

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

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Ростов-на-Дону – 2020

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования Южный федеральный университет.

**Научный руководитель:** доктор биологических наук, профессор  
**Шкурат Татьяна Павловна,**  
заведующая кафедрой генетики Академии  
биологии и биотехнологии ЮФУ

**Официальные оппоненты:** **Орлов Юрий Львович,**  
доктор биологических наук, профессор;  
ФГАОУ ВО Первый МГМУ имени И. М. Сече-  
нова Минздрава России (Сеченовский Универ-  
ситет), г. Москва.

**Базыкин Георгий Александрович,**  
доктор биологических наук, профессор;  
Сколковский институт науки и технологий;  
Институт проблем передачи информации  
РАН, г. Москва.

**Ведущая организация:** Федеральное государственное бюджетное  
учреждение науки Институт биологии разви-  
тия им. Н. К. Кольцова Российской академии  
наук, г Москва.

Защита состоится «\_\_\_\_\_» \_\_\_\_\_ 2020 г. в \_\_\_\_\_ часов на заседании дис-  
сертационного совета Д.002.214.01 в Федеральном государственном бюджетном  
учреждении науки Институте общей генетики им. Н. И. Вавилова Российской ака-  
демии наук по адресу: 119991, Москва, ул. Губкина, 3.

С диссертацией и авторефератом можно ознакомиться в библиотеке и на сайте  
Института [www.vigg.ru](http://www.vigg.ru), тел. 8-499-135-14-31.

Автореферат разослан «\_\_\_\_\_» \_\_\_\_\_ 2020 г.

Ученый секретарь диссертационного совета,  
доктор биологических наук

Горячева И. И.

# Общая характеристика работы

**Актуальность проблемы.** Проблема регуляции роста млекопитающих остается одной из самых давних загадок в биологии [19, 20, 21, 22].

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* [22]. На сегодняшний день остаются не известны механизмы, контролирующие скоординированное уменьшение экспрессии этих генов, и факторы, которые лежат в основе эволюционного модулирования соответствующей генетической программы [21, 22].

В системную регуляцию роста млекопитающих вовлечены гены соматотропной оси, из которых основными являются следующие 7 генов: *Gh1*, *Ghrh*, *Ghrl*, *Igf1*, *Sst*, *Igfbp3* и *Igfbp1* [23]. 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 [21, 23].

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 [24]. 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 [3].

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 [25, 26]. 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 [4].

**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 выяв-

лять 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 регулирующих 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 раз-

ных 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» SBB-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.

# Основное содержание работы

## Материалы и методы

### Виды млекопитающих

Для исследования были отобраны 36 видов млекопитающих, обладающих различными морфо-физиологическими характеристиками (таблица 1). В качестве морфо-физиологических характеристик были рассмотрены масса и размер тела взрослого животного, возраст полового созревания и продолжительность жизни.

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

### Гены, участвующие в регуляции роста млекопитающих

В исследование были включены следующие 17 генов: 10 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* (таблица 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 из базы данных NCBI Gene с помощью разработанного набора bash-скриптов *efunctions* для взаимодействия с базами данных NCBI.

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.

### Выявление консервативных элементов в окрестностях регулирующих рост генов млекопитающих

С помощью разработанной программы *dotolog* [1] для построения и анализа точечной матрицы гомологий было проведено попарное сравнение последовательностей окрестности каждого из рассмотренных генов у исследованных видов млекопитающих с последовательностью окрестности соответствующего гена *человека*.

На основании визуального анализа гомологий были выявлены *консервативные элементы* (КЭ) генома. 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 элемента, а -6893 — расстояние в н.п. от середины элемента до начала соответствующего гена в геноме *человека*.

Таблица 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 sabaeus</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

## 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

Таблица 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

*mblast* [2], реализующего поиск на основе алгоритма BLAST. Последовательности ИЭГ были получены из баз данных, приведенных в таблице 3.

Для каждого консервативного элемента был получен список перекрывающихся с ним ИЭГ.

Таблица 3: Базы данных известных элементов генома

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



## Корреляционный анализ

Был проведен корреляционный анализ зависимости между морфо-физиологическими характеристиками млекопитающих и расстоянием между консервативными элементами (КЭ) в окрестностях исследованных генов. 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. 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.

Оба вида анализа были также проведены на группе животных, из которой были исключены все приматы, кроме человека.

Расчеты были выполнены посредством реализации подпрограмм на языке Java с использованием функционала Java-библиотеки *Apache Commons Math*.

## Распределение по геному человека гомологов консервативных элементов

Для наиболее скоррелированных с морфо-физиологическими характеристиками консервативных элементов 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* из базы данных NCBI Gene для каждого гомолога был получен список генов в окрестности  $\pm 50000$  н.п.

Для полученного списка генов с помощью сервиса *PANTHER* [27] был произведен анализ сверхпредставленности категорий генов по всем доступным категориям Gene Ontology при уровне значимости  $p < 0.05$  с поправкой Бонферрони.

Был получен список CpG сайтов, входящих в состав эпигенетических часов Хорвата [28] и находившихся на расстоянии не более 65000 н.п. от положений гомологов консервативных элементов. Для перекрывававшихся с такими CpG сайтами генов была получена информация о принадлежности к категориям Gene Ontology.

## Результаты

### Корреляция между распределением консервативных элементов и морфо-физиологическими характеристиками млекопитающих

Значимая связь была обнаружена только между массой и длиной тела взрослого животного и расстоянием между некоторыми консервативными элементами (КЭ) в окрестностях генов *Мусп*

и *Plagl1* и между продолжительностью жизни и расстоянием между некоторыми КЭ в окрестности гена *Ezh2* (таблица 4).

Регуляторные элементы генома, перекрывавшиеся с данными КЭ, даны в таблице 5. Важно отметить, что для каждого гена наиболее скоррелированная пара КЭ включала элемент, перекрывавшийся с *промотором* соответствующего гена.

Таблица 4: 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	$1.45 \times 10^{-8}$	0.0007	MYCN(-6893)	<b>MYCN(74)</b>	Масса тела взрослого животного
-0.77	$1.32 \times 10^{-7}$	0.007	MYCN(-6893)	<b>MYCN(1273)</b>	
-0.79	$1.5 \times 10^{-7}$	0.008	MYCN(-6893)	MYCN(-1587)	
-0.78	$1.58 \times 10^{-7}$	0.008	MYCN(-6893)	MYCN(3711)	
-0.75	$6.51 \times 10^{-7}$	0.03	MYCN(-25368)	MYCN(-4493)	
-0.75	$7.31 \times 10^{-7}$	0.04	MYCN(-25368)	<b>MYCN(74)</b>	
0.73	$6.48 \times 10^{-7}$	0.03	<b>PLAGL1(79389)</b>	PLAGL1(112171)	
-0.8	$4.71 \times 10^{-8}$	0.002	MYCN(-6893)	<b>MYCN(74)</b>	Длина тела взрослого животного
-0.77	$2.12 \times 10^{-7}$	0.01	MYCN(-6893)	<b>MYCN(1273)</b>	
-0.77	$3.08 \times 10^{-7}$	0.02	MYCN(-6893)	MYCN(3711)	
-0.77	$3.79 \times 10^{-7}$	0.02	MYCN(-6893)	MYCN(-1587)	
-0.76	$5.9 \times 10^{-7}$	0.03	MYCN(-25368)	MYCN(-4493)	
0.74	$3.37 \times 10^{-7}$	0.02	<b>PLAGL1(79389)</b>	PLAGL1(112171)	
-0.79	$4.07 \times 10^{-7}$	0.02	EZH2(-8314)	EZH2(2278)	Продолжительность жизни
-0.78	$5.37 \times 10^{-7}$	0.03	EZH2(-8314)	<b>EZH2(194)</b>	
-0.78	$5.99 \times 10^{-7}$	0.03	EZH2(-8314)	<b>EZH2(1627)</b>	

Графическое представление в *полулогарифмической* шкале обнаруженных зависимостей для наиболее скоррелированных пар КЭ дано на рис. 1. Примечательно, что на каждом графике точки выстраиваются вдоль прямой линии, что указывает на наличие *экспоненциальной* зависимости между морфо-физиологической характеристикой и расстоянием между КЭ.

Наличие в выборке животных большого количества видов приматов могло повлиять на репрезентативность выборки и результаты. Тем не менее, результаты корреляционного анализа на группе млекопитающих, за исключением приматов, кроме человека, согласуются с вышеуказанными результатами.

## Представленность в геноме человека гомологов наиболее скоррелированных с морфо-физиологическими характеристиками консервативных элементов

Представленность в геноме последовательностей, гомологичных последовательностям наиболее скоррелированных с морфо-физиологическими характеристиками млекопитающих консервативных элементов (КЭ), может отражать функциональную значимость последовательностей

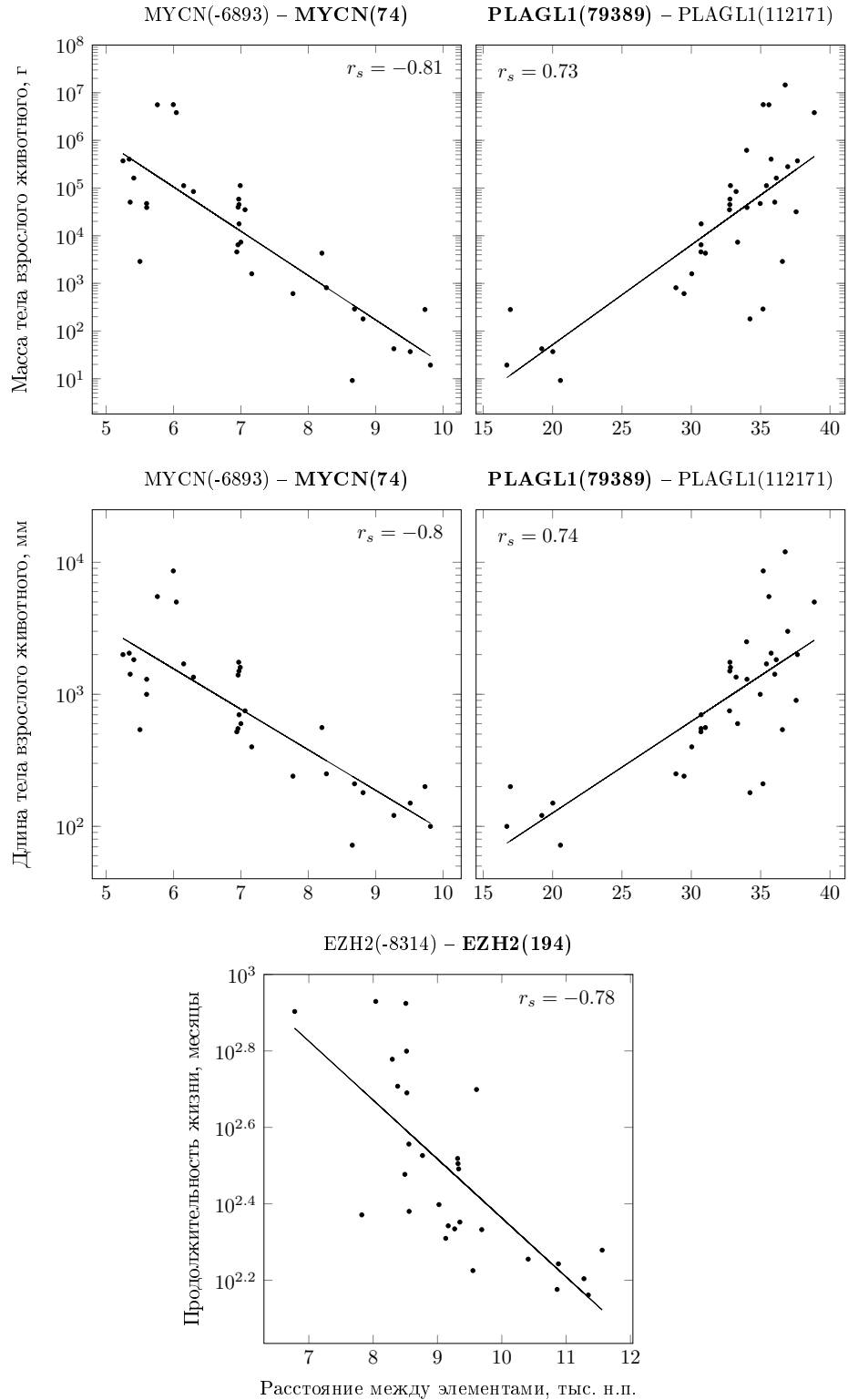


Рис. 1: Графическое представление в *полулогарифмической* шкале зависимости между морфофизиологическими характеристиками млекопитающих и геномным расстоянием между наиболее скоррелированными парами консервативных элементов в окрестностях генов *Mycn*, *Plagl1* и *Ezh2*. Элементы, перекрывавшиеся с промотором гена, отмечены жирным. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

Таблица 5: 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	MYCN_2, MYCN_1, Mycn_1	GATA2	ESRRB	STAT1								
MYCN(-6893)	98		GATA2										
MYCN(-4493)	157		<i>6 TFs</i>										
MYCN(-1587)	205												
<b>MYCN(74)</b>	307		<i>12 TFs</i>						PRDM14, TBX3 TFAP2C, SMARCA4	nsv997222	*		
<b>MYCN(1273)</b>	142		SIN3A, RAD21, EZH2, TAF1						nsv997222	*			
MYCN(3711)	1383		<i>12 TFs</i>						FOXA1, ESRRB, Klf4	FOXA1, CTCF, ETS1, ATOH1, RUNX1 Cdx1	nsv997222	*	
<b>PLAGL1(79389)</b>	130		Ezh2_1, EZH2_1, Ezh2_2						<i>62 TFs</i>	NFIC	Cdx1 E2F1		nsv997068, nsv510056 nsv997068
PLAGL1(112171)	297												
EZH2(-8314)	162												
<b>EZH2(194)</b>	729	EGR1, TCF7L2, NFYB, STAT3, E2F1, E2F3, E2F4, NRF1, E2F6		RB1, RBL2, EGR1, TFAP2C, STAT1, E2F4, FOS, ETS1, SMARCA4	rs550913824, rs886062082, rs545716282, rs886062081, rs886062083	*							
<b>EZH2(1627)</b>	510	EZH2_2		<i>12 TFs</i>	STAT1, ETS1, SMARCA4	*							
EZH2(2278)	509	<i>8 TFs</i>		STAT1, STAT3, JUND	STAT1, SMARCA4	*							

этих элементов.

Для последовательностей наиболее скоррелированных консервативных элементов был произведен поиск *гомологов* в геноме человека с помощью сервиса *Ensembl BLAST* (таблица 6).

Примечательно, что один из элементов каждой пары представлен лишь одной копией в геноме, т.е. является уникальной последовательностью генома, в то время как второй элемент пары — многими копиями. Последнее обстоятельство может указывать на потенциальную регуляторную

Таблица 6: Представленность в геноме человека гомологов последовательностей консервативных элементов, входивших в пары наиболее скоррелированных консервативных элементов. Консервативные элементы, перекрывававшиеся 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), <b>2(4)</b> , 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)
<b>MYCN(74)</b>	1	<b>2(1)</b>
<b>PLAGL1(79389)</b>	9	1(2), 3(1), 5(1), <b>6(2)</b> , 11(2), 12(1)
PLAGL1(112171)	1	<b>6(1)</b>
EZH2(-8314)	49	1(3), 2(8), 3(2), 4(1), 5(2), <b>7(7)</b> , 8(2), 9(3), 10(1), 11(3), 12(5), 14(3), 15(2), 16(2), 18(1), 20(1), X(3)
<b>EZH2(194)</b>	1	<b>7(1)</b>

роль таких последовательностей.

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

#### ***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  $\pm 50000$  н.п. и проведен анализ сверхпредставленности категорий Gene Ontology. Гены были сгруппированы по консервативному элементу и рассмотрены все вместе.

При группировке по консервативному элементу значимая сверхпредставленность категорий была обнаружена только для групп генов в окрестностях гомологов элементов MYCN(-6893) и EZH2(-8314) (таблица 7).

Наибольший интерес вызывают гены *ACVR1B* и *ACVRL1*, лежащие 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.

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 7: 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 function	PC00035 GO:0019838	TGF-beta receptor growth factor binding	0.1 0.0988	ACVR1B, ACVRL1 IL13RA1, ACVR1B, ACVRL1

### ***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*) [28].

В связи с этим был получен список из 353 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 8). Следует отметить, что из них два сайта накапливают метилирование с возрастом.

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.

Таблица 8: 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)

Хромосома	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	-

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

Значимая связь между морфо-физиологическими характеристиками млекопитающих и расстоянием от начала гена до ближайшей теломеры была обнаружена только для генов *Ghrh*, *Sst*, *C1s* и *Notch1* (таблицы 9 и 10).

Таблица 9: 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 $\times 200$	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 10: 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 $\times 200$	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>	

Графическое представление в *полулогарифмической* шкале обнаруженных зависимостей дано на рис. 2 и 3. Для гена *Ghrh* точки на графике выстраиваются вдоль прямой линии, что свидетельствует о наличии *экспоненциальной* зависимости.

Наличие в выборке животных большого количества видов приматов могло повлиять на репрезентативность выборки и результаты. Тем не менее, результаты корреляционного анализа на группе млекопитающих, за исключением приматов, кроме человека, согласуются с вышеуказанными результатами.

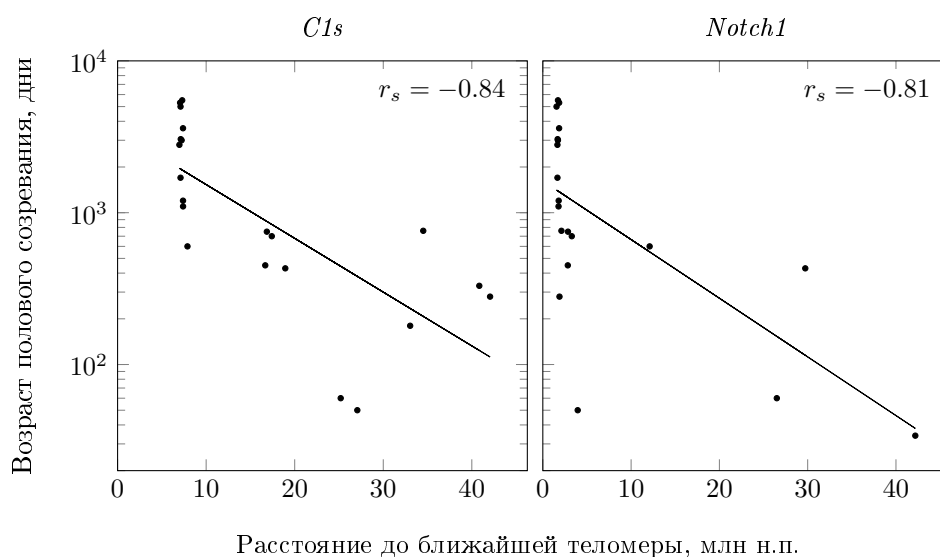


Рис. 2: Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *абсолютного* геномного расстояния (н.п.) между геном и ближайшей теломерой для генов *C1s* и *Notch1*, регулируемых механизмом TPE-OLD. Коэффициент корреляции Спирмена и прямая наилучшего приближения даны для каждого графика

## Обсуждение

### Эволюционное модулирование экспрессии регулирующих рост генов у разных видов млекопитающих

У млекопитающих показано существование единой эволюционно закрепленной генетической программы регуляции роста [22, 29, 30]. Эта программа реализуется сетью некоторых регулирующих рост генов, экспрессия которых градуально уменьшается с возрастом, а регуляция этих генов у разных млекопитающих может осуществляется одними и теми же *cis*-элементами, причем каждый ген может иметь свой набор *cis*-элементов [31, 32, 33].

Известно, что кривые роста млекопитающих сходны друг с другом [34, 35]. С другой стороны, временные паттерны уменьшения скорости соматического роста сходны с временными паттернами уменьшения уровня экспрессии регулирующих рост генов [36]. Отсюда следует, что скорость роста животного *пропорциональна* уровню экспрессии регулирующих рост генов.

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



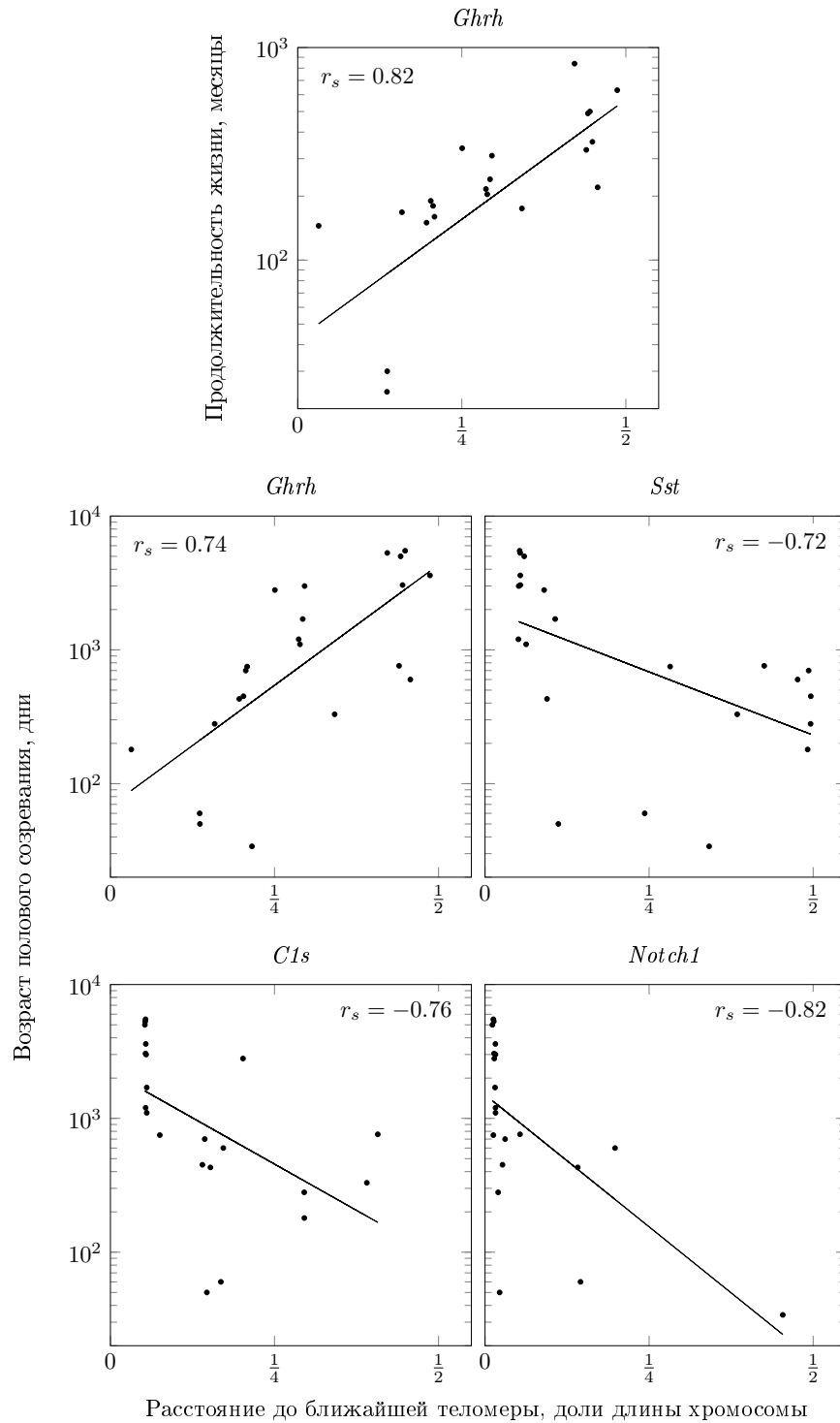


Рис. 3: Графическое представление в *полулогарифмической* шкале зависимости возраста полового созревания от *относительного* геномного расстояния (доли длины хромосомы) между геном и ближайшей теломерой для генов соматотропной оси *Ghrh* и *Sst* и генов *Cl<sub>s</sub>* и *Notch1*, регулируемых механизмом 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

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.

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

## **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 [20, 21, 36].

Постепенное замолкание может быть вызвано как эпигенетическими модификациями *промотора*, так и эпигенетическими модификациями *дистальных регуляторных сайтов* [37, 38].

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) [28].

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

## **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 *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.

Следует отметить, что для каждого из этих генов наиболее скоррелированная пара КЭ содержала элемент, перекрывающийся с *промотором* соответствующего гена, а зависимость носила *экспоненциальный* характер. Это позволяет предположить, что взаимодействие *промотора* и некоторого *дистального регуляторного сайта* может регулировать работу генов *Mycn*, *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.

### *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*.

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 [39].

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.

### *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*.

Единственным наследуемым репрессирующим механизмом взаимодействия на расстоянии является *эффект распространения неактивного состояния хроматина*. В частности, сайленсинг может быть обусловлен *эффектом распространения метилирования ДНК* [40, 41].

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.

## 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 *Cls* и *Notch1*, регулируемых механизмом TPE-OLD.

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 [42, 43]. 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.

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 [25, 26]. 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 *Cls* и *Notch1* значимо скоррелировано с возрастом полового созревания и продолжительностью жизни, причем в обоих случаях характер зависимости был сходен с таковым для 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 [4].

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 репрессирующие воздействие теломеры.

*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 *Cls* 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.

## 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.

## **Свидетельства о регистрации программ для ЭВМ**

1. 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.
2. 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.

## **Научные статьи по теме диссертации, опубликованные в журналах Scopus, WoS, RSCI**

3. 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.
4. 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.
5. Shkurat T. P., Romanov D. E., Shkurat M. A. Alu elements as source of microRNA sites in the human genome // Abstracts from the 51st European Society of Human Genetics Conference: Electronic Posters. — Vol. 27. — 2019. — P. 870–1041.

## **Другие научные работы, опубликованных по теме диссертации**

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. Romanov D. E., Shkurat T. P. Genome distance between regulatory elements of growth-related genes may determine morpho-physiological traits in mammals // Proceedings of the 12th International Multiconference “Bioinformatics of Genome Regulation and Structure/Systems Biology”. — Novosibirsk, 2020. — 7. — P. 95–96.
8. Романов Д. Е. Масса и размер тела млекопитающих скоррелированы с расстоянием между некоторыми элементами генома в окрестности генов регуляции роста // XXV Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов 2018». — г. Москва, 2018. — 4.
9. Романов Д. Е., Шкурат Т. П. Закономерности в расположении на хромосоме генов, регулирующих массу и размер тела млекопитающих // Материалы VII Международной научно-практической конференции с международным участием «Генетика — фундаментальная основа инноваций в медицине и селекции». — г. Ростов-на-Дону : Издательство ЮФУ, 2017. — 11. — С. 104–105.

10. Романов Д. Е., Шкурат Т. П. Закономерности в организации генома в окрестностях генов, регулирующих массу и размер тела млекопитающих // Материалы VII Международной научно-практической конференции с международным участием «Генетика — фундаментальная основа инноваций в медицине и селекции». — г. Ростов-на-Дону : Издательство ЮФУ, 2017. — 11. — С. 102–103.
11. Распространенность микро РНК в интронах и окрестностях генов CGA, FSHB, LHB, TSHB у млекопитающих с формированием различного числа доминирующих фолликулов / Т. П. Шкурат, Д. Е. Романов, Н. С. Пономарева и др. // Живые и биокосные системы. — 2015. — 12. — Т. 14.
12. Романов Д. Е. MSCANNER — a genome-wide motif finding tool // Материалы 7-й Международной Школы молодых ученых «Системная биология и биоинформатика» SBB-2015 / Под ред. А. В. Харкевич ; Институт цитологии и генетики СО РАН. — г. Новосибирск : ФГУП «Издательство СО РАН», 2015. — С. 40.
13. Localization of hsa-mir-5096 and hsa-mir-1268 around the Somatotropin Axis in Mammals / A. E. Panich, D. E. Romanov, N. S. Ponomareva et al. // Proceedings of the EMBO Conference “From Functional Genomics to Systems Biology”. — Heidelberg, Germany, 2014. — 11.
14. Аннотация некодирующей белок ДНК / А. А. Александрова, Д. Е. Романов, Г. Б. Бахтадзе и др. // Материалы VI съезда Вавиловского общества генетиков и селекционеров (ВОГиС) и ассоциированных генетических симпозиумов / Под ред. А. В. Харкевич. — г. Новосибирск : ФГУП «Издательство СО РАН», 2014. — 6. — С. 59.
15. Биоинформационные подходы к изучению структуры и функций регуляторных районов генов / Н. С. Пономарева, Д. Е. Романов, Е. С. Алешина и др. // Материалы VI съезда Вавиловского общества генетиков и селекционеров (ВОГиС) и ассоциированных генетических симпозиумов / Под ред. А. В. Харкевич. — г. Новосибирск : ФГУП «Издательство СО РАН», 2014. — 6. — С. 59.
16. Локализация микроРНК у млекопитающих в межгенных регионах и интронах генов оси соматотропина / Д. Е. Романов, А. А. Александрова, Г. Б. Бахтадзе и др. // Материалы VI съезда Вавиловского общества генетиков и селекционеров (ВОГиС) и ассоциированных генетических симпозиумов / Под ред. А. В. Харкевич. — г. Новосибирск : ФГУП «Издательство СО РАН», 2014. — 6. — С. 42.
17. Поиск зависимостей между некоторыми количественными признаками млекопитающих и характеристиками некодирующей ДНК вокруг генов, отвечающих за их формирование / А. И. Бутенко, Д. Е. Романов, М. Г. Адигеев и др. // Материалы V Междунар. науч.-практ. конф. «Актуальные проблемы биологии, нанотехнологий и медицины» / Под ред. Ю. А. Пехтеревой ; НИИ биологии ЮФУ. — г. Ростов-на-Дону : Издательство ЮФУ, 2013. — 10. — С. 23–24.
18. Анализ последовательностей геномов с помощью точечной матрицы гомологий / Н. С. Пономарева, Д. Е. Романов, А. И. Бутенко и др. // Материалы V Междунар. науч.-практ. конф. «Актуальные проблемы биологии, нанотехнологий и медицины» / Под ред. Ю. А. Пехтеревой ; НИИ биологии ЮФУ. — г. Ростов-на-Дону : Издательство ЮФУ, 2013. — 10. — С. 38–39.

## Список литературы

19. An extensive genetic program occurring during postnatal growth in multiple tissues / G. P. Finkielstein, P. Forcinito, J. C. Lui et al. // *Endocrinology*. — 2009. — Apr. — Vol. 150, no. 4. — P. 1791–1800.
20. 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.
21. Lui J. C., Baron J. Mechanisms limiting body growth in mammals // *Endocr. Rev.* — 2011. — Jun. — Vol. 32, no. 3. — P. 422–440.
22. 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.
23. Lui J. C., Garrison P., Baron J. Regulation of body growth // *Curr. Opin. Pediatr.* — 2015. — Aug. — Vol. 27, no. 4. — P. 502–510.
24. 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 Tassell et al. // *Genomics Proteomics Bioinformatics*. — 2008. — Dec. — Vol. 6, no. 3-4. — P. 129–143.
25. 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.
26. Misteli T. The long reach of telomeres // *Genes and Development*. — 2014. — Nov. — Vol. 28, no. 22. — P. 2445–2446.
27. 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.
28. Horvath S. DNA methylation age of human tissues and cell types // *Genome Biol.* — 2013. — Vol. 14, no. 10. — P. R115.
29. 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.
30. 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.
31. Monteiro A., Podlaha O. Wings, horns, and butterfly eyespots: how do complex traits evolve? // *PLoS Biol.* — 2009. — Feb. — Vol. 7, no. 2. — P. e37.
32. [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.
33. 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.
34. Филипченко Ю.А. Экспериментальная зоология. — Медгиз, 1932.
35. Bogin B. Evolutionary perspective on human growth // *Annu Rev Anthropol.* — 1999. — Vol. 28. — P. 109–153.
36. An imprinted gene network that controls mammalian somatic growth is down-regulated during post-



- natal 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.
37. Aran D., Hellman A. DNA methylation of transcriptional enhancers and cancer predisposition // *Cell.* — 2013. — Jul. — Vol. 154, no. 1. — P. 11–13.
  38. 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.
  39. 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.
  40. Turker M. S. Gene silencing in mammalian cells and the spread of DNA methylation // *Oncogene.* — 2002. — Aug. — Vol. 21, no. 35. — P. 5388–5393.
  41. 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.
  42. 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.
  43. 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.