

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Девятяров Руслан Мансурович

**Исследование регуляторных элементов генома, участвующих в
развитии и диапаузе эмбрионов амниот**

03.01.03 – Молекулярная биология

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2019

Работа выполнена в лаборатории «Экстремальная биология» института фундаментальной медицины и биологии Федерального государственного автономного образовательного учреждения высшего образования Казанского (Приволжского) федерального университета (КФУ) и института RIKEN, Center for Life Science Technologies (CLST), Йокогама, Япония.

Научный руководитель – Гусев Олег Александрович, кандидат биологических наук

Научный консультант – Абрамова Зинаида Ивановна, доктор биологических наук

Официальные оппоненты – Морозов Сергей Юрьевич, доктор биологических наук, заведующий лабораторией геномной инженерии вирусов, Научно-исследовательский институт физико-химической биологии имени А.Н. Белозерского ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова»

Воронежская Елена Евгеньевна, доктор биологических наук, руководитель лаборатории сравнительной физиологии развития, Федеральное государственное бюджетное учреждение науки Институт биологии развития имени Н.К. Кольцова Российской академии наук (ИБР РАН)

Мошковский Сергей Александрович, доктор биологических наук, профессор РАН, руководитель лаборатории медицинской протеомики, Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт биомедицинской химии имени В. Н. Ореховича» (ИБМХ)

Защита диссертации состоится « 6 » июня 2019 г. в 13:00 часов на заседании диссертационного совета МГУ.03.01 Московского государственного университета имени М.В.Ломоносова по адресу: 119234, Москва, Ленинские горы, д.1, стр. 12. Биологический факультет, ауд. М1

E-mail: mgu.03.01.dissovet@gmail.com

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на сайте ИАС «ИСТИНА»: <https://istina.msu.ru/dissertations/192245740/>

Автореферат разослан « ___ » _____ 20__ г.

Ученый секретарь диссертационного совета,
доктор биологических наук



Т.В.Комарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Точный пространственный и временной контроль транскрипции белок-кодирующих и некодирующих генов является фундаментальным процессом, лежащим в основе развития и дифференциации многоклеточных организмов. За последние десятилетия быстрое совершенствование омиксных технологий привело к более глубокому пониманию регуляции транскрипции на уровне промоторов, энхансеров, роли разнообразных транскрипционных факторов (ТФ) и микроРНК. Недавние исследования определяют промоторную область как важную детерминанту контроля транскрипции в эмбриональном развитии (Haberle et al., 2014). Промоторная область – относительно небольшой участок генома, окружающий точку инициации транскрипции (англ. transcription start site, TSS) и содержащий мотивы связывания ТФ, которые формируют комплекс пре-инициации. Было показано, что на разных стадиях развития стандартный комплекс пре-инициации или его отдельные субъединицы могут заменяться на альтернативные, что также видно из нуклеотидной последовательности промотора. Такое разнообразие промоторов связывают с регуляцией активности отдельных групп генов в процессе эмбриогенеза (Haberle et al., 2014, 2015; Nepal Chirag, Hadzhiev et al., 2013).

Подобных успехов в установлении механизмов регуляции на уровне отдельных промоторов удалось добиться в большей степени благодаря методу CAGE (англ. cap analysis of gene expression, кэп-анализ экспрессии генов), разработанного в рамках проекта по аннотации геномов млекопитающих и человека FANTOM японского исследовательского института RIKEN (Йокогама, Япония). В сочетании с секвенированием нового поколения на уровне отдельных молекул CAGE позволяет точно картировать точки инициации транскрипции с разрешением до 1 п.о., а также аккуратно оценить промоторную

активность в дифференцированных, стабильных клеточных популяциях, переходных состояниях под действием внешних стимулов, в процессе дифференцировки и прочих условиях (Forrest et al., 2014). Было показано, что помимо промоторов, CAGE позволяет определять активные энхансеры с большей точностью, чем другие полногеномные методы, в частности Chip-seq (Andersson et al., 2014; Arner et al., 2015). Созданные базы данных промоторов и энхансеров мыши и человека с применением метода CAGE являются ценным и востребованным ресурсом в области молекулярной биологии.

В связи техническими проблемами получения биоматериала, обусловленными имплантацией эмбриона и этическими проблемами промотором и энхансеры эмбрионального развития млекопитающих представлены фрагментарно. На данный момент ближайший к человеку вид с детально охарактеризованным промотором эмбриогенеза является рыбка *Danio rerio* (Haberle et al., 2014).

Куриный эмбрион является удобной моделью для изучения развития амниот, с хорошо аннотированным геномом, однако транскриптомных исследований эмбриогенеза не проводилось. Эмбриогенез домашней курицы *Gallus gallus* проходит по консервативному механизму с формированием трех зародышевых оболочек (Stern, 2005). Из 22000 белок кодирующих генов 60% имеют 1:1 ортолога в геноме человека (Hillier et al., 2014; Schmid et al., 2015). Интересной особенностью является то, что эмбрион домашней курицы способен останавливать свое развитие при понижении температуры, впадая в «температурный торпор» (англ. cold torpor), и занимает в своей адаптации промежуточное место между рептилиями и млекопитающими (Fasenko, 2007; Hiroshi and Hermann, 1986; Tazawa, 1973; Tazawa and Nakagawa, 1985; Whittow and Tazawa, 1991). Наконец, домашняя курица является важным сельскохозяйственным организмом, поэтому изучение адаптации ее эмбриона на

молекулярном уровне имеет не только фундаментальный интерес, связанный с эволюцией регуляторных элементов генома, но и потенциал применения таких данных на практике.

Цель и задачи

Целью работы явилось создание транскриптомной базы данных эмбрионального развития амниот на примере эмбрионов курицы, позволяющей оценить молекулярные механизмы процесса на уровне отдельных регуляторных элементов и расширить возможности функционального анализа *in vivo*.

В связи с поставленной целью решались следующие задачи:

1. Провести подготовку CAGE библиотек и секвенирование образцов куриных эмбрионов, находящихся на разных этапах развития и в различных состояниях метаболической активности
2. Создать интерактивную общедоступную транскрипционную базу данных для промоторов, активных в эмбриональном развитии курицы на платформе геномного браузера Zenbu
3. Определить гены, изменяющие свою экспрессию в ходе онтогенеза и в ответ на стресс экзогенного понижения температуры
4. Выявить специфичные промоторы, регулирующие процессы онтогенеза, ответ на стресс и индукцию гипометаболизма у эмбрионов
5. Провести валидацию идентифицированных промоторов путем активации соответствующих генов методом CRISPR-on *in vivo*
6. Определить транскрипционные факторы и сигнальные пути, участвующие в онтогенезе эмбрионов и в их адаптации к понижению температуры.

Научная новизна

Была впервые реконструирована динамика изменения транскрипционной активности на уровне отдельных промоторов в процессе эмбриогенеза амниот

на примере куриного эмбриона, покрывающая этапы от откладывания яйца до проклевывания. В сумме, 31863 TSS с уровнем экспрессии более 1 TPM (tag per million, тегов на миллион) были картированы на сборку генома galGal5, среди которых от 34 до 46% были активны на отдельных этапах развития. Геномный браузер Zenbu был использован для интеграции и интерактивной визуализации данных CAGE, аннотации генов, транскриптов, промоторов, энхансеров и других элементов генома курицы. TSS генов, важных для дифференцировки линий клеток были успешно определены в геноме, а также были получены профили их экспрессии, которые далее были подтверждены гибридизацией *in situ*. В данной работе были идентифицированы новые TSS генов «домашнего хозяйства» (housekeeping TSS), а также многочисленные специфичные для отдельных этапов развития TSS. Было продемонстрировано, что аккуратное картирование TSS методом CAGE позволяет эффективно предсказывать сайты связывания транскрипционных факторов в промоторных областях генов «домашнего хозяйства» и генов, меняющих свою активность в ходе эмбриогенеза. Впервые было показано, что метод CRISPR-on применим для изучения эмбрионального развития птиц.

Теоретическая и практическая значимость

Раннее эмбриональное развитие невозможно изучать на людях, поэтому анализ эмбриогенеза с использованием птиц как модели, филогенетически близкой к человеку, может помочь разобраться в понимании сложных регуляторных механизмах дифференцировки клеточных линий в процессе развития. Мы проследили весь трехнедельный период развития цыплёнка от откладывания яйца до проклевывания с использованием CAGE метода профилирования экспрессии и впервые собрали полногеномную карту точек инициации транскрипции, играющих ключевую роль в дифференцировке от ранних стадий прегаструляции до проклевывания. Путем анализа и сравнения

профилей экспрессии специфичных для отдельных этапов развития, были получены данные о транскрипционных факторах, отвечающих за дифференцировку, а также определены соответствующие им регуляторные модули. В дополнение был определен набор стабильно экспрессирующихся генов «домашнего хозяйства», которые можно использовать для нормализации образцов и калибровки. Впервые была продемонстрирована возможность применения метода CAGE для картирования TSS в эмбриональном развитии амниот. Далее мы показали на примере гена *Brachyury*, что технология геномного редактирования CRISPR/Cas9 может применяться для транскрипционной активации промоторов любых генов, для которых известно точное расположение сайта TSS. Полученные данные размещены в свободном онлайн доступе в браузере Zenbu в удобной для просмотра форме и представляют важный ресурс для дальнейшего исследования раннего развития амниот.

Степень достоверности результатов

Достоверность положений, выносимых на защиту, и выводов диссертации подтверждается выполнением экспериментов на современном надежном оборудовании с использованием стандартных протоколов и биоинформатических методов анализа. Для статистически точной оценки данных эксперименты были проведены в двух-трех повторностях и анализированы с учетом соответствующих критериев достоверности.

Личный вклад автора

Основные результаты работы были получены автором или при его непосредственном участии. Фильтрация и картирование чтений, полученных на платформе Heliscope, загрузка и визуализация в геномном браузере Zenbu осуществлялась совместно с Marina Lizio в институте RIKEN. Приготовление библиотек, секвенирование на платформе Illumina HiSeq2500 осуществлялось

автором. Дальнейший биоинформатический анализ, включая фильтрацию, картирование, кластеризацию сигнала CAGE, валидацию промоторов, в том числе микроРНК, нормализацию экспрессии, поиск специфичных TSS, TSS «домашнего хозяйства», ассоциация с генами и их транскриптами, поиск мотивов, характеристика основных свойств промоторных областей, анализ ко-экспрессии, поиск альтернативных промоторов, анализ дифференциальной экспрессии, функциональный анализ осуществлялись автором. Гибридизация *in situ* и CRISPR-on эксперименты проводились совместно с Hiroki Nagai. Обсуждение, основные положения и выводы диссертации сформулированы лично автором.

Основные положения, выносимые на защиту

1. Впервые собран промотором ранних этапов эмбрионального развития амниот, который представлен в интерактивной и общедоступной форме на базе геномного браузера.
2. Аннотированные промоторы имеют достоверный профиль экспрессии и точно идентифицированные точки инициации транскрипции, подтвержденные методами гибридизации *in situ* и CRISPR-on.
3. Охарактеризованы промоторные области специфичные для отдельных этапов эмбриогенеза, а также мотивы связывания транскрипционных факторов в этих промоторах.
4. Показано, что гипотермальная диапауза характеризуется циклической активностью транскриптома, регуляция которого опосредована специфическими факторами транскрипции.

Апробация работы

По материалам диссертации опубликовано 3 печатные работы в научных изданиях, входящих в список ВАК и в базы цитирования Scopus/WoS, из них 2 в высокорейтинговых иностранных рецензируемых журналах. 2 – в материалах

международных и всероссийских конференций.

Основные результаты данной работы были доложены и обсуждены на следующих конференциях:

1. Deviatiiarov R, Gusev O. Transcriptional response and structural features of chicken embryo promoterome under cold torpor, Information Technologies and Systems (ITaS), Kazan, Russia, 25 – 30 September 2018 (постерный доклад)

2. Deviatiiarov R, Gusev O. Promoters and enhancers landscape of embryonic development and hibernation in chicken, Bioinformatics: from Algorithms to Applications (BiATA), SPbU, Saint Petersburg, Russia, 16–19 July 2018 (устный доклад)

3. Deviatiiarov R, Gusev O. CAGE application for avian embryogenesis, Moscow Conference on Computational Molecular Biology (MCCMB), MSU, Moscow, Russia, July 27 – August 30, 2017 (постерный доклад)

4. Deviatiiarov R, Lizio M, Nagai H, Gusev O, Sheng G. Promoter landscape in chicken development and embryonic hibernation, Life of genomes, Kazan, Russia, August 31 – September 1, 2017 (устный доклад)

Структура и объём диссертации

Диссертационная работа изложена на 121 страницах и состоит из списка сокращений, введения, обзора литературы, материалов и методов, результатов и обсуждения в двух главах, выводов. Диссертация включает 40 рисунков и 2 таблицы, библиографический список из 141 наименований.

СОДЕРЖАНИЕ РАБОТЫ

ВВЕДЕНИЕ

Сформулированы цели и задачи исследования, обоснованы актуальность темы, ее научная новизна, теоретическое и практическое значение, обозначены положения, выносимые на защиту.

ОБЗОР ЛИТЕРАТУРЫ

В обзоре литературы приводятся примеры глобальных проектов, цель которых заключается в сборке геномов новых видов, созданию их аннотаций и

совершенствование предыдущих версий. Например, GOLD содержащий более 13500 секвенированных эукариотических геномов или проект «Genome 10K», собравший около 280 геномов позвоночных с конечной целью - 10 000 (Hausler et al., 2009; Koepfli et al., 2015). Указывается на тенденции последних лет, как практика больших геномных проектов ENCODE, Ensembl, RefSeq, VEGA, FANTOM по аннотации геномов модельных организмов стала распространяться на новые сборки геномов других видов благодаря снижению стоимости секвенирования.

Далее рассмотрены такие подходы по созданию аннотаций, среди которых отдельно отмечены транскриптомные методы (EST, SAGE, RNA-seq), их преимущества, недостатки и выгодное положение CAGE среди них (Forrest et al., 2014; Ozsolak and Milos, 2011; Parkinson and Blaxter, 2009; Saha et al., 2002). Обсуждаются проблемы, связанные с аналитической частью, включая выравнивание, уменьшение размерности, нормализацию. Описаны основные принципы подготовки библиотек CAGE и особенности их секвенирования на различных платформах (Illumina, HeliScope) (Kanamori-Katayama et al., 2011; Murata et al., 2014).

В обзоре на примере проекта FANTOM описана тенденция к переходу от аннотации отдельных элементов к функционированию систем в целом, стремление к пониманию процессов регуляции перехода клетки из одного состояния в другое при дифференцировке или под действием внешних факторов. Приводятся ключевые публикации проекта, в которых метод CAGE успешно применялся для характеристики таких процессов на уровне промоторов как кодирующих, так и не кодирующих РНК, и далее на уровне энхансеров (Andersson et al., 2014; Arner et al., 2015; De Rie et al., 2017; Takahashi et al., 2012).

Рассматривается проблема изучения эмбрионального развития, фрагментарность имеющихся данных, доступность материала для исследований.

Приводится описание эмбриогенеза куриного эмбриона на уровне морфологии и физиологии. Отдельный акцент сделан на особенности эмбриона останавливать свое развитие при понижении температуры окружающей среды. Дается определение гипотермальной диапаузы (температурный торпор, англ. cold torpor). Приводится обзор различных вариантов гибернации, указывается на промежуточное положение куриного эмбриона в такой адаптации между рептилиями, яйца которых могут месяцами храниться при охлаждении и млекопитающими, диапауза которых регулируется гормонально.

МАТЕРИАЛЫ И МЕТОДЫ

Материалы, использованные в работе

Оплодотворенные куриные яйца были получены с фермы Широяма (Канагава, Япония). Яйца инкубировали при 38.5° С до желаемых стадий. Для образцов HH1-HH26 (от 1,5 до 5 дней) РНК выделяли из эмбриональных и внеэмбриональных областей (HH1, HH2, HH3, HH4, HH5, HH7, HH11, HH14, HH19, HH23, HH26) китом miRNeasy (QIAGEN). Для образцов HH32, HH37, HH41 и HH45 (7-дневный, 10-дневный, 15-дневный и 20-дневный) РНК была выделена только из эмбриональной области. Выделение РНК из эмбрионов моложе трех дней проводили из нескольких эмбрионов (2-32), выделение РНК из эмбрионов возрастом от 3-х до 20-ти дней проводили из индивидуального организма. Экстраэмбриональные образцы РНК были также получены из внеэмбриональных тканей (амнион, хорион, аллантоис и желточный мешок), собранных из эмбрионов HH32 и HH41. Образцы зачатков конечностей были собраны из яиц (HH20, день 3), купленных у Granja Gibert (Камбрильс, Испания). Яйца инкубировали до стадии HH20. Почки передних и задних конечностей были расчленены отдельно в холодном PBS, и РНК экстрагировали с помощью Trizol (Gibco), согласно инструкции производителя. Образцы

стволовых клеток были приобретены в Cell Applications (клетки гладкой мускулатуры аорты: #CAC35405, гепатоциты: штамм Cornish cross, мезенхимальные стволовые клетки: штамм Rhode Island red).

Для проведения эксперимента с холодным торпором были выбраны точки НН1 (6 часов), НН4 (18 часов), НН16 (2.5 дня), НН28/29 (6 дней), что соответствует раннему, среднему и средне-позднему развитию цыплят. Две повторности для каждой из точек отбирались из эмбрионов, выращенных в нормальных условиях (группа А). Часть эмбрионов (по две повторности на каждую точку) помещалась на 15 градусов в течении 24 ч. (группа В). После охлаждения образцы, представляющие группу восстановления (группа С, 3 повторности) после индуцированной гибернации, возвращались на 38.5 °С (1 час восстановления). Таким образом было собрано 28 образцов (2[А]+2[В]+3[С] x 4 точки).

Библиотеки были приготовлены согласно стандартному протоколу для дальнейшего секвенирования на платформах Helicos и Illumina. Качество библиотек проверяли с помощью Agilent Bioanalyzer 2100. Затем библиотеки были секвенированы на одномолекулярном секвенаторе HeliScore, следуя указаниям производителя. Другая часть библиотек (эксперимент по гибернации) была секвенирована на платформе HiSeq 2500, Illumina в режиме одноконцевого чтения с длиной рида 50 п.о. по 8 образцов на дорожку.

Картирование, кластеризация сигнала CAGE

Данные секвенирования были картированы на сборку генома цыпленка galGal5 с использованием delve (fantom.gsc.riken.jp/software/), который использует скрытую Марковскую модель для выравнивания коротких чтений. Последовательности, которые однозначно выровнялись с текущей сборкой генома (Genbank ID: GCA_000002315.3; NCBI адрес https://www.ncbi.nlm.nih.gov/assembly/GCA_000002315.3/), с 85%

идентичностью и точностью 99%, были сохранены для всех последующих анализов. Картирование чтений эксперимента по гибернации проводилось с помощью bwa и hisat2. CAGE TSS определялись для каждого образца путем извлечения первой позиции всех картированных чтений и подсчета их количества, оценивая все перекрывающиеся положения относительно ориентации чтения. Пики (промоторы) определяли путем применения DPI-метода. Промоторы затем связывали с ближайшим геном в пределах 500 п.о. (как вверх, так и ниже по течению от TSS гена). Для достижения наивысшего охвата использовались модели генов Ensembl, RefSeq, NONCODE, ALDB, EST. Набор данных птичьего набора CAGE доступен в DDBJ (DNA Data Bank of Japan) с регистрационным номером DRA004812. Данные картирования CAGE были размещены на платформе браузера Zenbu в удобной интерактивной форме с общим доступом.

Последующий биоинформатический анализ проводился с использованием программ bedtools, MEME Suite, языков программирования python и R. В качестве основных пакетов для R были использованы edgeR, topGO, WGCNA, а также CAGEr и seqPattern.

CRISPR-on и гибридизация in situ

Единственный пик TSS был идентифицирован на 5'-конце куриного *Brachyury*. Промоторную область анализировали на наличие подходящих структур sgRNA и определили 4 последовательности sgRNA, расположенных в пределах 300 п.о. от TSS. Несколько других тестируемых генов (SNAI1, CDH1, MMP11 и DENND5B) были также выбраны для активации с помощью CRISPR-on. Для экспрессии sgRNA была использована конструкция «pAC154-dual-dCas9VP160-sgExpression» (addgene # 48240), несущая 4 sgRNA. Для электропорации векторов, экспрессирующих sgRNA и dCAS9-VP160, оплодотворенные яйца выращивали до стадии HH2-HH3 in ovo. Затем эмбрионы

выделяли, электропорировали экспрессионными векторами (конечная концентрация 1 мкг / мкл для каждой конструкции) и культивировали ex ovo в новой культуре, как описано ранее. В электропорированных эмбрионах определяли активность GFP, а после достижения HH10 эмбрионы фиксировали и обрабатывали для гибридизации РНК in situ как описано ранее. Для валидации экспрессии генов, связанных с зачатками конечностей, гибридизация РНК in situ выполнялась на всех стадиях с использованием меченых дигоксигенином антисмысловых РНК.

РЕЗУЛЬТАТЫ

Были получены образцы эмбриональной ткани цыпленка на 16 этапах развития, покрывая весь период со стадии Гамбургер и Гамильтон 1 (HH1) до HH45 (Рисунок 1). На этапе HH1 большая часть клеток плюрипотентны, тогда как на HH45 основная часть клеток достигает конечного этапа дифференцировки. Десять дополнительных образцов эмбриональных и стволовых клеток (2 зародышевые почки конечностей, 2 образца внезародышевой ткани, 3 гладкие мышцы аорты, 2 гепатоцитов, 1 из мезенхимальных стволовых клеток костного мозга) были включены для сравнения.

Картирование TSS

Средняя глубина прочтения для всех подготовленных образцов составила 2.7 миллиона, что сравнимо с 4 миллионами для образцов человека. Идентификация отдельных кластеров CAGE и соответствующих им TSS в геноме курицы проводилась с помощью метода DPI. В общем, было получено 142924 пика с мягким порогом фильтрации, предполагающим как минимум 3 чтения картированных в один TSS в одном из образцов. Среди них 31863 пика остались после фильтрации с «надежным» порогом, который требует 1 TPM

(tags per million). Подобно значениям, показанным при анализе TSS на основе CAGE в образцах человека и мыши, 60% (19172) всех устойчивых TSS для цыплят можно сопоставить с аннотированными TSS кодирующих или некодирующих генных моделей. Остальные 40% (12644) неаннотированных TSS представляют вероятные альтернативные промоторы и РНК некодирующих генов, которые еще предстоит охарактеризовать.

Около 31-46% всех устойчивых TSS в любом образце (медиана 10557 TSS, 33% от общих TSS) имели пиковые значения, превышающие 3 TPM. Было обнаружено, что 13701 из всех стабильных TSS (43%) перекрываются с CpG островками. Среди пиков TSS, идентифицированных в этой работе, 2660 (8.35%) перекрывали HCE (Highly Conserved Elements), а 110 (0.34%) перекрывали ASHCE (Avian-Specific HCE). Из 31863 TSS пиков было 3311 (11.4%) активных во всех 26 образцах (> 3 TPM), что можно считать повсеместно представленными транскриптами. Среди них 189 (0.59%) были высоко активны (> 100 TPM) и 79 (0.25%) были чрезвычайно высоко представлены во всех образцах (экспрессия > 300 TPM, и > 700 TPM по среднему значению). Большинство из наиболее часто используемых генов «домашнего хозяйства» (например, GAPDH и ACTB) также были представлены в последней группе. Согласно общей оценке генов «домашнего хозяйства» по Айзенбергу и Леванану мы также идентифицировали 1399 TSS «домашнего хозяйства».

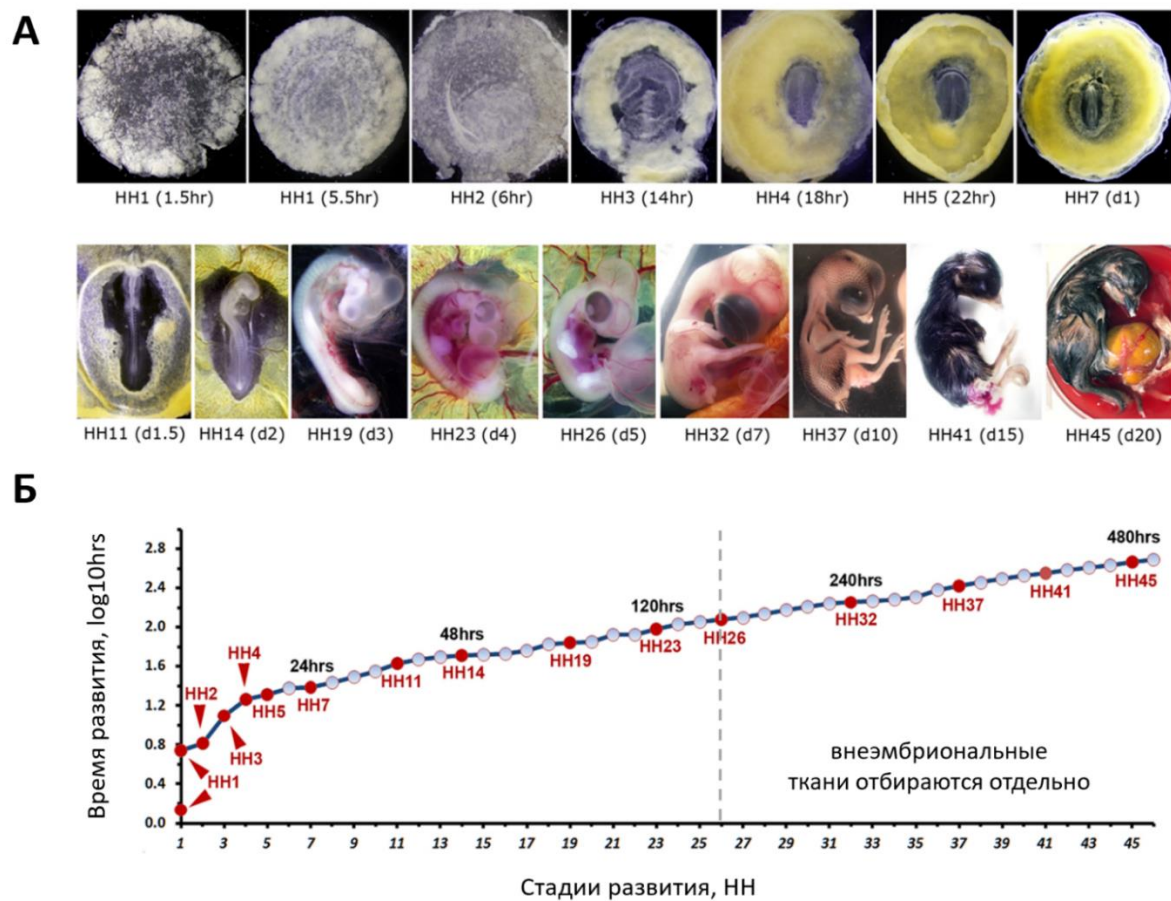


Рисунок 1. Этапы развития НН, выбранные для приготовления библиотек CAGE. А - Репрезентативные изображения куриных эмбрионов, используемых для выделения РНК; Б - Временное распределение выбранных стадий (красные точки) в эмбриогенезе курицы. Ось X: 46 этапов Гамбурга и Гамильтона (НН) представлены в числовом порядке. Ось Y: логарифмическая шкала времени инкубации, например, значение Y 2.0 = 100 часов.

Поиск и валидация специфичных TSS

Полученные данные позволили отслеживать изменения в положениях TSS на разных этапах эмбриогенеза и их количественную представленность. Половина сопоставленных аннотированных генов имела один связанный пик TSS во всех образцах, что указывает на то, что эти гены транскрибируются из стабильно расположенного TSS независимо от стадии развития или клеточного происхождения. Другая половина генов была ассоциирована с 2 или более пиками TSS, что, вероятно, отражает использование альтернативных промоторов.

Для оценки отношений между стадиями, основанной на их ландшафте TSS в геноме, был использован метод главных компонент (PCA, метод многомерного масштабирования). Образцы с первого дня эмбриогенеза курицы, покрывающие этапы от прегаструляции до ранних стадий образования сомитов (примерно соответствующие второй и третьей неделям развития человека), оказались плотно сгруппированы (Рисунок 2), что свидетельствует о том, что эти стадии обладают TSS с общими функциями.

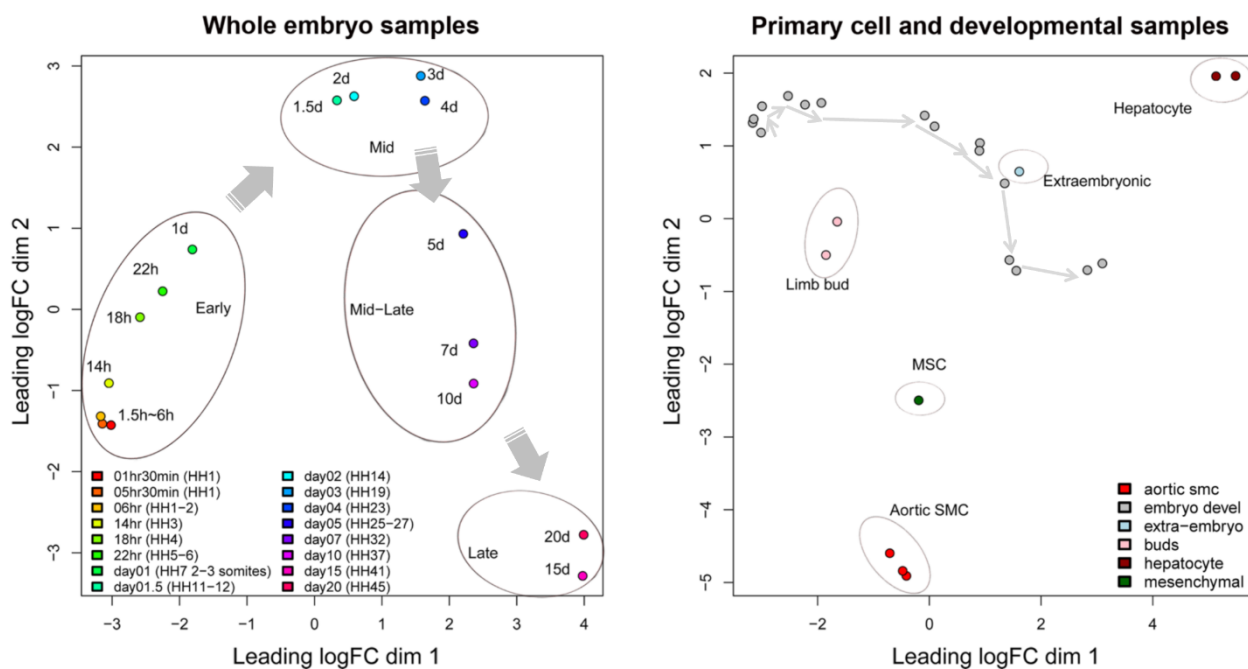


Рисунок 2. Анализ главных компонент PCA. А – Образцы развития цыплят. Мы выделили четыре группы образцов и обозначили их как ранние (Early), средние (Mid), средне-поздние (Mid-Late) и поздние (Late); Б – PCA для всех образцов, включая почки конечностей (Limb bud), внеэмбриональные ткани (Extraembryonic), гепатоциты (Hepatocyte), мезенхимальные стволовые клетки (MSC), клетки гладкой мускулатуры аорты (Aortic SMC).

Также мы определили группы, представляющие средние этапы развития (день 1.5-4, от сердечной до дифференциации печени), средне-позднее развитие (день 5-10, от окончательного гемопоэза до хондрогенеза) и поздние этапы (день 15 и 20, формирование костей и перьев, терминальная дифференциация большинства других клеточных линий), которые можно легко различить. Мы

показали, что профили экспрессии CAGE подтверждаются на примерах генов-маркеров, специфичных для эктодермы: *ASL1* (хрусталик), *GRIA4* (нейроны) и *PLP1* (олигодендроциты); маркеров, специфичных для мезодермы: *TNMD* (сухожилие), *TNNT3* (быстрая мышца), *MYH7B* (медленная мышца) и *LDB3* (поперечнополосатая мышца); для маркеров энтодермы: *ALB*, *F2* и *SPP2* (печень) и *GKN1* (желудок) на протяжении всего развития (Рисунок 3).

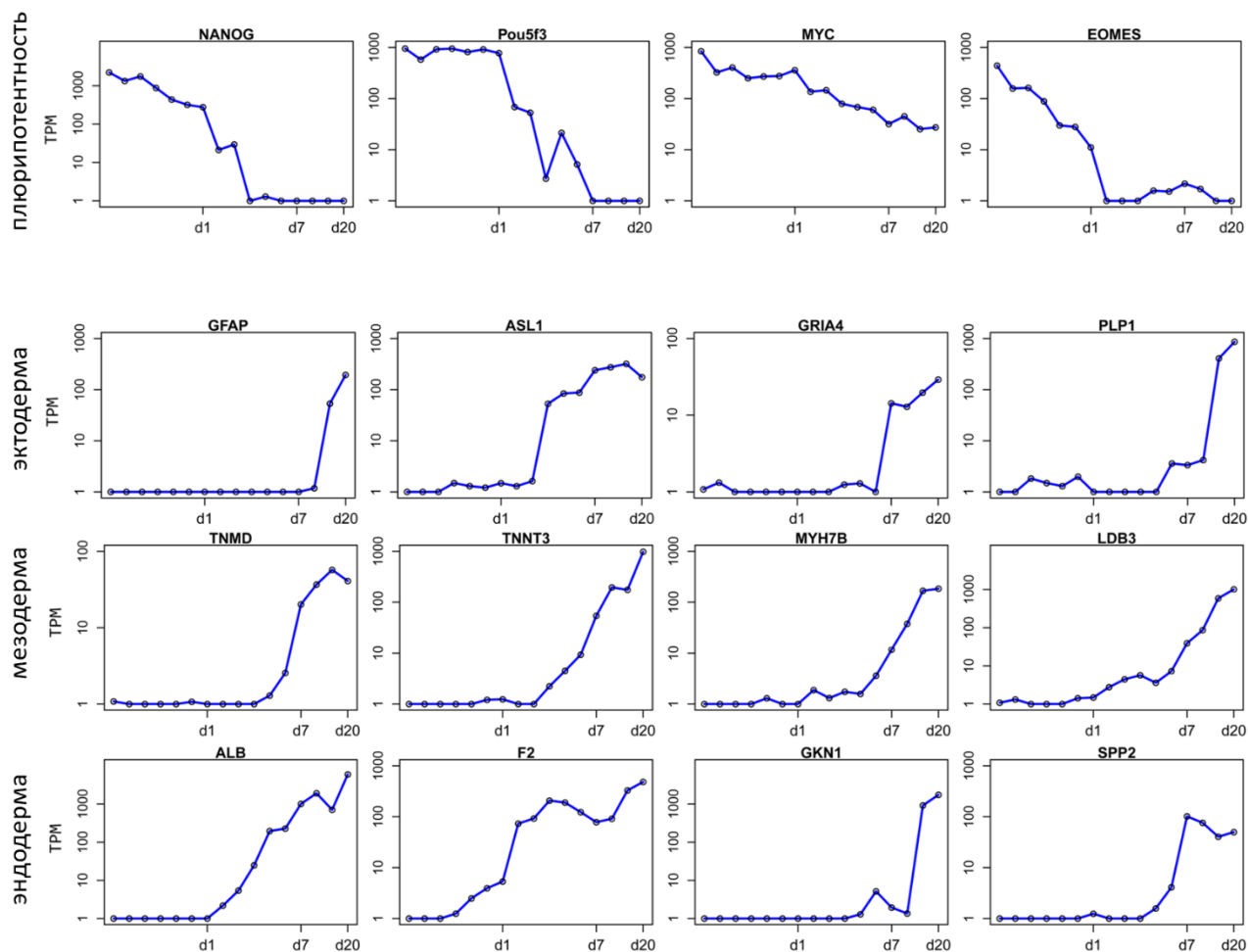


Рисунок 3. Профили экспрессии промоторов специфичных для определенных стадий развития и типов клеток. Ось X представляет собой стадии развития; ось y представляет значения экспрессии (TPM) в логарифмическом масштабе.

Для определения специфичных промоторов мы применили ограничение по активности в 10 TPM, а также условие, что они имеют более чем 10-кратное обогащение по сравнению с остальными образцами из всего эксперимента.

Было обнаружено, что 2296 TSSs были перепредставлены в исследованных клеточных линиях (1175 в гепатоцитах, 449 в клетках гладкой мускулатуры аорты и 672 в мезенхимальных стволовых клетках) и 608 TSS в образцах развития (172 в образцах ранней стадии, 118 в образцах средней, 184 в образцах средне-поздней стадии, 493 в образцах поздней стадии, 292 в экстраэмбриональных образцах и 81 в образцах из зачатков конечностей – рисунок 4).

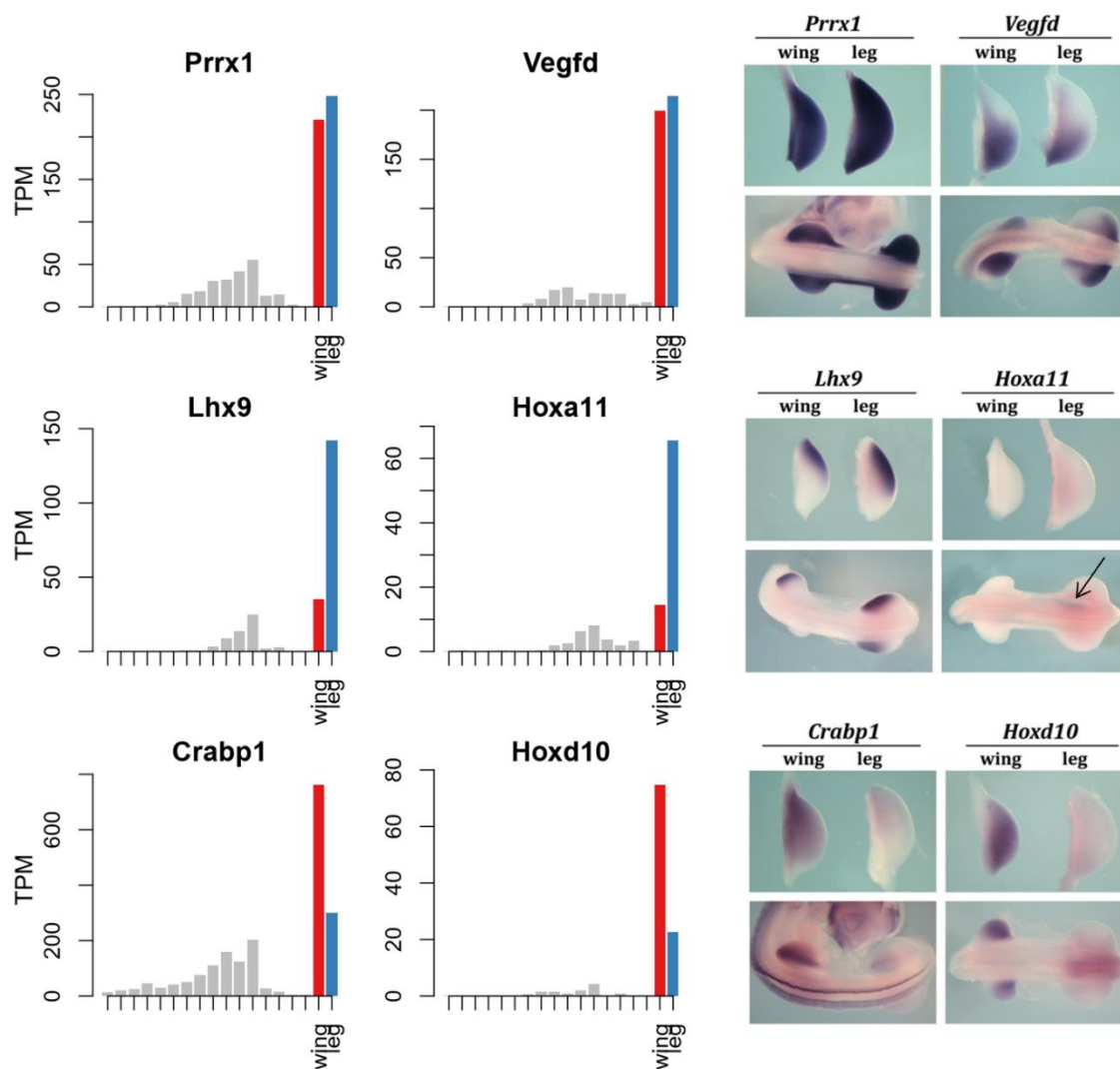


Рисунок 4. Примеры генов с профилем экспрессии специфичным для зачатков конечностей. А – профиль экспрессии соответствующих TSS для генов PRRX1 и VEGFD в зачатках крыльев и ног, CRABP1 и HOXD10 в крыльях, LHX9 и HOXA11 – в зачатках ног, серым отмечены образцы развития эмбриона; Б – гибридизация *in situ* для данных генов.

Далее мы продемонстрировали на примере гена *Brachyury* (Рисунок 5), а также некоторых других генах (*SNAI1*, *CDH1*, *MMP11* и *DENND5B*), возможность геномного редактирования с помощью CRISPR/Cas9 для транскрипционной активации промоторов любых генов, для которых известно расположение сайта TSS, что также подтверждает точность определения TSS методом CAGE. Точность определения специфично экспрессирующихся TSS и их положение в геноме имеет критическую роль для определения мотивов связывания транскрипционных факторов.

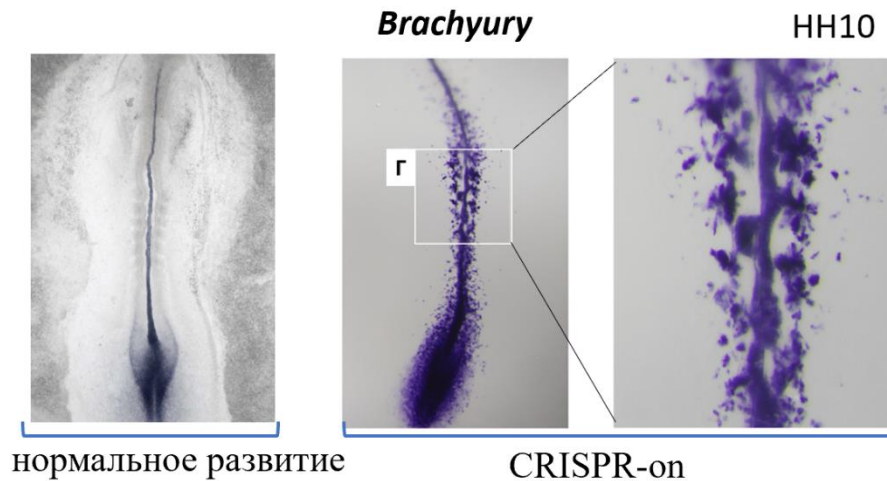
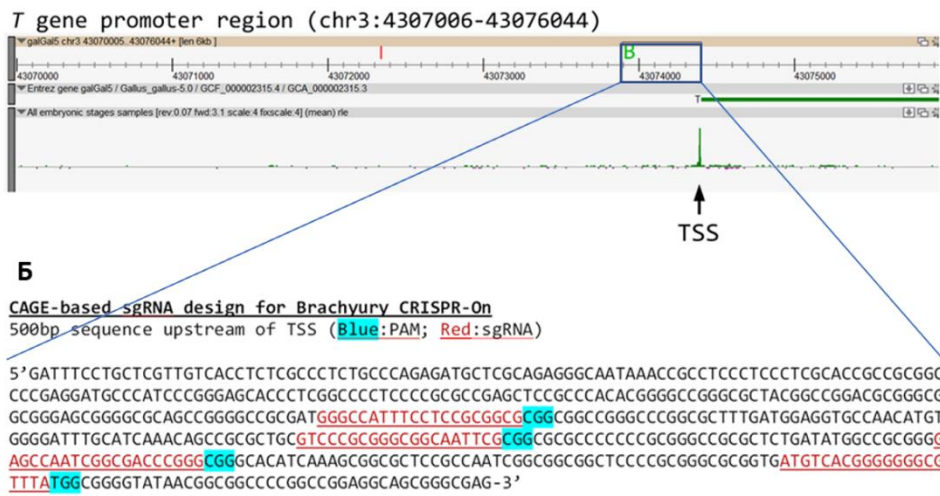


Рисунок 5. Индукция эндогенной экспрессии *Brachyury*, с помощью системы CRISPR-on. А – визуализация 5' концевой участка гена *Brachyury*, а также соответствующего ему TSS на 3 хромосоме в “Chicken-Zenbu”. Выделенный участок представлен в виде последовательности в Б. Б – визуализация выбранных гидовых РНК (gsRNA, отмечены

красным), расположенных в пределах 300 п.о. вверх от позиции TSS *Brachyury*. Бирюзовым отмечен PAM (protospacer adjacent motif); В – Электропорированные эмбрионы выращивались до стадии Гамбургера Гамильтона HH10, экспрессию гидовых РНК оценивали по уровню сигнала коэлектропорированного GFP. Далее проводилась *in situ* гибридизация для *Brachyury*; Г – приближенное изображение в средней области эмбриона, показывающая наличие продукта *Brachyury* в сомитах и нервной трубке.

Промотором гипотермальной диапаузы куриных эмбрионов

В данной работе мы проанализировали транскриптом CAGE раннего (HH1, HH4), среднего (HH16) и средне-позднего (HH28) эмбриогенеза цыплят при его охлаждении и восстановлении. Для 18304 идентифицированных TSS, среди которых 37% новых, мы провели анализ дифференциальной экспрессии и показали, что статистически значимо отвечают на понижение температуры от 803 до 2514 промоторов. Среди них от 50 до 151 имели обратимую экспрессию при восстановлении, кроме этапа HH1, на котором 301 промотор обратимо повышали экспрессию и 516 – понижали. Мы определили группу универсальных TSS, регуляция которых необходима вне зависимости от этапа эмбриогенеза и соответствующие им гены: SMS, HIST1H3H, GCSH, ATP6V0C, HP1BP3, MEAF6, TUBA1C, SLC25A5, PTCH2, кодирующие компоненты цитоскелета, нуклеосом, либо плохо охарактеризованные белки.

Для каждого этапа развития определили ключевые транскрипционные факторы, регулирующие процессы адаптации эмбриона. На рисунке 6 показаны точки связывания ТФ в промоторе гибернации на примере специфичных и общих мотивов. Всего обнаружено 25 общих ТФ: EGR4, ELK1, ELK3, ELK4, ETV1, ETV5, ERF, FLI1, Gabpa, KLF14, KLF16, KLF5, MZF1, NRF1, SP1, SP2, SP3, SP4, SP8, TBP, YY1, YY2, ZNF740, NFYB, ZBTB7A.

Нами был проведен анализ ко-экспрессии, на основе которого создана сеть всех активных промоторов эмбриона, внутри которой мы определили

ключевые модули, отвечающие за адаптацию на всех этапах развития, либо в конкретных точках. Например, были описаны модули стрессового ответа с GO тегами “ответ на стресс”, “клеточный ответ на стресс” и метаболическими путями “эксцизионная репарация нуклеотидов”, “убиквитин-опосредованный протеолиз” по идентификаторам KEGG. Также обнаружен примечательный модуль с сигнальным путем PPAR и работой пероксисом.

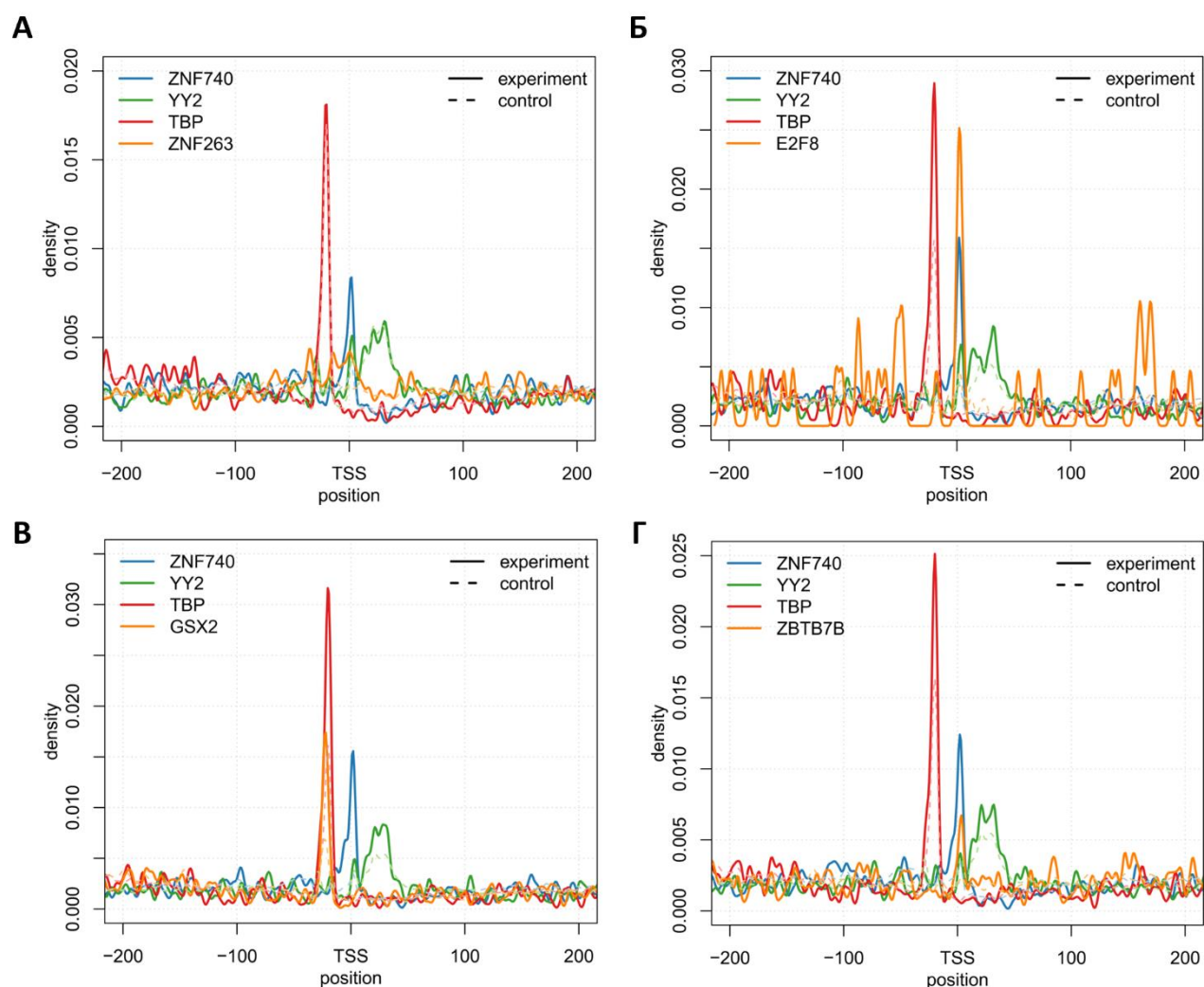


Рисунок 6. Визуализация позиционной перепредставленности мотивов в группах промоторных регионов (± 200 п.о.) имеющих дифференциальную экспрессию в эксперименте по охлаждению (А против В) на этапах развития HH1 (6 ч) – А, HH4 (18 ч) – Б, HH16 (2.5 д) – В, HH28 (6 д) – Г. Показаны 3 консервативных мотива – ZNF740, YY2, TBP, а также примеры специфичных наиболее перепредставленных мотивов с минимальным p-value.

Все данные CAGE размещены в интуитивно понятной интерактивной форме на платформе браузера Zenbu (<http://fantom.gsc.riken.jp/zenbu/gLyphs/#config=b1zZI1gUFZ6mHX6-4Gvxx>).

Пример визуализации использования альтернативного промотора на рисунке 7.

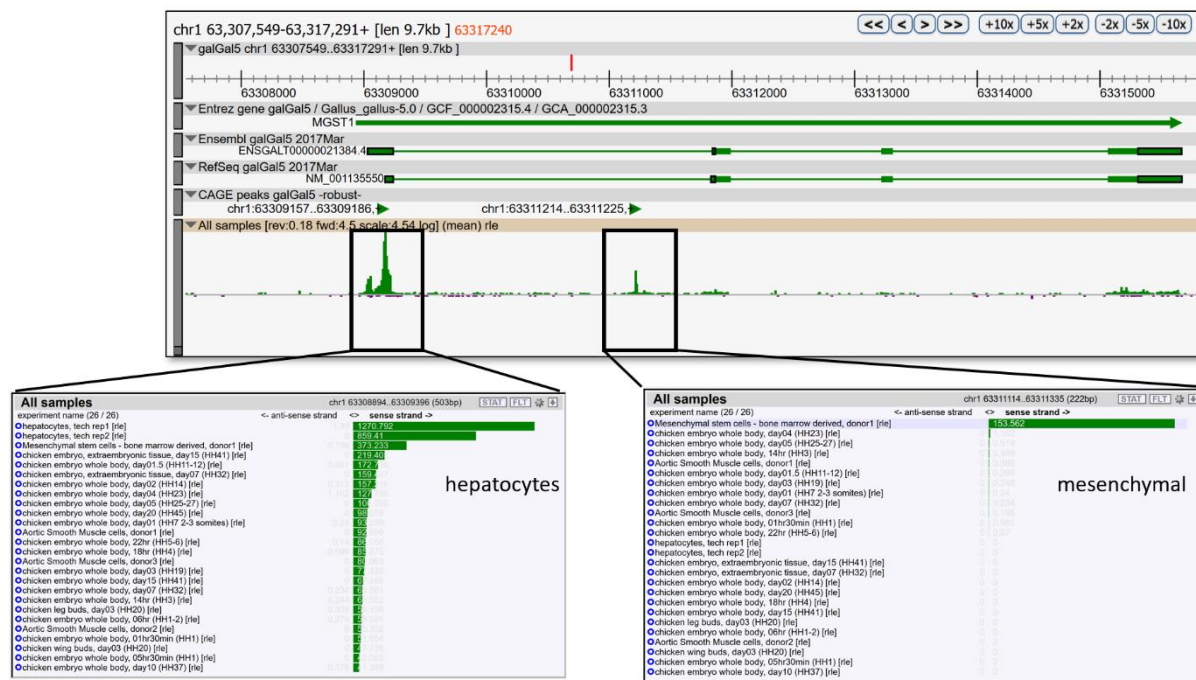


Рисунок 7. Пример визуализации промоторной области MGST1 (микросомальная глутатион-S-трансфераза 1) в Zenbu. TSS слева имеет специфичный для гепатоцитов профиль экспрессии, в то время как второй является альтернативным промотором с экспрессией, связанной с мезенхимальными стволовыми клетками.

ОБСУЖДЕНИЕ

Применяя технологию CAGE, мы создали полногеномные профили TSS для куриных эмбрионов, охватывающих весь период развития от прегаструляции до проклевывания. TSS развития были картированы на последнюю сборку генома цыплят и открыты для интерактивной визуализации в Chicken-ZENBU. Активность промоторов, отражающая сложные события спецификаций клеточной линии, а также функции генов «домашнего хозяйства», были определены и представляют богатый ресурс для анализа онтогенетической эволюции активности транскрипции в стабильных клеточных линиях. В

качестве первого генотипического TSS-картирования у птиц и первого всестороннего анализа TSS амниот, наши данные могут быть использованы для уточнения аннотации генома цыплят в сочетании с последним улучшением сборки генома и для анализа регуляции транскрипции 5' некодирующих областей генов, имеющих важное значение для развития.

Геномный проект по птицам (Avian Phylogenomics Project) уже содержит 48 собранных доступных сборок, каждая из которых может быть дополнительно аннотирована с использованием CAGE (Zhang et al., 2014). Кроме того, геномный проект 10К геномов включает большое количество секвенированных видов птиц и рептилий (около 5000 и 3000), которые также будут аннотированы в дальнейшем (Haussler et al., 2009; Koenfli et al., 2015). Подобные геномные проекты могут быть интересными с точки зрения перекрестных сопоставлений видов и поиска консервативных молекулярных механизмов, отвечающих за формирование специфических для птиц признаков, таких как вокальное обучение певчих видов, адаптация к полету, оперение и другие геномные особенности (Stern, 2005). Партеногенез некоторых рептилий, таких как ящерицы и змеи, делает эти организмы хорошей моделью для дальнейшего эволюционные исследования относительно регенерации, процессов определения пола или эффектов бесполого размножения (Sarre et al., 2011). Метод CAGE может помочь определить новые гены, участвующие в размножении или регуляторные сети, зависящие от температуры, объяснить некоторые специфические для вида признаки, такие как липучки на лапках, адаптации для полета или скольжения, формы языка, выработка ядов. Таким образом, база данных Reptilian Transcriptomes и проект птичьего филогенома могут быть дополнительно усовершенствованы путем интеграции с данными CAGE, для уточнения генных аннотаций, точно определяя положение TSS и энхансеров, и таким образом улучшить наше понимание эволюции признаков.

ВЫВОДЫ:

1. Впервые собран промотером птиц на основе эмбриона курицы, сопоставимый по полноте с атласом промоторов млекопитающих, насчитывающий всего около 30000 промоторов.

2. Впервые проведено полногеномное профилирование экспрессии генов и промоторов на ранних этапах развития амниот, начиная с точки 1.5 ч.

3. Определены промотеры и соответствующие гены активные на различных этапах развития, гены «домашнего хозяйства», а также обозначены ключевые транскрипционные факторы и сигнальные пути, отвечающие за регуляцию эмбриогенеза.

4. Метод CRISPR-on, впервые примененный нами на куриных эмбрионах, показал возможность активации генов *in vivo* у амниот при условии, что известна точка инициации транскрипции этих генов.

5. Регуляция перехода к состоянию гипотермальной диапаузы отличается от голодания, спячки и эмбриональной диапаузы - связана с известными метаболическими путями лишь фрагментарно.

6. Создана публично доступная база данных, включающая аннотацию и активность промоторов в эмбрионах курицы в ходе развития, дополненная данными по отдельным тканям и типам клеток

Публикации по теме диссертации

Научные статьи, опубликованные в журналах Scopus, WoS, RSCI, а также в изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности

1. Lizio M*, **Deviatiarov R***, Nagai H*, Galan L, Arner E, Itoh M, Lassmann T, Kasukawa T, Hasegawa A, Ros MA, Hayashizaki Y, Carninci P, Forrest ARR, Kawaji H, Gusev O, Sheng G. Systematic analysis of transcription start sites in avian development // PLoS Biol. – 2017. – 5;15(9):e2002887.

doi:10.1371/journal.pbio.2002887

2. **Deviatiiarov R**, Lizio M, Gusev O. Application of a CAGE Method to an Avian Development Study // *Methods Mol Biol.* – 2017. – V.1650 – P.101-109
doi:10.1007/978-14939-7216-6_6

3. Derek de Rie, Imad Abugessaisa, Tanvir Alam, Erik Arner, Peter Arner, Haitham Ashoor, Gaby Astrom, Magda Babina, Nicolas Bertin, A. Maxwell Burroughs, Carsten O. Daub, Michael Detmar, **Ruslan Deviatiiarov**, Alexandre Fort, Claudia Gebhard, Dan Goldowitz, Sven Guhl, Jayson Harshbarger, Akira Hasegawa, Kosuke Hashimoto, Peter Heutink, Chung Chau Hon, Edward Huang, Yuri Ishizu, Takeya Kasukawa, Peter Klinken, Timo Lassmann, Charles-Henri Lecellier, Weonju Lee, Marina Lizio, Vsevolod Makeev, Anthony Mathelier, Yulia A. Medvedeva, Christopher J. Mungall, Shohei Noma, Mitsuhiro Ohshima, Helena Persson, Filip Roudnicky, Pal Saetrom, Jessica Severin, Jay W. Shin, Hiroshi Tarui, Kristoffer Vitting-Seerup, Louise Winteringham, Yoko Yamaguchi, Kayoko Yasuzawa, Christine Wells, Kim M. Summers, Hideya Kawaji, Albin Sandelin, Michael Rehli, the FANTOM consortium, Yoshihide Hayashizaki, Piero Carninci, Alistair R. R. Forrest, Michiel J. L. de Hoon., An integrated expression atlas of miRNAs and their promoters in human and mouse // *Nature Biotechnology* – 2017. – V.35 – P.872–878
doi: 10.1038/nbt.3947

Тезисы конференций

1. Deviatiiarov R, Gusev O. Transcriptional response and structural features of chicken embryo promoterome under cold torpor, Information Technologies and Systems (ITaS), Kazan, Russia, 25 – 30 September 2018, P. 581-583
<http://itas2018.iitp.ru/papers/>

2. Deviatiiarov R, Gusev O. CAGE application for avian embryogenesis, Moscow Conference on Computational Molecular Biology (MCCMB 2017), MSU, Moscow, <http://mccmb.belozersky.msu.ru/2017/proceedings/abstracts/184.pdf>