослого животного (мм)	Продолжительность жизни (месяцы)	Расстояние между элементами пары (н.п.)		
				MYCN(-6893) – MYCN(74)	PLAGL1(79389) – PLAGL1(112171)	EZH2(-8314) – EZH2(194)
<i>Mus musculus</i>	19.3	100	24	9813	16700	
<i>Sorex araneus</i>	9.18	72	20	8653	20565	
<i>Cricetulus griseus</i>	37	150	28	9513	20016	
<i>Microtus ochrogaster</i>	42.5	121	18	9270	19221	
<i>Tupaia chinensis</i>	180	180	120	8813	34230	
<i>Rattus norvegicus</i>	282.89	200	30	9732	16954	
<i>Callithrix jacchus</i>	290.21	210	150	8686	35168	10860
<i>Pteropus alecto</i>	610.13	240	235	7772	29476	7823
<i>Otolemur garnettii</i>	811.17	250	200	8269	28891	
<i>Oryctolagus cuniculus</i>	1590.57	400	145	7160	30031	11345
<i>Felis catus</i>	2884.8	540	168	5501	36567	9551
<i>Canis lupus familiaris</i>	31756.51	900	220		37555	9167
<i>Capra hircus</i>	47386.47	1000	190	5600	34969	11560
<i>Acinonyx jubatus</i>	50577.92	1420	225	5356	36016	9348
<i>Homo sapiens</i>	58540.63	1750	840	6967	32782	8508
<i>Ovis aries</i>	39097.89	1300	160	5601	34013	11277
<i>Sus scrofa</i>	84471.54	1350	175	6295	33227	10880
<i>Lipotes vexillifer</i>	112138.32	1700	300	6150	35416	8493
<i>Panthera tigris altaica</i>	161914.66	1830	215	5410	36125	9688
<i>Ursus maritimus</i>	371703.81	2000	320	5249	37651	9319
<i>Equus caballus</i>	403598.53	2050	330	5342	35752	9313
<i>Tursiops truncatus</i>	281040.55	3000	250		36951	9023
<i>Bos taurus</i>	618642.42	2500	180		33986	10410
<i>Loxodonta africana</i>	3824539.93	5000	800	6041	38866	6780
<i>Balaenoptera acutorostrata</i>	5587093.59	5500	510	5760	35595	8381
<i>Orcinus orca</i>	5628758.66	8600	600	5996	35185	8297
<i>Physeter catodon</i>	14540959.82	12000	850		36767	8040
Корреляция с массой тела взрослого животного				$r_s = -0.78,$ $p = 1.02 \times 10^{-5}$	$r_s = 0.71,$ $p = 3.92 \times 10^{-5}$	
Корреляция с длиной тела взрослого животного				$r_s = -0.79,$ $p = 8.56 \times 10^{-6}$	$r_s = 0.72,$ $p = 2.73 \times 10^{-5}$	
Корреляция с продолжительностью жизни						$r_s = -0.84,$ $p = 2.97 \times 10^{-6}$

Таблица 3.7. Известные элементы генома, перекрывающиеся с консервативными элементами. Элементы, перекрывающиеся с промотором гена, отмечены жирным. Если некоторый консервативный элемент перекрывался со множеством известных элементов генома, то дано только суммарное их количество, отмеченное курсивом

Элемент	Длина (н.п.)	База данных						
		EPD	ENCODE	JASPAR	ORegAnno	Клин. знач. dbSNP	Клин. знач. dbVar	DENdb DHS
MYCN(-25368)	416		GATA2					
MYCN(-6893)	98		GATA2		STAT1			
MYCN(-4493)	157							
MYCN(-1587)	205		<i>6 TFs</i>	ESRRB	PRDM14, TBX3			
MYCN(74)	307	MYCN_2, MYCN_1, Mycn_1	<i>12 TFs</i>		TFAP2C, SMARCA4		nsv997222	*
MYCN(1273)	142		SIN3A, RAD21, EZH2, TAF1				nsv997222	*
MYCN(3711)	1383		<i>12 TFs</i>	FOXA1, ESRRB, Klf4	FOXA1, CTCF, ETS1, ATOH1, RUNX1		nsv997222	*
PLAGL1(79389)	130				Cdx1		nsv997068, nsv510056	
PLAGL1(112171)	297			NFIC	Cdx1		nsv997068	
EZH2(-8314)	162				E2F1			
EZH2(194)	729	Ezh2_1, EZH2_1, Ezh2_2	<i>62 TFs</i>	EGR1, TCF7L2, NFYB, STAT3, E2F1, E2F3, E2F4, NRF1, E2F6	RB1, RBL2, EGR1, TFAP2C, STAT1, E2F4, FOS, ETS1, SMARCA4	rs550913824, rs886062082, rs545716282, rs886062081, rs886062083		*
EZH2(1627)	510	EZH2_2	<i>12 TFs</i>		STAT1, ETS1, SMARCA4			*
EZH2(2278)	509		<i>8 TFs</i>	STAT1, STAT3, JUND	STAT1, SMARCA4			*

делены жирным шрифтом):

- MYCN(-6893) [STAT1, GATA2] и **MYCN(74)**;
- **PLAGL1(79389)** [CDX1] и PLAGL1(112171) [CDX1, NFIC];
- EZH2(-8314) [E2F1] и **EZH2(194)**.

Показано, что белок GATA2, по-видимому, индуцирует рост посредством активации гена *Mycn* [104]. В другом исследовании было высказано предположение, что фактор STAT1 опосредованно ингибирует экспрессию гена *Mycn* [105].

По данным литературы доказательств участия факторов CDX1 и NFIC в регуляции гена *Plagl1* обнаружено не было.

Показано, что фактор E2F1 напрямую регулирует работу гена *Ezh2* [106].

Таким образом, указанные белки действительно могут участвовать в регуляции исследуемых генов.

Для каждой пары значимо скоррелированных КЭ был также получен список *вариаций* между данными КЭ. Ни один такой интервал *не содержал полностью* клинически значимых вариаций; тем не менее, были обнаружены клинически значимые вариации, *перекрывавшиеся* с этими интервалами (таблица 3.8).

Вариация nsv997222 затрагивает весь ген *Mycn* и ассоциирована с синдромом Фейнгольда, характеризующимся в том числе задержкой роста [107].

Вариация nsv510056, перекрывавшаяся с геномным интервалом между элементами PLAGL1(79389) и PLAGL1(112171), ассоциирована с пузырьным заносом (MeSH D006828), который является нарушением нормального развития организма.

Таким образом, это может свидетельствовать о том, что регуляция генов *Mycn* и *Plagl1* может зависеть от взаимодействия регуляторных элементов, затрагиваемых указанными вариациями.

Таблица 3.8. Клинически значимые вариации, *перекрывававшиеся* с интервалами между парами значимо скоррелированных консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*.

Консервативные элементы, перекрывававшиеся с промотором гена, отмечены жирным

1-ой элемент	2-ой элемент	Клин. знач. dbVar
MYCN(-25368)	MYCN(-4493)	
MYCN(-25368)	MYCN(74)	
MYCN(-6893)	MYCN(-1587)	
MYCN(-6893)	MYCN(74)	
MYCN(-6893)	MYCN(1273)	nsv997222
MYCN(-6893)	MYCN(3711)	nsv997222
PLAGL1(79389)	PLAGL1(112171)	nsv997068, nsv510056
EZH2(-8314)	EZH2(194)	
EZH2(-8314)	EZH2(1627)	
EZH2(-8314)	EZH2(2278)	

3.4. Представленность в геноме человека гомологов консервативных элементов

Представленность в геноме последовательностей, гомологичных последовательностям наиболее скоррелированных с морфо-физиологическими характеристиками млекопитающих консервативных элементов (раздел 3.2, таблица 3.4), может отражать функциональную значимость последовательностей этих элементов.

Для данных последовательностей был произведен поиск *гомологов* в геноме человека с помощью сервиса *Ensembl BLAST* (таблицы 3.9 и 3.10).

Представленность гомологов в геноме человека приведена в таблице 3.11. Примечательно, что один из элементов каждой пары представлен лишь в одной копии в геноме, т.е. является уникальной последовательностью генома, в то время как второй элемент пары — многими копиями. Последнее обстоятельство может указывать на потенциальную регуляторную роль таких последовательностей.

Распределение в геноме человека гомологов каждого из элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314) приведено на рис. 3.4. Следует от-

Таблица 3.9. Результаты BLAST-поиска в геноме человека гомологов консервативных элементов MYCN(-6893) и PLAGL1(79389). Гомологи пронумерованы

Гомолог	Хро-мо-сома	Начало	Конец	Перекрываемые гены	Цепочка	Начало вхо-ждения	Конец вхо-ждения	Длина совпа-дения	BLAST Score	BLAST E-val	Сходство (%)
MYCN(-6893)_1	1	185092123	185092169	RNF2	-	29	78	50	42.5	0.19	80
MYCN(-6893)_2	1	14006975	14007017	KAZN	-	21	62	43	38.3	3.8	81.4
MYCN(-6893)_3	2	15933497	15933594	MYCNUT, MYCNOS	+	1	98	98	142	2.00E-31	100
MYCN(-6893)_4	2	219807320	219807371		-	21	72	53	41.1	0.52	79.25
MYCN(-6893)_5	2	80513927	80513974	CTNNA2	-	31	80	50	39.7	1.4	78
MYCN(-6893)_6	2	161638763	161638803	SLC4A10	-	21	60	41	38.3	3.8	82.93
MYCN(-6893)_7	3	86804970	86805012		+	21	62	43	38.3	3.8	81.4
MYCN(-6893)_8	3	120814424	120814461		-	25	63	39	38.3	3.8	84.62
MYCN(-6893)_9	3	175032619	175032667	NAALADL2	-	20	67	49	38.3	3.8	77.55
MYCN(-6893)_10	4	36502369	36502411	LINC02505	+	21	62	43	38.3	3.8	81.4
MYCN(-6893)_11	4	100438037	100438079	EMCN	-	21	62	43	38.3	3.8	81.4
MYCN(-6893)_12	4	114504142	114504182		-	22	62	41	38.3	3.8	80.49
MYCN(-6893)_13	4	167191435	167191474	SPOCK3	-	25	64	41	38.3	3.8	85.37
MYCN(-6893)_14	5	125881369	125881429		+	12	69	62	42.5	0.19	77.42
MYCN(-6893)_15	6	17017405	17017447		+	21	62	43	41.1	0.52	83.72
MYCN(-6893)_16	6	115761805	115761847		+	21	62	43	41.1	0.52	83.72
MYCN(-6893)_17	7	136628338	136628378		-	21	60	41	41.1	0.52	85.37
MYCN(-6893)_18	8	96554924	96555001	SDC2	-	17	97	81	38.3	3.8	67.9
MYCN(-6893)_19	9	96466519	96466596	HABP4	-	13	89	82	39.7	1.4	70.73
MYCN(-6893)_20	10	82827506	82827548	NRG3	+	29	69	43	38.3	3.8	81.4
MYCN(-6893)_21	11	105187684	105187732	AP003181.1	+	21	69	49	38.3	3.8	75.51
MYCN(-6893)_22	12	2403515	2403558	CACNA1C	-	21	65	45	38.3	3.8	80
MYCN(-6893)_23	12	76975139	76975177		-	19	57	39	38.3	3.8	82.05
MYCN(-6893)_24	14	49169857	49169900		-	21	63	44	39.7	1.4	81.82
MYCN(-6893)_25	17	66328402	66328470	PRKCA	+	27	97	72	39.7	1.4	72.22
MYCN(-6893)_26	X	66625113	66625155	EDA2R	+	4	50	47	38.3	3.8	80.85
MYCN(-6893)_27	X	142246477	142246517		+	21	61	41	38.3	3.8	80.49
MYCN(-6893)_28	X	147464980	147465023		-	21	64	45	38.3	3.8	82.22
MYCN(-6893)_29	X	154973482	154973520	F8	-	25	62	39	38.3	3.8	84.62
PLAGL1(79389)_1	1	110832970	110833013		-	1	44	44	48.3	0.009	86.36
PLAGL1(79389)_2	1	113272830	113272895		-	1	66	66	48.3	0.009	74.24
PLAGL1(79389)_3	3	102553213	102553302		+	2	92	92	51.1	0.001	71.74
PLAGL1(79389)_4	5	101327602	101327679		+	2	79	79	49.7	0.003	73.42
PLAGL1(79389)_5	6	143985145	143985274	PLAGL1	-	1	130	130	188	6.00E-45	100
PLAGL1(79389)_6	6	50489477	50489575		+	3	103	102	56.8	2.00E-05	72.55
PLAGL1(79389)_7	11	24455123	24455206		-	2	75	84	51.1	0.001	72.62
PLAGL1(79389)_8	11	87064254	87064334	TMEM135	-	2	83	82	48.3	0.009	70.73
PLAGL1(79389)_9	12	12046327	12046423		+	2	91	101	52.5	4.00E-04	72.28

Таблица 3.10. Результаты BLAST-поиска в геноме человека гомологов консервативного элемента EZH2(-8314). Гомологи пронумерованы

Гомолог	Хро-мо-сома	Начало	Конец	Перекрываемые гены	Цепочка	Начало вхождения	Конец вхождения	Длина совпадения	BLAST Score	BLAST E-val	Сходство (%)
EZH2(-8314)_1	1	158205026	158205180		+	4	153	155	95.4	9.00E-17	73.55
EZH2(-8314)_2	1	10270362	10270510	KIF1B	+	5	153	149	86.8	3.00E-14	69.8
EZH2(-8314)_3	1	178347886	178348037	RASAL2	-	7	153	153	86.8	3.00E-14	71.9
EZH2(-8314)_4	2	171738375	171738527	DYNC112	+	3	152	157	92.5	6.00E-16	74.52
EZH2(-8314)_5	2	133555859	133556009	NCKAP5, AC011243.1	-	3	153	153	89.7	5.00E-15	71.9
EZH2(-8314)_6	2	136691874	136692023		-	3	154	152	88.3	1.00E-14	70.39
EZH2(-8314)_7	2	64123874	64124029	PELI1	-	3	154	157	86.8	3.00E-14	70.7
EZH2(-8314)_8	2	134215049	134215202	MGAT5	+	4	156	161	86.8	3.00E-14	70.81
EZH2(-8314)_9	2	200544127	200544285	SGO2	+	1	154	159	86.8	3.00E-14	70.44
EZH2(-8314)_10	2	63739991	63740116	WDPCP	+	33	154	126	85.4	9.00E-14	74.6
EZH2(-8314)_11	2	147426700	147426847		-	3	153	154	85.4	9.00E-14	70.78
EZH2(-8314)_12	3	16104724	16104883		+	3	154	160	91.1	2.00E-15	71.25
EZH2(-8314)_13	3	69852446	69852596	MITF	+	4	160	158	88.3	1.00E-14	70.89
EZH2(-8314)_14	4	78710350	78710508		+	3	154	160	88.3	1.00E-14	71.25
EZH2(-8314)_15	5	15138214	15138362		-	3	148	150	88.3	1.00E-14	72
EZH2(-8314)_16	5	176753495	176753650		-	3	153	158	88.3	1.00E-14	72.78
EZH2(-8314)_17	7	148892895	148893056		-	1	162	162	233	2.00E-58	100
EZH2(-8314)_18	7	106654398	106654558	AC004917.1	-	3	154	163	98.3	1.00E-17	73.62
EZH2(-8314)_19	7	35159184	35159337	DPY19L2P1	+	4	154	158	94	2.00E-16	73.42
EZH2(-8314)_20	7	103251963	103252116	DPY19L2P2	+	4	154	158	91.1	2.00E-15	72.78
EZH2(-8314)_21	7	109174027	109174182		+	4	154	157	89.7	5.00E-15	71.97
EZH2(-8314)_22	7	29692913	29693066	DPY19L2P3	-	4	154	158	88.3	1.00E-14	72.15
EZH2(-8314)_23	7	32720696	32720849	DPY19L1P1	+	4	154	158	88.3	1.00E-14	72.15
EZH2(-8314)_24	8	100845513	100845672		-	1	156	161	89.7	5.00E-15	70.81
EZH2(-8314)_25	8	95284013	95284165	C8orf37-AS1	-	3	154	156	88.3	1.00E-14	71.79
EZH2(-8314)_26	9	33893337	33893491	UBE2R2	+	3	147	155	92.5	6.00E-16	73.55
EZH2(-8314)_27	9	68464017	68464145	PGM5	+	20	152	134	85.4	9.00E-14	75.37
EZH2(-8314)_28	9	97713467	97713623		+	4	156	164	85.4	9.00E-14	71.95
EZH2(-8314)_29	10	62365800	62365962		-	6	160	163	86.8	3.00E-14	70.55
EZH2(-8314)_30	11	92182833	92182970		-	17	152	139	92.5	6.00E-16	74.1
EZH2(-8314)_31	11	96040929	96041093	MAML2	+	3	152	166	88.3	1.00E-14	71.08
EZH2(-8314)_32	11	11186701	11186850		-	7	153	152	85.4	9.00E-14	71.71
EZH2(-8314)_33	12	51963571	51963715	ACVR1B	+	3	149	147	92.5	6.00E-16	72.11
EZH2(-8314)_34	12	63641584	63641737	DPY19L2, AC084357.2	+	4	154	158	91.1	2.00E-15	72.78
EZH2(-8314)_35	12	77053657	77053804	E2F7	-	3	153	151	89.7	5.00E-15	72.19
EZH2(-8314)_36	12	112560987	112561141		+	3	153	156	88.3	1.00E-14	71.15
EZH2(-8314)_37	12	67700728	67700890		+	4	154	164	85.4	9.00E-14	71.34
EZH2(-8314)_38	14	94182545	94182701	PPP4R4	+	3	154	157	98.3	1.00E-17	73.25
EZH2(-8314)_39	14	51073383	51073534	TRIM9	-	3	153	153	86.8	3.00E-14	71.24
EZH2(-8314)_40	14	74380516	74380682		-	3	162	173	86.8	3.00E-14	71.1
EZH2(-8314)_41	15	89707875	89708028	WDR93	-	3	153	159	89.7	5.00E-15	74.21
EZH2(-8314)_42	15	48092639	48092783		+	16	154	146	85.4	9.00E-14	72.6
EZH2(-8314)_43	16	73940369	73940507		-	13	146	139	89.7	5.00E-15	73.38
EZH2(-8314)_44	16	83760428	83760577	CDH13, AC009063.2	-	3	149	152	88.3	1.00E-14	73.03
EZH2(-8314)_45	18	47315444	47315582	MIR4527HG	+	23	153	140	85.4	9.00E-14	73.57
EZH2(-8314)_46	20	6118120	6118272	FERMT1	-	4	153	154	88.3	1.00E-14	71.43
EZH2(-8314)_47	X	118765439	118765598	IL13RA1	+	4	162	161	95.4	9.00E-17	72.05
EZH2(-8314)_48	X	128874948	128875104		+	3	153	157	86.8	3.00E-14	71.34
EZH2(-8314)_49	X	134140286	134140418		+	25	154	134	85.4	9.00E-14	73.13

Таблица 3.11. Представленность в геноме человека гомологов последовательностей консервативных элементов, входивших в пары наиболее скоррелированных консервативных элементов. Консервативные элементы, перекрывавшиеся с промотором гена, отмечены жирным. Хромосома, на которой находился исходный элемент, отмечена жирным. В скобках указано число гомологов на хромосоме

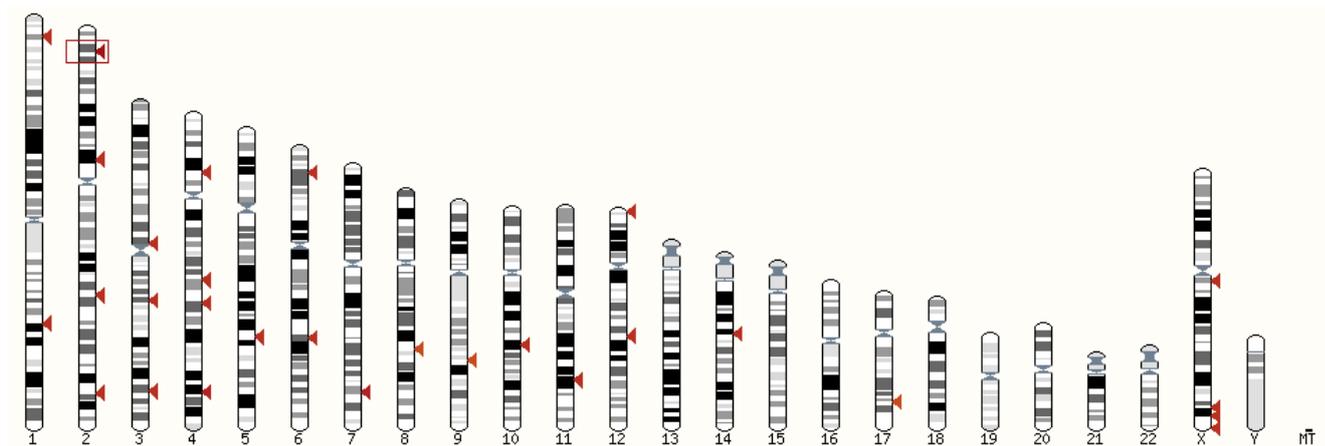
Элемент	Число гомологов в геноме человека	Число гомологов по хромосомам
MYCN(-6893)	29	1(2), 2(4) , 3(3), 4(4), 5(1), 6(2), 7(1), 8(1), 9(1), 10(1), 11(1), 12(2), 14(1), 17(1), X(4)
MYCN(74)	1	2(1)
PLAGL1(79389)	9	1(2), 3(1), 5(1), 6(2) , 11(2), 12(1)
PLAGL1(112171)	1	6(1)
EZH2(-8314)	49	1(3), 2(8), 3(2), 4(1), 5(2), 7(7) , 8(2), 9(3), 10(1), 11(3), 12(5), 14(3), 15(2), 16(2), 18(1), 20(1), X(3)
EZH2(194)	1	7(1)

метить, что гомологи первых двух элементов в основном представлены в регионах с высокой плотностью генов на хромосомах с 1-ой по 12-ую, а гомологи третьего элемента распределены по геному более равномерно.

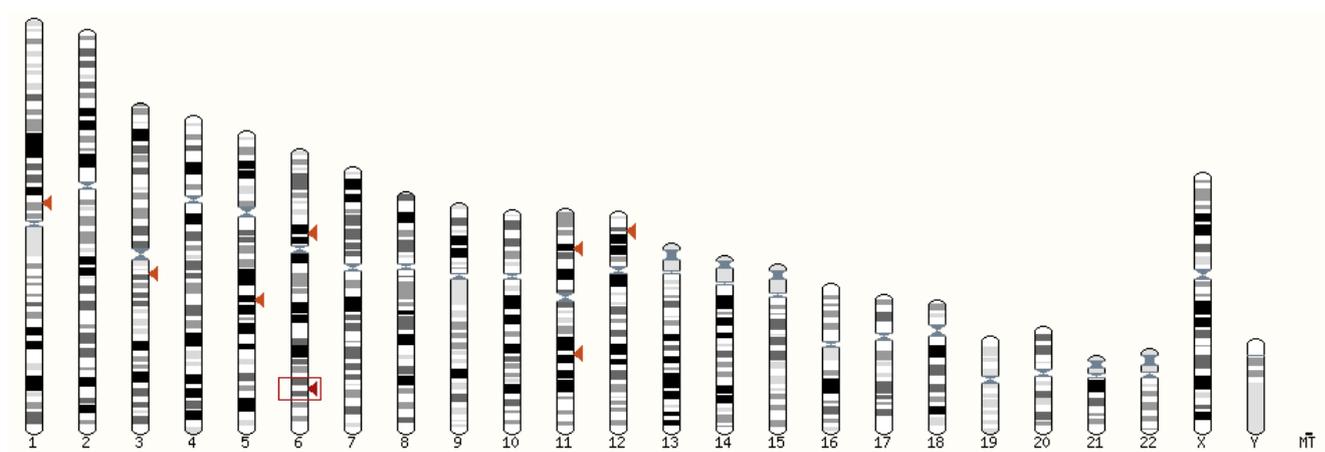
Выравнивание каждого из гомологов элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314) с исходной последовательностью показывает, что совпадения наиболее часто приходятся на среднюю часть элемента для элемента MYCN(-6893), на первую половину элемента для элемента PLAGL1(79389) и покрывают почти весь элемент в случае элемента EZH2(-8314) (рис. 3.5). Это может указывать на функциональную значимость только отдельных частей элементов.

Таким образом, последовательности элементов MYCN(74), PLAGL1(112171) и EZH2(194) можно рассматривать как уникальные, геноспецифичные регуляторные участки, в то время как последовательности элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314) могут быть отнесены к классу универсальных регуляторных последовательностей, способные контролировать любые соседние ге-

MYCN(-6893)



PLAGL1(79389)



EZH2(-8314)

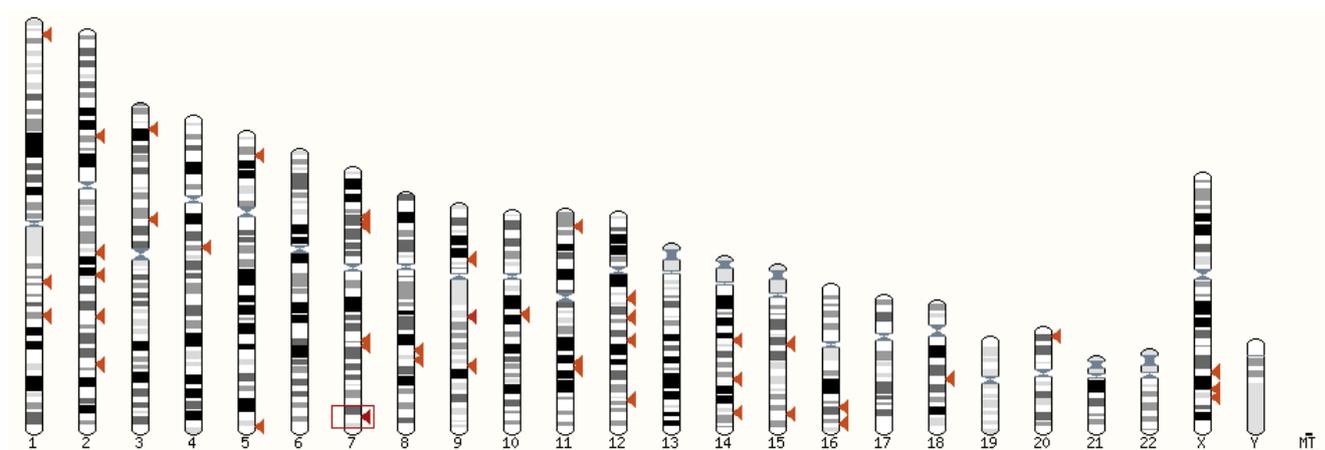
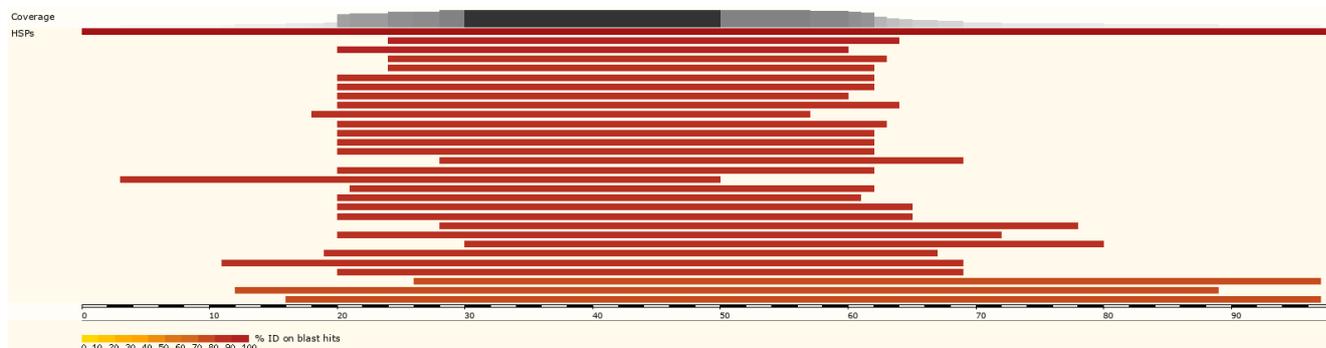
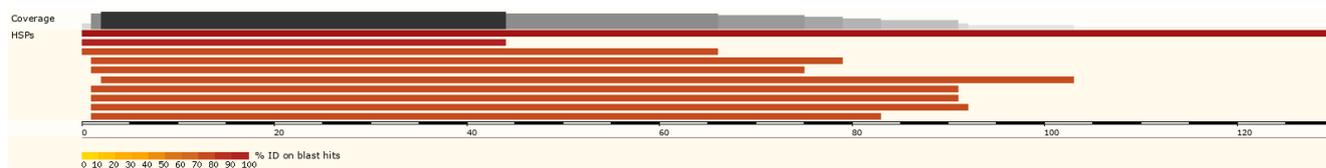


Рис. 3.4. Распределение по геному человека последовательностей, гомологичных последовательностям консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314). Красным прямоугольником выделено положение исходной последовательности

MYCN(-6893)



PLAGL1(79389)



EZH2(-8314)

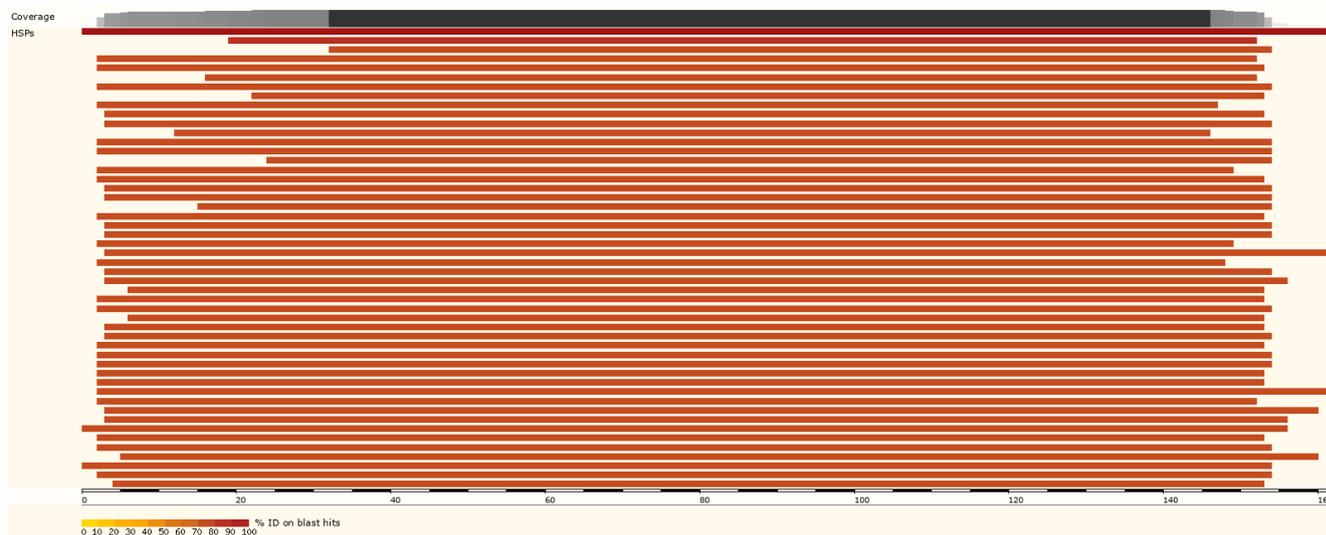


Рис. 3.5. Выравнивание гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314) в геноме человека с исходной последовательностью

ны.

3.4.1. Анализ сверхпредставленности категорий Gene Ontology среди генов в окрестностях гомологов консервативных элементов

Гены, лежащие в окрестностях гомологов элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314), могут быть также вовлечены в регуляцию роста благодаря наличию в их окрестности данных элементов.

Из базы данных NCBI Gene для каждого гомолога этих элементов с помощью набора скриптов *efunction* был получен список соседних генов в окрестности ± 50000 н.п. (таблицы 3.12 и 3.13).

Для полученного списка генов был проведен анализ сверхпредставленности категорий Gene Ontology. Гены были сгруппированы как по консервативному элементу, так и объединены все вместе в общую группу.

При группировке по консервативному элементу значимая сверхпредставленность категорий была обнаружена только для групп генов в окрестностях гомологов консервативных элементов MYCN(-6893) и EZH2(-8314) (таблица 3.14).

Значимая сверхпредставленность категорий Gene Ontology при рассмотрении объединенной группы генов дана в таблицах 3.15, 3.16, 3.17 и 3.18.

Наибольший интерес вызывают гены *ACVR1B* и *ACVRL1*, находящиеся в окрестностях гомологов элемента EZH2(-8314) и сгруппированные по категории *growth factor binding*, GO:0019838 ($p = 0.0988$). Эти же гены оказываются значимо представляющими категорию *response to growth factor*, GO:0070848 ($p = 0.0461$) при рассмотрении объединенной группы генов. Для данных генов была получена подробная информация о принадлежности к категориям Gene Ontology (таблицы 3.19 и 3.20).

Следует отметить, что оба этих гена принадлежат следующим категориям, связанных с ростом и замедлением экспрессии генов с возрастом:

- *in utero embryonic development*, GO:0001701,

Таблица 3.12. Гены в окрестностях гомологов в геноме человека консервативных элементов MYCN(-6893) (верхняя часть таблицы) и PLAGL1(79389) (нижняя часть таблицы)

Идентификатор	Ген	Описание	Хромо-сома	Положение в геноме
23254	KAZN	kazrin, periplakin interacting protein	1	1p36.21
6045	RNF2	ring finger protein 2	1	1q25.3
81627	TRMT1L	tRNA methyltransferase 1 like	1	1q25.3
100462799	FTH1P25	ferritin heavy chain 1 pseudogene 25	1	1q25.3
4613	MYCN	MYCN proto-oncogene, bHLH transcription factor	2	2p24.3
10408	MYCNOS	MYCN opposite strand	2	2p24.3
103752554	MYCNUT	MYCN upstream transcript	2	2p24.3
106480949	RN7SL104P	RNA, 7SL, cytoplasmic 104, pseudogene	2	2p24.3
57282	SLC4A10	solute carrier family 4 member 10	2	2q24.2
1496	CTNNA2	catenin alpha 2	2	2p12
2960	GTF2E1	general transcription factor IIE subunit 1	3	3q13.33
101926983	LINC02049	long intergenic non-protein coding RNA 2049	3	3q13.33
100130335	NAP1L1P3	nucleosome assembly protein 1 like 1 pseudogene 3	3	3q13.33
254827	NAALADL2	N-acetylated alpha-linked acidic dipeptidase like 2	3	3q26.31
100862679	NAALADL2-AS3	NAALADL2 antisense RNA 3	3	3q26.31
100421774	EEF1B2P8	eukaryotic translation elongation factor 1 beta 2 pseudogene 8	3	3q26.31
51705	EMCN	endomucin	4	4q24
107399301	SNORA101A	small nucleolar RNA, H/ACA box 101A	4	4q24
50859	SPOCK3	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 3	4	4q32.3
105374400	LINC02505	long intergenic non-protein coding RNA 2505	4	4p14
6383	SDC2	syndecan 2	8	8q22.1
8555	CDC14B	cell division cycle 14B	9	9q22.32-q22.33
22927	HABP4	hyaluronan binding protein 4	9	9q22.32
195828	ZNF367	zinc finger protein 367	9	9q22.32
10718	NRG3	neuregulin 3	10	10q23.1
59082	CARD18	caspase recruitment domain family member 18	11	11q22.3
79547	OR2AL1P	olfactory receptor family 2 subfamily AL member 1 pseudogene	11	11q22.3
775	CACNA1C	calcium voltage-gated channel subunit alpha1 C	12	12p13.33
144455	E2F7	E2F transcription factor 7	12	12q21.2
5578	PRKCA	protein kinase C alpha	17	17q24.2
106480529	RN7SL735P	RNA, 7SL, cytoplasmic 735, pseudogene	17	17q24.2
51438	MAGEC2	MAGE family member C2	X	Xq27.2
2157	F8	coagulation factor VIII	X	Xq28
60401	EDA2R	ectodysplasin A2 receptor	X	Xq12
963	CD53	CD53 molecule	1	1p13.3
55791	LRIF1	ligand dependent nuclear receptor interacting factor 1	1	1p13.3
81479	OR1111P	olfactory receptor family 11 subfamily I member 1 pseudogene	1	1p13.3
643232	NRBF2P3	nuclear receptor binding factor 2 pseudogene 3	1	1p13.3
5325	PLAGL1	PLAG1 like zinc finger 1	6	6q24.2
57061	HYMAI	hydatidiform mole associated and imprinted	6	6q24.2
153918	ZC2HC1B	zinc finger C2HC-type containing 1B	6	6q24.2
338645	LUZP2	leucine zipper protein 2	11	11p14.3
65084	TMEM135	transmembrane protein 135	11	11q14.2
100420509	XIAPP2	X-linked inhibitor of apoptosis pseudogene 2	11	11q14.2
100131842	HNRNPCP8	heterogeneous nuclear ribonucleoprotein C pseudogene 8	11	11q14.2
79370	BCL2L14	BCL2 like 14	12	12p13.2

Таблица 3.13. Гены в окрестностях гомологов в геноме человека консервативного элемента EZH2(-8314)

Идентификатор	Ген	Хромо-сома	Положение в геноме	Идентификатор	Ген	Хромо-сома	Положение в геноме	Идентификатор	Ген	Хромо-сома	Положение в геноме
23095	KIF1B	1	1p36.22	693134	MIR550A2	7	7p14.3	8445	DYRK2	12	12q15
107048981	RN7SL731P	1	1p36.22	441208	ZNRF2P1	7	7p14.3	101927901	LINC02421	12	12q15
106478920	RNU6-37P	1	1p36.22	100500830	MIR550B2	7	7p14.3	144455	E2F7	12	12q21.2
912	CD1D	1	1q23.1	401321	LINC00997	7	7p14.3	114088	TRIM9	14	14q22.1
909	CD1A	1	1q23.1	100129460	DPY19L1P1	7	7p14.3	646658	SYNDIG1L	14	14q24.3
729780	RPS10P8	1	1q23.1	57057	TBX20	7	7p14.2	55237	VRTN	14	14q24.3
646270	ELL2P1	1	1q23.1	554236	DPY19L2P1	7	7p14.2	326325	RPS2P2	14	14q24.3
9462	RASAL2	1	1q25.2	347701	S100A11P2	7	7p14.2	57718	PPP4R4	14	14q32.12-q32.13
344148	NCKAP5	2	2q21.2	106479596	RNU4-83P	8	8q22.3	283652	SLC24A5	15	15q21.1
106479136	RN7SKP93	2	2q21.2	100189248	TRS-AGA2-5	8	8q22.1	50804	MYEF2	15	15q21.1
4249	MGAT5	2	2q21.2-q21.3	55833	UBAP2	9	9p13.3	5346	PLIN1	15	15q26.1
100873375	RNA5SP104	2	2q21.2	54926	UBE2R2	9	9p13.3	374654	KIF7	15	15q26.1
106479871	RNU6-692P	2	2q22.3	101340252	SNORD121B	9	9p13.3	55897	MESP1	15	15q26.1
8604	SLC25A12	2	2q31.1	106481158	RNU4ATAC11P	9	9p13.3	8800	PEX11A	15	15q26.1
1781	DYNC112	2	2q31.1	5239	PGM5	9	9q21.11	56964	WDR93	15	15q26.1
316	AOX1	2	2q33.1	7507	XPA	9	9q22.33	359736	MRPL15P1	15	15q26.1
151246	SGO2	2	2q33.1	4686	NCBP1	9	9q22.33	100271646	RPL36AP43	15	15q26.1
130535	KCTD18	2	2q33.1	101928337	PTCSC2	9	9q22	106481219	RNU6-132P	15	15q26.1
100418973	CSP1	2	2p15	392371	KRT18P13	9	9q22.33	463	ZFH3	16	16q22.2-q22.3
391378	HNRNPA1P66	2	2p15	22891	ZNF365	10	10q21.2	645656	RPSAP56	16	16q22.3
57162	PEL11	2	2p14	106481084	RN7SL591P	10	10q21.2	1012	CDH13	16	16q23.3
4286	MITF	3	3p13	100874460	HMG1P22	11	11p15.4	3281	HSBP1	16	16q23.3
100505702	LINC01094	4	4q21.21	105376550	LINC02752	11	11p15.4	110175910	MIR4527HG	18	18q21.1
102577424	LINC01574	5	5q35.2	120114	FAT3	11	11q14.3	55612	FERMT1	20	20p12.3
9512	PMPCB	7	7q22.1	100271537	RPL7AP57	11	11q14.3	3597	IL13RA1	X	Xq24
349152	DPY19L2P2	7	7q22.1	84441	MAML2	11	11q21	100271393	RPL32P35	X	Xq25
729659	S100A11P1	7	7q22.1	22895	RPH3A	12	12q24.13	406899	MIR106A	X	Xq26.2
168455	CCDC71L	7	7q22.3	94	ACVRL1	12	12q13.13	574031	MIR363	X	Xq26.2
2146	EZH2	7	7q36.1	91	ACVR1B	12	12q13.13	574032	MIR20B	X	Xq26.2
6090	RNY5	7	7q36.1	160622	GRASP	12	12q13.13	407049	MIR92A2	X	Xq26.2
106481077	RN7SL569P	7	7q36.1	692159	GRASP-AS1	12	12q13.13	574033	MIR18B	X	Xq26.2
100616354	MIR550A3	7	7p14.3	106479820	RNU6-574P	12	12q13.13	406981	MIR19B2	X	Xq26.2
100271874	ZNRF2P2	7	7p14.3	283417	DPY19L2	12	12q14.2	100874417	TIMM8BP2	X	Xq26.2
442524	DPY19L2P3	7	7p14.3	100506938	S100A11P4	12	12q14.2				

Таблица 3.14. Сверхпредставленные категории Gene Ontology для генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893) и EZH2(-8314).

В окрестности гомолога какого элемента находятся	Категория Gene Ontology	Подкатегория	Описание подкатегории	Скорр. по Бонферрони р-значение	Гены в окрестности
MYCN(-6893)	protein class	PC00069	cell adhesion molecule	0.0101	SDC2, MAGEC2, F8, CTNNA2
EZH2(-8314)	cellular location	GO:0019866	organelle inner membrane	0.0934	DPY19L2P2, WDR93, PMPCB, DPY19L2
	protein class	PC00035	TGF-beta receptor	0.1	ACVR1B, ACVRL1
	function	GO:0019838	growth factor binding	0.0988	IL13RA1, ACVR1B, ACVRL1

- *negative regulation of gene expression, GO:0010629,*
- *negative regulation of cell growth, GO:0030308,*
- *cellular response to growth factor stimulus, GO:0071363,*
- *TGF-beta signaling pathway, P00052,*
- *Signaling by TGF-beta family members, R-HSA-9006936.*

Ген *ACVR1B* дополнительно принадлежит следующим категориям:

- *Gonadotropin-releasing hormone receptor pathway, P06664,*
- *Developmental Biology, R-HSA-1266738,*
- *growth factor binding, GO:0019838.*

Это указывает на то, что последовательность консервативного элемента EZH2(-8314) может являться регуляторной и участвовать в регуляции генов *ACVR1B* и *ACVRL1* одновременно с другими генами в геноме.

Таблица 3.15. Сверхпредставленные категории *Molecular function* генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)

Категория Gene Ontology	Описание	P-значение	Гены в окрестности
GO:0004675	transmembrane receptor protein serine/threonine kinase activity	0.00102	ACVR1B, ACVRL1
GO:0003774	motor activity	0.00211	KIF7, KIF1B, DYNC112
GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	0.00273	MYCN, ZNF367, PLAGL1, TBX20, E2F7, VRTN
GO:0008574	ATP-dependent microtubule motor activity, plus-end-directed	0.0035	KIF1B, DYNC112
GO:0001216	DNA-binding transcription activator activity	0.00447	MYCN, PLAGL1, TBX20
GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	0.00447	MYCN, PLAGL1, TBX20
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	0.00909	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:0001012	RNA polymerase II regulatory region DNA binding	0.00974	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:0019199	transmembrane receptor protein kinase activity	0.00999	ACVR1B, ACVRL1
GO:0000976	transcription regulatory region sequence-specific DNA binding	0.0125	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:1990837	sequence-specific double-stranded DNA binding	0.0158	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:0043565	sequence-specific DNA binding	0.0175	MYCN, ZFH3, PLAGL1, MESP1, TBX20, E2F7
GO:0140096	catalytic activity, acting on a protein	0.0185	CDC14B, ACVR1B, PELI1, NAALADL2, PRKCA, UBE2R2, RNF2, DYRK2, MGAT5, ACVRL1, PMPCB
GO:0000978	RNA polymerase II proximal promoter sequence-specific DNA binding	0.0189	MYCN, ZFH3, MESP1
GO:0003674	molecular function	0.0191	SLC24A5, PRKCA, ZFH3, KIF1B, CACNA1C, DYRK2, MYEF2, MAML2, KIF7, ACVR1B, UBE2R2, GTF2E1, XPA, AOX1, MGAT5, ACVRL1, E2F7, ZNF367, CDC14B, TRMT1L, NAALADL2, WDR93, PLAGL1, RNF2, EDA2R, SLC25A12, TBX20, DPY19L2, MYCN, NRG3, NCBP1, IL13RA1, PELI1, SLC4A10, PGM5, MESP1, CD1A, CD1D, DYNC112, VRTN, PMPCB
GO:0044212	transcription regulatory region DNA binding	0.0217	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:0001067	regulatory region nucleic acid binding	0.0217	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:0003690	double-stranded DNA binding	0.0231	MYCN, ZFH3, PLAGL1, MESP1, TBX20
GO:0005313	L-glutamate transmembrane transporter activity	0.0291	SLC25A12
GO:0003677	DNA binding	0.029	MYCN, ZFH3, XPA, PLAGL1, MESP1, TBX20, E2F7
GO:0005262	calcium channel activity	0.0312	SLC24A5, CACNA1C
GO:0140110	transcription regulator activity	0.0316	MYCN, ZNF367, PLAGL1, TBX20, E2F7, VRTN, MAML2
GO:0003700	DNA-binding transcription factor activity	0.0331	MYCN, ZNF367, PLAGL1, TBX20, E2F7, VRTN
GO:0003954	NADH dehydrogenase activity	0.0362	WDR93
GO:0048185	activin binding	0.0362	ACVR1B
GO:0015085	calcium ion transmembrane transporter activity	0.0384	SLC24A5, CACNA1C
GO:0016740	transferase activity	0.0397	ACVR1B, TRMT1L, PELI1, PRKCA, UBE2R2, RNF2, DYRK2, MGAT5, ACVRL1, DPY19L2
GO:0000339	RNA cap binding	0.0433	NCBP1
GO:0004674	protein serine/threonine kinase activity	0.0449	ACVR1B, PRKCA, DYRK2, ACVRL1
GO:0045504	dynein heavy chain binding	0.0468	DYNC112
GO:0008509	anion transmembrane transporter activity	0.0481	SLC24A5, SLC4A10, SLC25A12

Таблица 3.16. Сверхпредставленные категории *Biological process* генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314). Показаны вхождения с уровнем значимости $p < 0.01$

Категория Gene Ontology	Описание	P-значение	Гены в окрестности
GO:0044260	cellular macromolecule metabolic process	0.0006	ZNF367, TRMT1L, PRKCA, ZFHX3, RNF2, EDA2R, DYRK2, TBX20, DPY19L2, MAML2, MYCN, NCBP1, ACVR1B, PELI1, UBE2R2, GTF2E1, PGM5, XPA, MGAT5, ACVRL1, E2F7, VRTN, PMPCB
GO:0002456	T cell mediated immunity	0.00136	CD1A, CD1D
GO:0002821	positive regulation of adaptive immune response	0.00219	CD1A, CD1D
GO:0006357	regulation of transcription by RNA polymerase II	0.00224	MYCN, ZNF367, GTF2E1, ZFHX3, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:0006366	transcription by RNA polymerase II	0.00303	MYCN, ZNF367, GTF2E1, ZFHX3, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:0000122	negative regulation of transcription by RNA polymerase II	0.00475	ZFHX3, TBX20, E2F7, VRTN
GO:0006355	regulation of transcription, DNA-templated	0.00653	MYCN, ZNF367, GTF2E1, ZFHX3, EDA2R, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:1903506	regulation of nucleic acid-templated transcription	0.00653	MYCN, ZNF367, GTF2E1, ZFHX3, EDA2R, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:2001141	regulation of RNA biosynthetic process	0.00653	MYCN, ZNF367, GTF2E1, ZFHX3, EDA2R, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:0043412	macromolecule modification	0.0067	ACVR1B, TRMT1L, PELI1, PRKCA, UBE2R2, RNF2, EDA2R, DYRK2, MGAT5, ACVRL1, DPY19L2
GO:0007018	microtubule-based movement	0.00729	KIF7, CDC14B, KIF1B, DYNC1H2
GO:0034645	cellular macromolecule biosynthetic process	0.00808	ZNF367, ZFHX3, EDA2R, TBX20, DPY19L2, MAML2, MYCN, GTF2E1, PGM5, MGAT5, ACVRL1, E2F7, VRTN
GO:0009059	macromolecule biosynthetic process	0.00828	ZNF367, ZFHX3, EDA2R, TBX20, DPY19L2, MAML2, MYCN, GTF2E1, PGM5, MGAT5, ACVRL1, E2F7, VRTN
GO:0048646	anatomical structure formation involved in morphogenesis	0.0086	PGM5, MESP1, ACVRL1
GO:0006351	transcription, DNA-templated	0.00888	MYCN, ZNF367, GTF2E1, ZFHX3, EDA2R, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:0097659	nucleic acid-templated transcription	0.00888	MYCN, ZNF367, GTF2E1, ZFHX3, EDA2R, TBX20, ACVRL1, E2F7, VRTN, MAML2
GO:0032774	RNA biosynthetic process	0.00925	MYCN, ZNF367, GTF2E1, ZFHX3, EDA2R, TBX20, ACVRL1, E2F7, VRTN, MAML2

Таблица 3.17. Сверхпредставленные категории *Cellular component* генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)

Категория Gene Ontology	Описание	P-значение	Гены в окрестности
GO:0005875	microtubule associated complex	0.0011	KIF7, CDC14B, KIF1B, DYNC112
GO:0099513	polymeric cytoskeletal fiber	0.00246	KIF7, CDC14B, PGM5, KIF1B, NCKAP5
GO:0005874	microtubule	0.00374	KIF7, CDC14B, KIF1B, NCKAP5
GO:0099512	supramolecular fiber	0.00527	KIF7, CDC14B, PGM5, KIF1B, NCKAP5
GO:0099081	supramolecular polymer	0.00549	KIF7, CDC14B, PGM5, KIF1B, NCKAP5
GO:0099080	supramolecular complex	0.00549	KIF7, CDC14B, PGM5, KIF1B, NCKAP5
GO:0005856	cytoskeleton	0.00677	KIF7, CDC14B, PGM5, KIF1B, NCKAP5, DYRK2, DYNC112
GO:0044430	cytoskeletal part	0.0101	KIF7, CDC14B, PGM5, KIF1B, NCKAP5, DYNC112
GO:0005634	nucleus	0.0133	ZNF367, CDC14B, TRMT1L, ZFH3, PLAGL1, UBAP2, RNF2, DYRK2, TBX20, MYEF2, DPY19L2, MAML2, MYCN, NCBP1, HSBP1, GTF2E1, XPA, MESP1, E2F7
GO:0015630	microtubule cytoskeleton	0.014	KIF7, CDC14B, KIF1B, NCKAP5, DYNC112
GO:0030286	dynein complex	0.0157	CDC14B, DYNC112
GO:0019866	organelle inner membrane	0.0167	WDR93, DPY19L2, PMPCB
GO:0044291	cell-cell contact zone	0.0183	PGM5
GO:0043235	receptor complex	0.0199	IL13RA1, ACVR1B, ACVRL1
GO:0000109	nucleotide-excision repair complex	0.0219	XPA
GO:0044422	organelle part	0.0254	CDC14B, WDR93, PLAGL1, RNF2, KIF1B, DPY19L2, MAML2, KIF7, GTF2E1, C8orf37, PGM5, XPA, PEX11A, NCKAP5, DYNC112, E2F7, PMPCB
GO:0032991	protein-containing complex	0.0261	CDC14B, WDR93, RNF2, KIF1B, MYEF2, NCBP1, IL13RA1, KIF7, ACVR1B, GTF2E1, C8orf37, PGM5, XPA, ACVRL1, DYNC112, E2F7, PMPCB
GO:0043226	organelle	0.0262	ZFH3, KIF1B, DYRK2, MYEF2, MAML2, KIF7, HSBP1, GTF2E1, XPA, PEX11A, NCKAP5, MGAT5, E2F7, ZNF367, CDC14B, TRMT1L, WDR93, PLAGL1, UBAP2, RNF2, TBX20, DPY19L2, MYCN, NCBP1, C8orf37, PGM5, MESP1, DYNC112, PMPCB
GO:0005884	actin filament	0.0312	CDC14B, PGM5
GO:0043229	intracellular organelle	0.0354	ZFH3, KIF1B, DYRK2, MYEF2, MAML2, KIF7, HSBP1, GTF2E1, XPA, PEX11A, NCKAP5, MGAT5, E2F7, ZNF367, CDC14B, TRMT1L, WDR93, PLAGL1, UBAP2, RNF2, TBX20, DPY19L2, MYCN, NCBP1, PGM5, MESP1, DYNC112, PMPCB
GO:0044446	intracellular organelle part	0.0373	CDC14B, WDR93, PLAGL1, RNF2, KIF1B, DPY19L2, MAML2, KIF7, GTF2E1, PGM5, XPA, PEX11A, NCKAP5, DYNC112, E2F7, PMPCB
GO:0005868	cytoplasmic dynein complex	0.0433	DYNC112

Таблица 3.18. Сверхпредставленные категории *Protein class* генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)

Категория Gene Ontology	Описание	P-значение	Гены в окрестности
PC00205	serine/threonine protein kinase receptor	0.00118	ACVR1B, ACVRL1
PC00156	microtubule binding motor protein	0.0193	KIF7, KIF1B
PC00071	centromere DNA-binding protein	0.0219	SGO2
PC00157	microtubule or microtubule-binding cytoskeletal protein	0.0376	KIF7, KIF1B, DYNC112
PC00168	non-receptor tyrosine protein kinase	0.0397	DYRK2
PC00086	damaged DNA-binding protein	0.0397	XPA
PC00171	nucleic acid binding protein	0.0429	NCBP1, GTF2E1, HAPB4, XPA, SGO2, MYEF2, E2F7

3.4.2. Взаимное расположение гомологов консервативных элементов и CpG сайтов, входящих в состав эпигенетических часов Хорвата

Последовательности элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314) могут обладать регуляторными свойствами и в свою очередь изменять силу воздействия в зависимости от эпигенетических факторов таких, как метилирование ДНК. С другой стороны показано, что средневзвешенный уровень метилирования ДНК некоторых CpG сайтов в геноме человека является очень хорошим предиктором возраста (*эпигенетические часы Хорвата*) [103].

В связи с этим был получен список CpG сайтов, входящих в состав эпигенетических часов Хорвата, и находящихся на расстоянии не менее 65000 н.п. от любого из гомологов указанных элементов.

Только три CpG сайта из 353 удовлетворили данному условию (таблица 3.21). Следует отметить, что из них два сайта накапливают метилирование с возрастом.

Описание перекрываемых генов дано в таблице 3.22.

Общие онтологии перекрываемых генов даны в таблице 3.23.

Таблица 3.19. Категории *Biological process* (GO:0008150) и *Molecular function* (GO:0003674) генов

ACVRL1 и *ACVR1B*

Ген	Biological process (GO:0008150)	GO Acc	Molecular function (GO:0003674)	GO Acc
Общие для обоих генов	in utero embryonic development	GO:0001701	protein serine/threonine kinase activity	GO:0004674
	regulation of transcription, DNA-templated	GO:0006355	transmembrane receptor protein serine/threonine kinase activity	GO:0004675
	protein phosphorylation	GO:0006468	protein binding	GO:0005515
	signal transduction	GO:0007165	ATP binding	GO:0005524
	transmembrane receptor protein serine/threonine kinase signaling pathway	GO:0007178	activin receptor activity, type I	GO:0016361
	negative regulation of gene expression	GO:0010629	SMAD binding	GO:0046332
	positive regulation of pathway-restricted SMAD protein phosphorylation	GO:0010862	metal ion binding	GO:0046872
	negative regulation of cell growth	GO:0030308	activin binding	GO:0048185
	activin receptor signaling pathway	GO:0032924		
	positive regulation of transcription by RNA polymerase II	GO:0045944		
	cellular response to growth factor stimulus	GO:0071363		
	ACVRL1	angiogenesis	GO:0001525	transforming growth factor beta-activated receptor activity
response to hypoxia		GO:0001666	transforming growth factor beta receptor activity, type I	GO:0005025
regulation of endothelial cell proliferation		GO:0001936	protein kinase binding	GO:0019901
negative regulation of endothelial cell proliferation		GO:0001937	transforming growth factor beta binding	GO:0050431
positive regulation of endothelial cell proliferation		GO:0001938	BMP receptor activity	GO:0098821
lymphangiogenesis		GO:0001946		
blood vessel maturation		GO:0001955		
blood vessel remodeling		GO:0001974		
blood vessel endothelial cell proliferation involved in sprouting angiogenesis		GO:0002043		
endocardial cushion morphogenesis		GO:0003203		
regulation of DNA replication		GO:0006275		
negative regulation of cell adhesion		GO:0007162		
transforming growth factor beta receptor signaling pathway		GO:0007179		
heart development		GO:0007507		
blood circulation		GO:0008015		
regulation of blood pressure		GO:0008217		
negative regulation of cell population proliferation		GO:0008285		
negative regulation of endothelial cell migration		GO:0010596		
phosphorylation		GO:0016310		
negative regulation of cell migration		GO:0030336		
BMP signaling pathway		GO:0030509		
positive regulation of BMP signaling pathway		GO:0030513		
positive regulation of chondrocyte differentiation		GO:0032332		
wound healing, spreading of epidermal cells		GO:0035313		
dorsal aorta morphogenesis		GO:0035912		
regulation of blood vessel endothelial cell migration		GO:0043535		
negative regulation of blood vessel endothelial cell migration		GO:0043537		
negative regulation of endothelial cell differentiation		GO:0045602		
positive regulation of endothelial cell differentiation		GO:0045603		
positive regulation of angiogenesis		GO:0045766		
positive regulation of transcription, DNA-templated		GO:0045893		
negative regulation of focal adhesion assembly		GO:0051895		
lymphatic endothelial cell differentiation		GO:0060836		
artery development	GO:0060840			
venous blood vessel development	GO:0060841			
endothelial tube morphogenesis	GO:0061154			
retina vasculature development in camera-type eye	GO:0061298			
cellular response to transforming growth factor beta stimulus	GO:0071560			
cellular response to BMP stimulus	GO:0071773			
negative regulation of DNA biosynthetic process	GO:2000279			
ACVR1B	G1/S transition of mitotic cell cycle	GO:0000082	activin-activated receptor activity	GO:0017002
	hair follicle development	GO:0001942	growth factor binding	GO:0019838
	nervous system development	GO:0007399	ubiquitin protein ligase binding	GO:0031625
	central nervous system development	GO:0007417	inhibin binding	GO:0034711
	peptidyl-threonine phosphorylation	GO:0018107		
	positive regulation of activin receptor signaling pathway	GO:0032927		
	nodal signaling pathway	GO:0038092		
	positive regulation of erythrocyte differentiation	GO:0045648		
	development of primary female sexual characteristics	GO:0046545		
	protein autophosphorylation	GO:0046777		
	extrinsic apoptotic signaling pathway	GO:0097191		
	positive regulation of trophoblast cell migration	GO:1901165		

Таблица 3.20. Категории *Cellular component* (GO:0005575), *PANTHER pathway* и *REACTOME pathway* генов *ACVRL1* и *ACVR1B*

Ген	Cellular component (GO:0005575)	GO Acc	PANTHER pathway	GO Acc	REACTOME pathway	GO Acc
Общие для обоих генов	plasma membrane	GO:0005886	TGF-beta signaling pathway	P00052	Signal Transduction Signaling by TGF-beta family members	R-HSA-162582 R-HSA-9006936
	integral component of plasma membrane	GO:0005887				
	cell surface	GO:0009986				
	membrane	GO:0016020				
	receptor complex	GO:0043235				
ACVRL1	integral component of membrane	GO:0016021			Signaling by BMP	R-HSA-201451
	dendrite	GO:0030425				
	neuronal cell body	GO:0043025				
	BMP receptor complex	GO:0070724				
ACVR1B	cytosol	GO:0005829	Gonadotropin-releasing hormone receptor pathway	P06664	Signaling by NODAL	R-HSA-1181150
		GO:0048179	Wnt signaling pathway	P00057	Developmental Biology Regulation of signaling by NODAL Signaling by Activin	R-HSA-1266738 R-HSA-1433617 R-HSA-1502540

Таблица 3.21. CpG сайты, входящие в состав эпигенетических часов Хорвата и находящиеся на расстоянии не более 65000 н.п. от позиций гомологов в геноме человека консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)

Хро-мо-сома	CpG сайт	Позиция середины	Гомолог какого элемента	Позиция середины	Расстояние между серединами элементов (н.п.)	Ближайший ген	Накопление метилирования с возрастом
8	cg25070637	96493640	MYCN(-6893)	96554963	-61323	SDC2	+
2	cg08251036	134251352	EZH2(-8314)	134215126	36226	MGAT5	+
2	cg11388238	200510375	EZH2(-8314)	200544206	-33831	KCTD18	-

Таблица 3.22. Информация из базы данных NCBI Gene о генах, перекрывающихся одновременно с CpG сайтами из эпигенетических часов Хорвата и гомологами в геноме человека консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)

Идентификатор	Ген	Описание	Хромо-сома	Положение в геноме
151246	SGO2	shugoshin 2	2	2q33.1
4249	MGAT5	alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase	2	2q21.2-q21.3
6383	SDC2	syndecan 2	8	8q22.1

Таблица 3.23. Общие категории Gene Ontology для генов, перекрывающихся одновременно с CpG сайтами из эпигенетических часов Хорвата и гомологами в геноме человека консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)

Категория Gene Ontology	Подкатегория	Описание	Гены
Molecular function (GO:0003674)	GO:0005515	protein binding	SGO2, SDC2
Cellular component (GO:0005575)	GO:0016021	integral component of membrane	MGAT5, SDC2
REACTOME pathway	R-HSA-162582	Signal Transduction	SGO2, SDC2
	R-HSA-392499	Metabolism of proteins	MGAT5, SDC2
	R-HSA-597592	Post-translational protein modification	MGAT5, SDC2

Таким образом, наличие CpG сайтов в окрестности гомологов консервативных элементов MYCN(-6893) и EZH2(-8314) могло бы вызывать изменение эпигенетического статуса этих гомологов и тем самым оказывать регуляторное воздействие на данные последовательности. Тем не менее, геномные расстояния между CpG сайтами и гомологами достаточно велики для подобного рода регуляции.

3.5. Корреляция между положением на хромосоме регулирующих рост генов и морфо-физиологическими характеристиками млекопитающих

Был проведен корреляционный анализ зависимости между морфо-физиологическими характеристиками млекопитающих и расстоянием от начала каждого из исследуемых генов до ближайшей теломеры (раздел 2.5). Всего было произведено 192 сравнения. Для учета эффекта множественных сравнений была сделана поправка Бонферрони с коэффициентом 200.

Значимая связь между морфо-физиологическими характеристиками млекопитающих и расстоянием от гена до ближайшей теломеры была обнаружена только для генов *Ghrh*, *Sst*, *C1s* и *Notch1*, причем для генов *Sst*, *C1s*, *Notch1* корреляция была отрицательная, а для гена *Ghrh* — положительная (таблицы 3.24 и 3.25).

Таблица 3.24. Корреляция между морфо-физиологическими характеристиками млекопитающих и *абсолютным* геномным расстоянием (н.п.) от начала гена до ближайшей теломеры для генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD

Коэфф. корр. Спирмена	P-значение	Скорректированное по Бонферрони р-значение ×200	Ген	Морфо-физиологическая характеристика
-0.84	0.000002	0.0004	<i>C1s</i>	Возраст полового созревания
-0.81	0.00002	0.003	<i>Notch1</i>	
-0.76	0.00007	0.01	<i>C1s</i>	Продолжительность жизни
-0.74	0.0002	0.03	<i>Notch1</i>	

Таблица 3.25. Корреляция между морфо-физиологическими характеристиками млекопитающих и *относительным* геномным расстоянием (доли длины хромосомы) от начала гена до ближайшей теломеры для генов соматотропной оси *Ghrh* и *Sst* и генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD

Коэфф. корр. Спирмена	P-значение	Скорректированное по Бонферрони р-значение ×200	Ген	Морфо-физиологическая характеристика
0.82	0.000003	0.0006	<i>Ghrh</i>	Продолжительность жизни
0.74	0.00007	0.01	<i>Ghrh</i>	Возраст полового созревания
-0.72	0.0002	0.04	<i>Sst</i>	
-0.76	0.00006	0.01	<i>C1s</i>	Возраст полового созревания
-0.82	0.000009	0.002	<i>Notch1</i>	

Для каждого из этих генов расстояние от начала гена до ближайшей теломеры было максимально скоррелировано с возрастом полового созревания. Ген *Ghrh* демонстрировал еще более тесную связь с продолжительностью жизни.

Корреляция между массой или длиной тела взрослого животного и расстоянием от начала гена до ближайшей теломеры обнаружена не была.

Ожидаемо, что расстояние от начала гена до ближайшей теломеры для генов *Ghrh*, *C1s* и *Notch1* было одновременно скоррелировано и с возрастом полового созревания, и с продолжительностью жизни, т.к. данные морфо-физиологические характеристики сильно скоррелированы друг с другом.

Графическое представление в *полулогарифмической* шкале обнаруженных зависимостей между морфо-физиологическими характеристиками млекопитающих и расстоянием от начала гена до ближайшей теломеры (таблица 3.26) дано на рис. 3.6 и 3.7. Для гена *Ghrh* точки на графике выстраиваются вдоль прямой линии, что свидетельствует о наличии *экспоненциальной* зависимости между морфо-физиологической характеристикой и расстоянием от начала гена до ближайшей теломеры. Для генов *Sst*, *C1s* и *Notch1* эта закономерность была выражена слабее.

3.5.1. Статистический анализ на всей группе млекопитающих, за исключением приматов, кроме человека

Наличие в выборке животных большого количества видов приматов (раздел 2.5, таблица 2.7) может повлиять репрезентативность выборки и результаты. В связи с этим статистические исследования были также проведены на группе животных, из которой были исключены все приматы, кроме человека (раздел 2.5, таблица 2.6). Результаты корреляционного анализа приведены в таблице 3.27.

При рассмотрении данной группы животных для гена *Ghrh* выявляется зависимость между возрастом полового созревания и продолжительностью жизни и расстоянием до ближайшей теломеры в долях длины хромосомы. На всей выборке

Таблица 3.26. Возраст полового созревания млекопитающих и геномное расстояние от начала гена до ближайшей теломеры для генов соматотропной оси *Ghrh* и *Sst* и генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD

		Расстояние до ближайшей теломеры (н.п.)		Расстояние до ближайшей теломеры (доли длины хромосомы)			
Вид	Половая зрелость (дни)	<i>C1s</i>	<i>Notch1</i>	<i>Ghrh</i>	<i>Sst</i>	<i>C1s</i>	<i>Notch1</i>
<i>Mus musculus</i>	60	25194154	26503822	0.14	0.24	0.17	0.15
<i>Microtus ochrogaster</i>	34		42191646	0.22	0.34	0.15	0.45
<i>Rattus norvegicus</i>	50	27070747	3951015	0.14	0.11	0.15	0.02
<i>Callithrix jacchus</i>	430	18936796	29727454	0.2	0.09	0.3	0.14
<i>Oryctolagus cuniculus</i>	180	33037552		0.03	0.49		
<i>Felis catus</i>	280	42061844	1887402	0.16	0.5	0.3	0.02
<i>Macaca mulatta</i>	1100	7393034	1780171	0.29	0.06	0.06	0.02
<i>Macaca fascicularis</i>	1200	7387971	1795103	0.29	0.05	0.05	0.02
<i>Chlorocebus sabaeus</i>	1700	7106769	1656815	0.29	0.11	0.06	0.02
<i>Nomascus leucogenys</i>	2800	6979867	1654607	0.25	0.09	0.2	0.01
<i>Canis lupus familiaris</i>	600	7890864	12098110	0.46	0.48	0.17	0.2
<i>Papio anubis</i>	3000	7229728	1711713	0.3	0.05	0.05	0.02
<i>Capra hircus</i>	450	16675720	2829280	0.2	0.5	0.14	0.03
<i>Pan paniscus</i>	5000	7107229	1553675	0.44	0.06	0.05	0.01
<i>Pan troglodytes</i>	3600	7386605	1845129	0.49	0.05	0.05	0.02
<i>Homo sapiens</i>	5300	7060676	1848931	0.42	0.05	0.05	0.01
<i>Ovis aries</i>	750	16853139	2847499	0.21	0.28	0.08	0.01
<i>Sus scrofa</i>	330	40844541		0.34	0.38	0.39	
<i>Pongo abelii</i>	5500	7295038	1716210	0.45	0.05	0.05	0.01
<i>Gorilla gorilla</i>	3050	7131706	1676752	0.45	0.05	0.05	0.01
<i>Equus caballus</i>	760	34515253	2107195	0.44	0.43	0.41	0.05
<i>Bos taurus</i>	700	17412174	3279817	0.21	0.49	0.14	0.03
Корреляция с половой зрелостью		-0.84	-0.81	0.74	-0.72	-0.76	-0.82
Скорректированное по Бонферрони р-значение × 200		0.0004	0.003	0.01	0.04	0.01	0.002

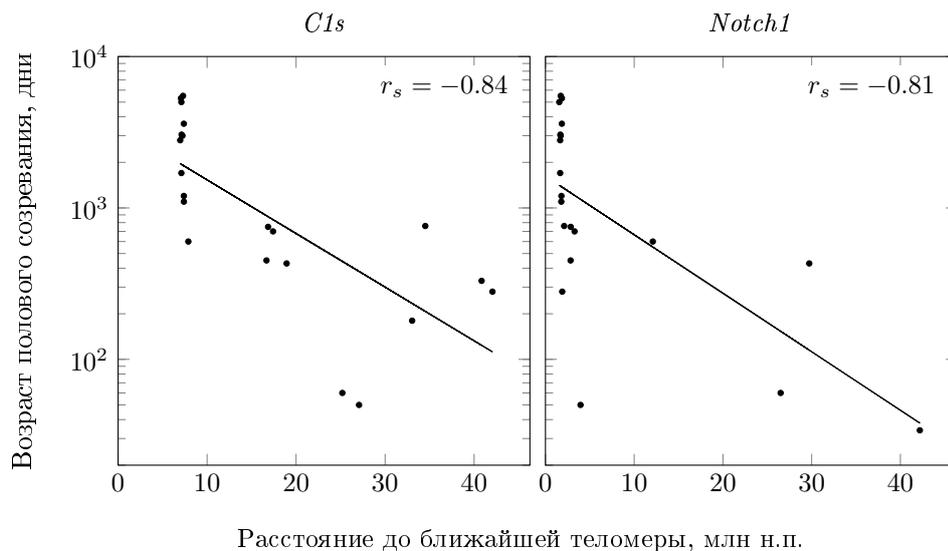


Рис. 3.6. Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *абсолютного* геномного расстояния (н.п.) между геном и ближайшей теломерой для генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

Таблица 3.27. Корреляция между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием от начала гена до ближайшей теломеры для генов соматотропной оси *Ghrh* и *Ghrl* и генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. **Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения**

Кoeff. корр. Спирмена	Р-значение	Ген	Морфо-физиологическая характеристика
Расстояние до ближайшей теломеры (н.п.)			
-0.52	0.08	<i>C1s</i>	Возраст полового созревания
-0.65	0.03	<i>Notch1</i>	
Расстояние до ближайшей теломеры (доли длины хромосомы)			
0.62	0.02	<i>Ghrh</i>	Возраст полового созревания
-0.73	0.01	<i>Ghrl</i>	
0.68	0.01	<i>Ghrh</i>	Продолжительность жизни
-0.67	0.02	<i>Ghrl</i>	

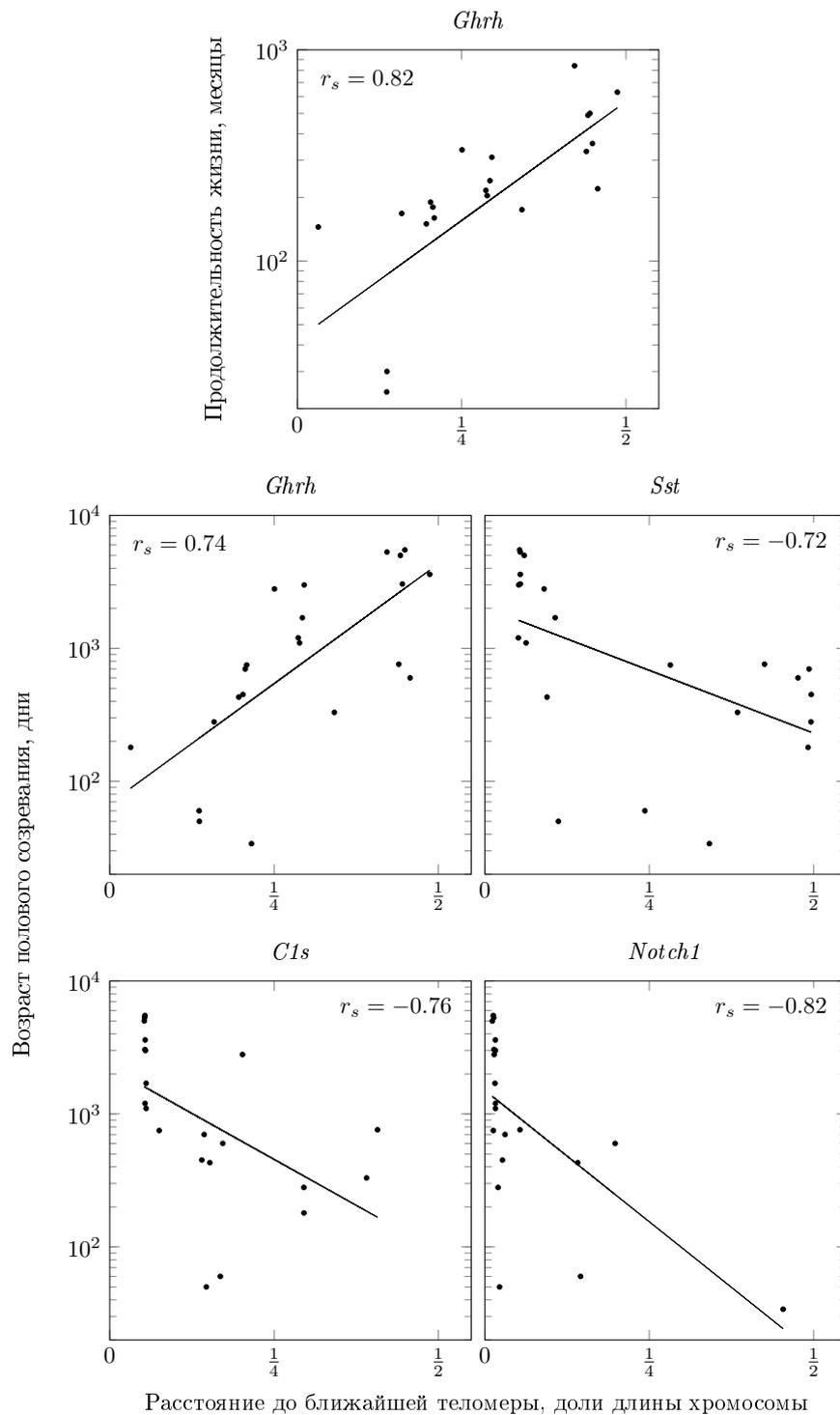


Рис. 3.7. Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *относительного* геномного расстояния (доли длины хромосомы) между геном и ближайшей теломерой для генов соматотропной оси *Ghrh* и *Sst* и генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. Для гена *Ghrh* дополнительно представлена зависимость продолжительности жизни от расстояния между геном и ближайшей теломерой. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

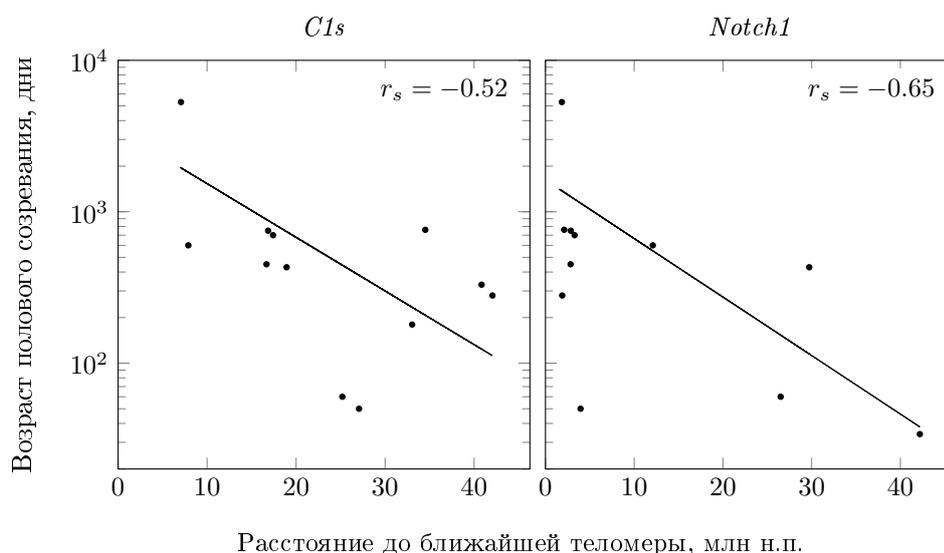


Рис. 3.8. Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *абсолютного* геномного расстояния (н.п.) между геном и ближайшей теломерой для генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. **Все виды приматов, кроме человека, исключены.** Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

животных данный ген не демонстрирует значимую корреляцию ни с одним из этих признаков ($r_s = -0.32$, $p = 0.17$ и $r_s = -0.43$, $p = 0.06$ соотв.; р-значения даны без поправки на множественные сравнения).

Графическое представление в *полулогарифмической* шкале обнаруженных зависимостей (таблица 3.28) дано на рис. 3.8 и 3.9.

Таким образом, гены, демонстрирующие максимальную корреляцию с морфо-физиологическими характеристиками на всей группе животных, также показывают максимальную корреляцию на выборке без приматов, что согласуется с вышеполученными результатами. Однако, в связи с уменьшившимся объемом выборки наблюдаемые р-значения оказываются значительно больше соответствующих р-значений на выборке со всеми животными.

Таблица 3.28. Морфо-физиологические характеристики млекопитающих и геномное расстояние от начала гена до ближайшей теломеры для генов соматотропной оси *Ghrh* и *Ghrl* и генов *CIs* и *Notch1*, регулируемых механизмом TPE-OLD. Все виды приматов, кроме человека, исключены.

R-значения даны без поправки на множественные сравнения

			Расстояние до ближайшей теломеры (н.п.)		Расстояние до ближайшей теломеры (доли длины хромосомы)	
Вид	Половая зрелость (дни)	Продолжительность жизни (мес)	<i>CIs</i>	<i>Notch1</i>	<i>Ghrh</i>	<i>Ghrl</i>
<i>Mus musculus</i>	60	24	25194154	26503822	0.14	0.24
<i>Microtus ochrogaster</i>	34	18		42191646	0.22	
<i>Rattus norvegicus</i>	50	30	27070747	3951015	0.14	0.21
<i>Callithrix jacchus</i>	430	150	18936796	29727454	0.2	0.37
<i>Oryctolagus cuniculus</i>	180	145	33037552		0.03	
<i>Felis catus</i>	280	168	42061844	1887402	0.16	0.31
<i>Canis lupus familiaris</i>	600	220	7890864	12098110	0.46	0.14
<i>Capra hircus</i>	450	190	16675720	2829280	0.2	0.07
<i>Homo sapiens</i>	5300	840	7060676	1848931	0.42	0.05
<i>Ovis aries</i>	750	160	16853139	2847499	0.21	0.1
<i>Sus scrofa</i>	330	175	40844541		0.34	0.34
<i>Equus caballus</i>	760	330	34515253	2107195	0.44	0.08
<i>Bos taurus</i>	700	180	17412174	3279817	0.21	0.11
Корреляция с половой зрелостью			$r_s = -0.52,$ $p = 0.08$	$r_s = -0.65,$ $p = 0.03$	$r_s = 0.62,$ $p = 0.02$	$r_s = -0.73,$ $p = 0.01$
Корреляция с продолжительностью жизни					$r_s = 0.68,$ $p = 0.01$	$r_s = -0.67,$ $p = 0.02$

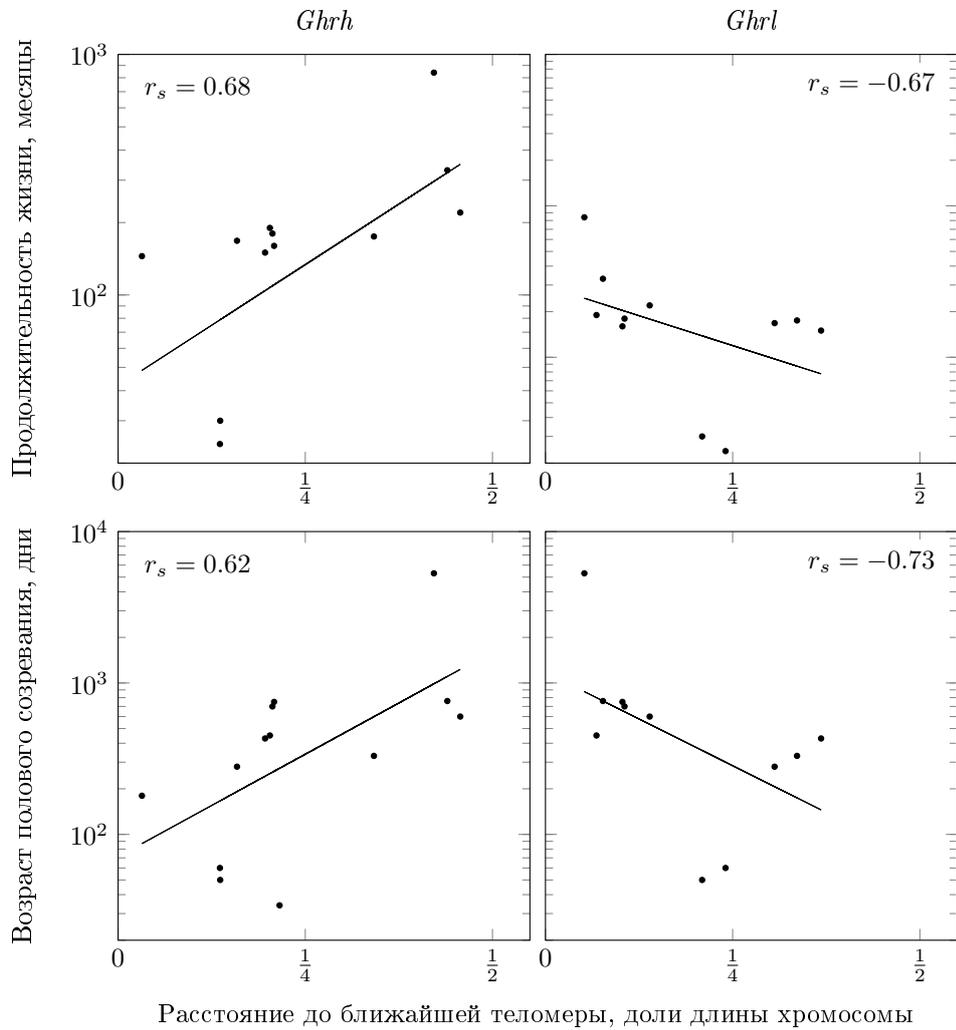


Рис. 3.9. Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания и продолжительности жизни от *относительного* геномного расстояния (доли длины хромосомы) между геном и ближайшей теломерой для генов соматотропной оси *Ghrh* и *Ghrl*. **Все виды приматов, кроме человека, исключены.** Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

Глава 4

Обсуждение

4.1. Эволюционное модулирование экспрессии регулирующих рост генов у разных видов млекопитающих

У млекопитающих показано существование единой эволюционно закрепленной генетической программы регуляции роста [4, 108, 109]. Эта программа реализуется сетью некоторых регулирующих рост генов, экспрессия которых градуально уменьшается с возрастом. Следует отметить, что регуляция этих генов у разных млекопитающих может осуществляться одними и теми же *cis*-элементами, причем каждый ген может иметь свой набор *cis*-элементов [110, 111, 112].

Известно, что кривые роста млекопитающих сходны друг с другом [21, 113]. С другой стороны, временные паттерны уменьшения скорости соматического роста сходны с временными паттернами уменьшения уровня экспрессии регулирующих рост генов [48]. Отсюда следует, что скорость роста животного может быть *пропорциональна* уровню экспрессии регулирующих рост генов.

Очевидно, масса взрослого животного есть *интеграл* по продолжительности жизни от скорости роста. Т.к. интеграл является *линейной* функцией, то модулирование экспрессии регулирующих рост генов будет приводить к *пропорциональному* изменению массы тела взрослого животного. И наоборот, межвидовые различия в массе тела взрослого животного могут указывать на тип модуляции экспрессии этих генов.

Следует отметить, что основной вклад в конечную массу тела животного вносит скорость роста, а продолжительность жизни оказывает лишь незначительное влияние, т.к. рост практически полностью останавливается на поздних этапах жизни, а основной рост приходится на период полового созревания [113].

Таким образом, могут быть введены следующие допущения:

- рост всех млекопитающих регулируется одной и той же эволюционно консервативной сетью генов;
- регуляция этих генов у разных млекопитающих осуществляется одними и теми же *cis*-элементами, причем каждый ген имеет свой набор *cis*-элементов;
- паттерны экспрессии регулирующих рост генов сходны у всех млекопитающих;
- скорость роста пропорциональна экспрессии регулирующих рост генов;
- модулирование экспрессии регулирующих рост генов ведет к пропорциональному изменению массы и размера тела взрослого животного.

При данных допущениях различие между видами млекопитающих в массе тела взрослого животного может быть вызвано несколькими причинами.

С одной стороны, может меняться *локальное* окружение гена, представляющее собой окрестность гена с его *cis*-элементами. В таком случае основным отличием между видами будет выступать различие в распределении *cis*-элементов в окрестности гена, обусловленное в основном различием в геномном *расстоянии* между *cis*-элементами.

С другой стороны, может меняться *глобальное* окружение гена, т.е. его положение в геноме, в частности, положение на хромосоме и *расстояние* до теломеры.

Известно, что регуляция некоторого параметра может осуществляться двумя путями: модуляцией и триггером. Модуляция выполняет тонкую настройку, в то время как триггер приводит к резкому изменению значения регулируемого параметра.

Эволюционная подстройка экспрессии некоторого гена может происходить по нескольким причинам. Во-первых, это может быть мутация в промоторе или регуляторном элементе. Во-вторых, это появление или удаление регуляторных элементов гена. Обе эти причины, как правило, ведут к *резкому* изменению паттерна

экспрессии. В-третьих, это транслокация гена, и если новое *локальное* окружение гена не будет оказывать влияние на экспрессию, то единственным фактором будет выступать *геномное* окружение гена.

Следует отметить, что геномное расстояние между регуляторными элементами как раз может является фактором, обеспечивающим *градуальное* изменение экспрессии гена. Во-первых, оно может меняться практически *непрерывно*, эволюционно модулируя генетическую программу. Во-вторых, последовательность нуклеотидов между функциональными элементами может свободно мутировать. Более того, геномное расстояние от гена до хромосомных элементов, в частности, до ближайшей теломеры, будет обладать теми же свойствами непрерывности изменения и независимости от промежуточной последовательности.

Таким образом, при введенных допущениях геномное расстояние будет выступать фактором модуляции экспрессии регулирующих рост генов, что в конечном итоге будет определять различия в фенотипе.

4.2. Уменьшение с возрастом экспрессии регулирующих рост генов млекопитающих

Было выдвинуто предположение, что уменьшение с возрастом экспрессии регулирующих рост генов вызвано эпигенетическими эффектами такими, как метилирование ДНК и модификация гистонов [2, 3, 48].

С одной стороны, эпигенетические модификации *промотора* могут вызывать замолкание гена. Продемонстрировано уменьшение с возрастом триметилирования гистона H3K4 в промоторных областях многих регулирующих рост генов [114]. Также была предложена гипотеза, что выключение этих генов связано с метилированием ДНК промоторной области гена, т.к. метилирование является хорошо известным механизмом замолкания генов, передающийся через цикл деления клеток. Однако, в частности, установлено, что уровень метилирования ДНК промоторных областей генов *Mest*, *Peg3* и *Plagl1* не изменяется с возрастом [48].

С другой стороны, эпигенетические модификации *дистальных регуляторных сайтов* могут быть связаны с замолканием гена. Показано, что степень метилирования ДНК некоторых дистальных регуляторных сайтов гена может оказывать более существенное влияние на уровень экспрессии гена, нежели степень метилирования промотора [115, 116].

Недавно было обнаружено, что средневзвешенный уровень метилирования ДНК 353 CpG сайтов в геноме человека (i) близок к нулю в эмбриональных клетках, (ii) имеет логарифмическую зависимость от возраста до периода полового созревания и линейную — для остальной жизни, причем (iii) эти закономерности справедливы и для тканей шимпанзе (т.н. эпигенетические часы Хорвата) [103]. Важно отметить, что метилирование одних из этих сайтов имело тенденцию к накоплению метилирования с возрастом, в то время как другая часть — наоборот, к потере метилирования.

Была показана связь между изменением уровня метилирования генома и функциональным изменением паттернов экспрессии генов с возрастом [117]. Была предложена гипотеза, что накопление с возрастом степени метилирования ДНК увеличивает вероятность выключения промотора или других регуляторных регионов таких, как энхансеры [118]. Для некоторых промоторов показано, что уровень их метилирования линейно увеличивается с возрастом [119].

Следует отметить, что эпигенетические модификации гистонов и метилирование ДНК могут являться взаимосвязанными явлениями [120, 121].

Таким образом, градуальное изменение с возрастом эпигенетического статуса регуляторных элементов гена может вызывать постепенное замолкание регулирующих рост генов млекопитающих.

4.3. Регуляторные механизмы взаимодействия на расстоянии

У эукариот показано три классических механизма взаимодействия на расстоянии регуляторных элементов генома: физическое сближение сайтов за счет

образования петель, скольжение вдоль хромосомы и растекание неактивного хроматина [122].

Следует отметить, что у дрожжей экспериментально установлено существование именно трех действующих на расстоянии различных механизмов модуляции экспрессии генов: два репрессирующих механизма — Sir-зависимый сайленсинг и транскрипционная интерференция, — и один активирующий — физическое объединение в пространстве множества регулируемых совместно генов, — причем было отмечено, что механизмы транскрипции у дрожжей сходны с таковыми у высших эукариот [123].

Хорошо известным механизмом взаимодействия на расстоянии является *энхансер-промоторное взаимодействие*, относящееся к активирующим типам взаимодействий и основанное на физическом сближении сайтов за счет образования петель [124, 125]. Показано, что сила влияния энхансера может зависеть от расстояния между энхансером и промотором [126, 127].

Другим примером взаимодействия на расстоянии является *механизм распространения гетерохроматина*, приводящий к инактивации генов [121, 122]. Сюда относятся *эффект теломерного замолкания* (telomere position effect, TPE) [128, 129, 130] и *эффект положения* (position effect variegation, PEV) [131].

Эффект распространения метилирования ДНК может также вызвать замолкание гена на расстоянии [132, 133]. Важно отметить, что механизм распространения метилирования ДНК у высших эукариот, по-видимому, схож с механизмом Sir-зависимого распространения неактивного хроматина у дрожжей [123].

4.4. Связь между морфо-физиологическими характеристиками млекопитающих и расстоянием между консервативными элементами в окрестностях регулирующих рост генов

Было обнаружено, что геномное расстояние между некоторыми консервативными элементами (КЭ) в окрестностях генов *Musp* и *Plagl1* значимо скоррелировано с массой и размером тела взрослого животного, а геномное расстояние между некоторыми КЭ в окрестности гена *Ezh2* — с продолжительностью жизни (раздел 3.2, таблица 3.2).

Следует отметить, что для каждого из этих генов наиболее скоррелированная пара КЭ содержала элемент, перекрывавшийся с *промотором* соответствующего гена, а зависимость носила *экспоненциальный* характер. Это позволяет предположить, что взаимодействие *промотора* и некоторого *дистального регуляторного сайта* может регулировать работу генов *Musp*, *Plagl1* и *Ezh2*, причем интенсивность влияния может зависеть как от эпигенетического состояния регуляторного сайта, так и от геномного расстояния от этого сайта до промотора.

Наличие как положительной, так и отрицательной корреляции может говорить о двух различных механизмах регуляции работы генов.

4.4.1. Случай отрицательной корреляции

Отрицательная корреляция была обнаружена между массой и размером тела взрослого животного и расстоянием между некоторыми КЭ в окрестности гена *Musp* и между продолжительностью жизни и расстоянием между некоторыми КЭ в окрестности гена *Ezh2* (раздел 3.2, таблица 3.2).

Близость регуляторного сайта и промотора у больших млекопитающих и отдаленность у маленьких указывает на наличие *активирующего* воздействия. Было отмечено, что масса взрослого животного падает практически экспоненциально при увеличении расстояния между КЭ. Таким образом, в основе данной регуля-

ции может лежать такое взаимодействие между регуляторным сайтом и промотором, при котором сила активирующего воздействия уменьшается при увеличении расстояния между элементами.

Одним из возможных типов взаимодействий может выступать энхансер-промоторное взаимодействие. Установлено, что сила воздействия энхансера действительно может падать экспоненциально при увеличении расстояния от энхансера до промотора [127]. Стоит отметить, что элемент MYCN(-6893), входящий в пару наиболее скоррелированных элементов в окрестности гена *Mycn* (пара элементов MYCN(-6893) и MYCN(74)), перекрывается с сайтом связывания фактора GATA2, который может являться компонентом транскрипционных энхансерных комплексов [134].

С другой стороны, показано, что степень метилирования ДНК энхансера может выступать основным фактором, влияющим на экспрессию гена, причем уровень экспрессии гена обратно-пропорционально связан со степенью метилирования энхансера [116]. В другом исследовании было установлено, что сила воздействия энхансера также падает с увеличением степени метилирования энхансера [135]. Важно отметить, что степень метилирования энхансера может гораздо сильнее отражаться на изменении уровня экспрессии гена, чем степень метилирования промотора [115].

Предполагается, что в основе эволюционного модулирования работы гена *Ezh2* могут лежать сходные механизмы, однако в данном случае зависимой морфофизиологической характеристикой будет выступать продолжительность жизни.

Таким образом, предлагается следующая регуляторная модель:

- дистальный регуляторный сайт является энхансером;
- сила воздействия энхансера экспоненциально падает при увеличении расстояния до промотора;
- сила воздействия энхансера линейно падает при увеличении степени метилирования энхансера;

- степень метилирования энхансера логарифмически увеличивается до периода полового созревания и затем замедляется до линейного роста в оставшейся жизни;
- в начале жизни энхансер не метилирован;
- скорость роста пропорциональна уровню экспрессии гена.

В результате, чем ближе будет находиться регуляторный элемент к промотору, тем дольше ген будет экспрессироваться, причем с возрастом экспрессия будет постепенно уменьшаться, а геномное расстояние между дистальным регуляторным элементом и промотором будет выступать фактором, оказывающим *модулирующее* воздействие на экспрессию гена у разных видов млекопитающих и таким образом определяющим фенотипические различия между видами.

4.4.2. Случай положительной корреляции

Положительная корреляция между массой и размером тела взрослого животного и расстоянием между КЭ была показана для одной пары КЭ в окрестности гена *Plagl1* (раздел 3.2, таблица 3.2).

Близость регуляторного сайта и промотора у маленьких млекопитающих и отдаленность у больших указывает на наличие *репрессивного воздействия*. Обнаружено, что масса взрослого животного медленно увеличивается при увеличении расстояния, а затем начинает стремительно возрастать при достижении некоторого порогового расстояния. Таким образом, в основе данной регуляции может лежать такое взаимодействие между регуляторным сайтом и промотором, при котором сила репрессивного воздействия уменьшается при увеличении расстояния между элементами.

Механизм *сайленсер-промоторного взаимодействия*, схожий с упомянутым выше энхансер-промоторным взаимодействием (раздел 4.4.1), по-видимому, не может лежать в основе данной регуляции, т.к. тогда репрессивное воздей-

ствие падало бы экспоненциально при увеличении расстояния и масса тела взрослого животного значительно бы увеличивалась при небольших изменениях расстояния между элементами.

Единственным наследуемым репрессирующим механизмом взаимодействия на расстоянии является *эффект распространения неактивного состояния хроматина*. В частности, сайленсинг может быть обусловлен *эффектом распространения метилирования ДНК*.

Показано, что уровень экспрессии гена может зависеть как от степени метилирования ДНК регуляторного сайта, так и от расстояния от этого регуляторного сайта — *центра метилирования* — до промотора, причем было выдвинуто предположение, что в основе этого взаимодействия лежит именно явление распространения метилирования ДНК [132, 133].

В другом исследовании было установлено, что растекание метилирования ДНК действует совместно с модификациями хроматина и может ингибировать близлежащий промотор, однако некоторое пороговое число модифицированных CpG сайтов требовалось для появления стабильной, способной к распространению хроматиновой структуры [120]. Следует подчеркнуть, что для выключения гена не обязательно требовалось метилирование ДНК промотора.

Известно, что степень эпигенетической модификации промотора отрицательно скоррелирована с уровнем экспрессии гена [136, 137, 138].

Таким образом, способность некоторого регуляторного сайта инактивировать близлежащий промотор посредством распространения метилирования ДНК может являться функцией как числа модифицированных CpG, так и расстояния между промотором и регуляторным сайтом.

Следует отметить, что ген *Plagl1* является импринтированным, а эффект распространения метилирования ДНК может участвовать в регуляции импринтированных генов [139]. Установлено, что промоторы некоторых импринтированных генов могут быть чувствительны к степени метилирования соседних участков ДНК [140, 141].

Показано, что наличие в окрестности гена участка контроля импринтинга (imprinting control region, ICR) может вызывать сайленсинг гена, причем указана роль близости ICR до целевого промотора в степени проявления репрессирующего эффекта [142]. Было отмечено, что для выключения гена не обязательно требовалось метилирование ДНК промотора, однако, было высказано предположение, что наиболее вероятной причиной сайленсинга является эффект распространения метилирования ДНК.

Была выдвинута гипотеза, что скоординированное с возрастом уменьшение экспрессии импринтированного гена *Plagl1* вызывается изменением метилирования ДНК промотора и последующим выключением аллеля, однако было продемонстрировано, что степень метилирования промотора гена *Plagl1* не меняется с возрастом [48]. Учитывая вышесказанное, опосредованное эффектом распространения метилирования ДНК выключение гена *Plagl1* могло бы не требовать метилирования ДНК промотора.

Таким образом, предлагается следующая регуляторная модель:

- дистальный регуляторный сайт является центром метилирования;
- степень метилирования соседних с ним участков медленно уменьшается с увеличением расстояния и затем значительно падает при достижении некоторого порогового значения расстояния;
- экспрессия гена линейно уменьшается с увеличением степени метилирования промотора гена;
- степень метилирования регуляторного сайта логарифмически увеличивается до периода полового созревания и затем замедляется до линейного роста в оставшейся жизни;
- в начале жизни регуляторный сайт не метилирован;
- скорость роста пропорциональна уровню экспрессии гена.

В результате, чем ближе будет находиться регуляторный элемент к промотору, тем дальше будет находиться регуляторный элемент от промотора, тем дольше ген будет экспрессироваться, причем с возрастом экспрессия будет постепенно уменьшаться, а геномное расстояние между дистальным регуляторным элементом и промотором будет выступать фактором, оказывающим *модулирующее* воздействие на экспрессию гена у разных видов млекопитающих и таким образом определяющим фенотипические различия между видами.

4.5. Связь между морфо-физиологическими характеристиками млекопитающих и положением на хромосоме регулирующих рост генов

Значимая корреляция между морфо-физиологическими характеристиками млекопитающих и расстоянием от начала гена до ближайшей теломеры была показана для двух регулирующих рост генов *Ghrh* и *Sst* и для двух генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD (раздел 3.5, таблица 3.24).

Было обнаружено, что гены *Ghrh* и *Sst* демонстрируют *противоположную* корреляцию между относительным расстоянием от начала гена до ближайшей теломеры и возрастом полового созревания, причем оба этих гена принадлежат к семейству генов соматотропной оси и оказывают *антагонистическое* воздействие на выработку гормона роста GH1. Показано, что уменьшение с возрастом релизинга GH1 связано с уменьшением релизинга GHRH и увеличению релизинга SST из гипоталамуса [143, 144]. Это позволяет предположить, что вариация между различными видами расстояния от генов *Ghrh* и *Sst* до ближайшей теломеры может выступать фактором, влияющим на регуляцию данных генов, что в конечном итоге определяет возраст полового созревания животного.

Для гена *Ghrh* было показано, что возраст полового созревания медленно увеличивается с увеличением расстояния от начала этого гена до ближайшей те-

ломеры и затем существенно начинает возрастать при достижении некоторого порогового значения расстояния. Для гена *Sst* наблюдается обратная ситуация, причем зависимость более грубая: возраст полового созревания был максимальным, если ген *Sst* находился около теломеры, а затем значительно уменьшается при увеличении расстояния до теломеры, достигая минимального значения, когда ген *Sst* находился в наиболее отдаленной от обеих теломер точке на хромосоме. Таким образом, в основе регуляции данных генов может лежать зависимый от расстояния механизм замолкания генов, основанный на взаимодействии между геном и теломерой.

Одним из возможных механизмов, объясняющих найденные закономерности, является *теломерный эффект положения* (telomere position effect, TPE), состоящий в инактивации генов вблизи теломер [128, 129, 130]. Это эффект вызван распространением неактивного хроматина от теломер. Однако, типовые расстояния действия этого механизма (~ 100 тыс. н.п.) существенно меньше наблюдаемых расстояний между генами и ближайшими теломерами у рассмотренных видов млекопитающих (~ 10 млн н.п.).

Недавние исследования показали, что теломера может регулировать гены на большом расстоянии. Этот эффект, названный *теломерным эффектом положения на длинных расстояниях* (telomere position effect over long distances, TPE-OLD), состоит в том, что теломера находится около гена, когда она длинная и в разделении этих локусов, когда теломера становится достаточно короткой, причем присутствие теломеры около гена подавляет экспрессию гена [9, 10]. Показано, что этот эффект может воздействовать на гены, находящиеся на расстоянии нескольких миллионов н.п. от теломеры.

Было выявлено, что расстояние от начала гена до ближайшей теломеры для двух регулируемых механизмом TPE-OLD генов *C1s* и *Notch1* значимо скоррелировано с возрастом полового созревания и продолжительностью жизни, причем в обоих случаях характер зависимости был сходен с таковым у гена *Sst*. Это может указывать на то, что расстояние до теломеры *модулирует* силу воздействия

эффекта TPE-OLD, причем модулирующим фактором может выступать как абсолютное, так и относительное расстояние от гена до теломеры [11].

Это позволяет предположить, что чем ближе ген находится к теломере, тем дольше он остается в молчащем состоянии, попадая в перmissive для экспрессии условия при достаточном укорочении теломеры, а удаленность от теломеры *градуально* уменьшает репрессирующее воздействие теломеры.

Интересно отметить, что у мышей обработка антагонистом GHRH ведет к увеличению активности гена теломеразы *Tert*, для которого также показана регуляция посредством TPE-OLD [145].

Ранее было показано (i) сокращение длины теломеры по мере роста животного, (ii) антагонистический характер воздействия продуктов генов *Ghrh* и *Sst* на выработку гормона роста GH1, (iii) уменьшение с возрастом уровня GHRH и увеличение уровня SST, и (iv) связь между геном *Ghrh* и активностью гена *Tert* теломеразы. *В данной работе* была выявлена (v) противоположенная по знаку корреляция между расстоянием от гена до теломеры и периодом полового созревания для генов *Ghrh* и *Sst*, и (vi) корреляция между расстоянием от гена до теломеры и периодом полового созревания для регулируемых механизмом TPE-OLD генов *C1s* и *Notch1*. *Учитывая вышесказанное*, можно предположить участие механизма TPE-OLD в регуляции генов *Ghrh* и *Sst*, причем сила воздействия эффекта TPE-OLD постепенно уменьшается с увеличением расстояния от теломеры до гена. Последнее обстоятельство может являться фактором, оказывающим *модулирующее* воздействие на экспрессию данных генов у разных видов млекопитающих.

Заключение

Были проанализированы окрестности 17 регулирующих рост генов *Ghl*, *Ghrh*, *Ghrl*, *Igf1*, *Sst*, *Igfbp3*, *Igfbp1*, *E2f3*, *Igf2*, *Ezh2*, *Gpc3*, *Mest*, *Peg3*, *Plagl1*, *Smo*, *Mdk* и *Mycn* у 36 видов млекопитающих.

Было обнаружено, что геномное расстояние между некоторыми консервативными элементами (КЭ) в окрестностях генов *Mycn* и *Plagl1* значимо скоррелировано с массой и размером тела взрослого животного, а геномное расстояние между некоторыми КЭ в окрестности гена *Ezh2* — с продолжительностью жизни. Для каждого из этих генов пара наиболее скоррелированных КЭ содержала промотор соответствующего гена, а зависимости морфо-физиологических характеристик от расстояния между КЭ как в случае отрицательной, так и в случае положительной корреляции имели экспоненциальный характер.

Различие в морфо-физиологических характеристиках между видами может быть обусловлено различием в модулировании экспрессии регулирующих рост генов. Впервые выдвинуто предположение, что *геномное расстояние* между дистальным регуляторным сайтом и промотором гена может выступать основным фактором эволюционного модулирования экспрессии регулирующих рост генов, что в конечном итоге определяет фенотип. Для обоих случаев отрицательной и положительной корреляции расстояния между КЭ и морфо-физиологическими характеристиками предложены соответствующие модели регуляции экспрессии генов.

У 21 вида млекопитающих, для которых была доступна полная версия генома, была проанализирована связь между расстоянием от гена до ближайшей теломеры и морфо-физиологическими характеристиками для 17 регулирующих рост генов. Было обнаружено, что геномное расстояние от начала гена до ближайшей теломеры для генов соматотропной оси *Ghrh* и *Sst*, а также для двух генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD, значимо скоррелировано возрастом полового созревания и продолжительностью жизни. Было выдвинуто предположе-

ние, гены *Ghrh* и *Sst* регулируются механизмом TPE-OLD, причем расстояние от этих генов до ближайшей теломеры выступает фактором модуляции экспрессии генов, выступая одновременно фактором эволюционного модулирования экспрессии этих генов и внося вклад в различие морфо-физиологических характеристик между видами.

Перспективой исследования является возможность влиять на массу и размер тела млекопитающих простым изменением расстояния между некоторыми консервативными элементами в окрестностях регулирующих рост генов *Musp* и *Plagl1*. В то же время, наличие связи между продолжительностью жизни и расстоянием между некоторыми консервативными элементами в окрестности гена *Ezh2* открывает перспективу влиять на продолжительность жизни, аналогично изменяя геномное расстояние между этими элементами.

Были определены наиболее вероятные геномные элементы, которые могут являться мишенями геномных манипуляций для подтверждения выдвинутых гипотез. Указана возможность изменять возраст полового созревания и продолжительность жизни посредством расположения на разном расстоянии до ближайшей теломеры генов соматотропной оси *Ghrh* и *Sst*.

Методика, примененная в данном исследовании, может быть использована для поиска новых соотношений генотипа и фенотипа. Предполагается, что если модуляция экспрессии генов ведет к согласованному изменению фенотипа, то и изменение фенотипа может указывать на тип модуляции экспрессии генов.

Наличие значимой корреляции между фенотипом и расстоянием между КЭ может также выступать дополнительным подтверждением для предсказанных регуляторных сайтов.

Выводы

1. Разработан набор bash-скриптов *efunctions*, позволяющий автоматизировать стандартный набор операций при взаимодействии с базами данных NCBI, и с помощью него получены геномные последовательности окрестностей регулирующих рост генов у различных видов млекопитающих.
2. С помощью разработанной программы *dotolog* для построения и анализа точечной матрицы гомологий выявлены консервативные элементы генома в окрестности 17 регулирующих рост генов у 36 видов млекопитающих.
3. С помощью разработанного программного конвейера *mblast* осуществлен поиск известных элементов генома в окрестностях этих генов.
4. Обнаружена значимая корреляция между массой, размером тела взрослого животного и геномным расстоянием между консервативными элементами в окрестностях регулирующих рост генов *Mucn* ($r_s = -0.81$, $p = 0.0007$) и *Plagl1* ($r_s = 0.73$, $p = 0.03$) и между продолжительностью жизни и геномным расстоянием между консервативными элементами в окрестности гена *Ezh2* ($r_s = -0.79$, $p = 0.02$).

Для каждого из этих трех генов наиболее скоррелированная пара консервативных элементов содержала *промотор* соответствующего гена, а зависимость носила *экспоненциальный* характер.

5. Произведен полногеномный поиск в геноме человека *гомологов* консервативных элементов, демонстрировавших максимальную корреляцию с морфо-физиологическими характеристиками, и получен список генов, находящихся в окрестности этих гомологов.

Анализ сверхпредставленности категорий Gene Ontology по этим генам выявил гены *ACVR1B* и *ACVRL1*, значимо представленные в категориях

growth factor binding, GO:0019838 ($p = 0.0988$) и *response to growth factor*, GO:0070848 ($p = 0.0461$).

6. Обнаружена значимая корреляция между периодом полового созревания и геномным расстоянием от гена до ближайшей теломеры для двух генов соматотропной оси *Ghrh* ($r_s = 0.76$, $p = 0.01$) и *Sst* ($r_s = -0.72$, $p = 0.04$) и для двух генов *Cls* ($r_s = -0.84$, $p = 0.0004$) и *Notch1* ($r_s = -0.81$, $p = 0.003$), регулируемых механизмом TPE-OLD.

Список литературы

1. An extensive genetic program occurring during postnatal growth in multiple tissues / G. P. Finkelstein, P. Forcinito, J. C. Lui et al. // *Endocrinology*. — 2009. — Apr. — Vol. 150, no. 4. — P. 1791–1800.
2. Coordinated postnatal down-regulation of multiple growth-promoting genes: evidence for a genetic program limiting organ growth / J. C. Lui, P. Forcinito, M. Chang et al. // *FASEB J.* — 2010. — Aug. — Vol. 24, no. 8. — P. 3083–3092.
3. Lui J. C., Baron J. Mechanisms limiting body growth in mammals // *Endocr. Rev.* — 2011. — Jun. — Vol. 32, no. 3. — P. 422–440.
4. Evolutionary conservation and modulation of a juvenile growth-regulating genetic program / A. Delaney, V. Padmanabhan, G. Rezvani et al. // *J. Mol. Endocrinol.* — 2014. — Jun. — Vol. 52, no. 3. — P. 269–277.
5. Lui J. C., Garrison P., Baron J. Regulation of body growth // *Curr. Opin. Pediatr.* — 2015. — Aug. — Vol. 27, no. 4. — P. 502–510.
6. Prevalence of miRNAs in Introns and Cis-Regulatory Regions of Genes of the Somatotropic Axis in Mammals / T. Shkurat, D. Romanov, E. Pshenichnyy et al. // *American Journal of Applied Sciences*. — 2015. — Vol. 12. — P. 1–7.
7. Identification of conserved regulatory elements in mammalian promoter regions: a case study using the PCK1 promoter / G. E. Liu, M. T. Weirauch, C. P. Van Tassel et al. // *Genomics Proteomics Bioinformatics*. — 2008. — Dec. — Vol. 6, no. 3-4. — P. 129–143.
8. Genome distance between conserved elements in neighborhoods of growth-regulating genes is correlated with morpho-physiological traits in mammals / D. E. Romanov, E. V. Butenko, G. B. Bakhtadze, T. P. Shkurat // *Gene Reports*. — 2019. — Vol. 17.
9. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances / J. D. Robin, A. T. Ludlow, K. Batten et al. // *Genes Dev.* — 2014. — Nov. — Vol. 28, no. 22. — P. 2464–2476.

10. Misteli T. The long reach of telomeres // *Genes and Development*. — 2014. — Nov. — Vol. 28, no. 22. — P. 2445–2446.
11. Romanov D. E., Butenko E. V., Shkurat T. P. Genome distance between growth-regulating genes and telomeres is correlated with morpho-physiological traits in mammals // *Gene Reports*. — 2019. — Vol. 14. — P. 124–128.
12. Penzo-Mendez A. I., Stanger B. Z. Organ-Size Regulation in Mammals // *Cold Spring Harb Perspect Biol*. — 2015. — Jul. — Vol. 7, no. 9. — P. a019240.
13. Jorgensen P., Tyers M. How cells coordinate growth and division // *Curr. Biol*. — 2004. — Dec. — Vol. 14, no. 23. — P. R1014–1027.
14. Winick M., Noble A. Quantitative changes in DNA, RNA, and protein during prenatal and postnatal growth in the rat // *Dev. Biol*. — 1965. — Dec. — Vol. 12, no. 3. — P. 451–466.
15. Changes in cell-cycle kinetics responsible for limiting somatic growth in mice / M. Chang, E. A. Parker, T. J. Muller et al. // *Pediatr. Res*. — 2008. — Sep. — Vol. 64, no. 3. — P. 240–245.
16. Growth fraction and cycle duration of hepatocytes in the three-week-old rat / B. Schultze, A. M. Kellerer, C. Grossmann, W. Maurer // *Cell Tissue Kinet*. — 1978. — May. — Vol. 11, no. 3. — P. 241–249.
17. Post J., Hoffman J. Changes in the replication times and patterns of the liver cell during the life of the rat // *Exp. Cell Res*. — 1964. — Oct. — Vol. 36. — P. 111–123.
18. Depletion of resting zone chondrocytes during growth plate senescence / L. Schrier, S. P. Ferns, K. M. Barnes et al. // *J. Endocrinol*. — 2006. — Apr. — Vol. 189, no. 1. — P. 27–36.
19. Widdowson E. M. Harmony of growth // *Lancet*. — 1970. — May. — Vol. 1, no. 7653. — P. 902–905.
20. Winick M., Noble A. Cellular response in rats during malnutrition at various ages // *J. Nutr*. — 1966. — Jul. — Vol. 89, no. 3. — P. 300–306.
21. Bogin B. Evolutionary perspective on human growth // *Annu Rev Anthropol*. — 1999. — Vol. 28. — P. 109–153.

22. Finkelstein G. P., Lui J. C., Baron J. Catch-up growth: cellular and molecular mechanisms // *World Rev Nutr Diet.* — 2013. — Vol. 106. — P. 100–104.
23. Roles of growth hormone and insulin-like growth factor 1 in mouse postnatal growth / F. Lupu, J. D. Terwilliger, K. Lee et al. // *Dev. Biol.* — 2001. — Jan. — Vol. 229, no. 1. — P. 141–162.
24. Igf1 gene disruption results in reduced brain size, CNS hypomyelination, and loss of hippocampal granule and striatal parvalbumin-containing neurons / K. D. Beck, L. Powell-Braxton, H. R. Widmer et al. // *Neuron.* — 1995. — Apr. — Vol. 14, no. 4. — P. 717–730.
25. Calikoglu A. S., Gutierrez-Ospina G., D'Ercole A. J. Congenital hypothyroidism delays the formation and retards the growth of the mouse primary somatic sensory cortex (S1) // *Neurosci. Lett.* — 1996. — Aug. — Vol. 213, no. 2. — P. 132–136.
26. Sankaran S., Kyle P. M. Aetiology and pathogenesis of IUGR // *Best Pract Res Clin Obstet Gynaecol.* — 2009. — Dec. — Vol. 23, no. 6. — P. 765–777.
27. Nijhout H. F. The control of body size in insects // *Dev. Biol.* — 2003. — Sep. — Vol. 261, no. 1. — P. 1–9.
28. Growth enhancement of transgenic mice expressing human insulin-like growth factor I / L. S. Mathews, R. E. Hammer, R. R. Behringer et al. // *Endocrinology.* — 1988. — Dec. — Vol. 123, no. 6. — P. 2827–2833.
29. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene / K. A. Woods, C. Camacho-Hubner, M. O. Savage, A. J. Clark // *N. Engl. J. Med.* — 1996. — Oct. — Vol. 335, no. 18. — P. 1363–1367.
30. Gluckman P. D., Grumbach M. M., Kaplan S. L. The neuroendocrine regulation and function of growth hormone and prolactin in the mammalian fetus // *Endocr. Rev.* — 1981. — Vol. 2, no. 4. — P. 363–395.
31. A single IGF1 allele is a major determinant of small size in dogs / N. B. Sutter, C. D. Bustamante, K. Chase et al. // *Science.* — 2007. — Apr. — Vol. 316, no. 5821. — P. 112–115.

32. Greer K. A., Hughes L. M., Masternak M. M. Connecting serum IGF-1, body size, and age in the domestic dog // *Age (Dordr)*. — 2011. — Sep. — Vol. 33, no. 3. — P. 475–483.
33. Laron Z., Lilos P., Klinger B. Growth curves for Laron syndrome // *Arch. Dis. Child*. — 1993. — Jun. — Vol. 68, no. 6. — P. 768–770.
34. Growth inhibition in giant growth hormone transgenic mice by overexpression of insulin-like growth factor-binding protein-2 / A. Hoeflich, S. Nedbal, W. F. Blum et al. // *Endocrinology*. — 2001. — May. — Vol. 142, no. 5. — P. 1889–1898.
35. Growth factors and intrauterine growth retardation. II. Serum growth hormone, insulin-like growth factor (IGF) I, and IGF-binding protein 3 levels in children with intrauterine growth retardation compared with normal control subjects: prospective study from birth to two years of age. Study Group of IUGR / J. Leger, M. Noel, J. M. Limal, P. Czernichow // *Pediatr. Res*. — 1996. — Jul. — Vol. 40, no. 1. — P. 101–107.
36. Serum free insulin-like growth factor I (IGF-I), total IGF-I, and IGF-binding protein-3 concentrations in normal children and children with growth hormone deficiency / N. Kawai, S. Kanzaki, S. Takano-Watou et al. // *J. Clin. Endocrinol. Metab*. — 1999. — Jan. — Vol. 84, no. 1. — P. 82–89.
37. Zapf J., Walter H., Froesch E. R. Radioimmunological determination of insulin-like growth factors I and II in normal subjects and in patients with growth disorders and extrapancreatic tumor hypoglycemia // *J. Clin. Invest*. — 1981. — Nov. — Vol. 68, no. 5. — P. 1321–1330.
38. Disruption of insulin-like growth factor-I expression in type II α 1 collagen-expressing cells reduces bone length and width in mice / K. E. Govoni, S. K. Lee, Y. S. Chung et al. // *Physiol. Genomics*. — 2007. — Aug. — Vol. 30, no. 3. — P. 354–362.
39. Maternal and child undernutrition: consequences for adult health and human capital / C. G. Victora, L. Adair, C. Fall et al. // *Lancet*. — 2008. — Jan. — Vol. 371, no. 9609. — P. 340–357.

40. Drewry M. M., Harris R. B., Martin R. J. Developmental changes in response to overfeeding: effect on composition of gain, liver metabolism and adipocyte cellularity in rats // *J. Nutr.* — 1988. — Feb. — Vol. 118, no. 2. — P. 194–198.
41. Pekas J. C. Animal growth during liberation from appetite suppression // *Growth.* — 1985. — Vol. 49, no. 1. — P. 19–27.
42. Influence of overfeeding on growth, obesity and intestinal tract in young chicks of light and heavy breeds / I. Nir, Z. Nitsan, Y. Dror, N. Shapira // *Br. J. Nutr.* — 1978. — Jan. — Vol. 39, no. 1. — P. 27–35.
43. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man / E. P. Smith, J. Boyd, G. R. Frank et al. // *N. Engl. J. Med.* — 1994. — Oct. — Vol. 331, no. 16. — P. 1056–1061.
44. Growth plate senescence is associated with loss of DNA methylation / O. Nilsson, R. D. Mitchum, L. Schrier et al. // *J. Endocrinol.* — 2005. — Jul. — Vol. 186, no. 1. — P. 241–249.
45. Fausto N., Campbell J. S., Riehle K. J. Liver regeneration // *J. Hepatol.* — 2012. — Sep. — Vol. 57, no. 3. — P. 692–694.
46. c-Myc regulates mammalian body size by controlling cell number but not cell size / A. Trumpp, Y. Refaeli, T. Oskarsson et al. // *Nature.* — 2001. — Dec. — Vol. 414, no. 6865. — P. 768–773.
47. Organization of the Indian hedgehog–parathyroid hormone-related protein system in the postnatal growth plate / M. Chau, P. Forcinito, A. C. Andrade et al. // *J. Mol. Endocrinol.* — 2011. — Aug. — Vol. 47, no. 1. — P. 99–107.
48. An imprinted gene network that controls mammalian somatic growth is down-regulated during postnatal growth deceleration in multiple organs / J. C. Lui, G. P. Finkelstein, K. M. Barnes, J. Baron // *Am. J. Physiol. Regul. Integr. Comp. Physiol.* — 2008. — Jul. — Vol. 295, no. 1. — P. R189–196.
49. Changes in gene expression associated with aging commonly originate during juvenile growth / J. C. Lui, W. Chen, K. M. Barnes, J. Baron // *Mech. Ageing Dev.* — 2010. — Oct. — Vol. 131, no. 10. — P. 641–649.

50. Evidence That Up-Regulation of MicroRNA-29 Contributes to Postnatal Body Growth Deceleration / F. Kamran, A. C. Andrade, A. A. Nella et al. // *Mol. Endocrinol.* — 2015. — Jun. — Vol. 29, no. 6. — P. 921–932.
51. Lui J. C. Regulation of body growth by microRNAs // *Mol. Cell. Endocrinol.* — 2016. — Oct.
52. Stanger B. Z., Tanaka A. J., Melton D. A. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver // *Nature.* — 2007. — Feb. — Vol. 445, no. 7130. — P. 886–891.
53. Olovnikov A. M. Telomeres, telomerase, and aging: origin of the theory // *Exp. Gerontol.* — 1996. — Vol. 31, no. 4. — P. 443–448.
54. Garcia C. K., Wright W. E., Shay J. W. Human diseases of telomerase dysfunction: insights into tissue aging // *Nucleic Acids Res.* — 2007. — Vol. 35, no. 22. — P. 7406–7416.
55. Harley C. B., Futcher A. B., Greider C. W. Telomeres shorten during ageing of human fibroblasts // *Nature.* — 1990. — May. — Vol. 345, no. 6274. — P. 458–460.
56. In vivo loss of telomeric repeats with age in humans / J. Lindsey, N. I. McGill, L. A. Lindsey et al. // *Mutat. Res.* — 1991. — Jan. — Vol. 256, no. 1. — P. 45–48.
57. Kipling D. Telomeres, replicative senescence and human ageing // *Maturitas.* — 2001. — Feb. — Vol. 38, no. 1. — P. 25–37.
58. Shay J. W. Role of Telomeres and Telomerase in Aging and Cancer // *Cancer Discov.* — 2016. — 06. — Vol. 6, no. 6. — P. 584–593.
59. Regulation of the Human Telomerase Gene TERT by Telomere Position Effect-Over Long Distances (TPE-OLD): Implications for Aging and Cancer / W. Kim, A. T. Ludlow, J. Min et al. // *PLoS Biol.* — 2016. — Dec. — Vol. 14, no. 12. — P. e2000016.
60. Venkatesan S., Khaw A. K., Hande M. P. Telomere Biology-Insights into an Intriguing Phenomenon // *Cells.* — 2017. — Jun. — Vol. 6, no. 2. — P. 1–17.
61. Long telomeres protect against age-dependent cardiac disease caused by NOTCH1 haploinsufficiency / C. V. Theodoris, F. Mourkioti, Y. Huang et al. // *J.*

- Clin. Invest. — 2017. — May. — Vol. 127, no. 5. — P. 1683–1688.
62. SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy / J. D. Robin, A. T. Ludlow, K. Batten et al. // *Genome Res.* — 2015. — Dec. — Vol. 25, no. 12. — P. 1781–1790.
 63. Fondon J. W., Garner H. R. Molecular origins of rapid and continuous morphological evolution // *Proc. Natl. Acad. Sci. U.S.A.* — 2004. — Dec. — Vol. 101, no. 52. — P. 18058–18063.
 64. Wang W., Kirkness E. F. Short interspersed elements (SINEs) are a major source of canine genomic diversity // *Genome Res.* — 2005. — Dec. — Vol. 15, no. 12. — P. 1798–1808.
 65. Richard G. F., Kerrest A., Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes // *Microbiol. Mol. Biol. Rev.* — 2008. — Dec. — Vol. 72, no. 4. — P. 686–727.
 66. A long AAAG repeat allele in the 5' UTR of the ERR- γ gene is correlated with breast cancer predisposition and drives promoter activity in MCF-7 breast cancer cells / C. L. Galindo, J. F. McCormick, V. J. Bubb et al. // *Breast Cancer Res. Treat.* — 2011. — Nov. — Vol. 130, no. 1. — P. 41–48.
 67. RUNX2 tandem repeats and the evolution of facial length in placental mammals / M. A. Pointer, J. M. Kamilar, V. Warmuth et al. // *BMC Evol. Biol.* — 2012. — Jun. — Vol. 12. — P. 103.
 68. Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia / E. Soragni, D. Herman, S. Y. Dent et al. // *Nucleic Acids Res.* — 2008. — Nov. — Vol. 36, no. 19. — P. 6056–6065.
 69. Newton A. H., Feigin C. Y., Pask A. J. RUNX2 repeat variation does not drive craniofacial diversity in marsupials // *BMC Evol. Biol.* — 2017. — 05. — Vol. 17, no. 1. — P. 110.
 70. The correlated evolution of Runx2 tandem repeats, transcriptional activity, and fa

- cial length in carnivora / K. E. Sears, A. Goswami, J. J. Flynn, L. A. Niswander // *Evol. Dev.* — 2007. — Vol. 9, no. 6. — P. 555–565.
71. Carroll S. B. Evolution at two levels: on genes and form // *PLoS Biol.* — 2005. — Jul. — Vol. 3, no. 7. — P. e245.
72. Highly conserved non-coding sequences are associated with vertebrate development / A. Woolfe, M. Goodson, D. K. Goode et al. // *PLoS Biol.* — 2005. — Jan. — Vol. 3, no. 1. — P. e7.
73. Initial sequencing and analysis of the human genome / E. S. Lander, L. M. Linton, B. Birren et al. // *Nature.* — 2001. — Feb. — Vol. 409, no. 6822. — P. 860–921.
74. Bulger M., Groudine M. Looping versus linking: toward a model for long-distance gene activation // *Genes Dev.* — 1999. — Oct. — Vol. 13, no. 19. — P. 2465–2477.
75. Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes / D. L. Gumucio, H. Heilstedt-Williamson, T. A. Gray et al. // *Mol. Cell. Biol.* — 1992. — Nov. — Vol. 12, no. 11. — P. 4919–4929.
76. Brassat E., Vaury C. Insulators are fundamental components of the eukaryotic genomes // *Heredity (Edinb).* — 2005. — Jun. — Vol. 94, no. 6. — P. 571–576.
77. Pennacchio L. A., Rubin E. M. Genomic strategies to identify mammalian regulatory sequences // *Nat. Rev. Genet.* — 2001. — Feb. — Vol. 2, no. 2. — P. 100–109.
78. Evolution of the autosomal chorion locus in *Drosophila*. I. General organization of the locus and sequence comparisons of genes *s15* and *s19* in evolutionary distant species / J. C. Martinez-Cruzado, C. Swimmer, M. G. Fenerjian, F. C. Kafatos // *Genetics.* — 1988. — Jul. — Vol. 119, no. 3. — P. 663–677.
79. Comparative genomics allows the discovery of cis-regulatory elements in mosquitoes / D. H. Sieglaff, W. A. Dunn, X. S. Xie et al. // *Proc. Natl. Acad. Sci. U.S.A.* — 2009. — Mar. — Vol. 106, no. 9. — P. 3053–3058.
80. Analysis of vertebrate SCL loci identifies conserved enhancers / B. Gottgens, L. M. Barton, J. G. Gilbert et al. // *Nat. Biotechnol.* — 2000. — Feb. — Vol. 18,

- no. 2. — P. 181–186.
81. Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional / K. A. Frazer, H. Tao, K. Osoegawa et al. // *Genome Res.* — 2004. — Mar. — Vol. 14, no. 3. — P. 367–372.
 82. Tautz D. Evolution of transcriptional regulation // *Curr. Opin. Genet. Dev.* — 2000. — Oct. — Vol. 10, no. 5. — P. 575–579.
 83. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons / G. G. Loots, R. M. Locksley, C. M. Blankespoor et al. // *Science.* — 2000. — Apr. — Vol. 288, no. 5463. — P. 136–140.
 84. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs) / E. T. Dermitzakis, A. Reymond, N. Scamuffa et al. // *Science.* — 2003. — Nov. — Vol. 302, no. 5647. — P. 1033–1035.
 85. Dermitzakis E. T., Reymond A., Antonarakis S. E. Conserved non-genic sequences - an unexpected feature of mammalian genomes // *Nat. Rev. Genet.* — 2005. — Feb. — Vol. 6, no. 2. — P. 151–157.
 86. Margulies E. H., Chen C. W., Green E. D. Differences between pair-wise and multi-sequence alignment methods affect vertebrate genome comparisons // *Trends Genet.* — 2006. — Apr. — Vol. 22, no. 4. — P. 187–193.
 87. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes / A. Siepel, G. Bejerano, J. S. Pedersen et al. // *Genome Res.* — 2005. — Aug. — Vol. 15, no. 8. — P. 1034–1050.
 88. Lunter G., Ponting C. P., Hein J. Genome-wide identification of human functional DNA using a neutral indel model // *PLoS Comput. Biol.* — 2006. — Jan. — Vol. 2, no. 1. — P. e5.
 89. Megabase deletions of gene deserts result in viable mice / M. A. Nobrega, Y. Zhu, I. Plajzer-Frick et al. // *Nature.* — 2004. — Oct. — Vol. 431, no. 7011. — P. 988–993.
 90. Ultraconserved elements in the human genome / G. Bejerano, M. Pheasant, I. Makunin et al. // *Science.* — 2004. — May. — Vol. 304, no. 5675. —

- P. 1321–1325.
91. Gibbs A. J., McIntyre G. A. The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences // *Eur. J. Biochem.* — 1970. — Sep. — Vol. 16, no. 1. — P. 1–11.
 92. Seibt K. M., Schmidt T., Heitkam T. FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses // *Bioinformatics.* — 2018. — Oct. — Vol. 34, no. 20. — P. 3575–3577.
 93. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals / Kate E. Jones, Jon Bielby, Marcel Cardillo et al. // *Ecology.* — 2009. — Vol. 90. — P. 2648.
 94. Kans J. Entrez Direct: E-utilities on the UNIX Command Line. — 2013. — Apr.
 95. Свидетельство о государственной регистрации программы для ЭВМ №2016663454. Dotolog: Программа для автоматизации визуального анализа дотплот-изображений нуклеотидных последовательностей ДНК / Д. Е. Романов, Т. П. Шкурат – Заявка №2016661011. Дата поступления 18 октября 2016 г. Зарегистрировано в Реестре программ для ЭВМ 07 декабря 2016 г.
 96. Defining the role of common variation in the genomic and biological architecture of adult human height / A. R. Wood, T. Esko, J. Yang et al. // *Nat. Genet.* — 2014. — Nov. — Vol. 46, no. 11. — P. 1173–1186.
 97. Quinlan A. R., Hall I. M. BEDTools: a flexible suite of utilities for comparing genomic features // *Bioinformatics.* — 2010. — Mar. — Vol. 26, no. 6. — P. 841–842.
 98. BioJava: an open-source framework for bioinformatics in 2012 / A. Prlic, A. Yates, S. E. Bliven et al. // *Bioinformatics.* — 2012. — Oct. — Vol. 28, no. 20. — P. 2693–2695.
 99. Свидетельство о государственной регистрации программы для ЭВМ №2016663454. Mscanner: Программа для автоматического поиска мотивов в последовательности ДНК / Д. Е. Романов, Н. С. Ксёиз – Заявка №2016661028. Дата поступления 18 октября 2016 г. Зарегистрировано в Реестре программ для ЭВМ 25 ноября 2016 г.

100. Shkurat T. P., Romanov D. E., Shkurat M. A. Alu elements as source of miRNA sites in the human genome // Abstracts from the 51st European Society of Human Genetics Conference: Electronic Posters. — Vol. 27. — 2019. — P. 870–1041.
101. Basic local alignment search tool / S. F. Altschul, W. Gish, W. Miller et al. // J. Mol. Biol. — 1990. — Oct. — Vol. 215, no. 3. — P. 403–410.
102. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools / H. Mi, A. Muruganujan, D. Ebert et al. // Nucleic Acids Res. — 2019. — Jan. — Vol. 47, no. D1. — P. D419–D426.
103. Horvath S. DNA methylation age of human tissues and cell types // Genome Biol. — 2013. — Vol. 14, no. 10. — P. R115.
104. Low-level GATA2 overexpression promotes myeloid progenitor self-renewal and blocks lymphoid differentiation in mice / S. K. Nandakumar, K. Johnson, S. L. Throm et al. // Exp. Hematol. — 2015. — Jul. — Vol. 43, no. 7. — P. 565–577.
105. Unveiling MYCN regulatory networks in neuroblastoma via integrative analysis of heterogeneous genomics data / C. L. Hsu, H. Y. Chang, J. Y. Chang et al. // Oncotarget. — 2016. — Jun. — Vol. 7, no. 24. — P. 36293–36310.
106. Activation of EZH2 and SUZ12 Regulated by E2F1 Predicts the Disease Progression and Aggressive Characteristics of Bladder Cancer / S. R. Lee, Y. G. Roh, S. K. Kim et al. // Clin. Cancer Res. — 2015. — Dec. — Vol. 21, no. 23. — P. 5391–5403.
107. Shaw-Smith C., Willatt L., Thalange N. Growth deficiency in oculodigitoesophagoduodenal (Feingold) syndrome—case report and review of the literature // Clin. Dysmorphol. — 2005. — Jul. — Vol. 14, no. 3. — P. 155–158.
108. Human growth is associated with distinct patterns of gene expression in evolutionarily conserved networks / A. Stevens, D. Hanson, A. Whatmore et al. // BMC Genomics. — 2013. — Aug. — Vol. 14. — P. 547.
109. Lynch V. J. Use with caution: developmental systems divergence and potential pitfalls of animal models // Yale J Biol Med. — 2009. — Jun. — Vol. 82, no. 2. —

- P. 53–66.
110. Monteiro A., Podlaha O. Wings, horns, and butterfly eyespots: how do complex traits evolve? // *PLoS Biol.* — 2009. — Feb. — Vol. 7, no. 2. — P. e37.
 111. [Role of genes and their cis-regulatory elements during animal morphological evolution] / B. Sun, J. Tu, Y. Li, M. Yang // *Yi Chuan.* — 2014. — Jun. — Vol. 36, no. 6. — P. 525–535.
 112. Gaunt S. J., Paul Y. L. Changes in Cis-regulatory Elements during Morphological Evolution // *Biology (Basel).* — 2012. — Oct. — Vol. 1, no. 3. — P. 557–574.
 113. Филипченко Ю.А. Экспериментальная зоология. — Медгиз, 1932.
 114. Broad shifts in gene expression during early postnatal life are associated with shifts in histone methylation patterns / J. C. Lui, W. Chen, C. S. Cheung, J. Baron // *PLoS ONE.* — 2014. — Vol. 9, no. 1. — P. e86957.
 115. Aran D., Hellman A. DNA methylation of transcriptional enhancers and cancer predisposition // *Cell.* — 2013. — Jul. — Vol. 154, no. 1. — P. 11–13.
 116. Aran Dvir, Sabato Sivan, Hellman Asaf. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes // *Genome Biology.* — 2013. — Vol. 14, no. 3. — P. R21.
 117. Genome-wide methylation profiles reveal quantitative views of human aging rates / G. Hannum, J. Guinney, L. Zhao et al. // *Mol. Cell.* — 2013. — Jan. — Vol. 49, no. 2. — P. 359–367.
 118. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context / B. C. Christensen, E. A. Houseman, C. J. Marsit et al. // *PLoS Genet.* — 2009. — Aug. — Vol. 5, no. 8. — P. e1000602.
 119. Issa J. P. Aging and epigenetic drift: a vicious cycle // *J. Clin. Invest.* — 2014. — Jan. — Vol. 124, no. 1. — P. 24–29.
 120. Molecular mechanisms of gene silencing mediated by DNA methylation / M. Curradi, A. Izzo, G. Badaracco, N. Landsberger // *Mol. Cell. Biol.* — 2002. — May. — Vol. 22, no. 9. — P. 3157–3173.
 121. Льюин Б., Ребриков Д.В., Кофиади И.А. Гены. Лучший зарубежный учеб-

- ник. — Бином. Лаборатория знаний, 2011.
122. Talbert P. B., Henikoff S. Spreading of silent chromatin: inaction at a distance // *Nat. Rev. Genet.* — 2006. — Oct. — Vol. 7, no. 10. — P. 793–803.
 123. Du M., Zhang Q., Bai L. Three distinct mechanisms of long-distance modulation of gene expression in yeast // *PLoS Genet.* — 2017. — Apr. — Vol. 13, no. 4. — P. e1006736.
 124. Krivega I., Dean A. Enhancer and promoter interactions-long distance calls // *Curr. Opin. Genet. Dev.* — 2012. — Apr. — Vol. 22, no. 2. — P. 79–85.
 125. Harmston N., Lenhard B. Chromatin and epigenetic features of long-range gene regulation // *Nucleic Acids Res.* — 2013. — Aug. — Vol. 41, no. 15. — P. 7185–7199.
 126. A quantitative model of transcriptional regulation reveals the influence of binding location on expression / K. D. MacIsaac, K. A. Lo, W. Gordon et al. // *PLoS Comput. Biol.* — 2010. — Apr. — Vol. 6, no. 4. — P. e1000773.
 127. Quintero-Cadena P., Sternberg P. W. Enhancer Sharing Promotes Neighborhoods of Transcriptional Regulation Across Eukaryotes // *G3 (Bethesda)*. — 2016. — Dec. — Vol. 6, no. 12. — P. 4167–4174.
 128. Doheny J. G., Mottus R., Grigliatti T. A. Telomeric position effect—a third silencing mechanism in eukaryotes // *PLoS ONE*. — 2008. — Vol. 3, no. 12. — P. e3864.
 129. Ottaviani A., Gilson E., Magdinier F. Telomeric position effect: from the yeast paradigm to human pathologies? // *Biochimie*. — 2008. — Jan. — Vol. 90, no. 1. — P. 93–107.
 130. Effect of telomere proximity on telomere position effect, chromosome healing, and sensitivity to DNA double-strand breaks in a human tumor cell line / A. Kulkarni, O. Zschenker, G. Reynolds et al. // *Mol. Cell. Biol.* — 2010. — Feb. — Vol. 30, no. 3. — P. 578–589.
 131. Elgin S. C., Reuter G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila* // *Cold Spring Harb Perspect Biol.* — 2013. — Aug. — Vol. 5, no. 8. — P. a017780.

132. Turker M. S. Gene silencing in mammalian cells and the spread of DNA methylation // *Oncogene*. — 2002. — Aug. — Vol. 21, no. 35. — P. 5388–5393.
133. Silencing of mouse *Aprt* is a gradual process in differentiated cells / P. A. Yates, R. Burman, J. Simpson et al. // *Mol. Cell. Biol.* — 2003. — Jul. — Vol. 23, no. 13. — P. 4461–4470.
134. GATA2 facilitates steroid receptor coactivator recruitment to the androgen receptor complex / B. He, R. B. Lanz, W. Fiskus et al. // *Proc. Natl. Acad. Sci. U.S.A.* — 2014. — Dec. — Vol. 111, no. 51. — P. 18261–18266.
135. Prediction of promoters and enhancers using multiple DNA methylation-associated features / Woonchang Hwang, Verity F. Oliver, Shannath L. Merbs et al. // *BMC Genomics*. — 2015. — Vol. 16, no. 7. — P. S11.
136. Correlation of MGMT promoter methylation status with gene and protein expression levels in glioblastoma / M. Uno, S. M. Oba-Shinjo, A. A. Camargo et al. // *Clinics (Sao Paulo)*. — 2011. — Vol. 66, no. 10. — P. 1747–1755.
137. DNA methylation is correlated with gene expression during early pregnancy in *Bos taurus* / C. G. Walker, M. D. Littlejohn, S. Meier et al. // *Physiol. Genomics*. — 2013. — Apr. — Vol. 45, no. 7. — P. 276–286.
138. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts / J. R. Wagner, S. Busche, B. Ge et al. // *Genome Biol.* — 2014. — Feb. — Vol. 15, no. 2. — P. R37.
139. Tissue-specific regulation of *Igf2r/Airn* imprinting during gastrulation / C. Marcho, A. Bevilacqua, K. D. Tremblay, J. Mager // *Epigenetics Chromatin*. — 2015. — Vol. 8. — P. 10.
140. Stelzer Y., Jaenisch R. Monitoring Dynamics of DNA Methylation at Single-Cell Resolution during Development and Disease // *Cold Spring Harb. Symp. Quant. Biol.* — 2015. — Vol. 80. — P. 199–206.
141. Tracing dynamic changes of DNA methylation at single-cell resolution / Y. Stelzer, C. S. Shivalila, F. Soldner et al. // *Cell*. — 2015. — Sep. — Vol. 163, no. 1. — P. 218–229.

142. H19ICR mediated transcriptional silencing does not require target promoter methylation / C. Gebert, Q. Rong, S. Jeong et al. // *Biochem. Biophys. Res. Commun.* — 2016. — 07. — Vol. 476, no. 3. — P. 121–126.
143. Decreased pulsatile release of growth hormone in old male rats / W. E. Sonntag, R. W. Steger, L. J. Forman, J. Meites // *Endocrinology.* — 1980. — Dec. — Vol. 107, no. 6. — P. 1875–1879.
144. Aging-related changes in release of growth hormone and luteinizing hormone in female rhesus monkeys / M. J. Woller, G. Everson-Binotto, E. Nichols et al. // *J. Clin. Endocrinol. Metab.* — 2002. — Nov. — Vol. 87, no. 11. — P. 5160–5167.
145. Effects of a growth hormone-releasing hormone antagonist on telomerase activity, oxidative stress, longevity, and aging in mice / W. A. Banks, J. E. Morley, S. A. Farr et al. // *Proc. Natl. Acad. Sci. U.S.A.* — 2010. — Dec. — Vol. 107, no. 51. — P. 22272–22277.

Список иллюстративного материала

- 2.1 Окрестность гена MYCN у человека и консервативные элементы внутри нее. Элемент MYCN(-6893) подсвечен красным, MYCN(74) — синим, ген *Mycn* подсвечен зеленым, остальные консервативные элементы — серым 34
- 2.2 Главное окно программы *dotolog*, демонстрирующее возможность построения ступенчатой *матрицы* из дотплотов 35
- 3.1 Изменение расстояния между консервативными элементами MYCN(-6893) (подсвечен красным) и MYCN(74) (подсвечен черным) в окрестности гена *Mycn* у различных видов млекопитающих при возрастании массы тела взрослого животного. Ген подсвечен зеленым, остальные консервативные элементы — серым. Последовательности окрестностей гена *Mycn* центрированы по элементу MYCN(74). Млекопитающие упорядочены по массе тела взрослого животного 49
- 3.2 Графическое представление в *полулогарифмической* шкале зависимости между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием между наиболее скоррелированными парами консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Элементы, перекрывавшиеся с промотом гена, отмечены жирным. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика . . . 51

3.3	Графическое представление в <i>полулогарифмической</i> шкале зависимости между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием между наиболее скоррелированными парами консервативных элементов в окрестностях генов <i>Mycn</i> , <i>Plagl1</i> и <i>Ezh2</i> . Все виды приматов, кроме человека, исключены. Элементы, перекрывавшиеся с промотором гена, отмечены жирным. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика	54
3.4	Распределение по геному человека последовательностей, гомологичных последовательностям консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314). Красным прямоугольником выделено положение исходной последовательности . . .	63
3.5	Выравнивание гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314) в геноме человека с исходной последовательностью	64
3.6	Графическое представление в <i>полулогарифмической</i> шкале зависимости возраста полового созревания от <i>абсолютного</i> геномного расстояния (н.п.) между геном и ближайшей теломерой для генов <i>C1s</i> и <i>Notch1</i> , регулируемых механизмом TPE-OLD. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика	79

- 3.7 Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *относительного* геномного расстояния (доли длины хромосомы) между геном и ближайшей теломерой для генов соматотропной оси *Ghrh* и *Sst* и генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. Для гена *Ghrh* дополнительно представлена зависимость продолжительности жизни от расстояния между геном и ближайшей теломерой. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика 80
- 3.8 Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *абсолютного* геномного расстояния (н.п.) между геном и ближайшей теломерой для генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. **Все виды приматов, кроме человека, исключены.** Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика 81
- 3.9 Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания и продолжительности жизни от *относительного* геномного расстояния (доли длины хромосомы) между геном и ближайшей теломерой для генов соматотропной оси *Ghrh* и *Ghrl*. **Все виды приматов, кроме человека, исключены.** Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика 83

Список таблиц

2.1	Виды млекопитающих и некоторые их морфо-физиологические характеристики	29
2.2	Информация из базы данных NCBI Gene об исследуемых генах человека	30
2.3	Информация из базы данных NCBI Gene об исследуемых генах человека, регулируемых механизмом TPE-OLD	31
2.4	Цвета подсветки различных элементов последовательности в программе <i>dotolog</i>	36
2.5	Базы данных известных элементов генома	37
2.6	Виды млекопитающих, за исключением приматов, кроме человека, и некоторые их морфо-физиологические характеристики	41
2.7	Вошедшие в исследование виды приматов, за исключением человека, и некоторые их морфо-физиологические характеристики	42
2.8	Категории Gene Ontology сервиса <i>PANTHER</i>	43
3.1	Количество и размер выявленных консервативных элементов в окрестностях регулирующих рост генов млекопитающих	46
3.2	Корреляция между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием между некоторыми консервативными элементами в окрестностях генов <i>Mycn</i> , <i>Plagl1</i> и <i>Ezh2</i> . Элементы, перекрывавшиеся с промотором гена, отмечены жирным	48

3.3	Морфо-физиологические характеристики млекопитающих и геномное расстояние между наиболее скоррелированными парами консервативных элементов в окрестностях генов <i>Mycn</i> , <i>Plagl1</i> и <i>Ezh2</i> . Везде дано скорректированное по Бонферрони р-значение × 50000. Элементы, перекрывавшиеся с промотором гена, отмечены жирным	50
3.4	Последовательности ДНК консервативных элементов, входивших в пары наиболее скоррелированных с морфо-физиологическими характеристиками консервативных элементов. Элементы, перекрывавшиеся с промотором гена, отмечены жирным	53
3.5	Корреляция между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием между некоторыми консервативными элементами в окрестностях генов <i>Mycn</i> , <i>Plagl1</i> и <i>Ezh2</i> . Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения. Элементы, перекрывавшиеся с промотором гена, отмечены жирным . . .	55
3.6	Морфо-физиологические характеристики млекопитающих и геномное расстояние между наиболее скоррелированными парами консервативных элементов в окрестностях генов <i>Mycn</i> , <i>Plagl1</i> и <i>Ezh2</i> . Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения. Элементы, перекрывающиеся с промотором гена, отмечены жирным . . .	56
3.7	Известные элементы генома, перекрывавшиеся с консервативными элементами. Элементы, перекрывавшиеся с промотором гена, отмечены жирным. Если некоторый консервативный элемент перекрывался со множеством известных элементов генома, то дано только суммарное их количество, отмеченное курсивом	57

3.8	Клинически значимые вариации, <i>перекрывавшиеся</i> с интервалами между парами значимо скоррелированных консервативных элементов в окрестностях генов <i>Mycn</i> , <i>Plagl1</i> и <i>Ezh2</i> . Консервативные элементы, перекрывавшиеся с промотором гена, отмечены жирным . . .	59
3.9	Результаты BLAST-поиска в геноме человека гомологов консервативных элементов MYCN(-6893) и PLAGL1(79389). Гомологи пронумерованы	60
3.10	Результаты BLAST-поиска в геноме человека гомологов консервативного элемента EZH2(-8314). Гомологи пронумерованы	61
3.11	Представленность в геноме человека гомологов последовательностей консервативных элементов, входивших в пары наиболее скоррелированных консервативных элементов. Консервативные элементы, перекрывавшиеся с промотором гена, отмечены жирным. Хромосома, на которой находился исходный элемент, отмечена жирным. В скобках указано число гомологов на хромосоме	62
3.12	Гены в окрестностях гомологов в геноме человека консервативных элементов MYCN(-6893) (верхняя часть таблицы) и PLAGL1(79389) (нижняя часть таблицы)	66
3.13	Гены в окрестностях гомологов в геноме человека консервативного элемента EZH2(-8314)	67
3.14	Сверхпредставленные категории Gene Ontology для генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893) и EZH2(-8314).	68
3.15	Сверхпредставленные категории <i>Molecular function</i> генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)	69

3.16	Сверхпредставленные категории <i>Biological process</i> генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314). Показаны вхождения с уровнем значимости $p < 0.01$	70
3.17	Сверхпредставленные категории <i>Cellular component</i> генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)	71
3.18	Сверхпредставленные категории <i>Protein class</i> генов, находящихся в окрестностях гомологов консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)	72
3.19	Категории <i>Biological process (GO:0008150)</i> и <i>Molecular function (GO:0003674)</i> генов <i>ACVRL1</i> и <i>ACVR1B</i>	73
3.20	Категории <i>Cellular component (GO:0005575)</i> , <i>PANTHER pathway</i> и <i>REACTOME pathway</i> генов <i>ACVRL1</i> и <i>ACVR1B</i>	74
3.21	СрG сайты, входящие в состав эпигенетических часов Хорвата и находящиеся на расстоянии не более 65000 н.п. от позиций гомологов в геноме человека консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)	74
3.22	Информация из базы данных NCBI Gene о генах, перекрывавшихся одновременно с СрG сайтами из эпигенетических часов Хорвата и гомологами в геноме человека консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)	74
3.23	Общие категории Gene Ontology для генов, перекрывавшихся одновременно с СрG сайтами из эпигенетических часов Хорвата и гомологами в геноме человека консервативных элементов MYCN(-6893), PLAGL1(79389) и EZH2(-8314)	75

3.24	Корреляция между морфо-физиологическими характеристиками млекопитающих и <i>абсолютным</i> геномным расстоянием (н.п.) от начала гена до ближайшей теломеры для генов <i>C1s</i> и <i>Notch1</i> , регулируемых механизмом TPE-OLD	76
3.25	Корреляция между морфо-физиологическими характеристиками млекопитающих и <i>относительным</i> геномным расстоянием (доли длины хромосомы) от начала гена до ближайшей теломеры для генов соматотропной оси <i>Ghrh</i> и <i>Sst</i> и генов <i>C1s</i> и <i>Notch1</i> , регулируемых механизмом TPE-OLD	76
3.26	Возраст полового созревания млекопитающих и геномное расстояние от начала гена до ближайшей теломеры для генов соматотропной оси <i>Ghrh</i> и <i>Sst</i> и генов <i>C1s</i> и <i>Notch1</i> , регулируемых механизмом TPE-OLD	78
3.27	Корреляция между морфо-физиологическими характеристиками млекопитающих и геномным расстоянием от начала гена до ближайшей теломеры для генов соматотропной оси <i>Ghrh</i> и <i>Ghrl</i> и генов <i>C1s</i> и <i>Notch1</i> , регулируемых механизмом TPE-OLD. Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения	79
3.28	Морфо-физиологические характеристики млекопитающих и геномное расстояние от начала гена до ближайшей теломеры для генов соматотропной оси <i>Ghrh</i> и <i>Ghrl</i> и генов <i>C1s</i> и <i>Notch1</i> , регулируемых механизмом TPE-OLD. Все виды приматов, кроме человека, исключены. Р-значения даны без поправки на множественные сравнения	82