

КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ ФУНДАМЕНТАЛЬНОЙ МЕДИЦИНЫ И БИОЛОГИИ

На правах рукописи

Девятяров Руслан Мансурович

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

06.06.01 – Биологические науки

03.01.03 – Молекулярная биология

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

Научный руководитель:

к.б.н., в.н.с., Гусев О.А.

Научный консультант:

д.б.н., профессор, Абрамова З.И.

Москва – 2018

СОДЕРЖАНИЕ

СПИСОК СОКРАЩЕНИЙ.....	4
ВВЕДЕНИЕ.....	7
ОБЗОР ЛИТЕРАТУРЫ.....	12
1. Аннотация генома.....	12
1.1 EST	13
1.2 SAGE.....	14
1.3 RNA-seq.....	15
1.4 Сети ко-экспрессирующихся генов.....	17
2. CAGE.....	17
3. Биоинформатический анализ данных CAGE.....	19
4. Эмбриогенез амниот.....	24
5. Гипотермальная диапауза	29
МАТЕРИАЛЫ И МЕТОДЫ	32
1. Сбор проб, культивирование, электропорация и гибридизация in situ.....	32
2. Подготовка и секвенирование библиотек CAGE	33
3. Картирование, определение пиков CAGE и ассоциация с известными генами.....	34
4. Нормализация, анализ экспрессии и свойств промоторных областей.....	35
4.1 Анализ характеристик промоторной области	35
4.2 Анализ мотивов связывания с факторами транскрипции	37
5. Подготовка образцов в эксперименте с гибернацией.....	38
5.1 Биоинформатическая обработка чтений, полученных на платформе Illumina	39
ПРОМОТЕРОМ ЭМБРИОГЕНЕЗА АМНИОТ	40
1. Картирование TSS	42
2. Анализ главных компонент	44
3. Chicken-ZENBU: интерактивная платформа для визуализации TSS в геноме.....	46
4. TSS специфичные для этапов развития и типов клеток.....	49

5. Полный список TSS для генов «домашнего хозяйства»	57
6. Общие особенности TSS цыплят	61
7. Картирование TSS как инструмент для систематического анализа промоторов и энхансеров	63
8. Картирование TSS позволяет эффективно активировать гены с помощью технологии CRISPR-on	68
Выводы по главе «Промотором эмбриогенеза амниот»	77
ГИПОТЕРМАЛЬНАЯ ДИАПАУЗА	81
1. Профилирование промоторов, регулирующих гипотермальную диапаузу куриных эмбрионов	81
2. Транскрипционные факторы, регулирующие переход к гипотермальной диапаузе.....	86
Выводы по главе «Гипотермальная диапауза»	100
ЗАКЛЮЧЕНИЕ	103
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	105

СПИСОК СОКРАЩЕНИЙ

- ALDB-lincRNA (a domestic-animal long noncoding RNA database) – база данных длинных некодирующих РНК домашних животных;
- BP (biological process) – биологический процесс;
- CAGE (cap analysis gene expression) – кэп-анализ экспрессии генов;
- CC (cellular component) – клеточный компонент;
- CRISPR (clustered regularly interspaced short palindromic repeats) – короткие палиндромные повторы, регулярно расположенные группами;
- CTSS (CAGE tag starting site) – точка начала CAGE тега;
- DPI (decomposition based peak identification) – кластеризация на основе декомпозиции;
- EST (expressed sequence tag) – экспрессируемые метки последовательностей;
- FANTOM (Functional Annotation of the Mouse/Mammalian Genome) – функциональная аннотация геномов мыши/млекопитающих;
- FC (fold change) – кратность изменения экспрессии;
- FDR (false discovery rate) – средняя доля ложных отклонений гипотез;
- FPKM (fragments per kilobase per million mapped reads) – фрагментов на тысячу оснований на миллион;
- GO (gene ontology) – генная онтология;
- HCE и ASHCE (highly conserved elements и avian-specific HCE) – высококонсервативные элементы и специфичны для птиц HCE;
- HH (Hamburger Hamilton) – Гамбургер Гамильтон;

lncRNA (long non-coding RNA) – длинная некодирующая РНК;

MDS (multidimensional scaling) – многомерное шкалирование;

MF (molecular function) – молекулярная функция;

mRNA/rRNA, мРНК/рРНК – матричная/рибосомальная РНК;

MSC (mesenchymal stem cells) – мезенхимальные стволовые клетки

PAM (protospacer adjacent motif) – мотив, смежный с протоспейсером;

PCA/PCoA (principal component analysis) – метод главных компонент;

RASE (rapid amplification of cDNA ends) – быстрая амплификация концов кДНК;

RLE (Relative Log Expression) – относительная логарифмическая экспрессия;

RPKM (reads per kilobase per million of mapped reads) – число прочтений на тысячу нуклеотидов на миллион картированных прочтений;

SAGE (serial analysis of gene expression) – серийный анализ генной экспрессии;

sgRNA, гРНК (single guide RNA) – одноцепочечная гидовая РНК;

SI (shape index) – индекс формы;

SMC (Smooth muscle cells) – гладкомышечные клетки;

SPRI (solid phase reversible immobilization) – твердофазная обратимая иммобилизация на парамагнитных частицах;

TF, ТФ (transcription factor) – транскрипционный фактор;

TFBS, ССТФ (transcription factor binding site) – сайт связывания транскрипционного фактора;

TMM (trimmed mean of M-value) – нормализованное среднее М значений;

TOM (topological overlap matrix) – матрица топологическая перекрытий;

TPM (tags per million) – тэгов на миллион;

TSS (transcription start site) – точка инициации транскрипции;

WGCNA (weighted correlation network analysis) – анализ взвешенных сетей коэкспрессии генов;

кДНК – комплементарная ДНК.

пн — пары нуклеотидов (единица измерения длины участка ДНК)

ВВЕДЕНИЕ

Актуальность проблемы. После презентации первых сборок человеческого генома в 2001 году, большой интерес научного сообщества был нацелен на создание функциональной аннотации, объясняющей различные особенности человека на клеточном, организменном и популяционном уровне. Появились большие геномные проекты как ENCODE, Ensembl, RefSeq, VEGA, FANTOM в рамках которых генерировались свои собственные данные и создавались новые базы данных.

FANTOM (функциональная аннотация геномов млекопитающих) — это международный исследовательский консорциум, запущенный на базе института RIKEN (Йокогама, Япония, <http://fantom.gsc.riken.jp/5/>) в 2000 году для присвоения функциональных тегов к полноразмерным комплементарным ДНК, которые были собраны во время проекта Mouse Encyclopedia Project – одной из первых крупных геномных инициатив Японии. Проект FANTOM с тех пор развивается (за 18 лет было осуществлено пять проектов серии) и расширяется, охватывая все большие транскриптомные области анализа. Область исследования консорциума неуклонно продвигается по уровням в системе живого, развиваясь таким образом из понимания «элементов» - расшифровки транскриптов - к пониманию «системы» - транскрипционной регуляторной сети, другими словами, системы индивидуальной формы жизни.

Основным подходом к аннотации геномов в проекте FANTOM является метод кэп-анализа экспрессии генов (CAGE, cap analysis gene expression) и его модификации, которые в сочетании с секвенированием нового поколения на уровне отдельных молекул позволяют точно картировать точки инициации транскрипции (TSS), а также аккуратно оценить промоторную активность в дифференцированных, стабильных клеточных популяциях. Намного меньше данных о промоторной активности и структуре имеется для нестабильных,

подверженных дифференцировке клеток, включая их переходные состояния, появляющиеся в процессе онтогенеза позвоночных животных. Однако, эмбриональное развитие невозможно изучать напрямую в человеке по этическим соображениям, тогда как в случае модельных млекопитающих, например, мышей, исследователи сталкиваются с непреодолимыми сложностями в процессе изоляции материала на ранних этапах развития.

Для понимания этих процессов мы использовали эмбрионы курицы в качестве модели для изучения динамики транскрипционной активности для отдельных промоторов в процессе развития позвоночных и осуществили профилирование экспрессии на уровне отдельных точек инициации транскрипции, применяя метод CAGE на образцах, покрывающих весь трехнедельный период эмбрионального развития. Данный подход позволил оценить активность промоторов как кодирующих, так и некодирующих РНК, включая микро РНК и энхансерную РНК.

Далее мы применили данную технологию и полученную базу данных для изучения транскрипционных механизмов, связанных с температурной регуляцией гипометаболизма амниот. Способность эмбрионов амниот замедлять метаболизм при понижении температуры и восстанавливать его после возвращения нормальных условий инкубации (температурный торпор), известна достаточно давно и, являясь практически неизученной с точки зрения молекулярной биологии, в то же время широко используется в сельском хозяйстве на птицефермах для оптимизации логистики и синхронизации развития и выклева цыплят. Процесс развития эмбриона *Gallus gallus* на физиологическом уровне достаточно подробно изучен и существует несколько классификаций разных стадий. При этом, вопрос о молекулярных механизмах и регуляции процесса остановки метаболизма и дальнейшего возвращения к нормальному развитию остается открытым. Кроме того, в отличие от, например, рептилий, температурное окно нормальной динамики развития у птиц

достаточно узкое. Примечательно, что гипометаболизм при температурных флуктуациях происходит без участия нервной системы [1], так как наблюдается уже на самых ранних стадиях развития эмбриона амниот. Высказывались предположения в пользу эпигенетической регуляции [2], однако строгого подтверждения эта теория не получила. Таким образом, применение современных методов оценки транскрипционной активности генов и их промоторов, предлагаемых в данном проекте, имеет огромный потенциал в установлении закономерностей регуляции обратимой «эмбриональной диапаузы» опосредованной температурой окружающей среды. Кроме того, исследование этого вопроса имеет огромное практическое значение, так как существует проблема снижения выживаемости эмбрионов при инициации «куриной спячки» на фермах, особенно в случаях более длительного хранения на стадии диапаузы или понижении температуры на определенных этапах эмбриогенеза.

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. В сумме, 31863 TSS с уровнем экспрессии более 1 TPM (tag per million, тегов на миллион) были картированы на новую сборку генома (galGal5), среди которых от 34 до 46% были активны на отдельных этапах развития. Геномный браузер Zenbu был использован для 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. 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.

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 проклеывания с использованием CAGE метода профилирования экспрессии и впервые собрали полногеномную карту точек инициации транскрипции, играющих ключевую роль в дифференцировке от ранних стадий прегаструляции до проклеывания. 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.

ОБЗОР ЛИТЕРАТУРЫ

В последние годы быстро развивающиеся и все более доступные в применении технологии секвенирования следующего поколения (NGS) резко ускорили накопление геномных данных. Например, база данных Genomes OnLine (GOLD) содержит более 13.5К секвенированных эукариотических геномов (<https://gold.jgi.doe.gov>) и проект «Геном 10К» уже собрал около 280 геномов позвоночных с конечной целью - 10 000 [3, 4]. Кроме создания базы данных как таковой, разработчики проекта рассчитывают, что такие геномные данные позволят разрешить вопросы эволюции позвоночных, а также генетические основы их многочисленных адаптаций.

1. Аннотация генома

Геномная сборка сама по себе обладает общими специфичными для отдельных видов особенностями, как например длина, состав ГЦ пар, наличие повторов, полиморфизмов, вставок и других. Может быть проведена дальнейшая аннотация кодирующих генов на основе гомологии с последовательностями, представленными в базах данных и оценена их активность в эксперименте в пределах этой группы идентифицированных генов, например, на микрочипах. Это ограниченный и устаревший подход непригодный для поиска новых или переменных генов, как длинные некодирующие РНК, роль которых в таком дизайне эксперимента также не будет определена. Были разработаны новые подходы для обнаружения сигналов РНК позволяющие создавать полноценный транскриптом, такие как EST (expressed sequence tag), SAGE (последовательный анализ экспрессии генов), RNAseq, CAGE, основанный на последовательности коротких 5' концевых транскриптов, DGE для точной оценки экспрессии на основе тегов и RACE (быстрая амплификация концов кДНК), предназначенные для

идентификации полноразмерных транскриптов РНК, что позволяет исследователям оценить экспрессию генов в масштабе всего генома, что также важно для понимания принципов генной регуляции [5–10].

1.1 EST

Исследования отдельных генов и их транскриптов велись задолго до появления высокопроизводительных методов секвенирования. С середины 70-х техника клонирования кДНК соответствующей отдельной мРНК начала быстро распространяться, была показана возможность сборки полной библиотеки клонов кДНК, включая даже случаи с низкой экспрессией или гомологией. Вскоре была собрана библиотека кДНК для хориона шелкопряда путем выделения мРНК и использованием обратной транскриптазы. Изначально для прочтения библиотек из случайных транскриптов использовали низкопроизводительный метод Сэнгера, в результате чего получали так называемые теги экспрессирующихся последовательностей (EST). EST оказались доступным и эффективным инструментом для обнаружения генов в различных масштабах. Было начато много небольших проектов, направленных на обнаружение генов, активных на ключевых этапах жизненного цикла. Например, для паразитов вроде клещей *Psoroptes ovis* или ленточного червя *Echinococcus granulosus* генерация всего нескольких сотен EST позволила обнаружить ряд антигенов для дальнейших детальных исследований. EST полученные из растений таких как арахис *Arachis hypogaea* или *Thellungiella salsuginea* позволили охарактеризовать гены устойчивости и стрессового ответа, что может быть использовано для улучшения программ выращивания. EST библиотеки для устриц (*Crassostrea virginica*) и земляных червей (*Lumbricus rubellus*) были использованы для поиска генов, которые могли бы быть использованы в качестве маркеров загрязненности. В проектах большего масштаба, когда собираются библиотеки для нескольких видов, например,

внутри *Apicomplexa*, *Coleoptera*, *Nematoda*, появляется возможность изучать консервативные и вариабельные гены. Например, у нематоды более половины собранных контигов из 265 тысяч EST были уникальны для типа, из которых 1262 были высоко консервативны для всех видов, что делает их хорошими мишенями для борьбы с нематоцидами. Такие проекты более экономичны по сравнению с полногеномным секвенированием, но также представляют собой хорошее дополнение к геномным проектам, где используются для подтверждения аннотации генов, предсказанных с помощью соответствующих алгоритмов, верификации границ экзонов и интронов. Количество индивидуальных транскриптов при таком подходе может определяться с помощью нозерн-блота или количественной ПЦР, которые являются трудоемкими и покрывают лишь часть транскриптома.

1.2 SAGE

Первым настоящим транскриптомным методом стал SAGE (serial analysis of gene expression), в основе которого лежит секвенирование методом Сэнгера сцепленных случайных транскриптов. Транскрипты далее подсчитывались при сопоставлении с известными генами. Метод оказался более продуктивным и чувствительным по сравнению с EST за счет сокращения длины чтений с 100-300 до 9-22 п.о. и перехода от концепции секвенирования индивидуальных транскриптов индивидуальных клонов к множественным транскриптам с индивидуальным маркером и клоном. SAGE протокол дешевле EST и дороже микрочипов, но может быть использован для определения экспрессии новых неаннотированных транскриптов, а также не ограничен уровнем флюоресценции, что делает его более аккуратным в количественном анализе. При этом данный метод сфокусирован на полиаденилированных мРНК, так как используются биотинилированные поли-

T-праймеры, за которые далее их связывают стрептавидиновые магнитные частицы. Далее DGE (digital gene expression), улучшенная версия SAGE, начала набирать популярность до тех пор, пока длина чтений не достигла 25 п.о., что позволило уникально картировать случайно фрагментированные кДНК чтения на геномы млекопитающих.

1.3 RNA-seq

Самым популярным транскриптомным методом стал RNA-seq на платформе Illumina (Genome Analyzer, Hi-Seq) и количество исследований с его использованием продолжает расти. Изначально это были короткие чтения 20-40 п.о., из-за чего возникала проблема со сдвигом при ПЦР-амплификации, аккуратностью картирования. Появились платформы и соответствующие методы преодоления этой проблемы, например, одномолекулярный секвенатор Helicos, и далее секвенаторы третьего поколения PacBio и IonTorrent, которые так и не смогли найти широкого применения из-за развития цепь-специфичных парноконцевых методов приготовления библиотек для дальнейшего секвенирования на платформах Illumina. Несмотря на решение проблемы с амплификацией в применении данного метода заложены другие проблемы, связанные с биоинформатической частью. Во время распространения метода микрочипов основными проблемами в аналитической части были вопросы связанные с уменьшением размерности различными алгоритмами (PCA, MDS, LTSA, LLE, ISOMAP, нейронными сетями RDP, BNN), с гибридизацией в различных образцах, с выбором стратегий нормализации и дальнейшего анализа дифференциальной экспрессии. Использование секвенаторов следующего поколения решило некоторые из этих вопросов, но появились новые, такие как выравнивание миллионов или десятков миллионов чтений на геном, неправильное выравнивание чтений похожих транскриптов, что может

приводить к “заглушению” генов, проблема с нормализацией осталась, но стала значительно проще. Со времен появления SAGE разработана метрика нормализации, решающая проблему разной глубины чтения библиотек в эксперименте - TPM (tags/transcripts per million, тегов/транскриптов на миллион), которой вполне достаточно в большинстве случаев. Далее появились новые метрики, позволяющие сравнивать экспрессию между образцами: RPKM, FPKM, UQUA, TMM. Искусственный сдвиг в количестве чтений, так называемый “batch” эффект, возникающий при сравнении образцов с разных запусков секвенатора может быть пересчитан при использовании соответствующих стратегий индексирования/мультиплексирования. Как правило технические повторности RNA-seq настолько похожи, что могут быть просто суммированы для дальнейшего анализа. Свойства чтений, получаемых с помощью секвенаторов нового поколения, потребовали развития новых алгоритмов для аккуратного анализа дифференциальной экспрессии. Например, во многих ранних статьях с RNA-seq часто использовалась модель Пуассона для определения генов с дифференциальной экспрессией, которая, как оказалось, не подходит для такой задачи. Наиболее часто применяемые сейчас подходы, относящиеся к параметрическим методам и использующие различные варианты отрицательного биномиального распределения: edgeR, HTseq, bayseq и NBPseq. Непараметрические методы такие как NOISeq и Samseq. В некоторых случаях точный тест Фишера может давать хорошие результаты сравнения. В данный момент нет универсального метода анализа.

В методе RNA-seq возникает проблема с выравниванием, сборкой и аннотацией транскриптов из-за альтернативного сплайсинга, который особенно характерен для различных тканей. Такая проблема разрешается при использовании длинных парноконцевых цепь-специфичных чтений. Также сейчас доступны специальные методы, позволяющие определить особенности сплайсинга и экспрессии экзонов, такие как DEXSeq и TopHat.

1.4 Сети ко-экспрессирующихся генов

Коэкспрессионные сети формируются на основе сравнения профилей экспрессии отдельных генов, их транскриптов или различных изоформ, в различных тканях или экспериментальных условиях. Цель такого подхода заключается в поиске коррелированных регуляторных модулей, что позволяет показать участие их элементов в определенных процессах, метаболических или сигнальных путях как в растениях, так и в млекопитающих. В данном случае проявляется еще одно из преимуществ метода RNA-seq по сравнению с микрочипами, так как такой анализ дает более надежные результаты в случае с наиболее полным транскриптомом. Также различное поведение изоформ гена может свидетельствовать об их различных функциональных особенностях. Для проведения анализа ко-экспрессии был разработан соответствующий пакет доступный внутри среды R - WGCNA (weighted gene co-expression network analysis) [11], который также позволяет связать полученные модули с определенными физическими или физиологическими особенностями исследуемого объекта.

Поскольку использование высокопроизводительных технологий секвенирования стали более доступными и доступными, многие исследовательские группы начали свои собственные геномные проекты для конкретных видов или в пределах целых классов [12] на подобие ENCODE [13], объединяя разные омиксные данные для аннотации генома и определения его функциональных элементов. Для таких целей, требуются соответствующие глобальные полногеномные решения в биоинформатике для каждого транскриптомного метода в отдельности с учетом их особенностей.

2. CAGE

Подобно популярному RNAseq, технология CAGE применима для сравнения экспрессии РНК в разных образцах, включая ткани и клеточные культуры в нормальных условиях или при внешней стимуляции, или в течение времени [7, 14, 15]. Среди транскрипционных методов, CAGE — это мощный подход, позволяющий идентифицировать сайты инициации транскрипции (TSS); например, он использовался для создания промоторного атласа млекопитающих на основе многочисленных человеческих и мышинных образцов [7]. CAGE также применялся в других проектах для улучшения аннотации генома дрозофилы (*D. melanogaster*) [16], рыбки Данио (*D. rerio*) [17, 18] и цыпленка (*G. gallus*) [19]. CAGE не только указывает позиции TSS с разрешением в 1 п.о. но также позволяет оценить уровень экспрессии промотора, и поэтому он может быть использован для определения особенностей регуляции генов, так как в промоторной области есть возможность идентифицировать специфические сайты связывания транскрипционных факторов (TF) [20].

Протоколы подготовки библиотеки CAGE продолжают улучшаться [20–23] и теперь также доступны коммерческие комплекты (https://cage-seq.com/cage_kit/index.html). Стандартный протокол для подготовки библиотеки CAGE применим к образцам, из которых можно извлечь достаточное количество РНК (1-5 мкг) [20, 21]. Оригинальный протокол подготовки библиотеки CAGE состоит из нескольких ферментативных реакции, ПЦР-амплификации и ключевой этап захват экзистированной РНК. Эта стадия включает окисление РНК, дальнейшее биотинилирование окисленных концов и захват на покрытых стрептавидином магнитных шариках [24].

Для технологии nanoCAGE [23], которая является вариацией CAGE для обработки образцов с небольшим количеством РНК, данный этап заменен на «замену матрицы» (template switching, Рисунок 1), что подразумевает использование специфических адаптеров с гуанозином, которые лигируют

только с молекулами экзистированных РНК, и тем самым позволяет получать библиотеку из небольших количеств исходного материала (около 1000 клеток). Однако такой протокол включает этап ПЦР амплификации, что может приводить к известному эффекту смещения при подсчете уровня экспрессии.

Оригинальный протокол CAGE [24] был адаптирован для секвенирования на платформе HeliScore, которая была в своем роде революционной, так как позволяла секвенировать отдельные молекулы исключая этап амплификации при подготовке ДНК образцов (Рисунок 2). Секвенатор поочередно читает отдельные молекулы ДНК, позволяя направленно и аккуратно определять соответствующие последовательности. В статье Kanamori-Katayama было подтверждено, что такой подход снижает сдвиг между образцами и генерирует высокопроизводимые данные при количестве тотальной РНК от 5 микрограмм до всего 100 нанограмм [20, 25].

CAGE метод был также адаптирован для других популярных секвенаторов нового поколения, таких как Illumina, 454, SOLiD. В настоящее время широко используется автоматизированный стандартный протокол приготовления библиотек для дальнейшего секвенирования на платформе Illumina (HiSeq2500), основанный на твердофазной обратимой иммобилизации (SPRI) и AMPure [21]. Такой протокол также не использует ПЦР и пропускает этап отрезания 5'-концевых участков последовательностей nAnTi-CAGE [26].

3. Биоинформатический анализ данных CAGE

Полученные чтения CAGE 27-70 п.о. демультимплексируются согласно заданным индексам на адаптерах, индексы удаляются, чтения выравниваются по фиксированной длине, например, 50 п.о. Далее удаляются последовательности с низким качеством прочтения (порог Q33) и содержащие неопределенные нуклеотиды - N. Эти стандартные этапы обработки сырых

данных, а также удаление последовательностей рибосомальной РНК из библиотеки, с дальнейшим картированием и генерированием общих результатов выравнивания может проводиться в среде Moirai, специально разработанной в рамках проекта FANTOM. Другие варианты также возможны, например, использование комбинации bwa и delve, или hisat2. Как правило, достаточно только bwa с дальнейшей TPM нормализацией и кластеризацией DPI. Как и другие транскрипционные методы результаты картирования CAGE библиотек представлены в универсальных форматах SAM/BAM, которые можно визуализировать в геномном браузере, например, IGV, UCSC или специально разработанном браузере Zenbu.



Рисунок 1. Дизайн адаптеров при подготовке nanoCAGE библиотек, этап лигирования «template switching» [23].

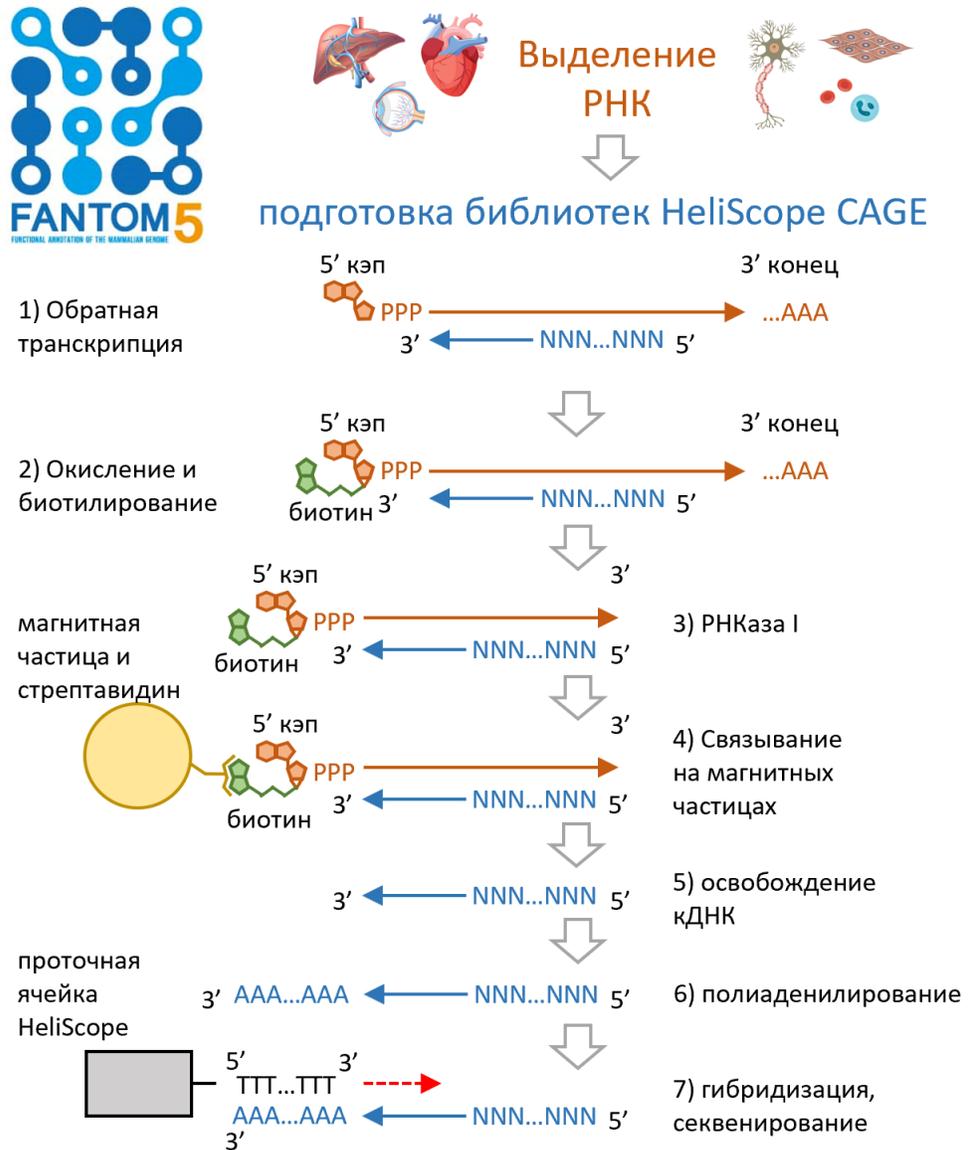


Рисунок 2. Основные этапы подготовки библиотек CAGE в проекте FANTOM5 [25]. 1) кДНК синтезируется с помощью набора SuperScript III и случайных праймеров N15; 2) Кэп окисляется пероксидом натрия и биотинилируется (длинной цепью) биотин гидролазой; 3) одноцепочечная РНК расщепляется РНКазой I; 4) захват биотинилированных РНК/кДНК гибридных молекул на магнитных стрептавидиновых частицах; 5) отмывка РНК/кДНК гибридов не связавшихся с частицами и освобождение одноцепочечной кДНК при помощи РНКазы H, РНКазы I и нагревания; 6) полиаденилирование с помощью терминальной дезоксирибонуклеотидил трансферазы и дАТФ, которое останавливается биотин-ддАТФ; 7) при загрузке кДНК в ячейках HeliScore она связывается с поли(50)-Т подложкой, добавляется ДНК полимеразы, дТТФ и терминатор.

Менее ресурсоемкий формат хранения CAGE данных - в виде CTSS или OSC таблиц, в которых для каждой позиции в геноме подсчитано количество картированных чтений CAGE по 5' концу. Весь последующий статистический анализ не имеет особых отличий от других транскриптомных методов, за исключением этапа ассоциации известных аннотированных транскриптов с полученными кластерами CAGE, который может меняться в зависимости от вида или задачи.

Комбинация технологий секвенирования нового поколения отдельных молекул с транскриптомным методом CAGE, позволяет провести картирование и количественную оценку активности точек инициации транскрипции (TSS) с разрешением в одну пару оснований [25, 27]. Применение такой технологии на большом наборе образцов человека и мыши, покрывающем почти все известные стабильные типы клеток, позволило провести всеобъемлющий обзор использования промоторов млекопитающих, а также собрать сложные сети взаимодействия промоторов, их генов и транскрипционных факторов, специфичных для отдельных типов клеток [7, 14, 15]. Комбинация такого высокопроизводительного полногеномного профилирования TSS позвоночных, беспозвоночных животных и грибов позволило по-новому взглянуть на структуру эукариотических промоторов и их эволюцию [28].

Высокая чувствительность метода CAGE открывает новые возможности оценки транскрипционной активности, как в случае с энхансерной РНК или микро РНК. Двухнаправленная сбалансированная экспрессия кэпированных транскриптов из данных CAGE является более аккуратным маркером активных энхансеров по сравнению с методом ChIPseq. Кроме того, данные по транскрипционной активности энхансеров и промоторов позволяют находить соответствующие регуляторные связи с большей точностью чем данные по организации хроматина. Была показана связь GWAS полиморфизмов с

промоторными и энхансерными областями предсказанными по данным CAGE и их экспрессией [14].

Другой особенностью данных CAGE является возможность оценки экспрессии различных типов некодирующей РНК, в том числе микро РНК, которая принимает участие в регуляции эмбрионального развития, дифференцировки, иммунного ответа, размножения и продолжительности жизни. Нарушения в экспрессии микро РНК могут приводить к заболеваниям, в том числе раковым. Поэтому профилирование их экспрессии имеет большое значение с точки зрения понимания функциональной роли отдельных микро РНК. Так как большинство первичной микроРНК (pri-микроРНК) транскрибируются РНК полимеразой II, то имеют кэп на 5' конце и могут быть анализированы с помощью CAGE. Нами было показано, что CAGE не только показывает точку инициации транскрипции микро РНК, но и точку рестрикции Drosha на 3' конце пре-микроРНК, что может быть использовано для валидации и классификации микро РНК [29].

Понимание онтогенеза тканей и клеток взрослого организма требует последовательного анализа в ходе дифференцировки и модификациях эмбриональных и зародышевых тканей. Однако, из-за этических проблем, такой анализ невозможно осуществить на эмбрионах человека, но из-за технических это также практически невыполнимо на ранних этапах развития мыши. Среди 975 образцов человека, представленных в проекте FANTOM5 основанном в значительной части на CAGE методе профилирования TSS нет образцов эмбрионов. Среди образцов мыши (всего 399) примерно одна четвертая относится к эмбриональному развитию, но все пробы получены не ранее стадии E11, то есть с середины развития [7]. Такие ограничения могут быть преодолены с использованием модельных организмов, не относящихся к млекопитающим и чьи эмбриональные образцы проще получить. В недавнем исследовании с использованием рыбки *Danio rerio* в качестве модели был

проанализирован ранний эмбриогенез - процесс перехода использования материнской РНК к собственной зиготической с использованием CAGE профилирования TSS, что позволило показать наличие двух принципиально разных механизмов регуляции на уровне промоторных областей, которые определяют материнскую и зиготическую транскрипцию [17, 18].

Среди модельных организмов с хорошо собранным и аннотированным геномом, подходящих для изучения эмбрионального развития можно выделить домашнюю курицу *Gallus gallus* [30, 31]. И птицы, и человек относятся к позвоночным амниотам. Их развитие эволюционно консервативно и происходит по универсальным шаблонам [32]. Среди всех амниотических моделей, кроме млекопитающих, домашняя курица является лучшей экспериментальной системой для изучения развития эмбриона человека. Около 60% из 22000 белок кодирующих генов имеют ортолога в геноме курицы [30, 33, 34]. Программа развития, ведущая к образованию трех эмбриональных слоев и их последующая специализация крайне консервативны в обоих видах [31]. Важно отметить, что один из наиболее критических периодов развития – гастрюляция и стадии после нее происходят после имплантации, что затрудняет получение материала из млекопитающих, но находится в относительно легком доступе в эмбрионах курицы.

4. Эмбриогенез амниот

Амниоты включают завропсид *Sauropsida* (класс пресмыкающиеся *Reptilia* и птицы *Aves*) и синапсидообразных *Synapsidomorpha* (класс синапсид *Synapsida* и млекопитающих *Mammalia*). Амниоты в отличие от анамний имеют внутреннее оплодотворение и их размножение происходит на суше за редким исключением. Зоологи частично характеризуют амниот по их эмбриональному развитию, которое включает формирование нескольких зародышевых мембран -

амнион, хорион (сероза) и аллантаис. У млекопитающих хорион представлен в виде зародышевой части плаценты. Бластомеры плацентарных млекопитающих (эутериев) на стадии бластоцисты дифференцируются на два слоя - трофобласт и эмбриобласт. При переходе зародыша в матку он освобождается от яйцевой оболочки и прикрепляется к стенке матки с помощью трофобласта - имплантация. На этом этапе эмбрион врастает ворсинками в эпителий матки, что сильно затрудняет получение материала. Яйцевая оболочка яйцекладущих амниот заменяется на скорлупу, свободно пропускающей воздух и позволяющей амниотам откладывать относительно большие яйца. Большое содержание желтка в таких полилецитальных яйцах позволяет развиваться напрямую без переходных стадий. Зародышевые оболочки формируют амниотическую полость, которая заполнена жидкостью с относительно постоянными условиями среды, что позволяет эмбриону развиваться на суше. Из стенки задней кишки возникает аллантаис, который собирает жидкие продукты обмена. Амниоты, как правило, развиваются непосредственно в конечную наземную форму с конечностями и толстым многослойным эпителием, исключая кормовую стадию головастика с последующим метаморфозом, как в случае амфибий [35]. У амниот переход от двуслойной перидермы к стратифицированному плоскоклеточному эпителию запускается гормоном щитовидной железы во время эмбрионального развития, а не метаморфозом [36]. Такие характерные черты амниот отражают их специализацию к размножению в условиях суши.

Куриные эмбрионы являются удобной и легкодоступной моделью для изучения эмбрионального развития. Было несколько попыток создания атласа эмбрионального развития цыплят, но наиболее часто цитируемой стала работа Гамбургера и Гамильтона [32], где они описали 46 визуально различимых этапов от эмбрионального диска до проклевывания (Таблица 1).

Эмбриональный диск — это структура, которая образуется в бластуле на ранних стадиях эмбрионального развития птиц, рептилий и млекопитающих. Большая часть данных о диске получена для эмбриона курицы. Он образуется на спинной (задней) поверхности развивающегося эмбриона, к каудальному или заднему концу. Эмбриональный диск устанавливает билатеральную симметрию, определяет место гаструляции и инициирует образование зародышевого слоя [37]. Для формирования диска рептилии, птицы и млекопитающие распределяют мезенхимальные клетки вдоль предполагаемой центральной оси, определяется как клетки будут мигрировать в процессе гаструляции. Формирование диска происходит из эпибласта и при участии нижележащего гипобласта, а также различных секреторных белков Vg1, Nodal, Wnt8C, FGF8 и хордин (Chordin), транскрипционных факторов Brachyury и Goosecoid [38]. Регуляция процесса «выбора клеткой своей судьбы» (cell fate decision) на молекулярном уровне остается плохо изученной и остается одним из основных вопросов биологии развития на данный момент. Известно несколько ключевых сигнальных белков морфогенов, такие как Wnt, Shh (sonic hedgehog) или BMP (костный морфогенетический белок). В зависимости от концентрации морфогенов в клетках определяются наборы транскрипционных факторов, регулирующих экспрессию всего множества остальных генов, что далее влияет на судьбу клетки [39]. Например, градиент концентрации BMP и Shh распределяется в дорзо-вентральном направлении для формирования нескольких отделов нервной трубки, из которых будут далее формироваться нейроны разных типов. Среди ТФ можно отметить Pax6, Lhx2, Emx2, Emx1. Судьба нейрогенных стволовых клеток также зависит от цитокинов семейства IL-6 (цилиарный нейротрофический фактор CNTF, лейкомиический ингибиторный фактор LIF, кардиотропин-1 CT-1) [40].

Таблица 1. Стадии развития куриного эмбриона по Гамбургеру и Гамильтону.

НН стадии	Возраст эмбриона	Особенности строения
1	-	Эмбриональный диск
2	6-7 ч	Начальный эмбриональный диск, длиной 0.3-0.5 мм
3	12-13 ч	Средний эмбриональный диск
4	18-19 ч	Окончательный эмбриональный диск, длиной ± 1.88 мм
5	19-22 ч	Формирование головы (хорды)
6	23-25 ч	Сгибание головы
7	23-26 ч	1 сомит; нейронные складки
7-8	23-26 ч	1-3 сомита; целом
8	26-29 ч	4 сомита; кровяные островки
9	29-33 ч	7 сомитов; первичные везикулы
9-10	33 ч	8-9 сомитов; первичная амниотическая складка
10	33-38 ч	10 сомитов; 3 первичных мозговых везикулы
11	40-45 ч	13 сомитов; 5 нейромеров заднего мозга
12	45-49 ч	16 сомитов; мозговой пузырь
13	48-52 ч	19 сомитов; атриовентрикулярный канал
13-14	50-52 ч	20-21 сомит; хвостовая почка
14	50-53 ч	22 сомита; изгиб ствола; висцеральные дуги I и II, щели 1 и 2
14-15	50-54 ч	23 сомита; предпандибулярные полости головы
15	50-55 ч	24-27 сомитов; висцеральная арка III, щель 3
16	51-56 ч	26-28 сомитов; почка крыла; задняя амниотическая складка
17	52-64 ч	29-32 сомита; почка ноги; эпифиз
18	3 д	30-36 сомитов, простирающиеся за пределы уровня ножных почек; аллантоис
19	3.0-3.5 д	37-40 сомитов, переходящие в хвост; верхнечелюстной процесс
20	3.0-3.5 д	40-43 сомитов; завершение вращения; глазной пигмент
21	3.5 д	43-44 сомита; висцеральная арка IV, расщелина 4
22	3.5-4.0 д	Сомиты простираются до кончика хвоста
23	4 д	Спинной контур от заднего мозга до хвоста - изогнутая линия

24	4.5 д	Пальцевая пластинка
25	4.5-5.0 д	Локоть и коленные суставы
26	5 д	Первые 3 пальца задних конечностей
27	5.0-5.5 д	Клюв
28	5.5-6.0 д	3 пальца передних конечностей, 4 пальца на задних
29	6.0-6.5 д	Рудимент пятого пальца
30	6.5-7.0 д	Зачатки оперения; склеральные сосочки; яйцевой зуб
31	7.0-7.5 д	Перепонка между 1-й и 2-й пальцами передних конечностей
32	7.5 д	Передний конец нижней челюсти достиг клюва
33	7.5-8.0 д	Перепонка на краю крыла и первого пальца
34	8 д	Мигательная перепонка
35	8.5-9.0 д	Фаланги пальцев задних конечностей
36	10 д	Длина 3-го пальца от кончика до середины плюсневых суставов = $5,4 \pm 0,3$ мм; длина клюва от ноздри до кончика = $2,5$ мм; зачаток гребешка; губная канавка; копчиковая железа
37	11 д	Длина третьего пальца ноги - $7,4 \pm 0,3$ мм; длина клюва = $3,0$ мм
38	12 д	Длина третьего пальца ноги - $8,4 \pm 0,3$ мм; длина клюва = $3,1$ мм
39	13 д	Длина третьего пальца ноги - $9,8 \pm 0,3$ мм; длина клюва = $3,5$ мм
40	14 д	Длина клюва = $4,0$ мм; длина третьего пальца ноги = $12,7 \pm 0,5$ мм
41	15 д	Длина клюва от ноздри до кончика = $4,5$ мм; длина третьего пальца ноги = $14,9 \pm 0,8$ мм
42	16 д	Длина клюва = $4,8$ мм; длина третьего пальца ноги = $16,7 \pm 0,8$ мм
43	17 д	Длина клюва = $5,0$ мм; длина третьего пальца ноги = $18,6 \pm 0,8$ мм
44	18 д	Длина клюва = $5,7$ мм; длина третьего пальца - $20,4 \pm 0,8$ мм
45	19-20 д	Желточный мешок, закрытый в полости тела; хорио-аллантоидная мембрана содержит меньше крови и «липкая» в живом эмбрионе
46	20-21 д	Проклевывание

5. Гипотермальная диапауза

Кроме изучения развития как такового, куриные эмбрионы интересны с точки зрения их адаптации к пониженной температуре окружающей среды через переход в состояние гипотермического анабиоза. Среди яйцекладущих видов рептилий большинство проходят через стадию эмбрионального ареста как минимум один раз до откладки яиц, но бывают случаи с облигатной диапаузой уже после откладки перед формированием серозной оболочки. Для яиц рептилий, остающихся без родительского присмотра, чтобы выжить в случае похолодания критически важно осуществлять регуляцию своего метаболизма в зависимости от условий окружающей среды. В таком состоянии гибернации яйца рептилий могут находиться несколько недель [41]. Яйца птиц также сохранили способность впадать в краткосрочный (несколько дней) анабиоз, на случай если родительская особь отправилась на поиски пищи или не может обеспечить должный уход по каким-то другим причинам. При этом такая задержка развития принципиально отличается от эмбриональной диапаузы позвоночных и задержкой имплантации млекопитающих, которая происходит один раз на стадии бластоцисты. В случае куриных эмбрионов она напрямую регулируется температурой окружающей среды, тогда как в случае млекопитающих и других позвоночных диапауза может регулироваться внешними факторами как световой день, количество доступной еды и температурой, но посредством материнского гормонального контроля [42]. Например, ящерицы могут регулировать диапаузу прогестероном, секретлируемым надпочечниками или еще неизученными механизмами как в случае *Calotes versicolor* до откладки яиц [43].

Состояние гибернации, другими словами спячка или торпор, характерно для гетеротермных животных, которые имеют постоянную температуру в

активном состоянии и непостоянную при в состоянии покоя. Ярким примером может служить зимняя спячка у сонь полчков *Glis glis* и ежедневный торпор у фруктовых летучих мышей *M. minimus* [44]. В короткий ночной торпор могут впадать некоторые виды птиц. При этом, откладываемые птицами яйца по своей физиологии, как правило, схожи с пойкилотермными животными, так как энергии от работы сердца, циркуляции крови и метаболизма в целом не достаточно для поддержания температуры яйца, даже при снижении температуры окружающей среды всего на 2°C, включая поздние стадии перед проклевыванием [45]. Таким образом, отклонения температуры окружающей среды от нормы оказывают стрессовое воздействие на эмбрионы, нарушая температуру их тела, скорость метаболизма, рН крови и сердечный выброс [46]. Несмотря на наличие буферной амниотической жидкости, окружающей эмбрион на более поздних стадиях развития, температура эмбриона в ответ на изменение температуры окружающей среды меняется вразрез с правилом Вант-Гоффа, что указывает на наличие слаженных регуляторных механизмов [47]. В отличие от взрослых особей такая регуляция не может осуществляться посредством нервной системы, так как ее полная активация происходит непосредственно перед проклевыванием [1]. В настоящее время такой регуляторный путь остается неизученным, но, по некоторым предположениям, происходит на эпигенетическом уровне и включает работу определенных генов [2].

Адаптация к изменению температуры инкубации связана с тем, что в природе яйца могут оставаться без нагрева, когда птица уходит за кормом. Короткого периода охлаждения (с 38°C до 30°C, 2 ч) достаточно для резкого снижения метаболизма [48], при более длительном воздействии замедляется частота сердечных сокращений вплоть до полной остановки сердца [49]. Было сделано наблюдение, что если инкубировать яйца домашней курицы при 8°C до остановки сердца, то большинство эмбрионов смогут восстановить его ритм при

дальнейшем прогревании до 38°C [49]. При этом эмбрионы на ранних стадиях развития успешнее переживали остановку сердца, чем на поздних [50]. Авторы также отметили, что при низкой температуре и замедленном метаболизме простой диффузии O₂ или его запасов в эмбриональных жидкостях и крови должно быть достаточно для обеспечения тканей на некоторое время. Домашние курицы обычно не оставляют яйца на несколько часов, поэтому такая толерантность к низкой температуре является необычной и примечательной. Для полной характеристики генов, вовлеченных в эмбриогенез у птиц, был инициирован проект по профилированию экспрессии генов на разных стадиях развития – eChickAtlas [51]. Однако, в этом проекте нет задачи оценки экспрессии на стадии остановки метаболизма. Работы по изучению диапаузы куриных эмбрионов в основном носили чисто практический характер, главной целью которых было установить оптимальное время индукции спячки без последующей потери выживаемости, а значит экономических потерь [50]. На сегодняшний день актуальной является гипотеза, что одним из механизмов поддержания эмбриона в состоянии диапаузы является предотвращение апоптоза. Присутствуют фрагментарные работы, указывающие на механизм репрессии апоптоза белками теплового шока: была показана роль белка Hps70 в ингибировании апоптоза [52]. Недавно Hwang с соавторами продемонстрировали, что экспрессия гена hsp25 значительно повышается в бластодерме во время хранения эмбриона при пониженной температуре. Эксперименты по инактивации hsp25 показали, что функция этого гена связана с антиапоптотическим, антиоксидантным и про-аутофагическим эффектами в бластодерме зародыша [53]. Системных исследований включающих полногеномные подходы в изучении генетической экспрессии холодной диапаузы у куриных эмбрионов и рептилий не опубликовано.

МАТЕРИАЛЫ И МЕТОДЫ

В данной работе не используются ткани человека, млекопитающих или эмбрионов. Все эксперименты с использованием эмбрионов курицы в этой работе проводились в соответствии с принципами проведения экспериментов на животных КФУ, Центра биологии развития RIKEN и Университета Кумамото.

1. Сбор проб, культивирование, электропорация и гибридизация *in situ*

Оплодотворенные куриные яйца были получены с фермы Широама (Канагава, Япония). Яйца инкубировали при 38.5° С до желаемых стадий. Для образцов НН1-НН26 (от 1,5 до 5 дней) РНК выделяли из эмбриональных и внеэмбриональных областей (НН1, НН2, НН3, НН4, НН5, НН7, НН11, НН14, НН19, НН23, НН26) китом miRNeasy (QIAGEN). Для образцов НН32, НН37, НН41 и НН45 (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 НН32 и НН41. Экстраэмбриональный образец НН41 имел относительно низкое количество чтений и не был включен в РСoA. Образцы почек и задних конечностей были собраны из яиц (НН20, день 3), купленных у Granja Gibert (Камбрильс, Испания). Яйца инкубировали до стадии НН20. 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, неизвестный штамм, гепатоциты: стандартная подготовка, штамм Cornish cross, мезенхимальные стволовые клетки: стандартная подготовка, штамм Rhode Island red). 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 [16, 57]. В электропорированных эмбрионах определяли активность GFP, а после достижения HH10 эмбрионы фиксировали и обрабатывали для гибридизации РНК *in situ* как описано ранее [16]. Для валидации экспрессии генов, связанных с зачатками конечностей, гибридизация РНК *in situ* выполнялась на всех стадиях с использованием меченых дигоксигенином антисмысловых РНК [54]. В качестве маркеров использовались PRRX1, VEGFD, LHX9, HOXA11, CRABP1 и HOXD10, любезно предоставленные А. Nieto, М. Davey и J. Fallon или сгенерированные с помощью ПЦР в лаборатории Мартина Роса из Института Биомедицины и Биотехнологии Кантабрии (Сантандер, Испания).

2. Подготовка и секвенирование библиотек CAGE

Библиотеки CAGE были подготовлены по стандартному протоколу с использованием 5 мкг общей РНК в качестве исходного материала [20, 21]. Вкратце, тотальная РНК сначала использовалась в качестве матрицы для синтеза первой цепи кДНК, которая затем биотинилировалась на 5'-конце, что

позволяет провести ее селективный отбор при помощи стрептавидиновых частиц. РНК, которая не содержит 5'-кэп (например, рРНК) элиминируется на этом этапе. Далее последовательно пришиваются линкеры на 5' и 3' концы. 5'-линкер содержит сайт узнавания эндонуклеазы рестрикции XmaI и сайт узнавания рестриктазы класса II MmeI, а также баркод для мультиплексирования. 3'-линкер содержит сайт узнавания XbaI рестриктазой. Обработка указанными рестриктазами позволяет получить короткие CAGE-теги, к которым лигируется сиквенсовый праймер. На финальном этапе синтезировалась вторая цепь кДНК, качество библиотек проверяли с помощью Agilent Bioanalyzer 2100. Затем библиотеки были секвенированы на одномолекулярном секвенаторе HeliScore, следуя указаниям производителя [55]. Другая часть библиотек (эксперимент по гибернации) была секвенирована на платформе HiSeq 2500, Illumina одноконцевые чтения, 50 п.о., 8 образцов на дорожку (Таблица 2).

3. Картирование, определение пиков CAGE и ассоциация с известными генами

Полученные короткие CAGE-теги картировали на геном с последующей идентификации сайтов инициации транскрипции (TSS), оценивали активность промоторов, проводили поиск мотивов вовлеченных в регуляцию элементов, а также проводили анализ дифференциальной экспрессии генов.

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. 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, как описано ранее [7]. 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 и RefSeq. Набор данных птичьего набора CAGE доступен для загрузки в DDBJ (DNA Data Bank of Japan) с регистрационным номером DRA004812.

4. Нормализация, анализ экспрессии и свойств промоторных областей

Нормализация экспрессии была сделана для каждого образца отдельно в виде TPM. Дополнительный этап нормализации выполнялся в среде R bioconductor edgeR [56], по протоколу «RLE» (относительная логарифмическая экспрессия). Для поиска тканеспецифичных и специфичных для отдельных этапов развития промоторов был рассчитан \log_{10} от нормализованной средней экспрессии во всех образцах эксперимента. Несколько функций в пакете edgeR также использовались для анализа дифференциальной экспрессии (DE) между повторностями образцов, анализа обогащения GO для дифференциально экспрессированных генов и графиков PCoA, показывающих прогрессирование развития и разделение образцов. Поиск альтернативных промоторов проводился с помощью анализа ко-экспрессии на уровне идентифицированных DPI промоторов с помощью пакета WGCNA для R [11].

4.1 Анализ характеристик промоторной области

Границы CpG островков были загружены из геномного браузера UCSC, а перекрытия с TSS были вычислены с использованием функции `crossinterBed` `bedtools`. Расчет SI и GC отношения был основан на методе, описанном ранее [16].

В зависимости от метода кластеризации доступны различные подходы к классификации промоторов. Для методов на основе межквартильной кластеризации промоторов возможно разделение по принципу: менее 10 п.о. “острые”, а остальные “широкие” [7, 17]. Другой простой способ основан на вычислении индекса формы - SI [57]. Преимущество последнего метода заключается в том, что оно применимо к пользовательским областям промотора, например, 100 п.о. вокруг TSS.

Промоторы были фиксированы в диапазоне 100 п.о. (± 50 п.о.) от позиции TSS (позиция с наивысшей экспрессией) и были классифицированы в зависимости от активности на той или иной стадии эмбриогенеза, тканеспецифичности, различий по форме и составу GC. Группировка в суперкластеры была сделана путем применения функции «`mergeBed`» из набора `bedtools`, требуя слияния между TSS в пределах 100 п.о. Это расстояние гарантирует, что одиночные пиковые TSS остаются таковыми, в то время как соседние TSS становятся частью более широкого мультимодального промотора. Чтобы оценить сдвиги размеров промотора, суперкластеры были определены для всех устойчивых TSS, активных в 26 образцах (приблизительно 31000 пиков), а также для каждой группы развития и клеточных типов отдельно (клетки аорты, печени, мезенхимальные клетки, образцы раннего, среднего, средне-позднего, позднего развития, зачатки конечностей, внеэмбриональная ткань). Для создания суперкластеров для каждой отдельной группы учитывались только TSS, которые имели значение 1 TPM и выше, по меньшей мере в 1 образце из группы, чтобы сделать их сопоставимыми со всем набором надежных TSS.

4.2 Анализ мотивов связывания с факторами транскрипции

Для анализа мотивов ССТФ области (-300, +100) п.о. вокруг TSS были извлечены с помощью bedtools v2.25. Затем экстрагированные последовательности проверяли на наличие обогащения мотивов с использованием инструмента AME (MEME Suite 4.11.2) [58] отдельно для каждой из следующих подгрупп промоторов: аортальной, мезенхимальной группы, гепатоцитов, ранней, средней, средне-поздней и поздней эмбриональной ткани. TSS генов «домашнего хозяйства» также тестировались как отдельная группа. Мотивы, с которыми сравнивались области TSS, были взяты из базы данных JASPAR CORE для позвоночных (2016). Были отобраны и сопоставлены 50 лучших обогащенных мотивов (скорректированное значение $p \leq 0.05$) для эмбриона (все подгруппы развития эмбрионов), первичные клетки (аорта, MSC и гепатоциты) и другие ткани (почки конечностей, внеэмбриональные) чтобы определить общие мотивы. Общая схема представлена на Рисунке 3 [59].

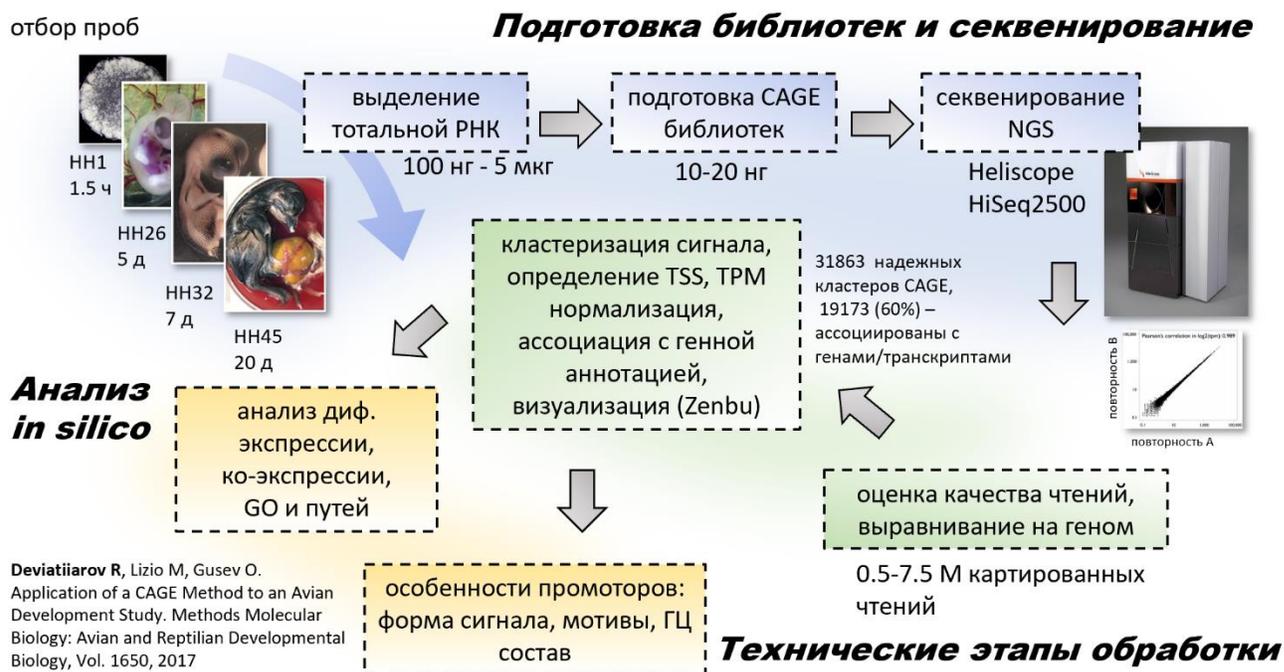


Рисунок 3. Основные этапы в проекте “Chicken FANTOM”. Этапы, отмеченные синим, представляют этапы «мокрой» части эксперимента, зеленым и желтым – этапы вычислительной обработки данных. Подготовка библиотек и секвенирование – относительно консервативные этапы [25].

5. Подготовка образцов в эксперименте с гибернацией

Для оценки молекулярных механизмов, отвечающих за адаптацию эмбриона к снижению температуры окружающей среды был также выбран метод CAGE. Данный метод успешно использован нами для исследования эмбриогенеза курицы. Впервые была создана аннотация промоторов, активных в эмбриогенезе амниот, сопоставимая с аннотацией, сделанной для тканей и клеток человека и мыши. Так как домашняя курица является модельным организмом с хорошей геномной сборкой и аннотацией, это позволило провести биоинформатический анализ представленности известных метаболических и

сигнальных путей. Для проведения данного эксперимента 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 [60]. 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 течение 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[A]+2[B]+3[C] x 4 точки).

5.1 Биоинформатическая обработка чтений, полученных на платформе Illumina

Оценка качества чтений проведена программой FastQC [61]. Далее проведен тримминг (очистка) чтений низкого качества и удаление чтений, соответствующих рибосомальной РНК [62]. Выравнивание на геном проводили в программной среде bwa (burrows-wheeler aligner) [63]. Кластеризация картированных чтений для идентификации промоторных областей и сайтов инициации транскрипции, подсчет чтений для каждого кластера и нормализация полученных значений осуществлены с помощью протокола PromoterPipeline (<https://github.com/Population-Transcriptomics/C1-CAGE-preview>) разработанного в рамках проекта FANTOM [29]. Далее проведена оценка корреляции между библиотеками. Ассоциация промоторных областей с аннотированными генами с дальнейшим анализом дифференциальной экспрессии генов проведены в программной среде R edgeR [56], поиск преобладающих GO тегов в topGO. Расчет дифференциальной экспрессии на уровне отдельных промоторов и поиск мотивов среди промоторов со значимыми изменениями уровня экспрессии проведен с помощью MEME Suite

[64]. Дополнительно данные CAGE были использованы для определения некоторых базовых свойств промоторов, таких как GC состав, форма сигнала CAGE (острые/широкие промоторы), их расположение (однонаправленные/двунаправленные, интронные/экзонные/5' концевые). Такой подробный анализ проведен с помощью стандартных биоинформатических программ, таких как bedtools (Quinlan and Hall., 2010), а также пакетами для R – CAGEr [65] и seqPattern.

ПРОМОТЕРОМ ЭМБРИОГЕНЕЗА АМНИОТ

Были получены образцы эмбриональной ткани цыпленка на 16 этапах развития, покрывая весь период со стадии Гамбургер и Гамильтон 1 (HH1) до HH45 (Рисунок 4). 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. Тотальная РНК выделенная из образцов переводилась в кДНК, из которой готовились библиотеки CAGE согласно стандартному протоколу. Чтения были картированы на последнюю сборку генома курицы galGal5 (Таблица 2).

Таблица 2. Дизайн экспериментов по эмбриональному развитию и гибридации куриного эмбриона с результатами секвенирования и картирования

Sample name (development)	Platform	Library	Reads	Processed reads	TSS (>3 TPM)	% TSS
<i>chicken embryo whole body, 01hr30min (HH1)</i>	Heliscope	CNhs12735	15091585	12539660	14127	44%
<i>chicken embryo whole body, 05hr30min (HH1)</i>	Heliscope	CNhs13027	4476045	2293924	10873	34%
<i>chicken embryo whole body, 06hr (HH1-2)</i>	Heliscope	CNhs13028	5488700	3261195	12583	40%
<i>chicken embryo whole body, 14hr (HH3)</i>	Heliscope	CNhs13029	6447740	3687204	11940	38%
<i>chicken embryo whole body, 18hr (HH4)</i>	Heliscope	CNhs13030	4990595	2917397	12155	38%
<i>chicken embryo whole body, 22hr (HH5-6)</i>	Heliscope	CNhs12736	14370500	12467423	13408	42%
<i>chicken embryo whole body, day01 (HH7 2-3 somites)</i>	Heliscope	CNhs14200	5659005	4145006	13684	43%
<i>chicken embryo whole body, day01.5 (HH11-12)</i>	Heliscope	CNhs14201	6322335	4275300	13803	43%
<i>chicken embryo whole body, day02 (HH14)</i>	Heliscope	CNhs14202	6324370	4400888	14259	45%
<i>chicken embryo whole body, day03 (HH19)</i>	Heliscope	CNhs14203	8174870	5469480	13582	43%
<i>chicken embryo whole body, day04 (HH23)</i>	Heliscope	CNhs14204	6912555	4712551	14116	44%
<i>chicken embryo whole body, day05 (HH25-27)</i>	Heliscope	CNhs14208	3730910	2514840	14686	46%
<i>chicken embryo whole body, day07 (HH32)</i>	Heliscope	CNhs14205	9424445	4561746	12944	41%
<i>chicken embryo whole body, day10 (HH37)</i>	Heliscope	CNhs10485	8299955	4551336	13677	43%
<i>chicken embryo whole body, day15 (HH41)</i>	Heliscope	CNhs14206	1596115	702803	11633	37%
<i>chicken embryo whole body, day20 (HH45)</i>	Heliscope	CNhs14207	5420935	2914064	13950	44%
<i>chicken embryo, extraembryonic day07 (HH32)</i>	Heliscope	CNhs14210	3469985	3185960	13738	43%
<i>chicken embryo, extraembryonic day15 (HH41)</i>	Heliscope	CNhs14211	669820	231516	9743	31%
<i>chicken leg buds, day03 (HH20)</i>	Heliscope	CNhs14213	12732465	8009765	11778	37%
<i>chicken wing buds, day03 (HH20)</i>	Heliscope	CNhs14212	1025285	436274	11312	36%
<i>Aortic Smooth Muscle cells, donor1, Primary cell</i>	Heliscope	CNhs11296	7357720	6110496	13822	43%
<i>Aortic Smooth Muscle cells, donor2, Primary cell</i>	Heliscope	CNhs11299	5444475	4317255	13374	42%
<i>Aortic Smooth Muscle cells, donor3, Primary cell</i>	Heliscope	CNhs11314	7315315	6380588	13586	43%
<i>hepatocytes, tech_rep1, Primary cell</i>	Heliscope	CNhs11922	2313280	1952569	11848	37%
<i>hepatocytes, tech_rep2, Primary cell</i>	Heliscope	CNhs11306	5838168	5416885	11733	37%
<i>Mesenchymal stem cells - bone marrow derived, Primary cell</i>	Heliscope	CNhs11292	7505955	5797152	13818	43%
Sample name (cold torpor)	Platform	Library	Reads	Processed reads	TSS (>3 TPM)	% TSS
<i>chicken embryo whole body, day2.5 (HH16), control 1</i>	HiSeq2500	CNhi10707	22131624	17347150	12520	29%
<i>chicken embryo whole body, day2.5 (HH16), control 2</i>	HiSeq2500	CNhi10707	18167525	14234392	13028	30%

<i>chicken embryo whole body, day2.5 (HH16), cold torpor 1</i>	HiSeq2500	CNhi10707	19934880	16041694	9247	21%
<i>chicken embryo whole body, day2.5 (HH16), cold torpor 3</i>	HiSeq2500	CNhi10707	14679081	11559239	12043	28%
<i>chicken embryo whole body, day2.5 (HH16), recovery 1</i>	HiSeq2500	CNhi10707	19761283	15401684	11446	26%
<i>chicken embryo whole body, day2.5 (HH16), recovery 2</i>	HiSeq2500	CNhi10707	17974364	13791501	14320	33%
<i>chicken embryo whole body, day2.5 (HH16), recovery 3</i>	HiSeq2500	CNhi10707	14420950	11384208	14045	32%
<i>chicken embryo whole body, day6 (HH28-29), control 1</i>	HiSeq2500	CNhi10708	22788998	18157952	15599	36%
<i>chicken embryo whole body, day6 (HH28-29), control 2</i>	HiSeq2500	CNhi10708	12961615	10334419	15473	36%
<i>chicken embryo whole body, day6 (HH28-29), cold torpor 1</i>	HiSeq2500	CNhi10708	18562925	14492438	13073	30%
<i>chicken embryo whole body, day6 (HH28-29), cold torpor 3</i>	HiSeq2500	CNhi10708	21445974	17171122	14232	33%
<i>chicken embryo whole body, day6 (HH28-29), recovery 1</i>	HiSeq2500	CNhi10708	19928861	15828193	13567	31%
<i>chicken embryo whole body, day6 (HH28-29), recovery 2</i>	HiSeq2500	CNhi10708	17311336	13616345	15816	36%
<i>chicken embryo whole body, day6 (HH28-29), recovery 3</i>	HiSeq2500	CNhi10708	16560258	13185948	16477	38%
<i>chicken embryo whole body, 6 hrs (HH1), control 1</i>	HiSeq2500	CNhi10827	9399913	7288812	15115	35%
<i>chicken embryo whole body, 6 hrs (HH1), control 2</i>	HiSeq2500	CNhi10827	7904381	6447728	15447	36%
<i>chicken embryo whole body, 6 hrs (HH1), cold torpor 1</i>	HiSeq2500	CNhi10827	19702234	15726218	13575	31%
<i>chicken embryo whole body, 6 hrs (HH1), cold torpor 3</i>	HiSeq2500	CNhi10827	754344	619472	10943	25%
<i>chicken embryo whole body, 6 hrs (HH1), recovery 1</i>	HiSeq2500	CNhi10827	29973282	23055845	13025	30%
<i>chicken embryo whole body, 6 hrs (HH1), recovery 2</i>	HiSeq2500	CNhi10827	20966616	16865265	15444	36%
<i>chicken embryo whole body, 6 hrs (HH1), recovery 3</i>	HiSeq2500	CNhi10827	22289059	17977625	14564	34%
<i>chicken embryo whole body, 18 hrs (HH4), control 1</i>	HiSeq2500	CNhi10828	10712235	9241148	15668	36%
<i>chicken embryo whole body, 18 hrs (HH4), control 2</i>	HiSeq2500	CNhi10828	16881135	14558545	15627	36%
<i>chicken embryo whole body, 18 hrs (HH4), cold torpor 1</i>	HiSeq2500	CNhi10828	8698690	7524430	14301	33%
<i>chicken embryo whole body, 18 hrs (HH4), cold torpor 3</i>	HiSeq2500	CNhi10828	13893332	12136966	12994	30%
<i>chicken embryo whole body, 18 hrs (HH4), recovery 1</i>	HiSeq2500	CNhi10828	10844815	9385235	9082	21%
<i>chicken embryo whole body, 18 hrs (HH4), recovery 2</i>	HiSeq2500	CNhi10828	8206011	7022513	16370	38%
<i>chicken embryo whole body, 18 hrs (HH4), recovery 3</i>	HiSeq2500	CNhi10828	4362610	3815010	15875	37%

Средняя глубина прочтения для всех подготовленных образцов составила 2.7 миллиона, что сравнимо с 4 миллионами для образцов человека.

1. Картирование TSS

Идентификация отдельных кластеров CAGE и соответствующих им TSS в геноме курицы проводилась с помощью метода DPI (decomposition-based peak identification, определение пиков на основе разложения). 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 TSS с более чем 10 чтениями в 1 образце (эквивалент нормализации тегов на миллион [TPM, tags per million]).

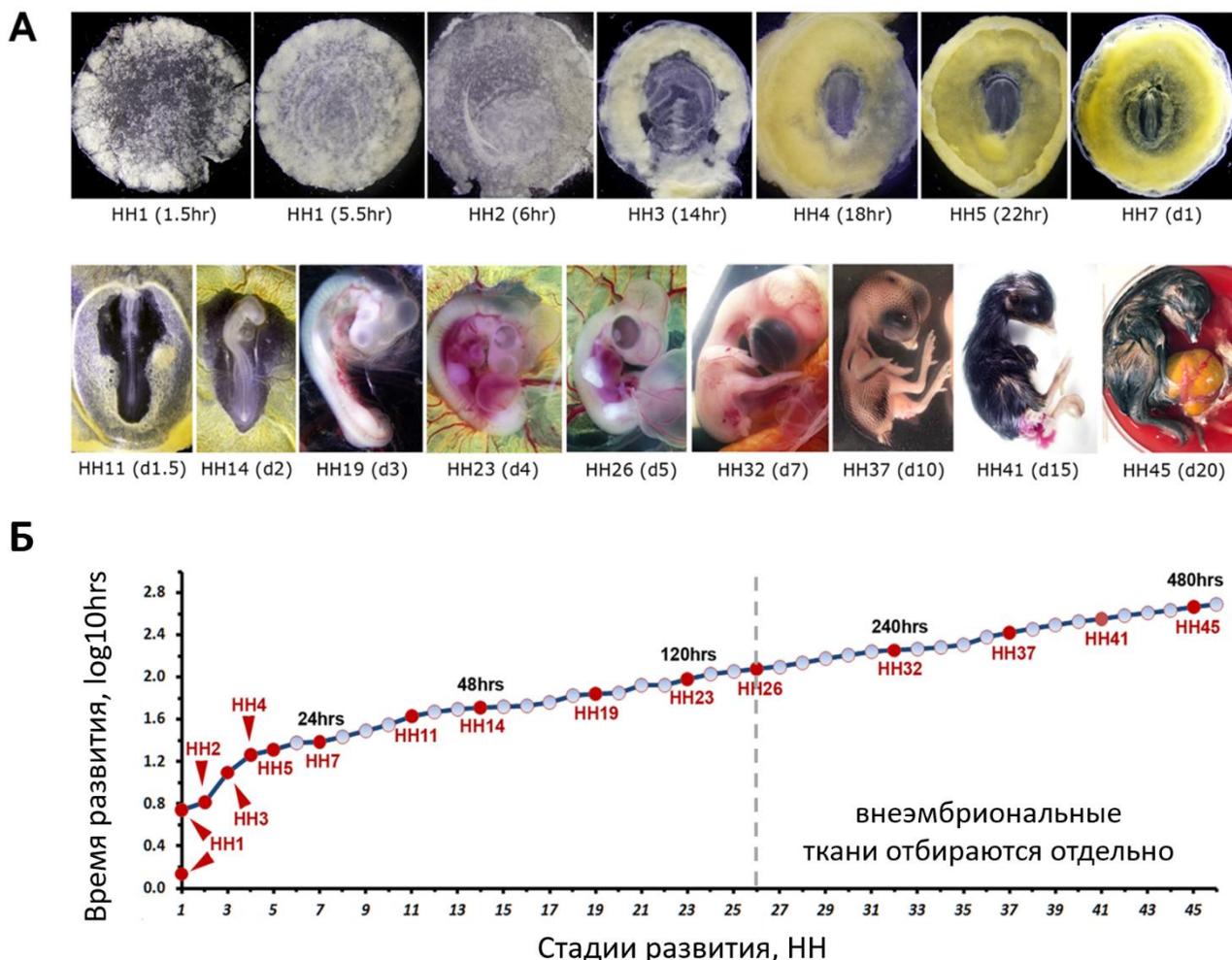


Рисунок 4. Этапы развития НН, выбранные для приготовления библиотек CAGE. А - Репрезентативные изображения куриных эмбрионов, используемых для выделения РНК (подготовка изображений согласно [54]); Б - Временное распределение выбранных стадий (красные точки) в эмбриогенезе курицы. Ось X: 46 этапов Гамбурга и Гамильтона (НН) представлены в числовом порядке. Ось Y: логарифмическая шкала времени инкубации, необходимого для достижения этих этапов (на основе зарегистрированных оценок), например, значение Y 2.0 = 100 часов.

Подобно значениям, показанным при анализах TSS на основе CAGE в образцах человека и мыши [7], 60% (19172) всех устойчивых TSS для цыплят можно сопоставить с аннотированными TSS кодирующих или некодирующих генных моделей в пределах 500 пар оснований (на основе аннотаций RefSeq, Ensembl и EST для кодирующих генов цыплят, а также NONCODE, ALDB-lincRNA и других баз данных lincRNA для некодирующих генов). Остальные 40% (12644) неаннотированных TSS представляют вероятные альтернативные промоторы и РНК некодирующих генов, которые еще предстоит охарактеризовать (Рисунок 5). Около 31-46% всех устойчивых TSS в любом образце (медиана 10557 TSS, 33% от общих TSS) имели пиковые значения, превышающие 3 TPM. Следуя общей практике в анализе CAGE, ассоциированные с картированными пиками гены классифицированы как активные.

2. Анализ главных компонент

Поскольку развитие является непрерывным процессом, профили активности промоторов на разных этапах причинно связаны. Для оценки отношений между стадиями, основанной на их ландшафте TSS в геноме, был использован метод главных компонент (PCoA, метод многомерного масштабирования). PCoA всех образцов развития выявила тесную корреляцию между данными на основе TSS и этапами развития на основе эмбриональной стадии (Рисунок 6). 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, что свидетельствует о том, что эти стадии обладают TSS с общими функциями. Среди них 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.

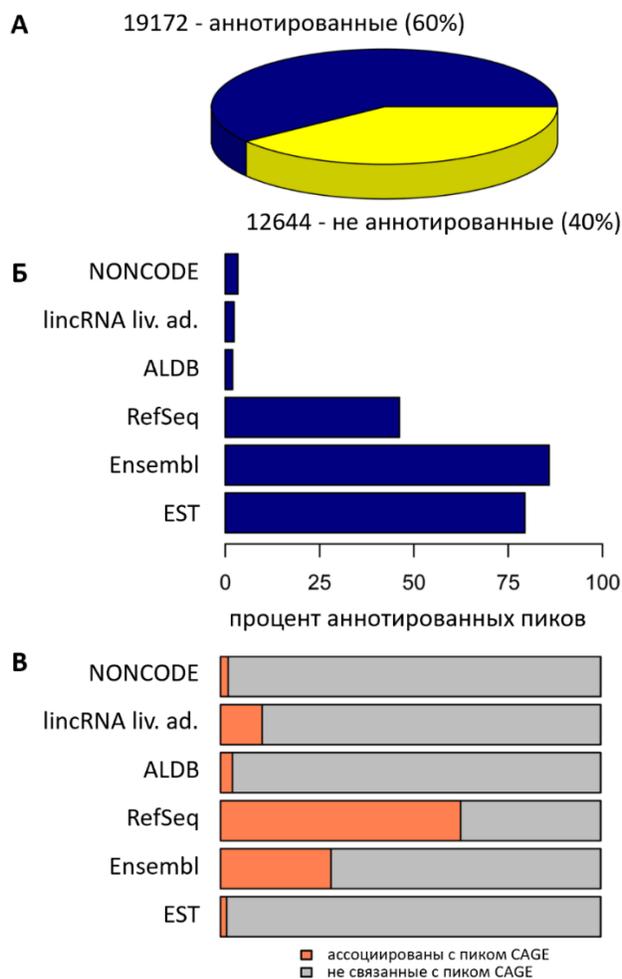


Рисунок 5. CAGE CTSS ассоциированные с аннотацией для геномной сборки galGal5. А - Круговая диаграмма, представляющая долю пиков, которые аннотируются с любой из используемых моделей генов (синий), а также те, которые остаются не связанными (желтый); Б – процент ассоциированных CTSS для различных аннотаций (срез синей группы из А); В - Пропорции генов, связанных с пиками CAGE, по сравнению с не связанными. Несмотря на то, что набор генов RefSeq намного меньше, чем Ensembl, гены RefSeq правильнее аннотируются, что показано более высокой ассоциацией с CTSS.

При сравнении всех образцов между собой стволовые клетки образовывали отдельные кластеры, отражающие гомогенность транскрипции в стабильных типах клеток, независимо от того, представляют ли они дифференцированные или мультипотентные линии. Было показано, что экстраэмбриональные и конечные образцы почек, представляющие субпопуляции развивающихся тканей, более тесно связаны с целыми эмбрионами, чем со стволовыми клетками.

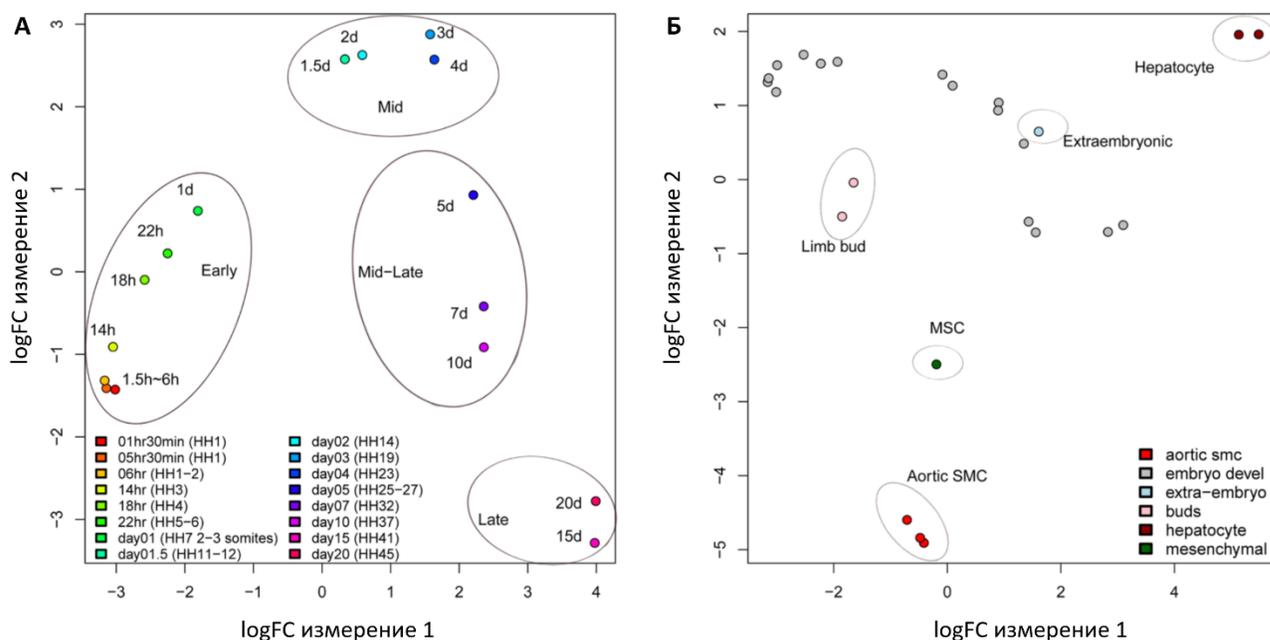


Рисунок 6. Анализ главных компонент PCA. А – Все 16 образцов развития цыплят. Мы выделили четыре группы образцов и обозначили их как ранние (Early), средние (Mid), средне-поздние (Mid-Late) и поздние (Late); Б – PCA для всех образцов, включая почки конечностей (Limb bud), внеэмбриональные ткани (Extraembryonic), гепатоциты (Hepatocyte), мезенхимальные стволовые клетки (MSC), клетки гладкой мускулатуры аорты (Aortic SMC).

3. Chicken-ZENBU: интерактивная платформа для визуализации TSS в геноме.

Геномный браузер ZENBU был разработан для визуализации данных CAGE человека и мыши. Мы использовали эту платформу для создания конфигурации ZENBU на основе куриного эмбриона (далее называемую Chicken-ZENBU, пример на Рисунке 7) для интерактивного и веб-исследования данных CAGE для курицы (доступно по адресу <http://fantom.gsc.riken.jp/zenbu/gLyphs/#config=b1zZI1gUFZ6mHX6-4Gvvr>).

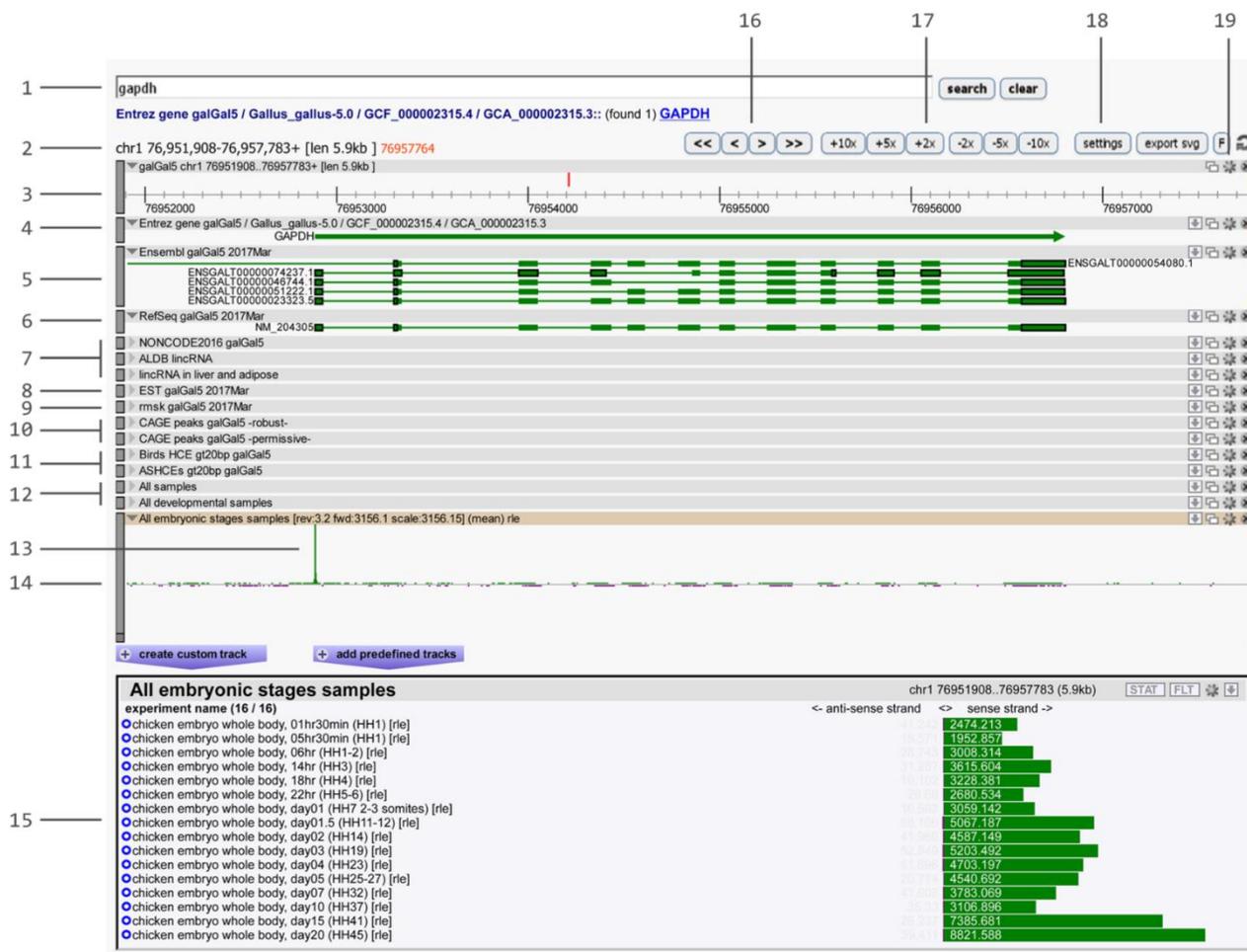


Рисунок 7. Вид терминала Chicken-Zenbu в локусе гена GAPDH с подробным описанием: 1 – окно поиска (генов, локусов по координатам, тегов); 2 – координаты просматриваемой области в геноме; 3 – трек для визуализации хромосом (с возможностью сохранения последовательностей выбранных областей); 4 – трек, представляющий генную аннотацию Entrez (стрелка обозначает направление от 5' к 3' концу, также зеленый цвет соответствует «+» цепи, фиолетовый «-»); 5 и 6 – треки для моделей Ensembl и RefSeq,

соответственно; 7 – генные модели некодирующих РНК [66–68]; 8 и 9 – аннотация EST и повторов; 10 – кластеры CAGE (с мягким и жестким порогом фильтрации); 11 – треки для HCE (высококонсервативные элементы) и ASHCE (высококонсервативные элементы птиц, после удаления элементов характерных для млекопитающих); 12 – профили экспрессии CAGE TSS для всех образцов и отдельно для образцов развития эмбриона; 13 – типичный профиль экспрессии CAGE в области TSS; 14 – Профили экспрессии всех 16 образцов развития эмбрионов; 15 – гистограмма, показывающая уровень экспрессии в каждом образце; 16-19 – настройки и средства визуализации.

Chicken-ZENBU дает возможность визуального представления позиций TSS и оценки их количественных значений при картировании на последней версии сборки генома galGal5. Браузер также был полностью интегрирован с последними ресурсами по аннотации цыплят, включая Ensembl, RefSeq, EST, NONCODE, ALDB-lincRNA и высококонсервативные элементы (HCE, ASHCE) из 48 представленных видов птиц [69, 70]. Гены и TSS можно искать и визуализировать при желаемых геномных разрешениях, от уровня одного нуклеотида до целых хромосом. Это позволило получить интуитивную оценку согласованности TSS во всех доступных точках выборки. Например, проверка результатов картирования TSS для гена «домашнего хозяйства» GAPDH (Рисунок 7) показала во всех образцах надежный и одиночный пик на 5'-конце транскрипта RefSeq NM_204305 и транскрипта Ensembl ENSGALT00000023323.5, охватывающий область в 5 пар оснований с центром пика, отображаемым с точностью до одного нуклеотида на Chr1: 76,952,888 (Рисунок 7). Это подтверждает точность как генных моделей RefSeq, так и Ensembl для GAPDH. Однако во многих случаях модели генов RefSeq и Ensembl расходятся с предсказанием TSS, и наше CAGE отображение TSS может разрешать такие расхождения (например, для бета-актина гена ACTB, Рисунок 8, Б). Если ни одна из двух моделей генов не соответствует нашему картированию

TSS на основе CAGE (например, для рибосомного белка RPL32, Рисунок 9, В), рекомендуется повторная редакция аннотации.

4. TSS специфичные для этапов развития и типов клеток

Полученные данные позволили отслеживать изменения в положениях TSS 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 (Рисунок 10), что, вероятно, отражает использование альтернативных промоторов. Это соотношение сопоставимо с использованием альтернативных промоторов, описанным для геномов человека и дрозофилы [71, 72].

Для поиска и аннотации функционально значимых случаев использования альтернативных промоторов мы провели анализ ко-экспрессии на уровне отдельных промоторов с помощью пакета WGCNA (Рисунок 11). Примерно в половине случаев с наличием более одного промотора, были обнаружены отличия в профиле экспрессии, что выразалось в группировке промоторов в разные кластеры/модули (Рисунок 12).

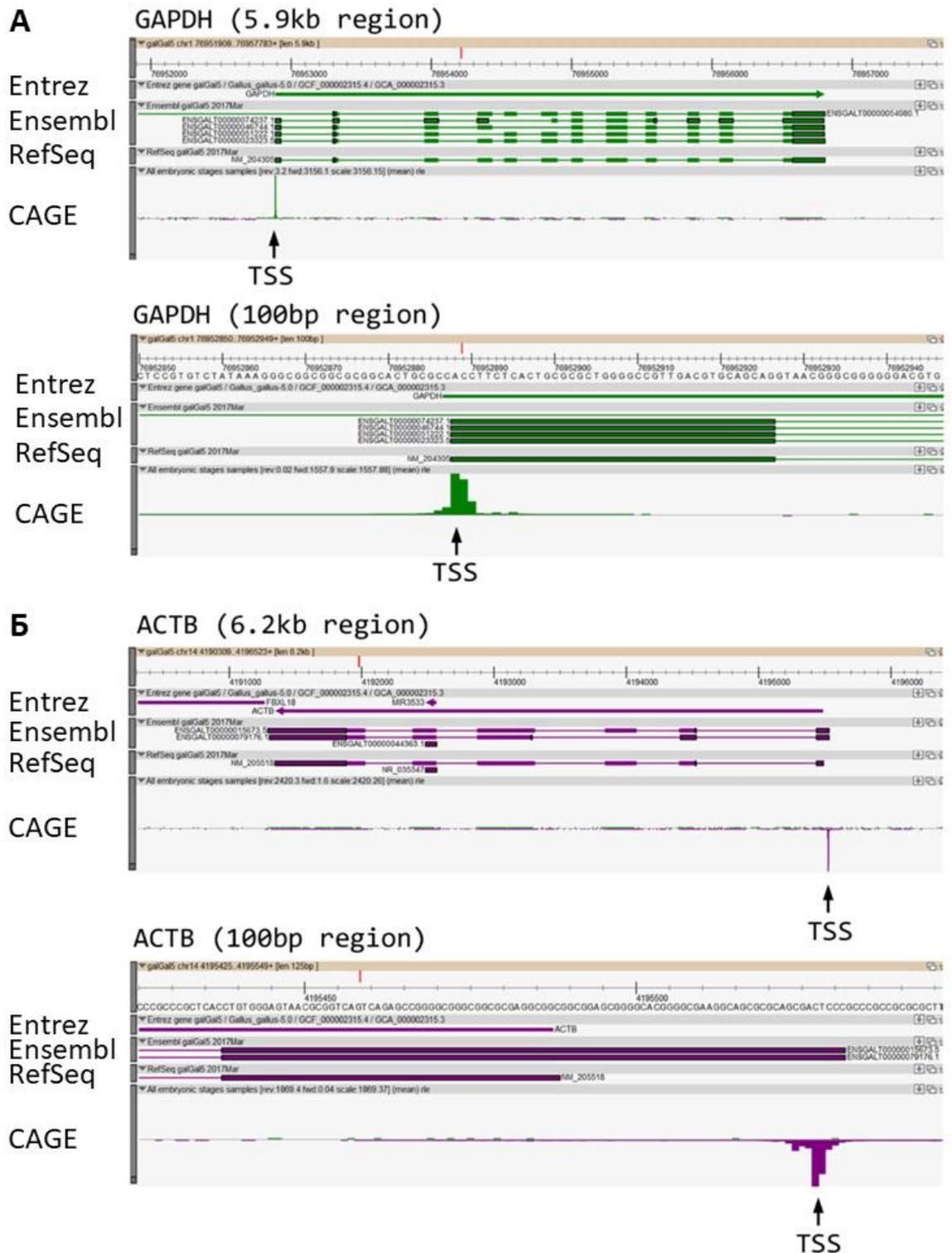


Рисунок 8. Визуализация генной аннотации и сигнала CAGE в Chicken-Zenbu. А – одиночный пик TSS на 5' конце аннотации Ensembl и RefSeq для гена GAPDH (ген целиком и область TSS отдельно); Б – неоднозначная аннотация гена ACTB может быть разрешена с помощью данных CAGE.

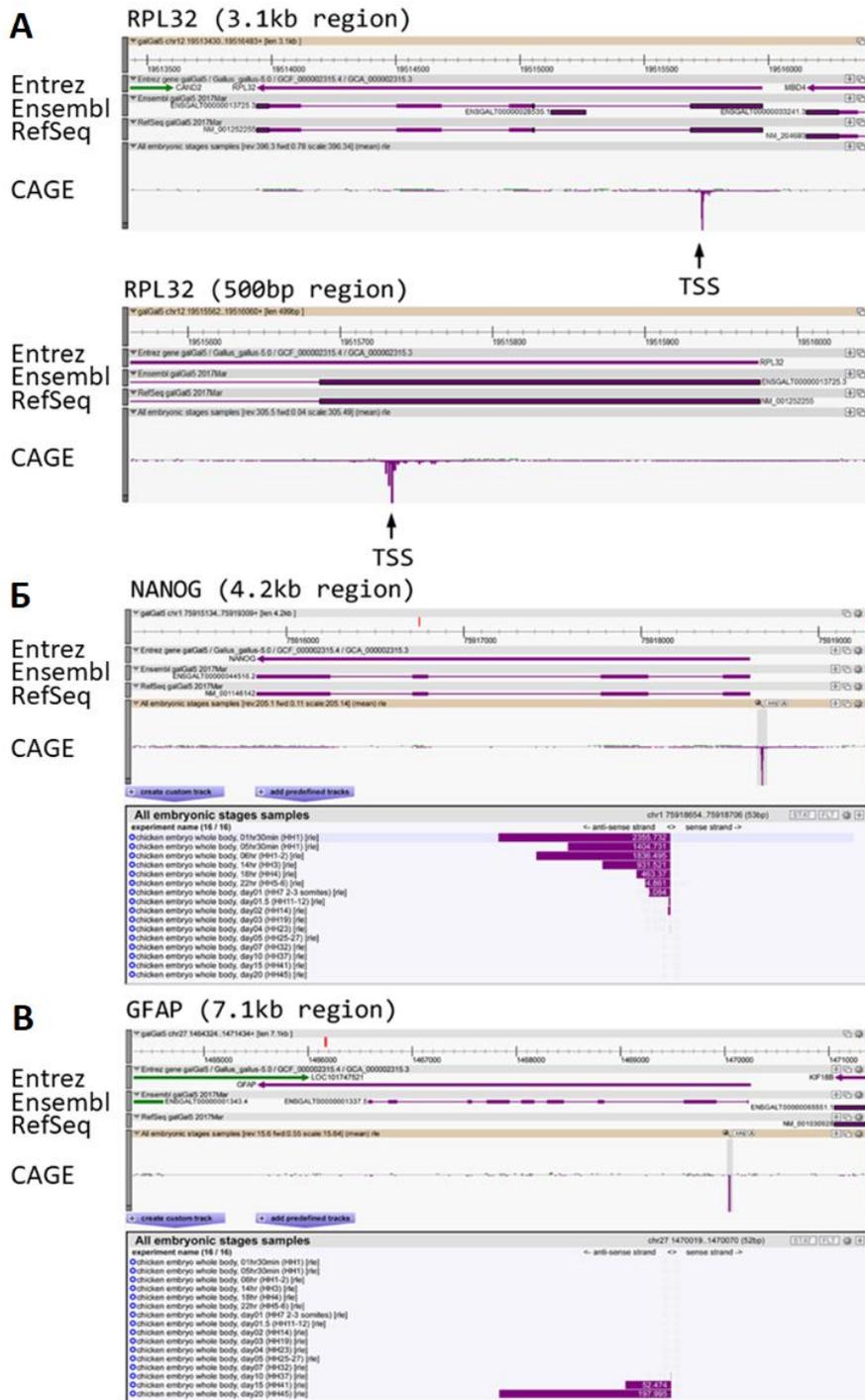


Рисунок 9. Визуализация генной аннотации и сигнала CAGE в Chicken-Zenbu. А – в случае гена RPL32 ни одна из аннотаций не оказалась верной; Б – визуализация локуса и экспрессии гена NANOG, связанного с плюрипотентностью, что также видно из профиля экспрессии его промотора; В – вид на ген GFAP с профилем экспрессии специфичным для самых поздних стадий развития (H41 и 45). Гистограммы отсортированы по стадиям развития.

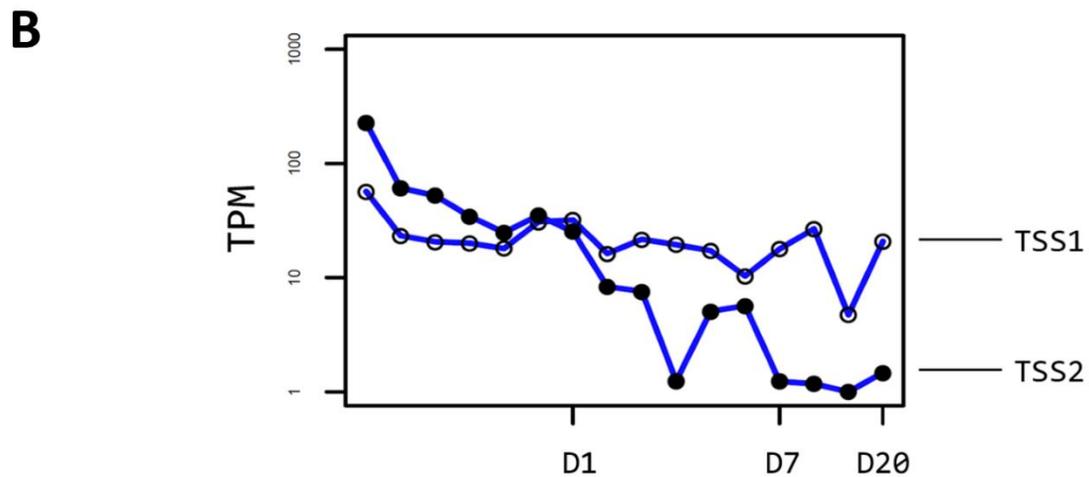
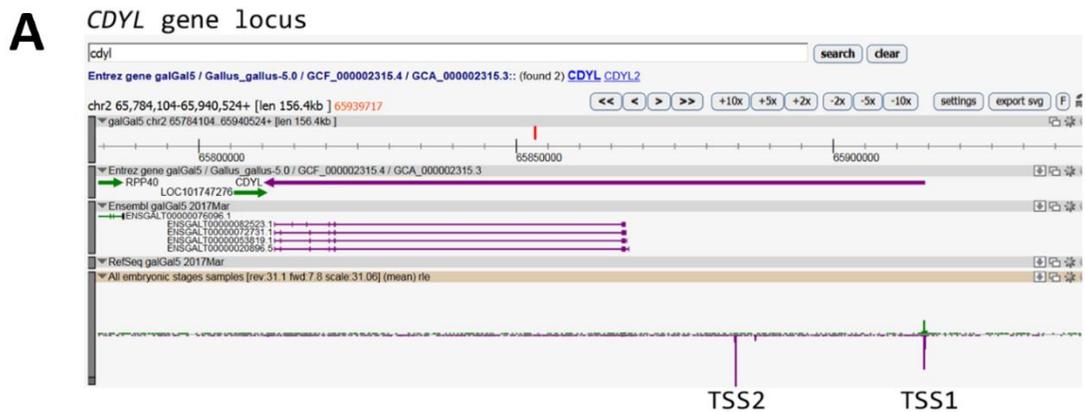


Рисунок 10. Визуализация локуса гена *CDYL* в геномном браузере “Chicken-ZENBU”.
 А – визуализация сигнала CAGE позволяет определить два отдельных пика TSS – 1 и 2 на «-»
 цепи (фиолетовый цвет); Б – гистограммы, показывающие, что TSS1 имеет стабильный

профиль экспрессии на протяжении всего эмбрионального развития, тогда как TSS2 специфичен для ранних стадий; В – тренд изменения экспрессии TSS1 и TSS2 на всех этапах развития отображенные в логарифмической шкале.

С другой стороны, пиковые значения TSS были стабильны только для небольшого процента всех зарегистрированных TSS. Для большинства TSS их пиковые значения демонстрировали заметные различия в количестве чтений среди образцов развития. Такое изменение может быть обнаружено в нашем анализе с высокой чувствительностью, оценивая значения экспрессии в кластерах CAGE в пределах 5 порядков.

Такое профилирование экспрессии также позволило провести количественную оценку активности промотора в гетерогенных клеточных популяциях эмбриона. Например, потеря плюрипотентности эпибласта в течение первого дня эмбрионального развития была четко продемонстрирована путем отслеживания экспрессионных изменений пика TSS гена NANOG (на рисунке 9, Б и рисунке 13). Аналогичные изменения наблюдались и для многих других генов, ассоциированных с плюрипотентностью (например, POU5F3, MYC и EOMES) (Рисунок 13) [73–75].

Аналогичным образом, после прогрессирования развития можно было легко оценить специфическую дифференциацию линии в каждом из трех основных зародышевых слоев (эктодерма, мезодерма и энтодерма). Например, один единственный пик TSS был обнаружен в локусе GFAP (на рисунке 9, В и количественное определение на рисунке 13), а его быстрый рост активности во второй половине эмбриогенеза означает дифференцировку астроцитов, производных от эктодермального зародышевого слоя.

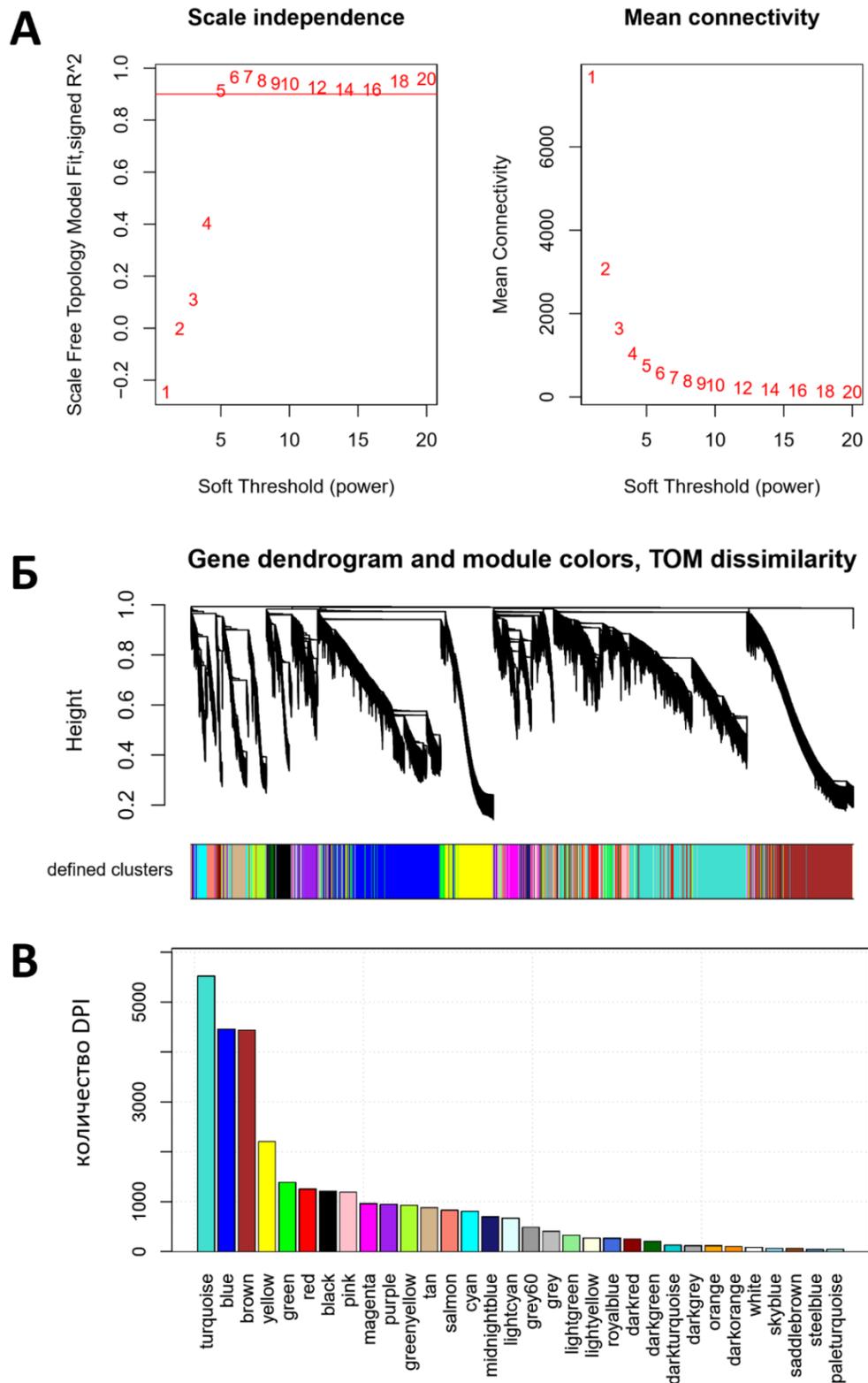


Рисунок 11. Основные этапы анализа ко-экспрессии на уровне отдельных промоторов в пакете WGCNA. А – обобщенные индексы сети (ось Y) в зависимости от мощности мягкого

порога (“soft threshold”, ось X). Числа на графиках указывают на соответствующие уровни мягкого порога. Графики показывают, что не масштабированная топология (“scale-free topology”) достигает желаемого уровня вокруг мощности мягкого порога 5. Связанность сети резко снижается в повышении порога, поэтому в данном случае выгодно ориентироваться на топологический индекс [11, 76]. Б - дендрограмма и соответствующие модули на основе индекса топологического перекрытия (ТОМ, topological overlap measure); В – количество DPI промоторов в идентифицированных ко-экспрессионных модулях.

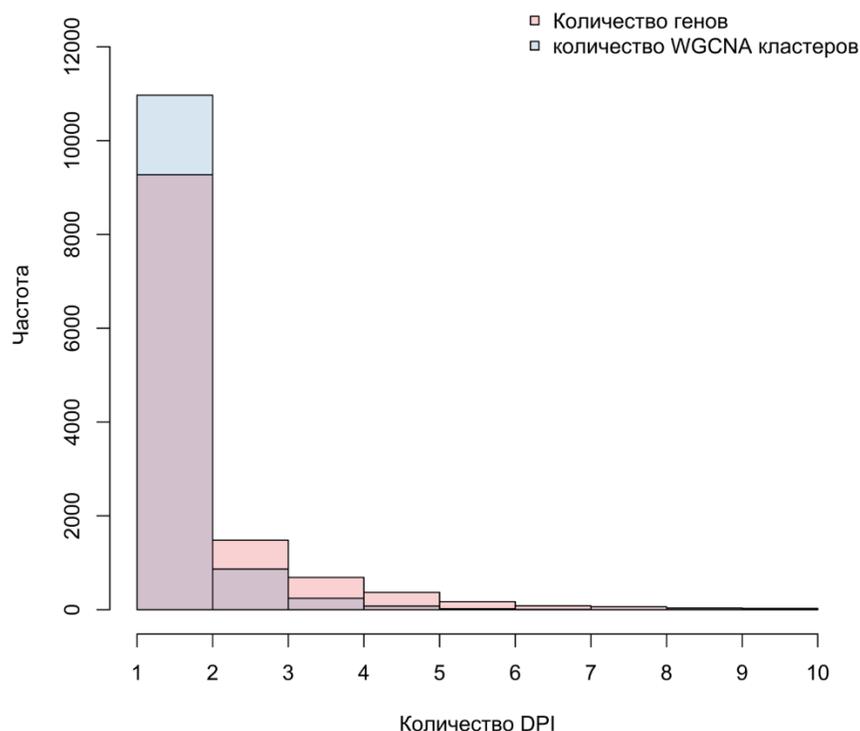


Рисунок 12. Использование альтернативных промоторов в эмбриональном развитии цыплят. Большинство генов имеют один промотор и естественным образом относятся к одному экспрессионному кластеру. Количество WGCNA кластеров посчитано для промоторов. Далее при $DPI > 1$ видно, что примерно половина из ассоциированных промоторов имеют различные профили экспрессии, так как относятся к различным кластерам.

Кроме того, TSS-профили для многих других маркеров, специфичных для линии эктодермы, таких как *ASL1* (хрусталик), *GRIA4* (нейроны) и *PLP1* (олигодендроциты), а также для маркеров, специфичных для линии мезодермы

и энтодермы (например, TNMD [сухожилие], TNNT3 [быстрая мышца], MYH7B [медленная мышца] и LDB3 [поперечнополосатая мышца] для мезодермы и ALB, F2 и SPP2 [печень] и GKN1 [желудок] в энтодерме) также могут точно отслеживаться на протяжении всего развития (Рисунок 13).

Интересно отметить, что в группе генов с множественными пиками TSS каждый пик демонстрировал уникальные профили развития, которые, вероятно, отражают действия альтернативных промоторов согласно специфической регуляции их транскрипции. Например, ген CDYL, кодирующий не охарактеризованный хромопротеин, имел 2 устойчивых пика TSS (Рисунок 10 А). Один из них (TSS1) был экспрессирован повсеместно, тогда как другой (TSS2) показал высокие уровни экспрессии только в раннем развитии (Рисунок 10 Б, В).

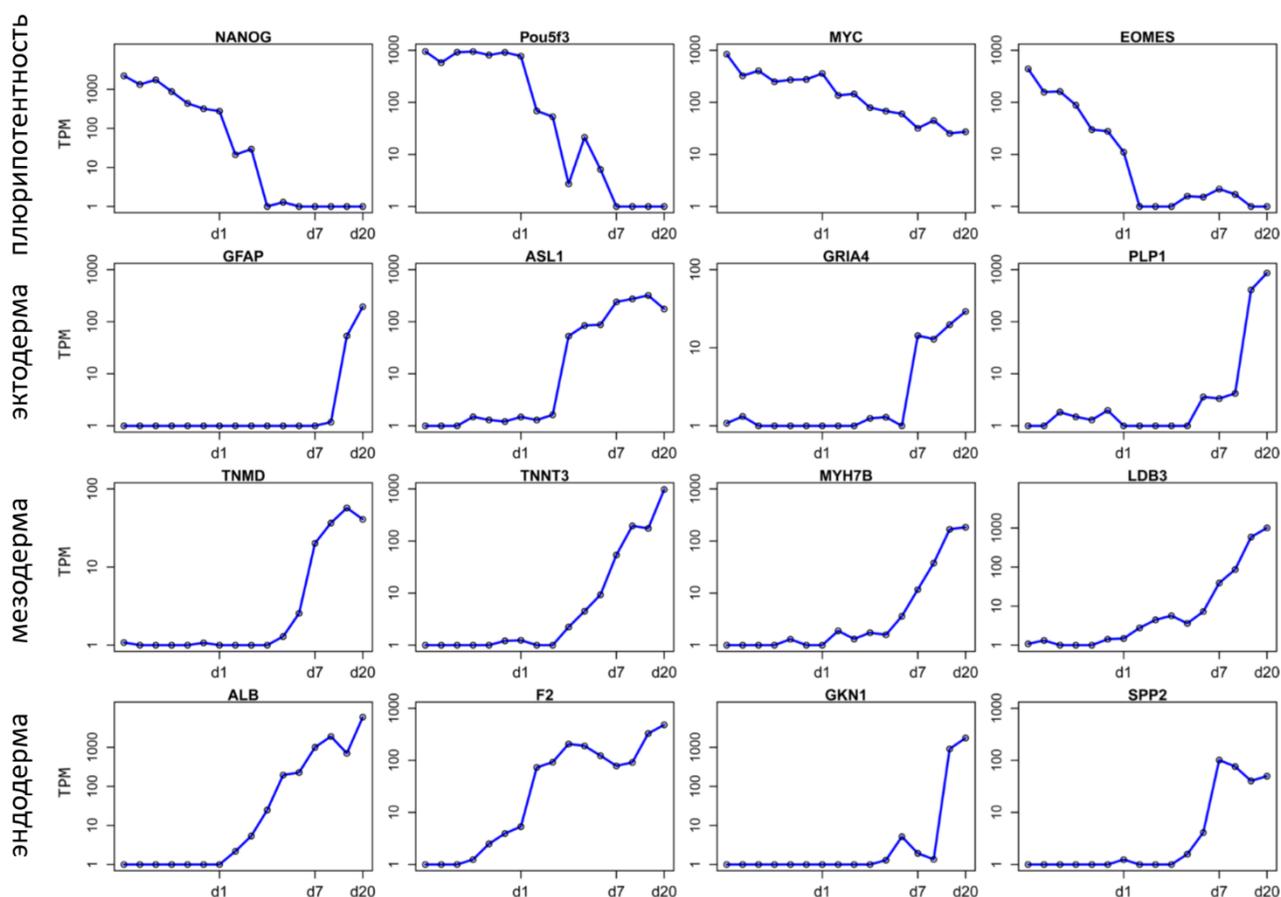


Рисунок 13. Профили экспрессии промоторов специфичных для определенных стадий развития и типов клеток. TSS ассоциированные с плюрипотентностью и различными зародышевыми листками имеют различные профили экспрессии в процессе развития. Гены связанные с плюрипотентностью NANOG, POU5F3, MYC, EOMES активны на ранних этапах, гены связанные с формированием эктодермы, мезодермы и энтодермы - наоборот активируются ближе к поздним стадиям развития. Ось X представляет собой стадии развития; ось y представляет значения экспрессии (TPM) в логарифмическом масштабе.

Для систематической оценки TSS на уровне отдельных этапов развития и клеток, мы применили ограничение по активности промотора 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 конечностей). Стоит отметить, что многие из этих промоторов, специфичных для какой-либо стадии или дифференцированного типа клеток, ранее не были зарегистрированы или плохо охарактеризованы (Рисунок 14).

5. Полный список TSS для генов «домашнего хозяйства»

Многие гены участвуют в поддержании основного клеточного аппарата и экспрессируются повсеместно во всех типах клеток. Некоторые из них являются высокоэкспрессируемыми (например, гены, кодирующие GAPDH, тубулин, актин и рибосомальные белки). Эти гены обычно используются в качестве контроля для нормирования при анализе дифференциальной экспрессии. Однако недавние исследования показали, что большинство таких генов

«домашнего хозяйства» не экспрессируются ни повсеместно, ни стабильно в разных типах тканей/клеток. В нашем исследовании из 31863 TSS пиков, идентифицированных в геноме цыплят, было 3311 (11.4%) активных во всех 26 образцах (> 3 TPM), что можно ~~было~~ 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 АСТВ) также были представлены в последней группе.

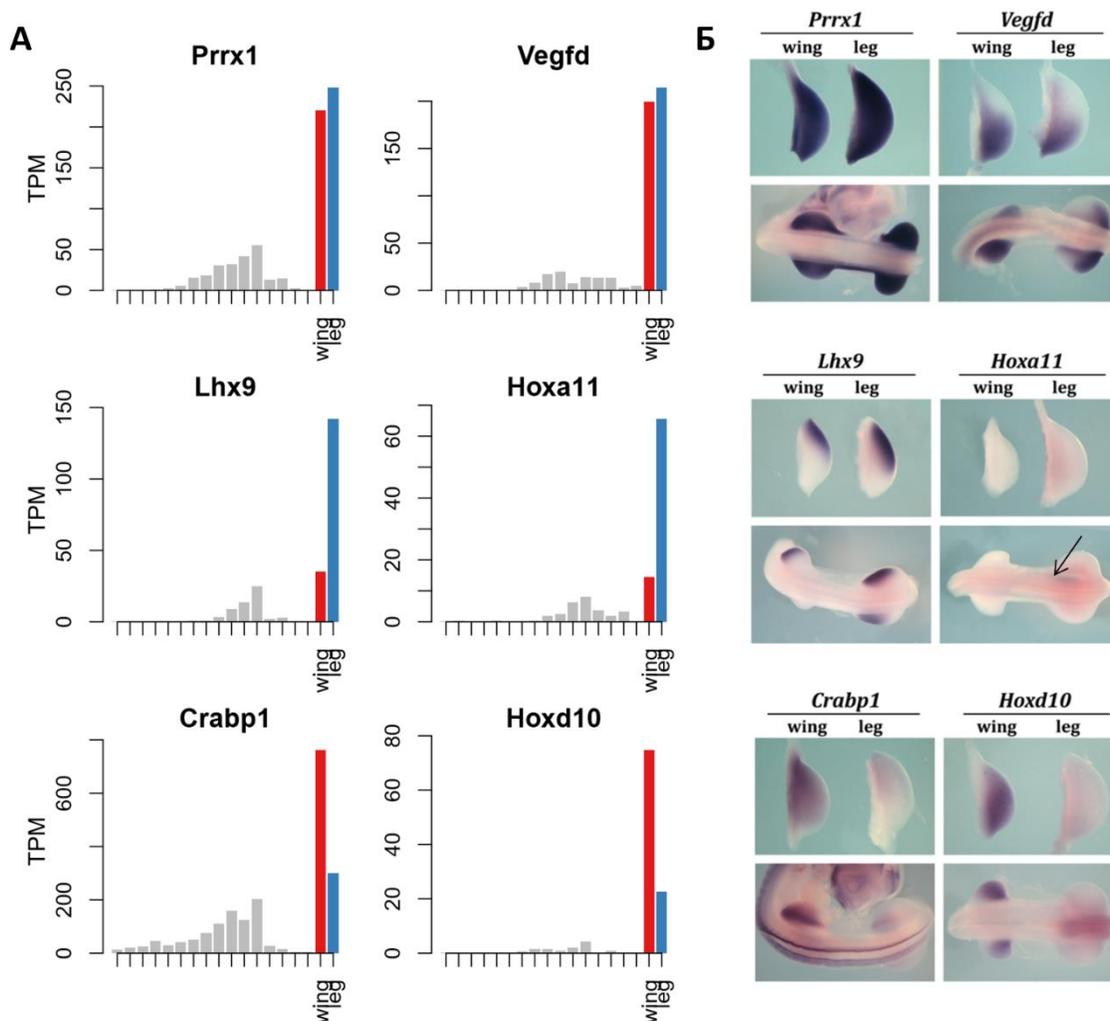


Рисунок 14. Примеры генов с профилем экспрессии специфичным для зачатков конечностей. А – профиль экспрессии соответствующих TSS для генов PRRX1 и VEGFD в

зачатках крыльев и ног, CRABP1 и HOXD10 в крыльях, LHX9 и HOXA11 – в зачатках ног; Б – гибридизация *in situ* для данных генов.

Однако, несмотря на повсеместный и высокий уровень экспрессии, для таких генов были отмечены очень значительные вариации активности в экспериментах (высокая дисперсия), согласно общей оценке генов «домашнего хозяйства» по Айзенбергу и Леванану [77]. Используя аналогичный подход (>10 TPM, $\sigma(\log_2\text{TPM}) < 1$), мы идентифицировали 1254 генов «домашнего хозяйства» цыплят (1399 TSS), среди которых 787 (63%) являются ортологами между человеком и курицей. Анализ представленности генных онтологий (GO) для выбранных 1254 генов показал их исключительную функциональную связь с основными клеточными процессами, включая трансляцию, транскрипцию, субклеточную организацию, метаболизм и транспорт белка. Подавляющее большинство этих генов «домашнего хозяйства» были выражены гораздо более стабильно, чем GAPDH (в рейтинге 1077) и ACTB (который не прошел фильтрацию в соответствии с выбранным критерием). Среди самых стабильных генов в топ 10 по стабильности экспрессии вошли IK, EIF3K, EIF3I, FBXW2, MED21, RPL7, SAP18, RALGAPB, DERL2 и SNX4 (Рисунок 15, А); а в топ 10 по уровню экспрессии были включены RPL7, H3F3C, VAPA, EIF3K, EIF3I, NAA20, PSMC6, SAP18, ZMAT2 и POLDIP3 (Рисунок 15, Б). Эти гены рекомендуются нами в качестве надежных генов «домашнего хозяйства» для нормализации экспрессии в исследованиях птиц.

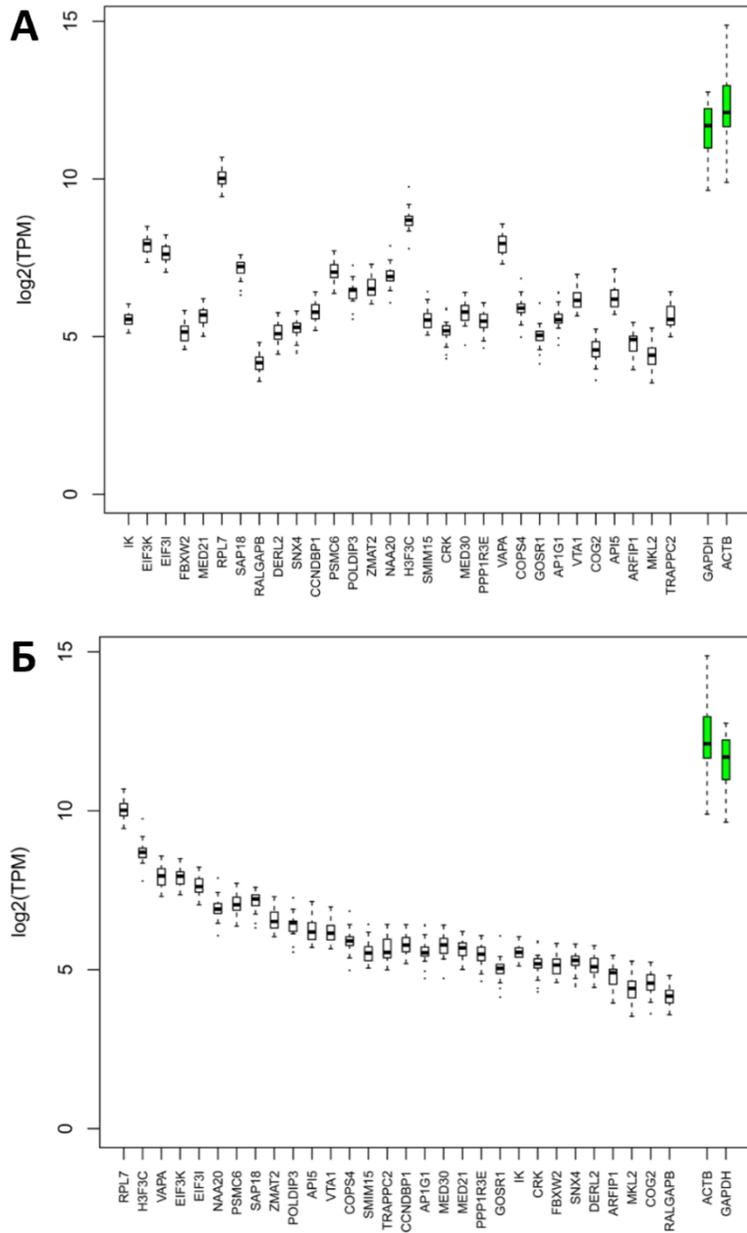


Рисунок 15. Разброс уровня экспрессии топ генов «домашнего хозяйства» во всех 26 образцах развития и типов клеток с наименьшей дисперсией в сравнении с классическими GAPDH и ACTB. А - Гены упорядочены по возрастанию дисперсии, Б – гены упорядочены по уровню экспрессии TPM.

6. Общие особенности TSS цыплят

Полногеномное картирование TSS на сборку генома курицы позволило нам проанализировать общие особенности куриных промоторов, такие как перекрытие с CpG островками, содержание GC, длина и форма сигнала, а также их использование на разных этапах и типах клеток. На основе биоинформатического анализа генома курицы было предсказано приблизительно 70000 CpG островков, среди которых 48% были связаны с соответствующим геном [33, 78]. Было обнаружено, что 13701 из всех стабильных TSS (43%) перекрываются с CpG островками, что выше, чем указано для человеческих TSS (33%) [7]. Это согласуется с общим более высоким содержанием GC в геноме курицы (куриный 52%, человек 41%) [79]. Мы также проанализировали соотношение формы пика (как показатель точности определения TSS) с другими функциями промотора. Форма сигнала CAGE в промоторной области (широкая или острая) оценивалась по «индексу формы» (SI) [16]. Острый пик определялся как имеющий $SI > -1$ и широкий пик как имеющий $SI < -1$, причем максимально возможное значение SI равно 2, представляющее 100% однопозиционное отображение положения всех тегов. Широкие пики демонстрировали более высокое содержание GC (в пределах области 100 п.о., ± 50 bp вокруг соответствующего положения TSS) (Рисунок 16, А) и, следовательно, показали более высокое присутствие CG-богатых мотивов (Рисунок 16, Б). Кроме того, они перекрывают CpG островки с более высокой частотой (Рисунок 16, В), хотя общий процент CpG-ассоциированных TSS (среди всех экспрессирующихся TSS > 3 TPM) не имеет значимых изменений среди анализируемых образцов (Рисунок 16, Е). Широкие пики также были связаны по большей части с генами без тканеспецифичного профиля экспрессии (Рисунок 16, Г, Д). Острые пики, с другой стороны, были явно связаны с

мотивом ТАТА в области -25 п.о. от TSS, что видно из позиционной визуализации матрицы весов (Рисунок 16, Б).

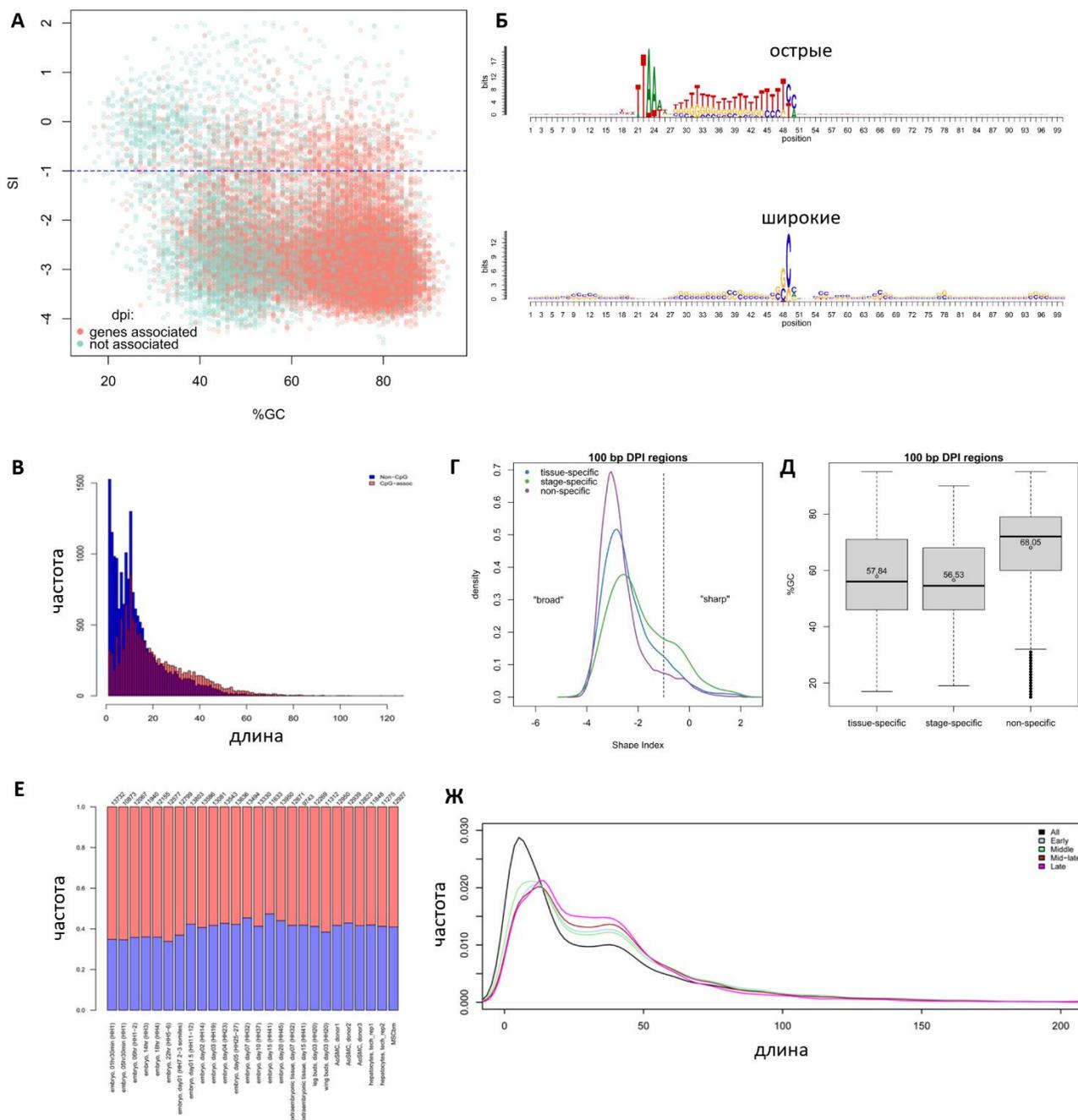


Рисунок 16. Общие свойства промоторов эмбрионального развития курицы. Сигнал CAGE в промоторной области (+/-50 п.о. вокруг TSS) охарактеризован с помощью индекса формы (SI), который отмечен как «острые» (SI>-1) или «широкие» (SI<-1) и был сопоставлен с содержанием ГЦ пар в той же области. А – широкие пики TSS (SI <-1) связаны с более

высоким содержанием ГЦ пар. Красным отмечены промоторы, ассоциированные с генами, зеленым - не связанные с генами; Б – визуализация матрицы весов для промоторных областей 100 п.о. отдельно для острых и широких пиков. Позиция 50 соответствует TSS; В – гистограмма, показывающая размер CpG-ассоциированных (красные) и не связанных с CpG (синие) пиков TSS. CpG-ассоциированные пики имеют тенденцию быть длиннее; Г – распределение сигнала CAGE в промоторах, активных специфично на определенном этапе развития или тканях, смещено в сторону острой формы; Д – в среднем содержание GC промотора выше для неспецифических TSS, чем для TSS, со специфичным профилем экспрессии; Е – процент TSS, ассоциированных с CpG островками (отмечены красным) и не ассоциированных (синие) в каждом из образцов. Количество TSS с экспрессией более 3 TPM для каждого образца указано сверху; Ж – график плотности, показывающий изменение размеров суперкластеров (по оси x) в разных группах развития. All - все образцы объединены.

Полученные результаты согласуются с предыдущими исследованиями по структуре промотора млекопитающих [7] и дрозофилы [16]. Мы дополнительно исследовали использование промоторов путем слияния пиков TSS (> 1 TPM) в пределах 100 п.о. в суперкластеры. Всего было получено 21632 суперкластера; 14422 (71%) из них имели 1 TSS, а остальные - 2 или более TSS. Общий сдвиг к более широкому размеру суперкластера наблюдался в образцах развития (Рисунок 16, Ж), что указывает на общую тенденцию к увеличению количества TSS в процессе развития. На рисунке 17 показана активность специфичных суперкластеров.

7. Картирование TSS как инструмент для систематического анализа промоторов и энхансеров

Системные исследования развития птиц дают беспрецедентную информацию о том, как транскрипционные регулирующие сети контролируют спецификацию клеточных линий [80–84]. Подобные исследования также могут способствовать пониманию промоторной и энхансерной архитектуры, которые еще не были описаны для генома курицы. Так, в недавней работе по

сравнительной геномике 48 видов птиц [69, 70] было показано, что более 99% (примерно 250 тысяч) высоко консервативных геномных элементов птиц (ASHCE - Avian-Specific Highly Conserved Elements, подмножество всех консервативных геномных элементов HCE) были расположены в некодирующих областях генома, указывая на их потенциальную роль в регуляции транскрипции. Среди пиков TSS, идентифицированных в этой работе, 110 (0.34%) непосредственно перекрывали ASHCE, а 2660 (8.35%) непосредственно перекрывали HCE. Неудивительно, что анализ GO топ 500 генов, связанных с ASHCE, предполагал сильный сдвиг в сторону функций развития [70]. Для облегчения дальнейшего анализа ASHCE в регулировании процессов развития у куриц, мы интегрировали все ASHCE со сборкой galGal5 и добавили эту информацию в Chicken-ZENBU (как дорожки HCE и ASHCE, метка 11 на рисунке 7). Другие ресурсы сравнительной геномики позвоночных, такие как 0.6 миллиона консервативных не экзонных элементов (CNEE), идентифицированные путем выравнивания всего генома широкого спектра видов позвоночных [85], в будущем могут быть сопоставлены и визуализированы на Chicken-ZENBU.

В дополнение к интеграции консервативных регуляторных элементов и активности TSS наши данные также позволили систематически профилировать сайты связывания факторов транскрипции (ССТФ) в любом промоторе, представляющем интерес. Для подтверждения этого мы выбрали промоторные участки от -300 до +100 пар оснований вокруг идентифицированных TSS. TSS, обогащенные в каждой группе образцов (> 10 TPM), были просканированы на наличие перепредставленных ССТФ с использованием анализа последовательности мотивов (Analysis of Motif Enrichment, AME), основанного на сравнении весовых матриц с последовательностью генома (MemeSuite) [58, 64]. 50 лучших мотивов для обогащенных TSS в каждой из 10 групп и для генов «домашнего хозяйства» показаны на рисунках 18-20. Небольшой процент из

первых 50 перепредставленных мотивов был общим между различными группами образцов (0 для эмбриональных образцов, 3/50 для стволовых клеток (рисунок 21, А') и 0 для других тканей) и для генов «домашнего хозяйства» (0 общих) (рисунок 21). Большинство идентифицированных топ 50 мотивов ССТФ были весьма специфичными для отдельных категорий образцов (31/50 для ранней эмбриональной группы, 32/50 для группы среднего развития, 20/50 для средне-поздней группы, 8/50 для поздней эмбриональной группы, 24/50 для аортальных SMC, 35/50 для MSCs, 26/50 для гепатоцитов, 47/50 для внеэмбриональной ткани, 6/6 для почек крыла и 13/17 для почек ног) (рисунок 21), предполагая, что они в основном были вовлечены в специфическое регулирование транскрипции. Аналогичные результаты были получены, когда сравнение было выполнено с использованием топ 100 мотивов ССТФ (рисунок 21). Более того, когда анализ ССТФ был распространен на все TSS с острыми или широкими пиками, были выделены отдельные группы обогащенных мотивов для острых (богатых TA мотивов, включая сайт связывания TBP – ТАТААА) и широких (CG-богатые мотивы) промоторов (рисунок 22). Эти данные показывают, что картирование TSS в сочетании с профилированием TFBS позволяет нам идентифицировать обогащенные мотивы с потенциально специфическими ролями в регуляции транскрипции, связанной с развитием.

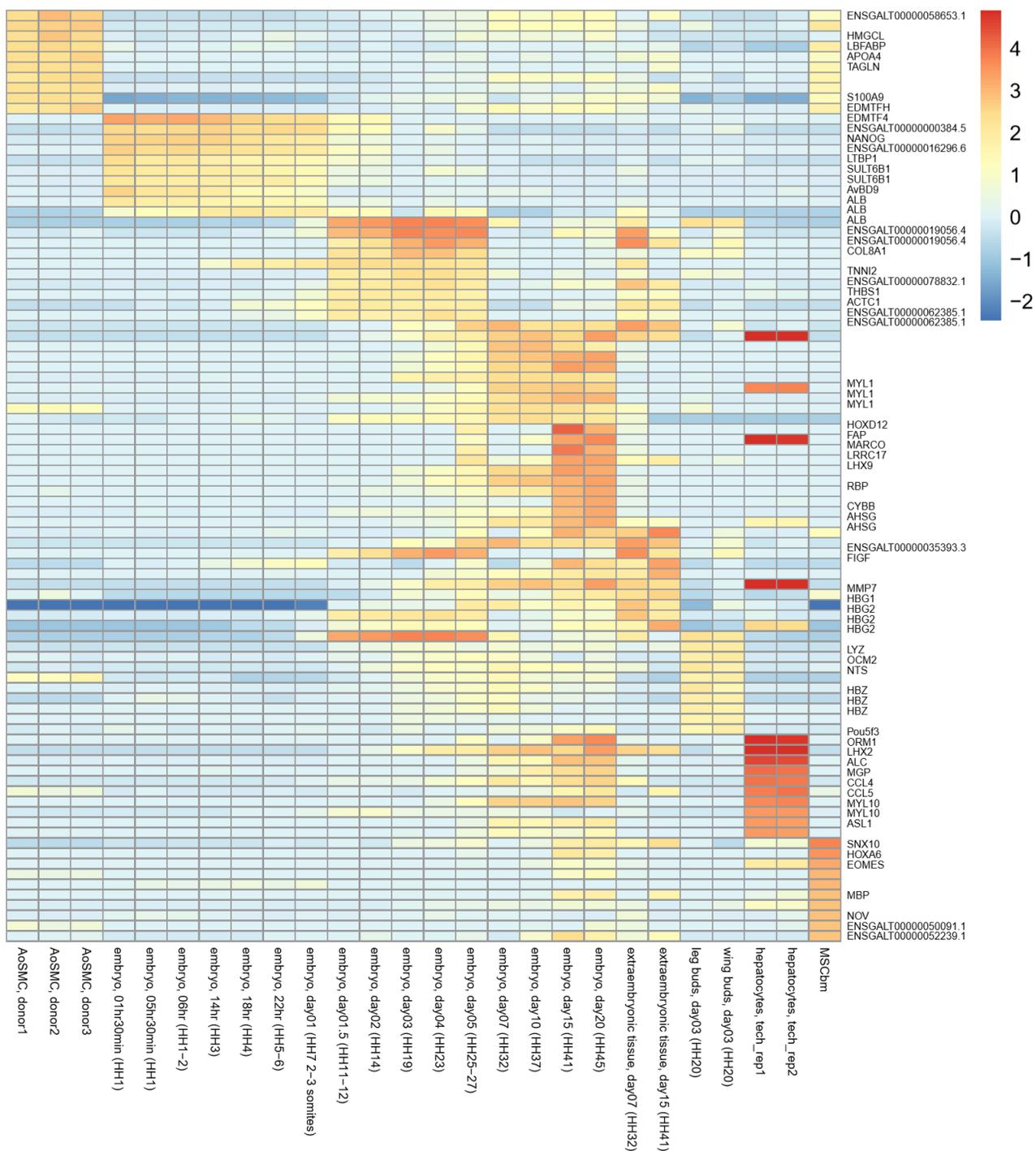


Рисунок 17. Тепловая карта активности суперкластеров имеющих специфичный профиль экспрессии на определенных этапах развития, тканях и клетках цыплят. Показаны 10 самых экспрессируемых кластера. Экспрессия представлена в виде нормализованных значений $\log_2(\text{TPM})$. На оси Y указаны названия соответствующих ассоциированных с суперкластерами транскриптов и генов.

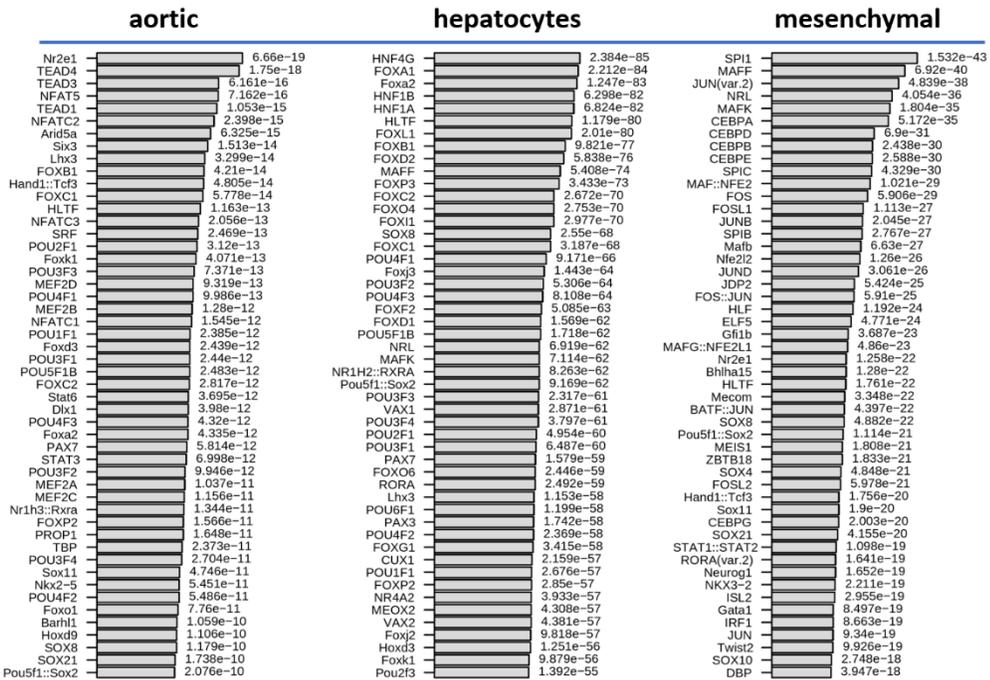


Рисунок 18 Топ 50 перепредставленных мотивов связывания транскрипционных факторов в области промоторов специфично экспрессирующихся в клетках гладкой мускулатуры аорты, гепатоцитах и стволовых мезенхимальных клетках.

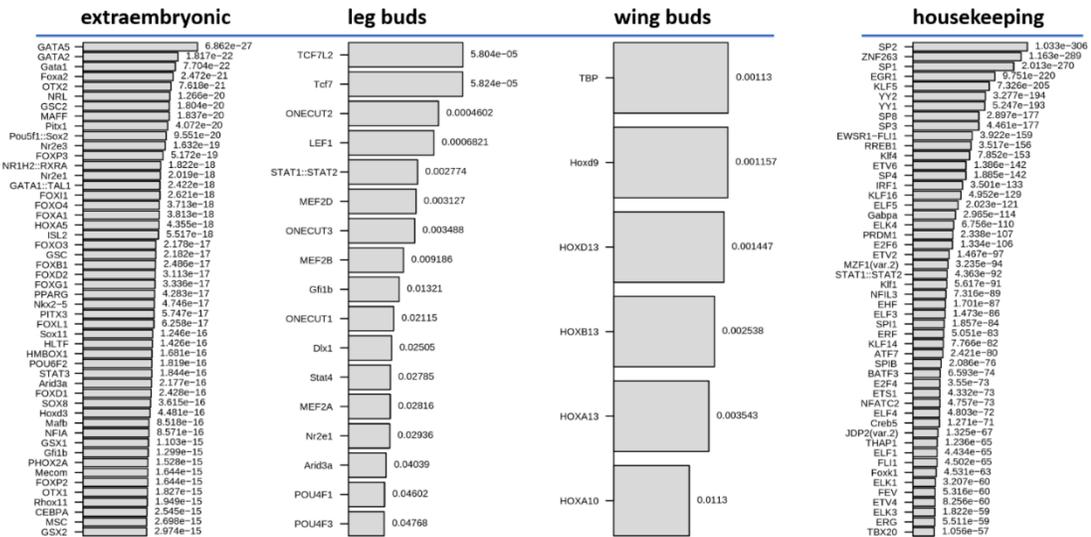


Рисунок 19 Обогащенные мотивы (50) в промоторах активных во внеэмбриональных тканях, зачатках нижних и верхних конечностей цыплят, а также мотивы, характерные для идентифицированных в данном исследовании генов «домашнего хозяйства».

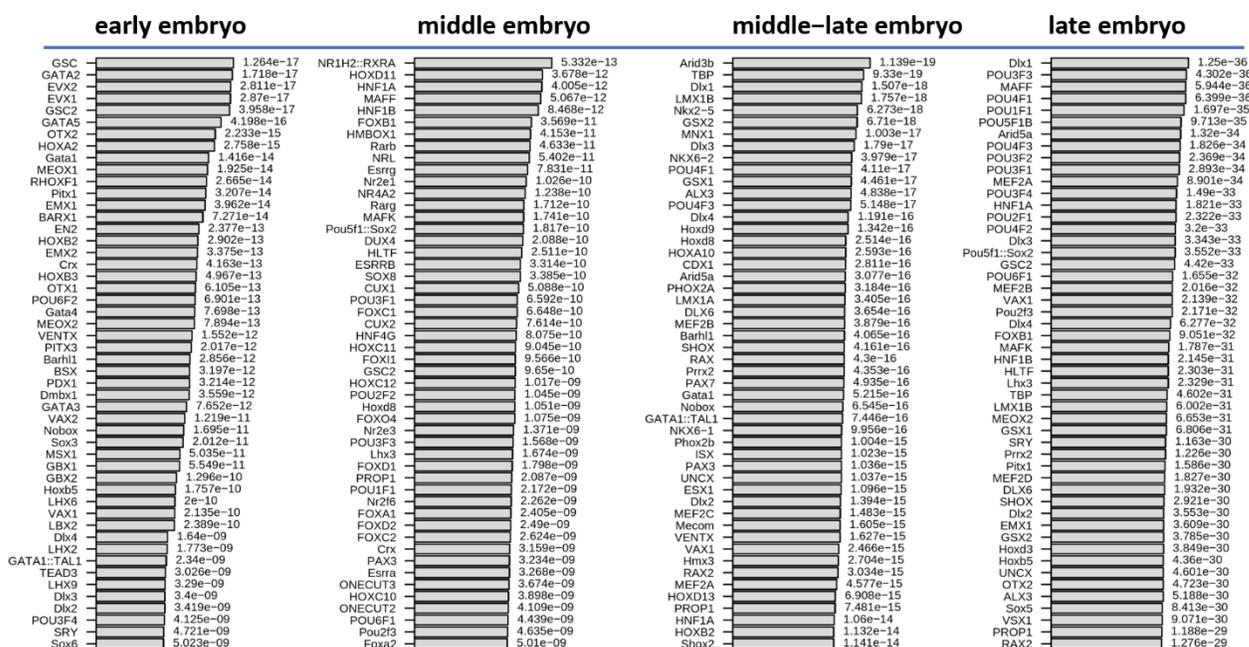


Рисунок 20. Топ 50 мотивов, связывающих транскрипционные факторы регулирующие процессы развития куриного эмбриона на ранних, средних, средне-поздних и поздних стадиях.

8. Картирование TSS позволяет эффективно активировать гены с помощью технологии CRISPR-on

Система CRISPR/Cas9 используется бактериями и археями для распознавания и расщепления вирусных последовательностей ДНК. Как показано в многочисленных моделях млекопитающих, эта технология редактирования генома также применима к птичьим клеткам и эмбрионам [86, 87].

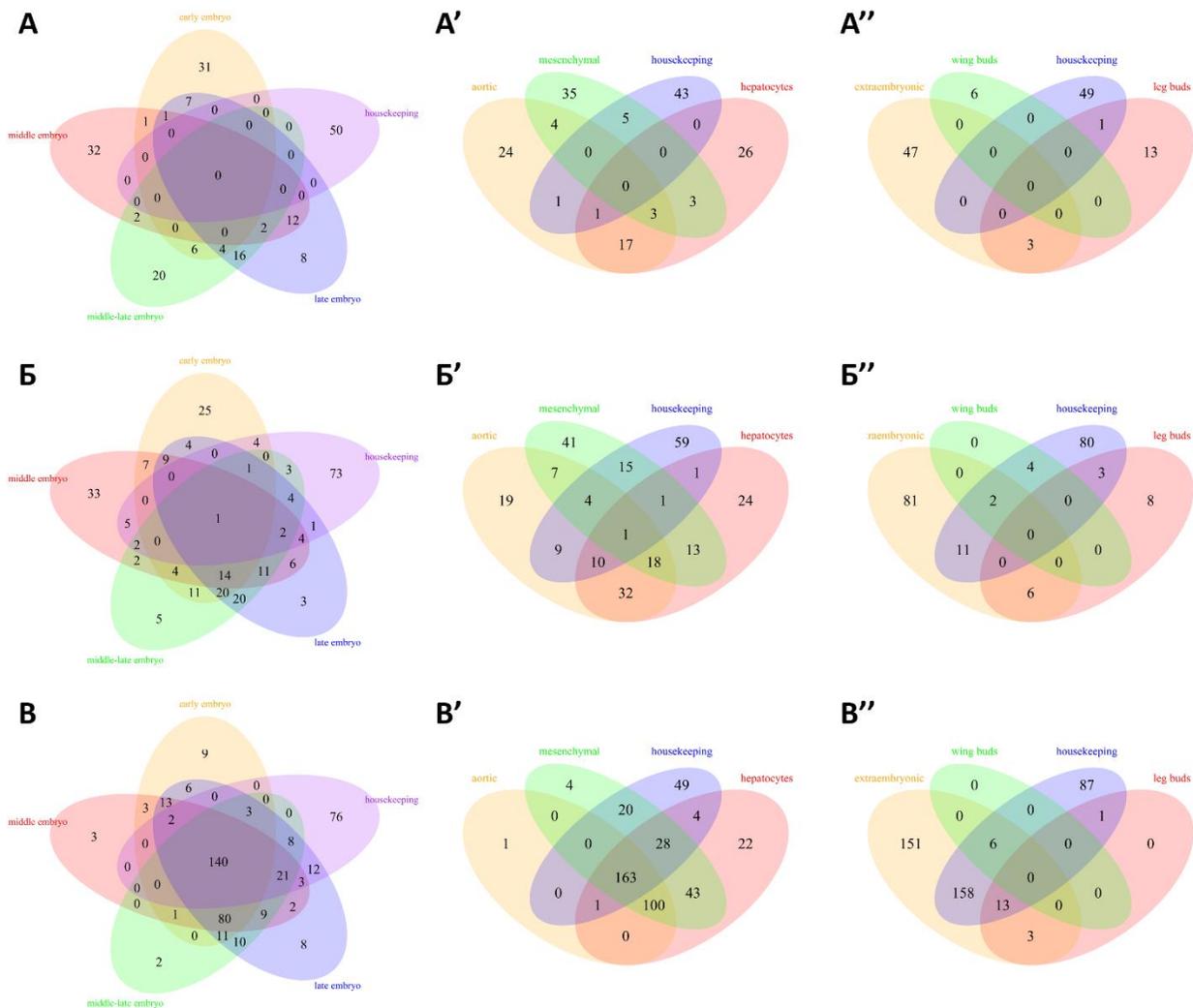


Рисунок 21 Диаграммы Венна для перепредставленных мотивов на различных этапах развития цыплят (А, Б, В), клетках (А', Б', В') и тканях (А'', Б'', В''). А – топ 50 мотивов, упорядоченных по E-value [64]. Б – топ 100; В – все мотивы с порогом E-value > 0.05.

Модифицированный белок Cas9 с его отключенной нуклеазной активностью и слитый с 10 копиями домена транскрипции VP16 смог запускать механизм инициации транскрипции в промоторах генов, для которых были подобраны гидовые РНК (sgRNA) [88] – метод «CRISPR-on» для CRISPR/dCas9 зависимой активации целевых генов. Предыдущие исследования показали, что максимальная эффективность CRISPR-on была достигнута за счет

использования нескольких sgRNAs, нацеленных на первые 200-300 пар оснований выше TSS, тогда как sgRNAs, нацеленные на последовательности, расположенные более удаленно, имели низкую эффективность, а те, которые были нацелены на последовательности после TSS, имели отрицательный эффект [88]. Метод CRISPR может быть использован для активации любых генов, многие из которых плохо подходят для обычного клонирования на основе плазмид и сверхэкспрессии. Его успех, однако, в значительной степени зависит от точного знания соответствующих TSS, и до сих пор этот метод не использовался ни в одном исследовании развития птиц или млекопитающих. Чтобы проверить, может ли наше картирование TSS в геноме в целом ускорить применение CRISPR в анализе развития цыплят, мы использовали ген *Brachyury* (*T*) для подтверждения принципа CRISPR-on опосредованной активации. Один единственный пик TSS был идентифицирован на 5'-конце куриного *Brachyury* (рис. 6A), ген, который является маркером предшественников мезодермы и аксиальных клеток мезодермы на ранних стадиях развития. Промоторную область анализировали на наличие подходящих структур sgRNA (crispr.mit.edu) [89] и определили 4 последовательности sgRNA, расположенных в пределах 300 п.о. от TSS (рисунок 23, А, Б). CRISPR-активированные сигналы *Brachyury* + широко и эктопически наблюдались в сомитах и клетках нервной трубки (рисунок 23, В, Г) в дополнение к своим нормальным доменам экспрессии в предшественниках первичной мезодермы и нотокордных клетках [90]. Несколько других тестируемых генов (*SNAI1* [91], *CDH1* [92], *MMP11* и *DENND5B*) были также надежно активированы с помощью CRISPR-on (рисунок 24-27), что указывает на то, что метод CRISPR (с помощью картирования TSS) может быть эффективно применен к исследованиям в области развития птиц.

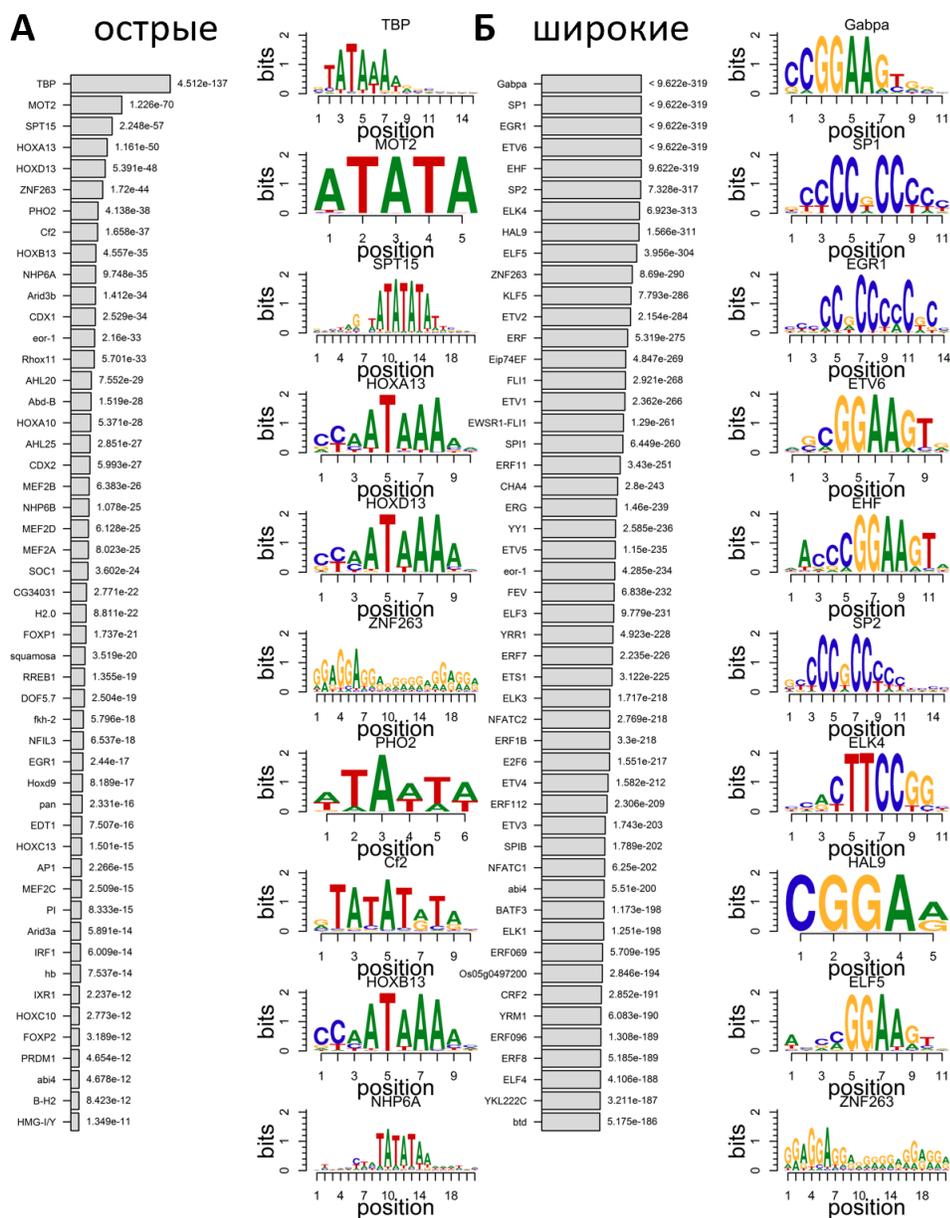
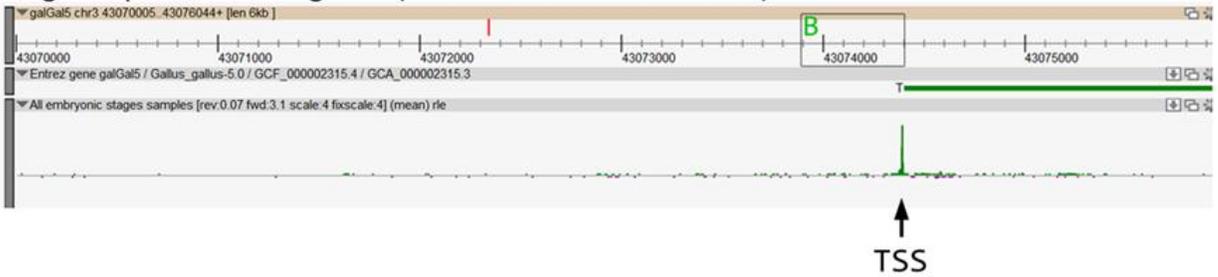


Рисунок 22. Сравнение топ 50 мотивов связывания транскрипционных факторов в промоторах с “широким” типом сигнала CAGE по сравнению с промоторами, имеющими “острую” форму сигнала. Также показаны весовые матрицы топ 10 мотивов для каждой группы. Острые промоторы связаны с А/Т богатыми мотивами, тогда как для широких в большей степени характерны Г/Ц мотивы.

A**T gene promoter region (chr3:4307006-43076044)****Б****CAGE-based sgRNA design for Brachyury CRISPR-On**

500bp sequence upstream of TSS (Blue:PAM; Red:sgRNA)

5'GATTTCTGCTCGTTGTACCTCTCGCCCTCTGCCAGAGATGCTCGCAGAGGGCAATAAACCGCTCCCTCCCTCGCACCGCCGCGGC
 CCCGAGGATGCCCATCCCGGGAGCACCTCGGCCCTCCCGCGCGAGCTCCGCCACACGGGGCCGGGCGCTACGGCCGGACGCGGGCG
 GCGGGAGCGGGGCGCAGCCGGGGCCGCGATGGGCCATTTCTCTCGGGCGCGGGCGGGCCCGGGCGCTTTGATGGAGGTGCAACATGT
 GGGGATTTGCATCAAACAGCCGCGCTGCGTCCCGCGGGCGGCAATTCCGGCGCGCCCCCGGGGCGCGCTCTGATATGGCCGCGGGG
 AGCCAATCGGGCACC CGGGCGGCACATCAAAGCGCGCTCCGCCAATCGGCGGGCGGCTCCCGCGGGCGCGGTGATGTACGGGGGGGCG
 TTTATGGCGGGTATAACGGCGGGCCCGGCCGGAGGCAGCGGGCGAG-3'

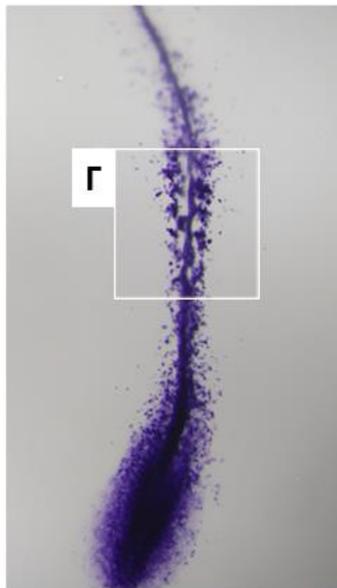
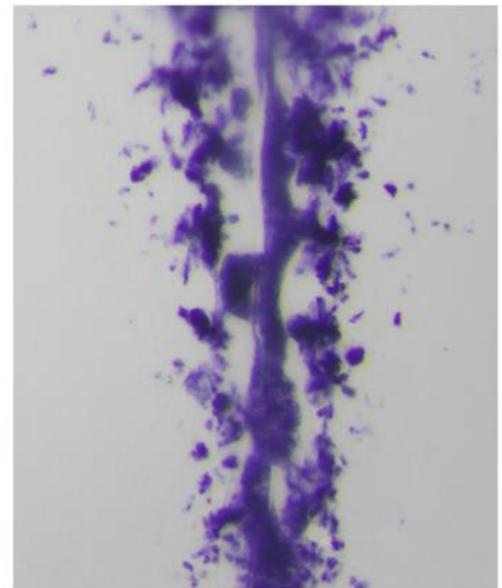
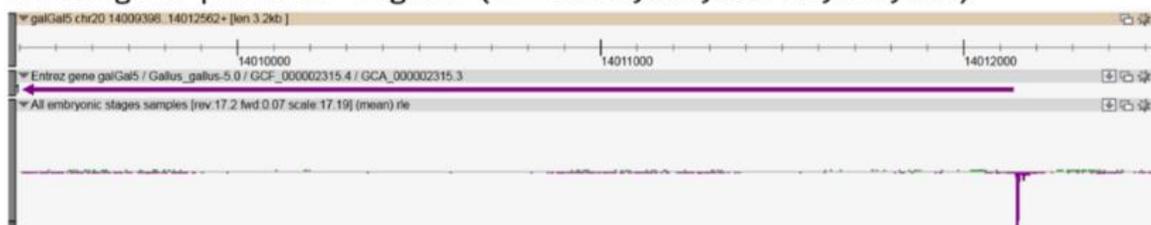
В**Г**

Рисунок 23. Индукция эндогенной экспрессии *Brachyury*, с помощью системы CRISPR-on. А – визуализация 5' концевого 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 коэлектротпорированного GFP. Далее проводилась *in situ* гибридизация для *Brachyury*; Д – приближенное изображение в средней области эмбриона, показывающая наличие продукта *Brachyury* в сомитах и нервной трубке.

А *SNAI1* gene promoter region (chr20:14,009,398-14,012,562)



поиск sgRNA в области 500 п.о. в области TSS гена *SNAI1*
(бирюзовый PAM, красный sgRNA)

5' GC**GAAGTGGGGCGAATCGCCTG**CGGGCCCGCCGCGACGACGACGTGCAACAAGGAGGAACGGGGCAGGGCCGGGCCCGG
CAACACGGGGCTCGAGAAGATTGCCAACCGCTCTCTTTGAGAGGGGGAGCGGGCCGCTCGAGCTGCAGCCGCGGGCAGTA
TTGCCGCGTTTCCCACT**TCTGTACTCGAGGCGGGTTG**CGGCCACCCAGGCTGGTGCCGAGGCAAGTGCGTGGCGCCATGCCC
CGGGCGGGAGAGCTGTCCGGCTCCGCCACCCCCCGCACCC**CCGGGGTCCGCTCGTAACGCAGG**GGAGGGCGAAGGTGGCTGTG
AGCTCGTGAGGACGGAGCAATGGAGGAGGACCGGCAGCGCTCTGGACCCTTTTCCAGCAGACGCTCCC**GAGCCGGGAGCAAC**
GATGCCCGGGCCGGCACCTGTAGTCTGCCCGCCCGGGCCGTCCCATTTGGCTCCGGGCATGGGGGGCGGCCCTGCACCGCCC
TCCTCGCCCC-3'

Б *CDH1* gene promoter region (chr11:19,113,996-19,116,013)

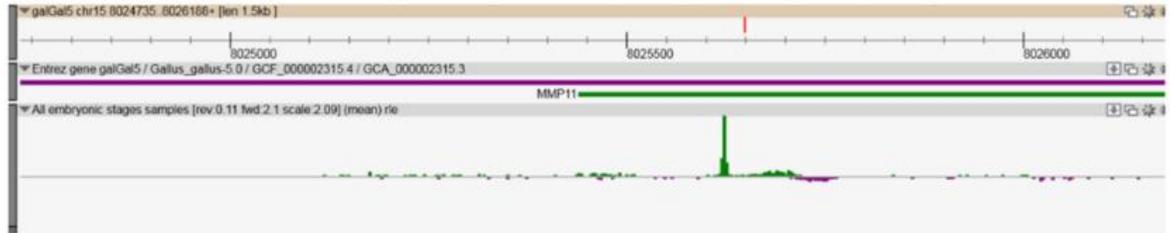


поиск sgRNA в области 500 п.о. в области TSS гена *CDH1*
(бирюзовый PAM, красный sgRNA)

5' CACCCCCAGGGCTCTCTGCCCCCTTGGCGCTCAGCAATAAGTGCATTAAACA**CTTAACTGCATTAAACCACGC**CGGCTGA
GTGCTGTGCTGAGGGGAAACGAGCAGCGCTGCGGCTGCACCAGGCCGGGGTAGAGCCGCGTCCAGCGCGGTAGGGG**CA**
GCCCGAACGTCGCGAGTGCGGACTCTGCCCGCTGCGGCCGCGGGGGCGGCGGCGGCACTGAGGGTTCGCCGCTATCCCGGG
TCGGAGCG**GAGCGCCGTAGGTTCGCGGAA**GGGACGGCAGCCGGAGGGAGAGGACCGACCCCG**CGCCCCCTCATCCCCCGT**
GGGCCGCGTGCCCGTGAAGGCGCGGTGGGACGGGGCGGGCGGGACGGGACGGCCCCGGTGAGGGGAGCGGGGGCGGGGAGG
GGCCCCGGCGGGGCCAACACCT-3'

Рисунок 24 Дизайн гидовых РНК в CRISPR-оп эксперименте для генов *SNAI1*, *CDH1*, *MMP11* и *DENND5B*. А - *SNAI1*; Б - *CDH1*.

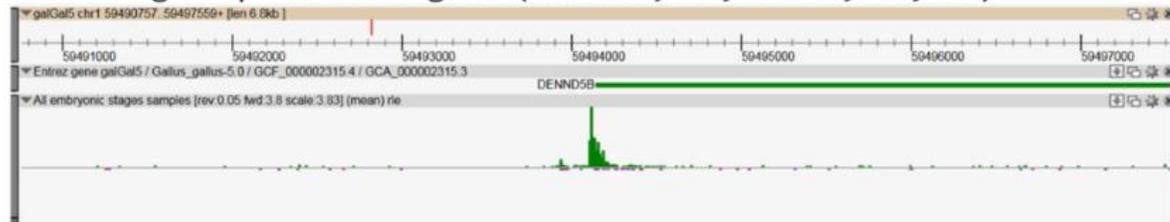
В *MMP11* gene promoter region (chr15:8,024,735-8,026,188)



поиск sgRNA в области 500 п.о. в области TSS гена *MMP11*
(бирюзовый PAM, красный sgRNA)

5' CGCGTGTGCAACTTGCAGCCAGGTGCTGCCTCACTGTACATGCAGACGGCCGGAGGCGAAACAGCATCGTCAGCCCGAGG
GCAGCAGCGCACAGCCCCGACCTCCTTTTCTCCCTTCTCCCTCGTCTCCTCCTCCTCCTCCTCCTCCGCTCCGGCCGGGGCTG
CAGCCTCCGGCGGCCCGCGCCGTTCGGCCGCGGCTCCCGCGCCTCACTCCCTCCCTCCCTCCACGGGGGCATGAAGCCC
GGGGGGCCGGGAGCGCGGGGGCGCCCCGTCCCGGGCGGCTCCGCCCCGATACCGACGGGCCGCATCCCTCGGCCCG
CAGGGCTGTAACCCGACGAGGGGAGAACA AAAAGGCAGGAGCGCGGGCGCGCGGGCCGAGAGGGAGGGGAAGGGAGGGCGC
GGGGGAGGGCAGGCGCCGCGGGGGGATAAAAAGGCGCTGCGGGACCCCCCGCCCCATGTCGAGCCCGCCCGCCCGG
CGCTCCTCGC-3'

Г *DENND5B* gene promoter region (chr1:59,490,757-59,497,559)



поиск sgRNA в области 500 п.о. в области TSS гена *DENND5B*
(бирюзовый PAM, красный sgRNA)

5'ACTTTGTATAAATACCTGGGATTGCCCGACCCAAAGTGCAGCACCTGCACCTTGATCTTGTGAACCTCATTAGGTTTTCA
TGGCCCCACTTCTAGCCTGAATGACAAAGTGCTTTGGATTTTGCCTCGTGTATGCAACAAGGAGACAGCTGTCGGTATG
ATAACCGTGATATGATTAAACCATTTATGGTAGGTCAGCGGAGACCTCCTCTGCCACGCACATCCATCACCCGCCCTC
CGCGGCTCATTCTGTCGTTCCGGTCCTCAGCCTAACAAAAACCCGCCCTCAGCGCCATTCCACAGGAAGACGCTAC
GGGACGCCGAGCGCTGCTCCCGCCCGCTGTTTTTGTCTGGAAGCCGCCCGCCCCAGTCACGCCGCCCGCCCAAAACCAA
GCACGCCGGGGTCCCGACGGGGCGGTGCCGACCCGCGAGCGTAAAAAGCCCCGC-3'

Рисунок 25. Дизайн гидовых РНК в CRISPR-оп эксперименте для генов *SNAI1*, *CDH1*, *MMP11* и *DENND5B*. В - *MMP11*; Г - *DENND5B*. Последовательности гидовых РНК (выделены красным) были выбраны в пределах 500 п.о. вверх от позиции соответствующих TSS. Бирюзовым отмечен PAM.

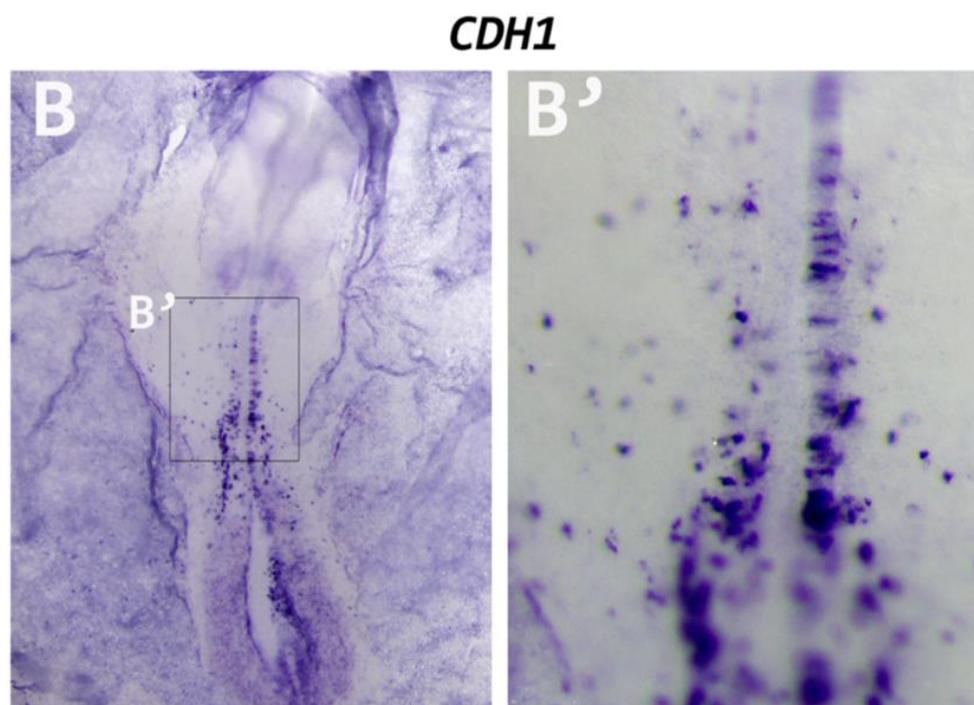
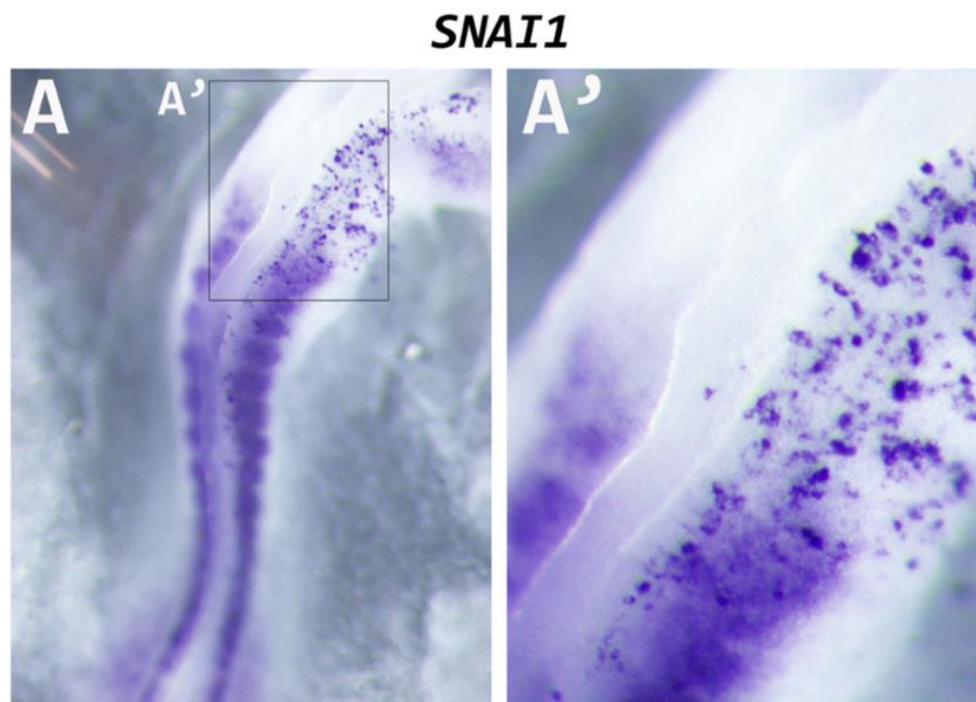
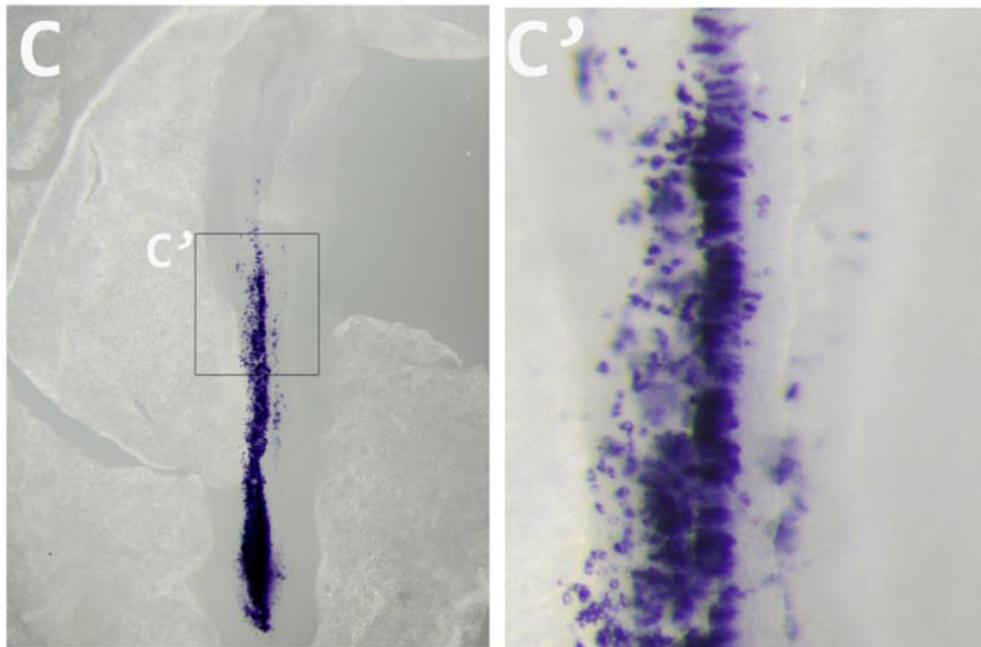


Рисунок 26. CRISPR-оп опосредованная индукция экспрессии генов *SNAI1* (A, A') и *CDH1* (B, B'). Хорошо видна эктопическая экспрессия в выделенных областях, определенных по коэлектрпорированному GFP.

MMP11



DENND5b

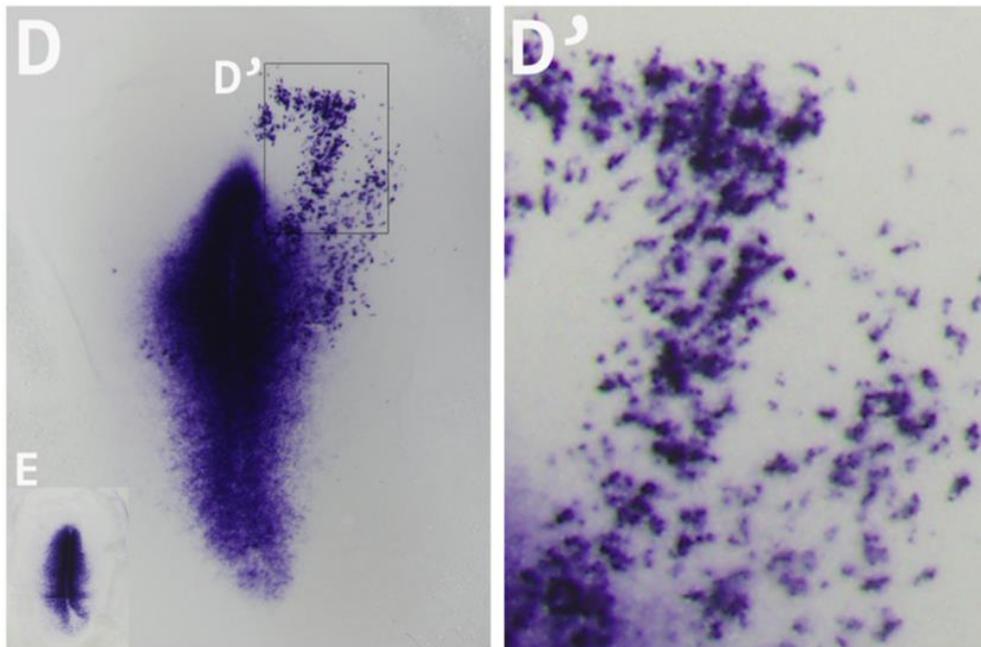


Рисунок 27. CRISPR-оп опосредованная индукция экспрессии генов *MMP11* (C, C') и *DENND5B* (D, D'). Эктопическая экспрессия наблюдалась для всех выбранных генов. E – эндогенная экспрессия *DENND5B*.

Выводы по главе «Промотором эмбриогенеза амниот»

В прошлом TSS были исследованы с помощью защиты нуклеазы, расширения праймера и 5' концевой анализа RACE [93–95]. В эпоху высокопроизводительного секвенирования профилирование транскриптомов с помощью RNA-seq значительно расширило способность исследователей определять структуру транскриптов в полногеномном масштабе. Однако обычные методы РНК-seq нацелены на случайные фрагменты РНК и не подходят для точной идентификации TSS [96]. Технология CAGE, нацеленная на 5' концевые фрагменты РНК, стала наиболее широко распространенным методом картирования TSS на геном по сравнению с другими подходами, такими как dRNA-seq [96], TagRNA-seq [97], RAMPAGE [98] и SMRT [99]. Применяя технологию CAGE, мы создали полногеномные профили TSS для куриных эмбрионов, охватывающих весь период развития от прегастрюляции до вылупления. 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 прокладывают путь для уточнения аннотации генома цыплят в сочетании с последним улучшением сборки генома [30] и для анализа системного уровня (на уровне регуляции транскрипции) и 5' некодирующих областей генов, имеющих важное значение для развития. В качестве подтверждения, мы продемонстрировали, что профилирование TSS позволило нам найти перепредставленные специфичные TFBS-мотивы и добиться высокой эффективности в CRISPR-опосредованной

транскрипционной активации регулятора мезодермы. Таким образом, наша работа — это первый пример использования метода CRISPR в исследованиях развития, применимый, в принципе, к любому гену. CRISPR-on и аналогичное CRISPR-опосредованному ингибированию транскрипции [100] и CRISPR-опосредованным методам эпигенетической модификации [101], каждый из которых требуют точного знания TSS, могут революционизировать исследования в области развития в модельных системах, не поддающихся традиционному генетическому анализу.

Повторяющейся проблемой при высокопроизводительном анализе является гетерогенность клеток независимо от источника образца (например, опухоли или эмбриона). С другой стороны, эта проблема решается в области разработки алгоритмов вычислительной деконволюции [89, 102, 103], применения нано и пикограммных платформ РНК-seq, в том числе CAGE-адаптированных nanoCAGE и C1 CAGE [104–106]. В отличие от опухолевых тканей, которые неоднородны по своей природе и переменны в их относительном клеточном составе, эмбрионы состоят из высоко стереотипических и воспроизводимых популяций клеток, причем их дальнейшие пути развития и роли хорошо обозначены. В нашем текущем анализе ранние стадии эмбрионов содержат относительно немного различных популяций клеток, тогда как эмбрионы поздней стадии состоят из сотен типов устойчивых клеток. Поэтому ранние эмбриональные ткани идеально подходят для тестирования и оптимизации алгоритмов деконволюции. Чувствительность в идентификации пика CAGE также подразумевает, что популяции мелких клеток можно легко обнаружить и количественно определить, как в простых, так и в сложных смешанных популяциях. Например, клетки куриных зародышей представляют собой крайний случай, на некоторых стадиях развития процент стволовых клеток может достигать 0.1% в отдельной ткани [107]. Тем не менее TSS, представляющие маркеры, специфичные для зародышевых

клеток, такие как DDX4 (CVH), все еще могут быть обнаружены со стадии HH1 до стадии HH6, что указывает на то, большая часть событий специализации линии, даже во время позднего эмбриогенеза (когда сложность ткани резко возрастает), фиксируются в нашем анализе. В совокупности мы делаем вывод о том, что наши исследования TSS, основанные на CAGE, а также верификация TSS, будут значительно способствовать исследованиям развития птиц и млекопитающих. Вместе с будущим улучшением методов деконволюции, создания библиотек CAGE и РНК-секвенирующих платформ, эмбрионы птиц представляют редкий физиологический сценарий для системного изучения онтогенетической эволюции спецификации клеточных линии.

Геномный проект по птицам (Avian Phylogenomics Project) уже содержит 48 собранных доступных для скачивания сборок [98], каждая из которых может быть дополнительно аннотирована с использованием CAGE. Кроме того, геномный проект 10К геномов включает большое количество секвенированных видов птиц и рептилий (около 5000 и 3000, соответственно), которые также будут аннотированы в дальнейшем [27, 55]. 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. Партеогенез 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 [27]. 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 [77] и проект птичьего

филогенома могут быть дополнительно усовершенствованы путем интеграции с данными CAGE, для уточнения генных аннотаций, точно определяя положение TSS и энхансеров [17, 78], и таким образом улучшить наше понимание эволюции признаков.

Первые данные CAGE цыплят, полученные в FANTOM5 в проекте, выявили ключевые регуляторные гены на различных этапах развития и связанные с ними пути.

ГИПОТЕРМАЛЬНАЯ ДИАПАУЗА

1. Профилирование промоторов, регулирующих гипотермальную диапаузу куриных эмбрионов

Существующий в настоящее время рептилии (чешуйчатые, крокодилы, черепахи и гаттерия) обладают самым разнообразным набором стратегий для продления стадии яйца, тогда как у млекопитающих и птиц эти механизмы заменяются родительской заботой. Так частота остановки развития снижается у видов млекопитающих, обеспечивающих большую родительскую опеку, а эмбрионы птиц обычно лишь замедляют развитие, когда родители не ухаживают за ними. Богатое изобилие стратегий для остановки развития у рептилий позволяет эмбрионам противостоять изменяющейся инкубационной среде в различных экологических условиях. Большинство видов рептилий подвергаются по меньшей мере одному периоду остановки развития во время пребывания в амниотическом яйце, а в некоторых случаях как до, так и после яйцекладки [41]. Задержка эмбрионального развития у млекопитающих, опосредованная большим количеством внешних и внутренних факторов, в том числе гормонально, была хорошо изучена [108], тогда как понимание задержки развития у яйцекладущих рептилий и птиц остается крайне ограниченным. Более того, птицы, занимающие промежуточное положение между рептилиями и млекопитающими в плане регуляции диапаузы и родительской заботы, могут открыть эволюционные особенности такой адаптации.

В данной работе мы проанализировали транскриптомы CAGE раннего (НН1, НН4), среднего (НН16) и средне-позднего (НН28) эмбриогенеза цыплят, а также оценили транскрипционный ответ эмбриона при его охлаждении и восстановлении (рисунок 28). Кластеризация суммарного сигнала CAGE для всех экспериментов позволила идентифицировать 18304 промоторных региона с

соответствующими TSS, 63% из которых перекрываются с промоторными областями определенными в рамках проекта Chicken-FANTOM [60]. Значительное количество CAGE кластеров (82.5%) было связано с транскриптами и соответствующими генами (всего 8052). Используя предсказанные промоторные участки, мы смогли показать, что куриные эмбрионы активно реагируют на уровне изменения транскрипции в ответ на охлаждение на всех исследованных точках эмбриогенеза. Количество промоторов, классифицированных как дифференциально экспрессирующиеся ($FDR < 0.05$) варьировало от 488 до 1535 в группе понижающих активность (*down-regulated*, $\log FC < 0$), и от 259 до 979 в группе повышающих активность в ответ на охлаждение промоторов (*up-regulated*, $\log FC > 0$) (рисунок 29). Среди промоторов, статистически значимо реагирующих на понижение температуры 329 оказались универсальными (рисунок 30, A), то есть меняющими уровень своей экспрессии в ответ на данный стресс на всех этапах развития куриного эмбриона. Этим промоторам соответствуют 317 ассоциированных генов. Анализ перепредставленности функциональных тегов GO для специфичных генов, связанных с гипотермальной диапаузой представлен на рисунке 35 и 40, из которых видно, что процессы происходящие на разных этапах развития при охлаждении эмбриона пересекаются между собой незначительно. Среди общих тегов можно отметить “mitochondrion” (митохондрия) и “aryl hydrocarbon receptor complex” (рецептор ароматических углеводородов, Ahr).

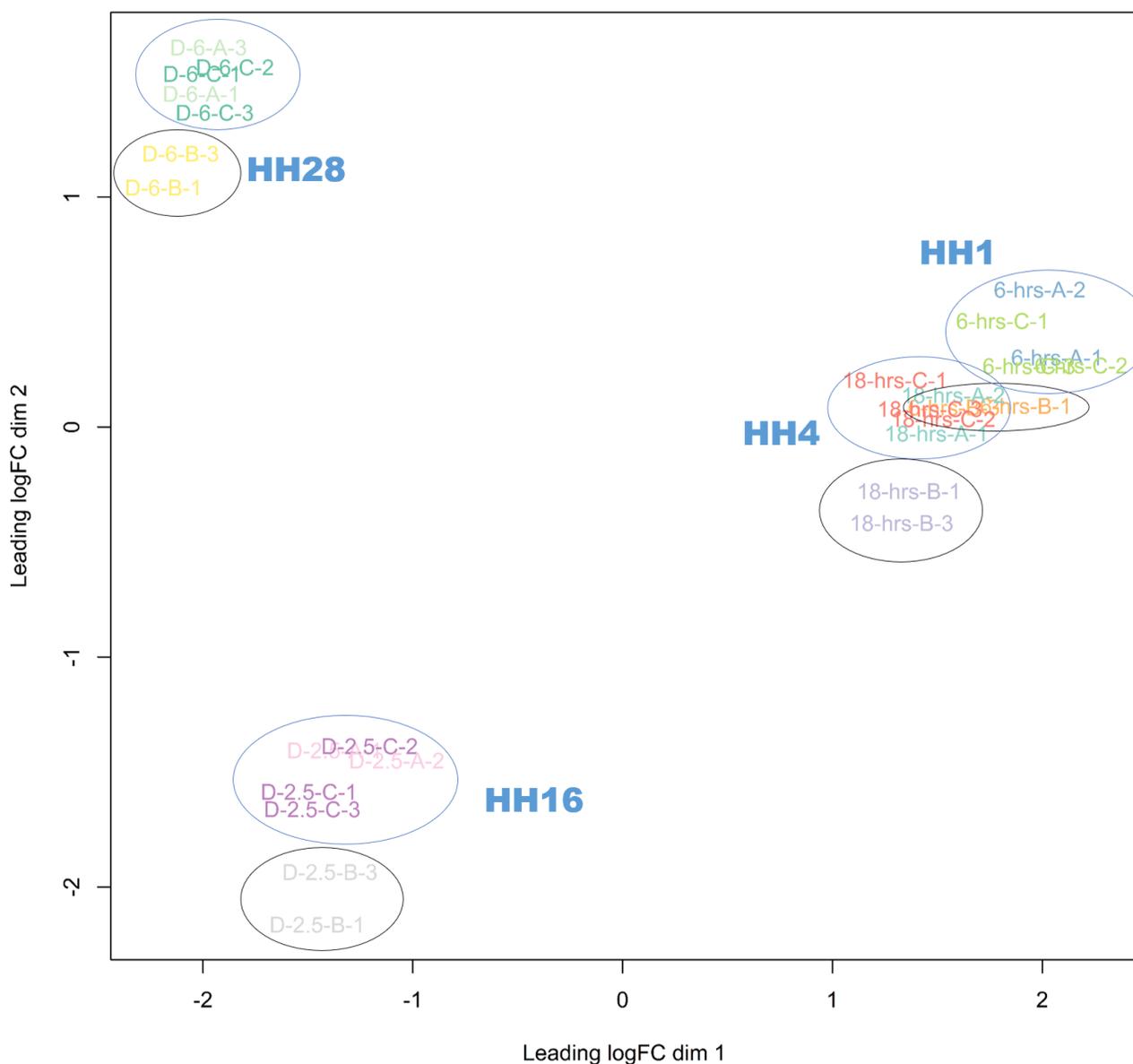


Рисунок 28. График многомерного шкалирования MDS [56] для образцов в эксперименте с индукцией гибернации. Образцы А – контроль (нормальное развитие эмбриона), В – охлажденные до 15 °С эмбрионы (24 ч индуцированная гибернация), С – эмбрионы после гибернации (восстановление при нормальных условиях 1 ч). Видно, что разница между различными состояниями эмбриона (А, В, С) значительно меньше, чем между различными этапами эмбрионального развития, даже относительно близкими (6-18 ч).

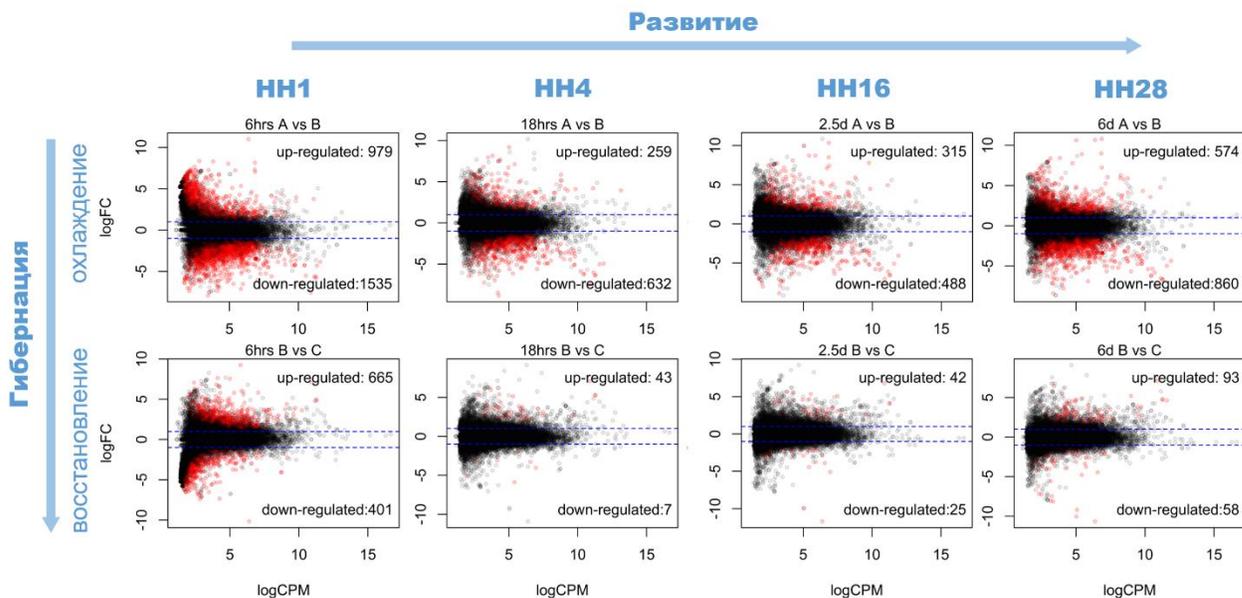


Рисунок 29. Анализ дифференциальной экспрессии промоторов (18304 CAGE кластера) между различными состояниями эмбриона: контроль-А против индуцированной гибернации-В (охлаждение), и охлажденные эмбрионы-В против вновь нагретых до нормальной температуры-С (восстановление), на ранних (HH1, HH4), средних (HH16) и средне-поздних (HH28) этапах развития. Красным отмечены промоторы с $FDR < 0.05$. Значение \log_{FC} (логарифм от разницы изменения экспрессии в эксперименте против контроля) более 0 соответствует случаям повышения активности (up-regulated), менее 0 – снижению (down-regulated). Видно, что разница между образцами В и С (восстановление) небольшая по сравнению с экспериментом по охлаждению А против В, что вероятно связано временем, необходимым для запуска всех путей, отвечающих за возобновление процессов эмбрионального развития.

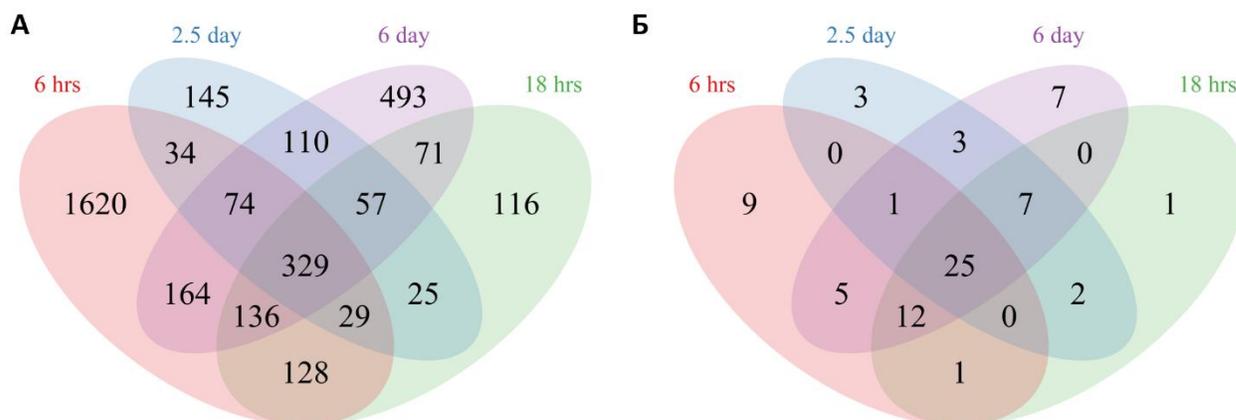


Рисунок 30. Особенности регуляции транскрипции в эмбрионах курицы в ответ на охлаждение на различных этапах эмбриогенеза. А - Диаграмма Венна для дифференциально экспрессирующихся промоторов в эксперименте по охлаждению эмбрионов (А против В). Выбраны статистически значимые случаи $FDR < 0.05$, повышающие и понижающие уровень экспрессии. Имеются как специфичные для отдельных этапов развития промоторы (от ~100 до 500), так и те, которые ведут себя одинаково в ответ на охлаждение на всех этапах – 329; Б - Диаграмма Венна для мотивов связывания транскрипционных факторов, позиционно перепредставленных [109] в промоторах меняющих экспрессию в ответ на охлаждение на различных этапах развития эмбриона. Промоторы, имеющие специфичную регуляцию на отдельных этапах 6, 18 ч и 2.5, 6 дней, имеют относительно небольшое количество специфичных транскрипционных факторов – 9, 1, 3 и 7, соответственно. При этом видно 25 универсальных транскрипционных фактора, которые участвуют в адаптации на всех исследованных этапах развития.

2. Транскрипционные факторы, регулирующие переход к гипотермальной диапаузе

Из анализа обогащения мотивов мы обнаружили 25 общих транскрипционных факторов: EGR4, ELK1, ELK3, ELK4, ETV1, ETV5, ERF, FLI1, Gabra, KLF14, KLF16, KLF5, MZF1, NRF1, SP1, SP2, SP3, SP4, SP8, TBP, YY1, YY2, ZNF740, NFYB, ZBTB7A (рисунок 30, Б; рисунок 31), большинство из которых имеют GC богатые последовательности. Особенно примечателен SP1, который, как известно, играет важную роль в дифференциации, реагирует на повреждение ДНК и окислительный стресс [110–112], в комплексе с MZF1 регулирует гемопоэз [113]. Эмбрионы мышей лишённые Sp1 умирают в течение 10 дней, а без Sp3 – при рождении [114, 115]. Другие факторы семейства Sp регулирующие активность различных процессов, включая дифференциацию, гемопоэз, морфогенез, развитие скелета эмбриона [116–118], оказались вовлечены в регуляцию гипометаболизма на всех этапах развития (Рисунок 31). В группу Sp факторов также входят другие ГЦ/ГТ связывающие белки KLF (Krüppel-like factor), регулирующие дифференцировку, пролиферацию, развитие клеток эпителия, жировой ткани, клеток скелетной мускулатуры [119]. Gabra регулирующий активность цитохром оксидазы митохондрий также образует комплексы с вышеупомянутыми факторами Sp и выступать в роли активаторов транскрипции [120]. YY2 – многофункциональный транскрипционный фактор, осуществляющий как положительную, так и отрицательную регуляцию активности большого количества клеточных и вирусных генов, путем связывания области инициации транскрипции. Влияние на регуляцию транскрипции зависит от контекста, в котором он связывается транскрипционный фактор – это может быть прямая активация или репрессия, но также опосредованная за счет дополнительных кофакторов или метилирования, изменения конформации ДНК. YY2 может играть важную роль

в развитии и дифференциации, участвует в репарации ДНК [121, 122]. EGR4 - транскрипционный регулятор, распознающий и связывающий последовательность 5'-GCGGGGGCG-3' (GSG) активирует транскрипцию генов-мишеней, продукты которых необходимы для митогенеза и дифференцировки [123]. Промоторы, чувствительные к понижению температуры также часто находятся под регуляцией транскрипционных факторов ELK и ETV, участвующих в дифференцировке [124, 125]. Интересно отметить, что функциональная роль транскрипционного фактора ZNF740, наиболее представленного в данном анализе, еще не установлена.

Кроме GC богатых мотивов, в группу универсальных мотивов связывания транскрипционных факторов в промоторах, отвечающих за адаптацию эмбрионов к охлаждению также вошел ТАТА бокс, который характерен для промоторов с острой формой сигнала CAGE в области TSS. Такие промоторы имеют специфичную по времени и по тканям экспрессию в эмбриональном развитии [16], поэтому присутствие ТВР связывающего мотива в данной группе промоторов закономерно. Еще более примечательно то, что его представленность возрастает в ходе развития, то есть непосредственно связана с усложнением строения эмбриона, появлением новых типов тканей (Рисунок 32).

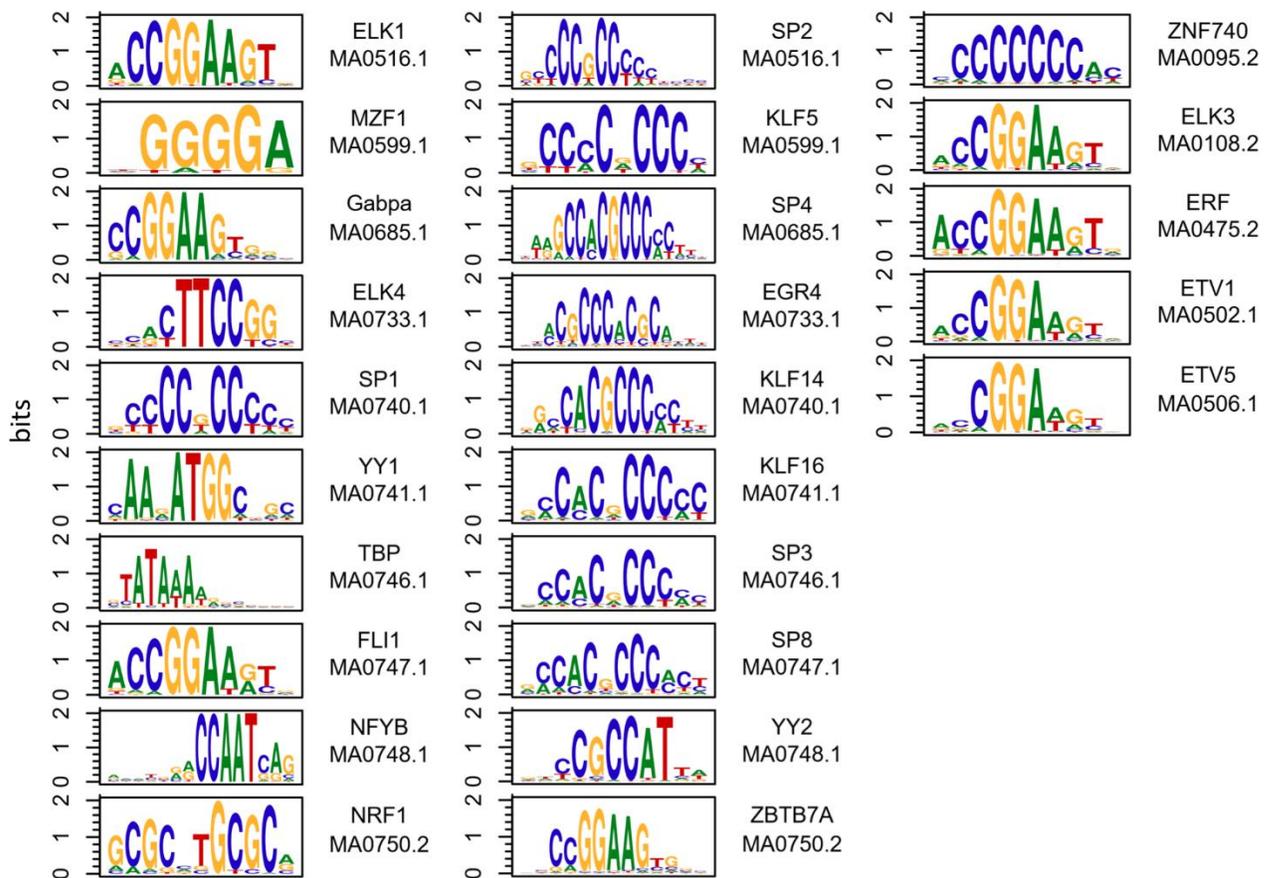


Рисунок 31. Визуализация весовых матриц 25 универсальных мотива из рисунка 30, Б, связывающие соответствующие транскрипционные факторы (указаны справа, а также ID мотива из базы JASPAR2018 [126]).

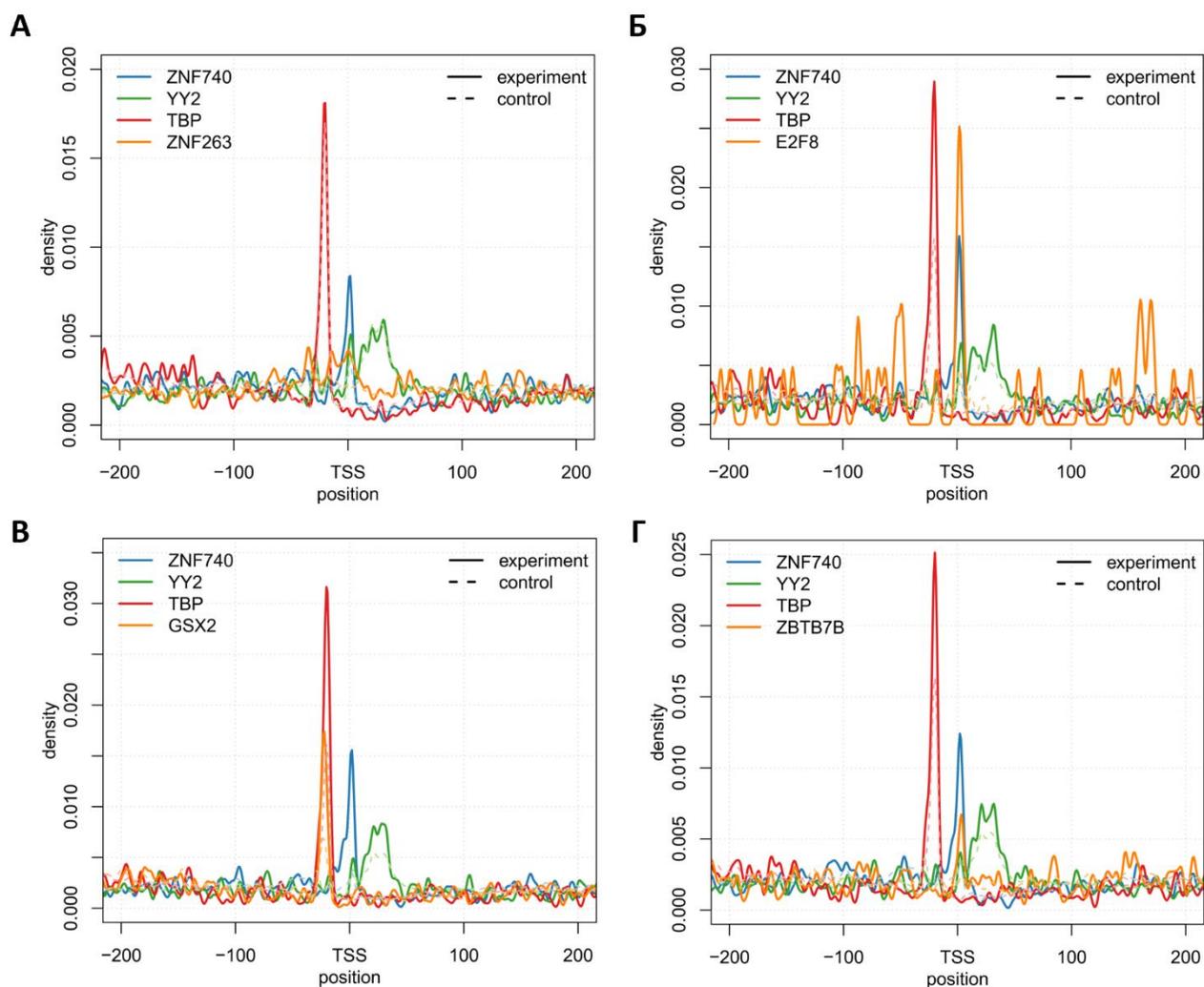


Рисунок 32. Визуализация позиционной перепредставленности мотивов в группах промоторных регионов (± 200 п.о.) имеющих дифференциальную экспрессию в эксперименте по охлаждению (А против В) на этапах развития HH1 (6 ч) – А, HH4 (18 ч) – Б, HH16 (2.5 д) – В, HH28 (6 д) – Г. Показаны 3 консервативных мотива – ZNF740, YY2, TBP, а также наиболее перепредставленные специфичные кандидаты с минимальным p -value, в случае А и В это 1 (ZNF263) и 2 (PLAG1 и MEF2B) специфичных мотивов соответственно. Для Б и Г мотивов 82 и 10, поэтому показан один в качестве примера.

1. Обратимый ответ промоторов в цикле норма-охлаждение-восстановление

Восстановление в течение 1 часа приводило к изменению экспрессии относительно небольшого количества промоторов на стадиях после НН4: от 50 до 151 (рисунок 29), но этого было достаточно для обнаружения генов с обратимым профилем экспрессии (рисунок 33). При этом на самом раннем исследованном этапе – 6 ч (НН1), одного часа восстановления при нормальной температуре было достаточно для активации равноценного количества генов. Промоторы и соответствующие им гены с циклическим профилем экспрессии (активация – рисунок 33, Д; инактивация – рисунок 33, Е) представлены на рисунке 34. Среди двух, повышающих экспрессию промоторов, один связан с неизвестным транскриптом (рисунок 34, Б), тогда как другой – с геном *Igfbp5*, кодирующим IGF-связывающий белок, который регулирует множество процессов, связанных с дифференцировкой, делением, миграцией клеток, а также старением. Было также показано, что данный белок участвует в обратном ингибировании пути PI3K-Akt-mTORC1 через воздействие на каскад ИФР1, отвечающий за рост органов и организма в целом [127].

Группа генов, понижающих экспрессию представлена генами *SMS*, *HIST1H3H*, *GCSH*, *ATP6V0C*, *HP1BP3*, *MEAF6*, *TUBA1C*, *SLC25A5*, *PTCH2* (Рисунок 34, В), кодирующие компоненты цитоскелета, нуклеосом, либо плохо охарактеризованные белки.

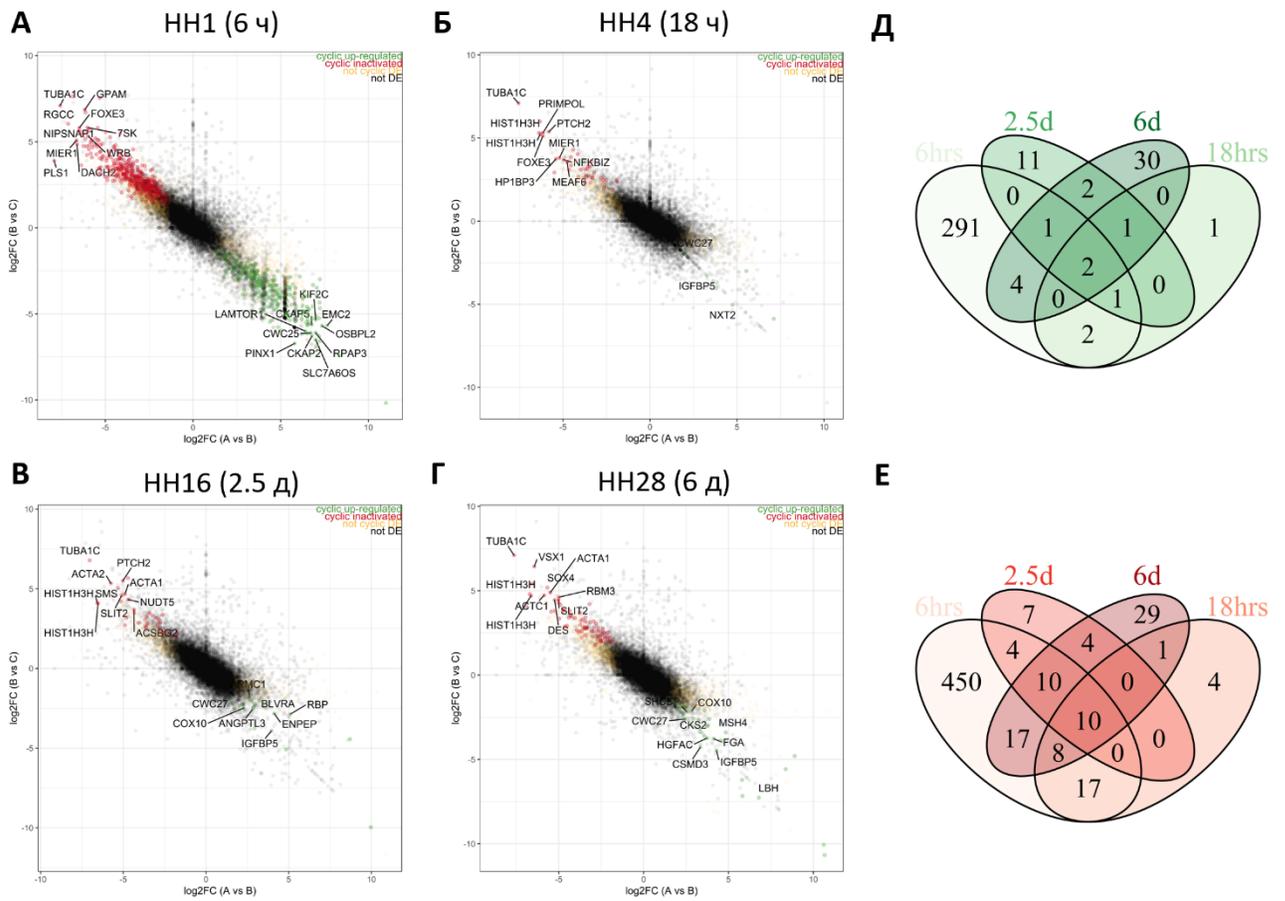


Рисунок 33. Обратимая дифференциальная экспрессия идентифицированных CAGE кластеров и соответствующих им генов. По оси X восстановление (recovery), по оси Y - охлаждение (cooling). Зеленым отмечены промоторы, повышающие активность в ответ на снижение температуры и снижающие при восстановлении, красным – наоборот, желтым – промоторы с однонаправленной дифференциальной экспрессией, черным – не меняющие статистически значимо своего уровня экспрессии (FDR>0.05).

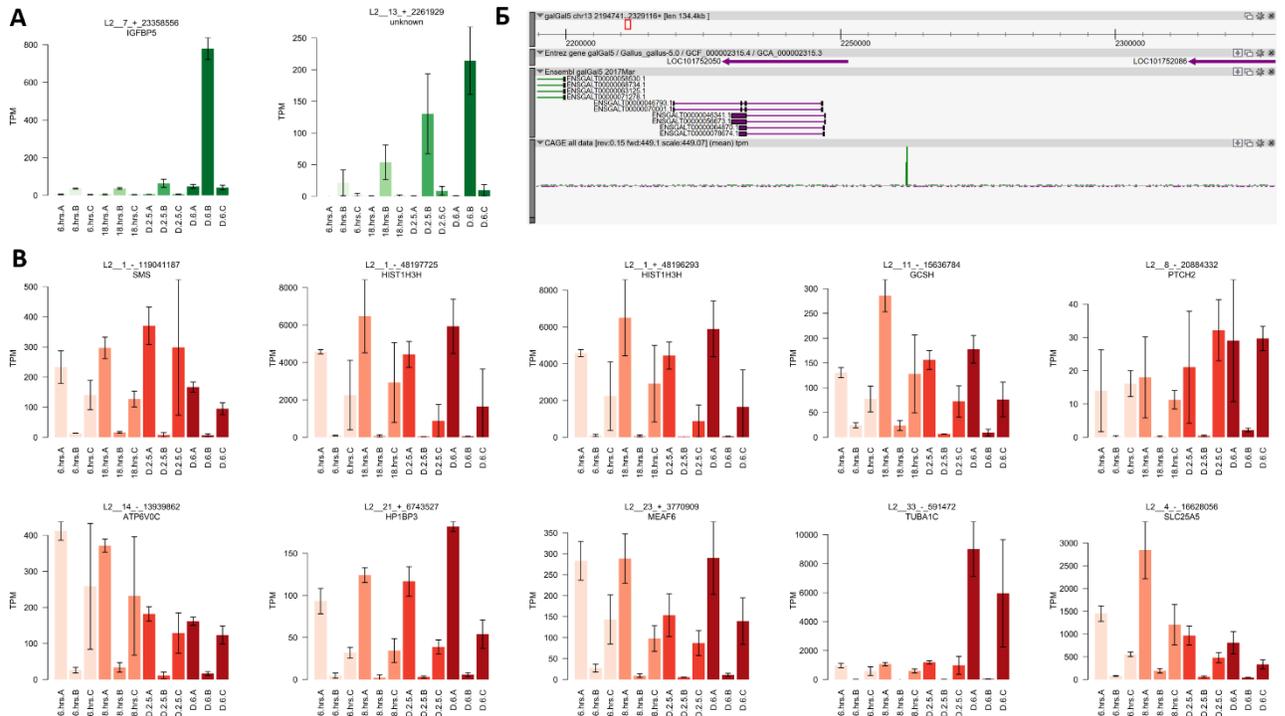


Рисунок 34. Промоторы с универсальным обратимым профилем экспрессии в экспериментах по индукции гибернации куриных эмбрионов. Зеленым (А) отмечены промоторы, повышающие активность в ответ на охлаждение, красным (В) – понижающие (см. рис. 33). Б – визуализация локуса промотора L2_13+_2261929, для которого не было обнаружено соответствующего аннотированного транскрипта/гена.

2. Функциональный и ко-экспрессионный анализ

Анализ перепредставленности функциональных GO тегов в экспериментах с охлаждением выявил универсальные компоненты, связанные с митохондриями и комплексом рецептора Ahr. Топ специфичных тегов представлен на рисунке 35 и характеризует особенности регуляции при адаптации к охлаждению на разных этапах развития.

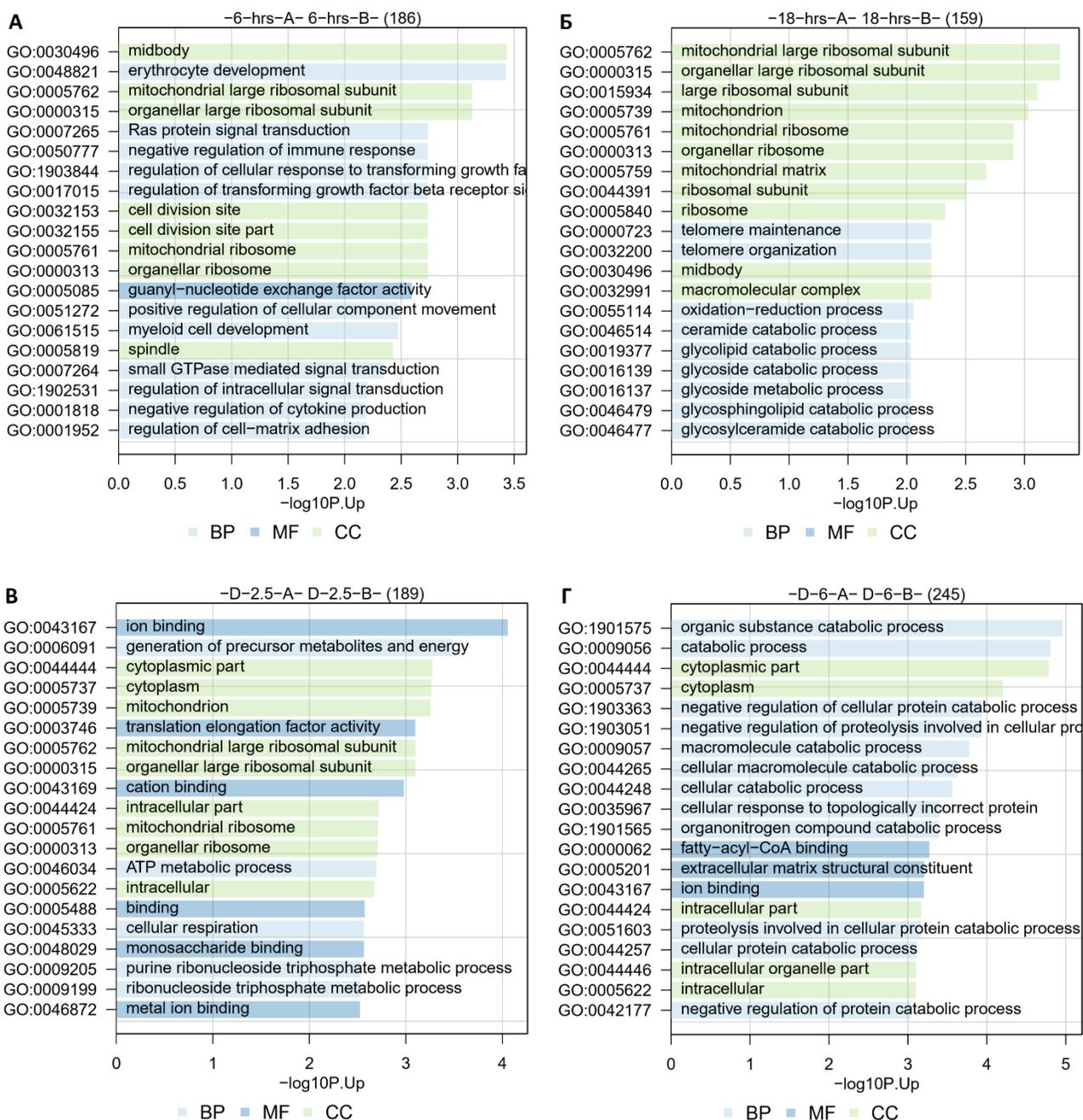


Рисунок 35. Топ 20 функциональных GO тегов, перепредставленных в генах, активирующихся в ответ на охлаждение куриных эмбрионов (сравнение группы А против В), на разных этапах развития: А – НН1 (6 ч), Б – НН4 (18 ч), В – НН16 (2.5 д), Г – НН28 (6 д). В круглых скобках указано общее количество GO тегов с p-value < 0.05. BP – биологический процесс, MF – молекулярная функция, CC – клеточный компонент.

Анализ ко-экспрессии на уровне генов позволил классифицировать их по профилю экспрессии на 21 кластер/модуль (Рисунок 36, А). Чтобы определить модули, имеющие отношение к адаптации эмбриона ~~при~~ к понижению температуры, мы оценили представленность дифференциально экспрессирующихся генов в этих модулях (Рисунок 36, Б). Два модуля оказались универсальными для всех стадий развития – желтый (*yellow*) и оранжево-розовый (*salmon*), тогда как другие оказались вариабельными, с небольшим количеством генов, вероятно, отражающие специфику адаптации на конкретных этапах.

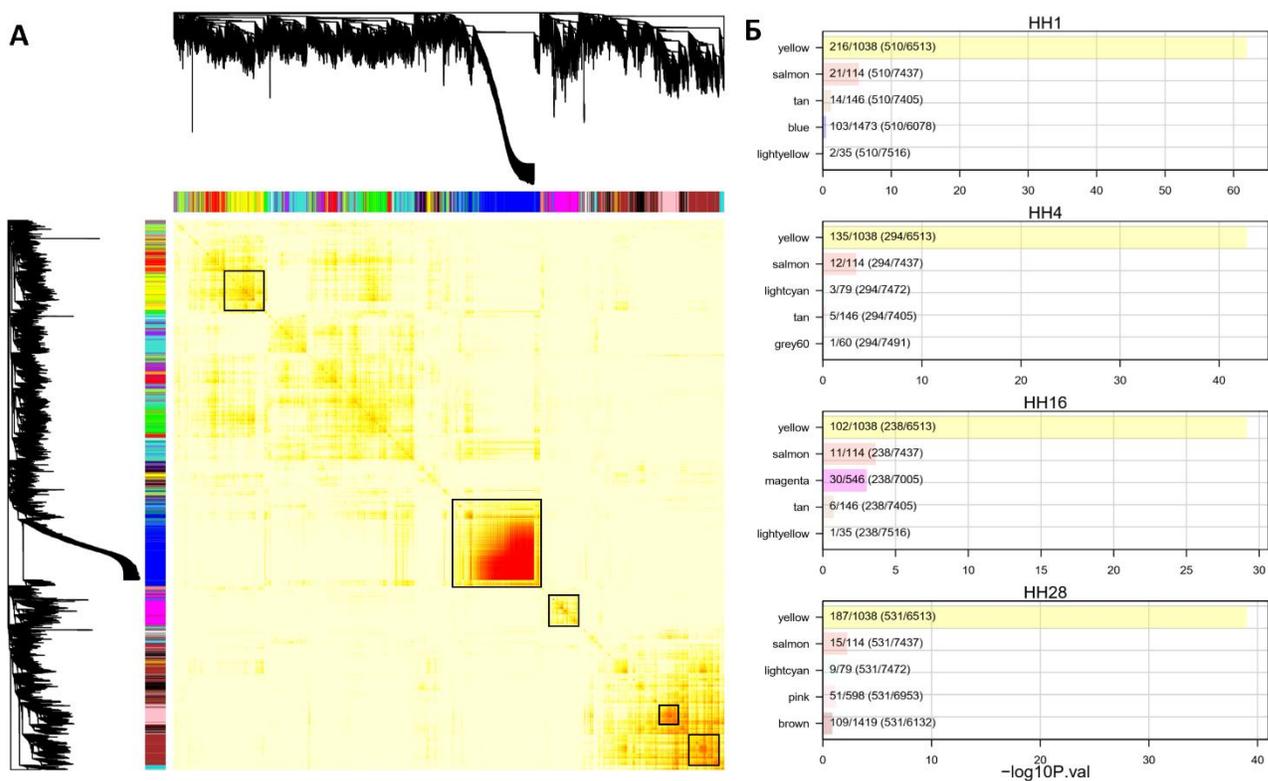


Рисунок 36. Сеть ко-экспрессирующихся генов в эксперименте с охлаждением эмбрионов курицы на разных этапах развития (HH1,4,16,28). А – тепловая карта на основе топологического перекрытия профилей экспрессии; Б – топ 5 кластеров в которых сконцентрированы дифференциально экспрессирующиеся гены из сравнения А против В на этапах HH1-HH28, цифрами указано количество генов значимо меняющих активность и относящиеся к конкретному кластеру/количество генов в кластере (количество генов меняющих значимо активность/количество всех активных генов вне кластера).

Для оценки функциональной роли данных кластеров был проведен анализ представленности GO тегов в следующих модулях: желтый (yellow), пурпурный (purple), синий (blue), розовый (pink), коричневый (brown), оранжево-розовый (salmon), желтовато-коричневый (tan) (Рисунок 37). Для двух последних значимых тегов не обнаружено. Для желтого модуля, собравшего наибольшее количество генов, отвечающих на стресс, было обнаружено два принципиальных тега - внутреннее пространство органелл (люмен) GO:0043233 и ядерные спеклы GO:0016607, что связывает данный модуль с регуляцией транскрипции на уровне альтернативного сплайсинга. Модули пурпурный и синий (Рисунок 37 Б, В) связаны с регуляцией энергетического метаболизма и ответом на стрессовые факторы, соответственно. Розовый и коричневый модули отвечают за различные аспекты эмбрионального развития и дифференцировки (Рисунок 37 Г, Д).

В качестве подтверждения функциональных особенностей модулей мы провели анализ регуляторных промоторных областей на наличие обогащенных транскрипционных факторов, связанных с обозначенными GO тегами (Рисунок 38). Например, для промоторов из желтого кластера кроме SP2, SP1, в топ мотивов связывания попадает NRF1, который, как известно, представляет собой мембранный белок эндоплазматического ретикулума, реагирующий на внутриклеточные стрессы и выступающий в качестве ТФ в ядре. HNF4a, HNF4g, HNF1b – регулируют активность генов из пурпурного модуля, что также согласуется с тегами метаболизма карбоновых кислот (Рисунок 37, Б). Синий модуль включает относительно большое количество генов со сложной регуляцией, но можно отметить Foxo1, связанный с ответом на окислительный стресс. Модули коричневый и розовый, отвечающие за дифференцировку отличаются представленностью ТФ, в первом это SP, а во втором – Znf белки.

Внутри WGCNA модулей мы определили ключевые сигнальные/метаболические пути, которые характеризуют функциональные особенности генов внутри этих модулей (Рисунок 39).

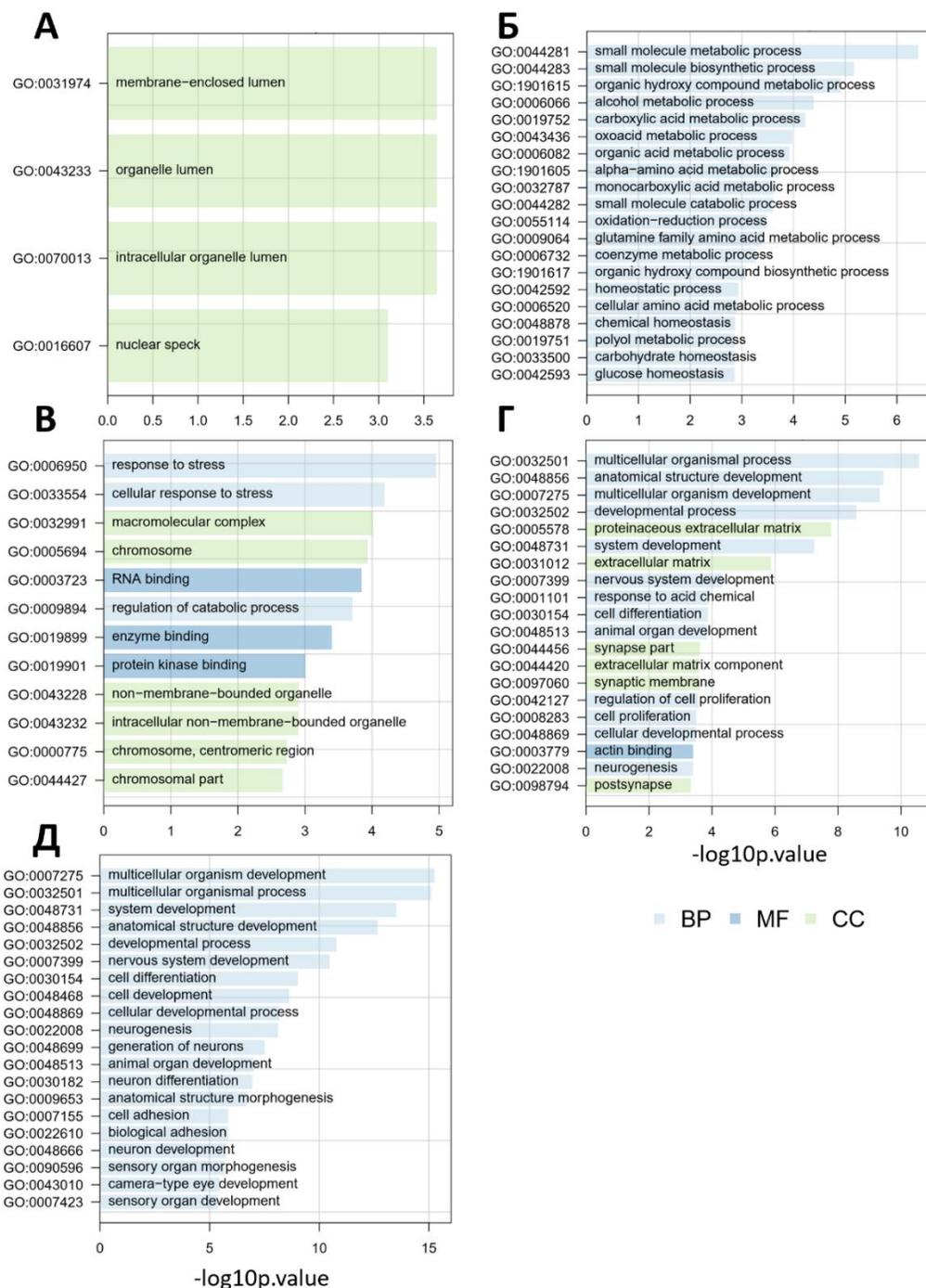


Рисунок 37. Обогащенные функциональные GO теги генов из выбранных WGCNA кластеров: А – желтый, Б – пурпурный, В – синий, Г – розовый, Д – коричневый.

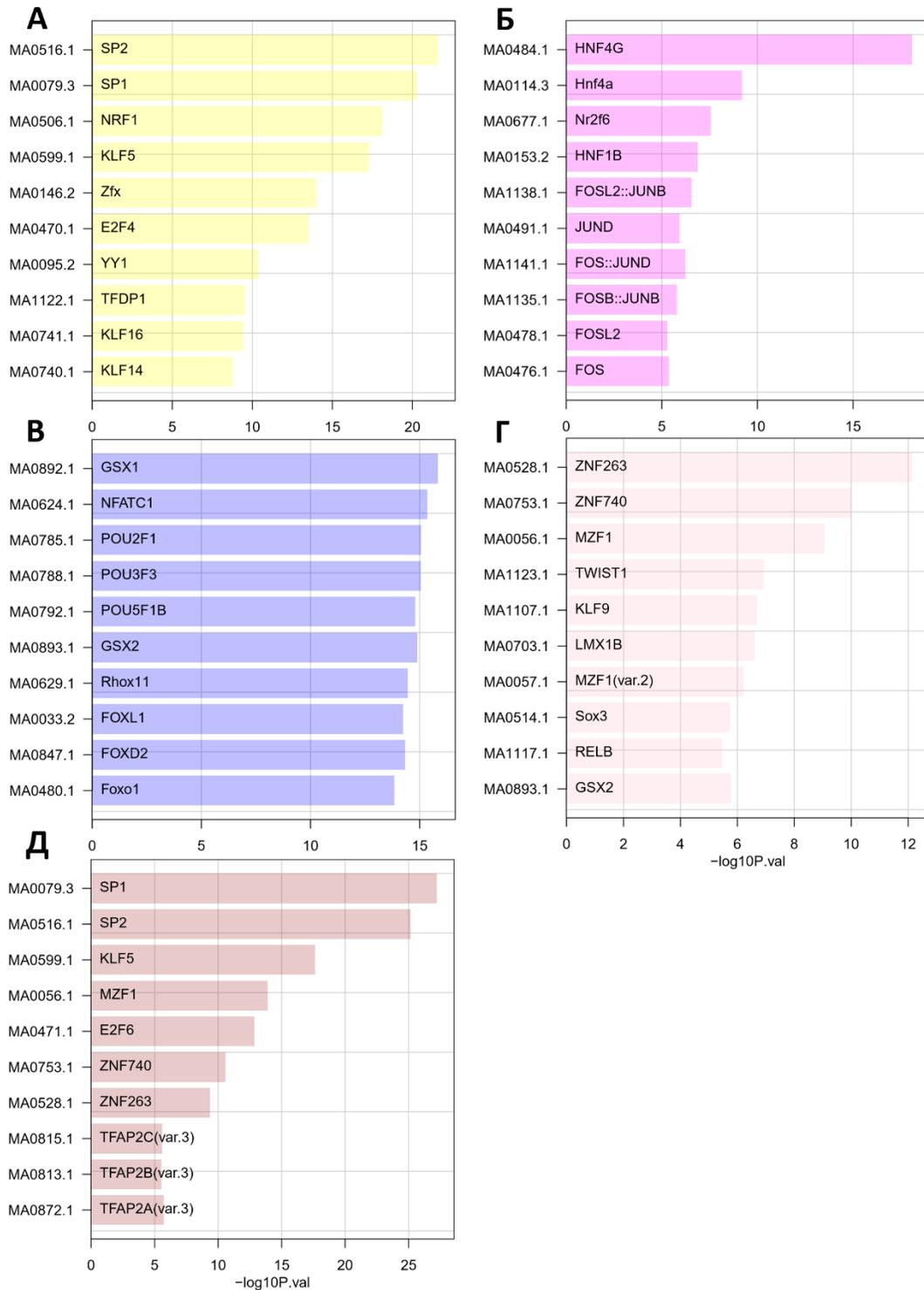


Рисунок 38. Перепредставленные мотивы связывания ТФ промоторов, относящихся к генам из выбранных WGCNA кластеров: А – желтый, Б – пурпурный, В – синий, Г – розовый, Д – коричневый.

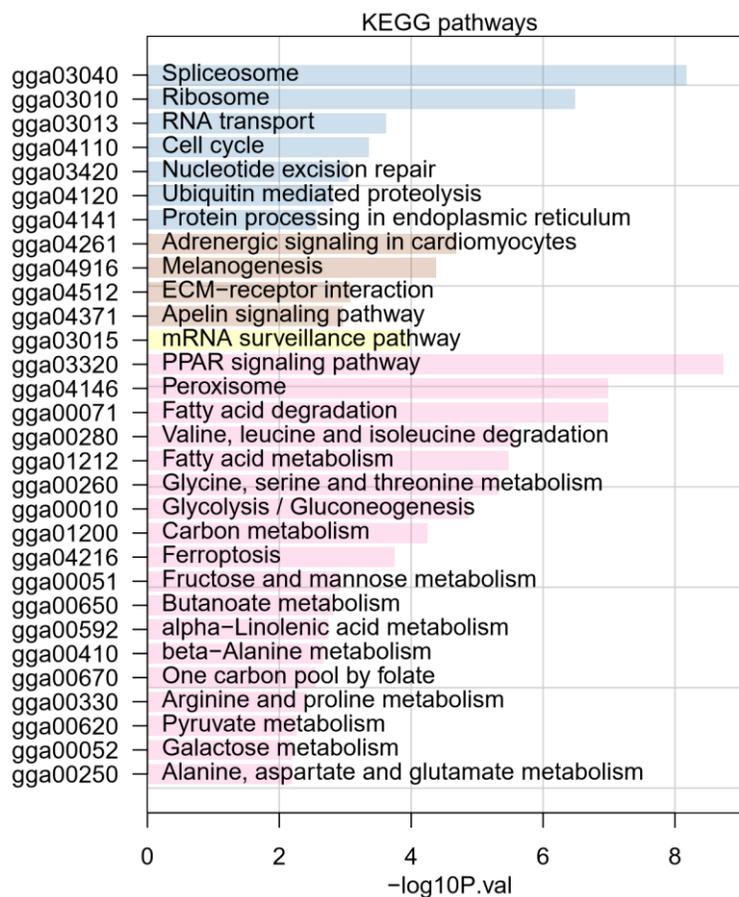


Рисунок 39. Перепредставленные метаболические пути на основе базы данных KEGG. Цветами обозначены кластеры, для которых характерны данные пути. В кластере обозначенном розовым значимо перепредставленных путей не обнаружено.

Например, «контроль качества мРНК» в желтом модуле, что сочетается с ядерными спеклами и также указывает на роль сплайсинга в гибернации цыплят. Также следует отметить, что гены с универсальным ответом (329, рисунок 28, А) представлены в основном в желтом кластере (Рисунок 40, Б), но в виде отдельной группы не характеризуются более специфичными тегами, чем митохондрии, рибосомы и убиквитинилирование (Рисунок 40, А).

Кроме того, сплайсосома, убиквитинилирование и эксцизионная репарация нуклеотидов, подтверждают связь синего модуля со стресс ответом.

Деградация жирных кислот, пероксисома и PPAR путь хорошо согласуются с транскрипционными факторами HNF и метаболизмом карбоновых кислот (Рисунок 37, Б).

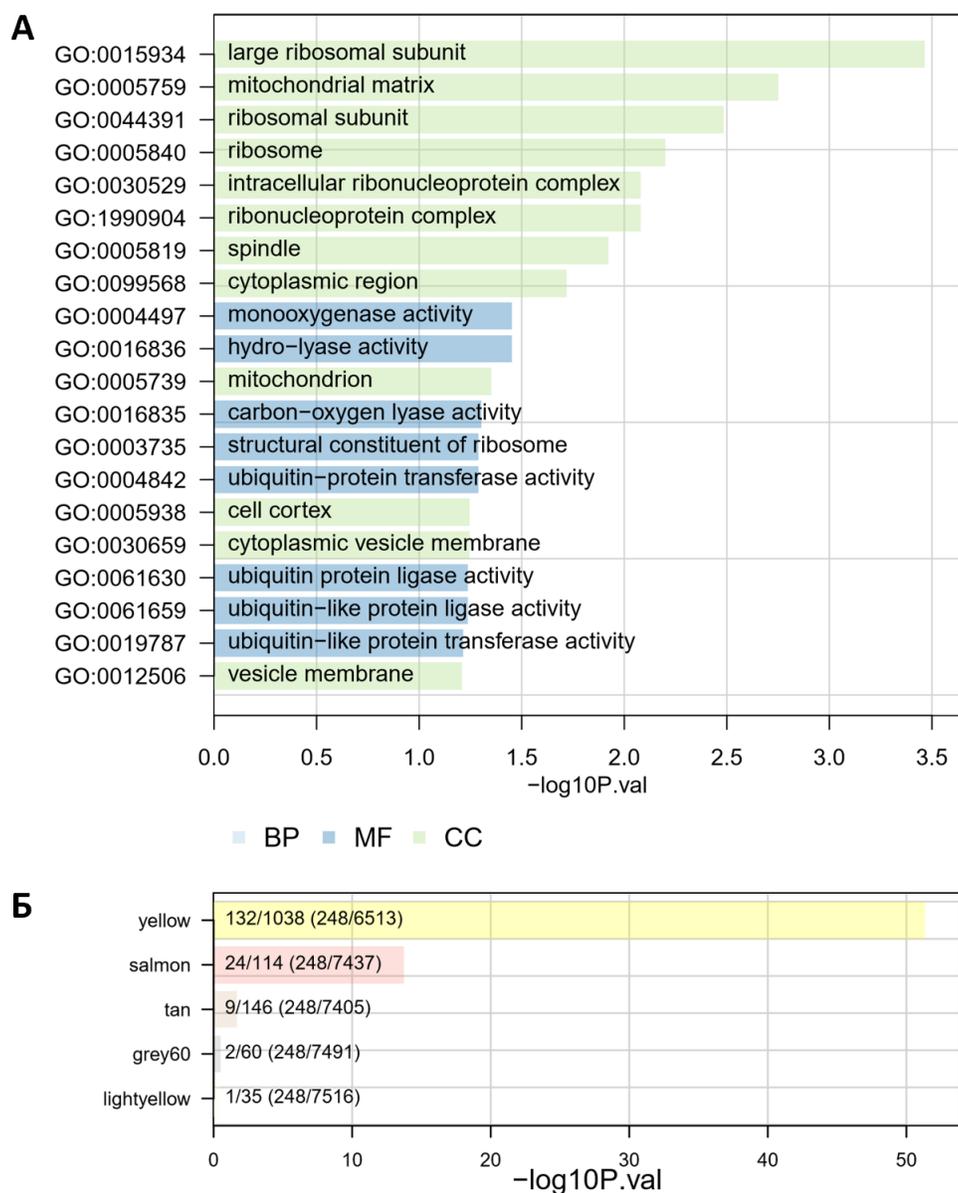


Рисунок 40. Функциональная характеристика генов универсального ответа (n=329, рисунок 30, А). А – перепредставленные GO категории в данной группе; Б – топ 5 ко-экспрессионных модулей, в которых эти гены наиболее представлены.

Выводы по главе «Гипотермальная диапауза»

Известны работы, в которых указывается сходство между голодом во время гибернации или суточным торпопромом и ограничением калорий у млекопитающих, которые не впадают в спячку [128]. Во время длительного оцепенения в спячке у таких видов млекопитающих углеводный метаболизм переключается на использование липидов. Многие исследования предполагают, что необходима активация AMPK для индукции состояния торпора или гибернации [129–131]. Однако в другом исследовании была показана активация AMPK только в белой жировой ткани, а не в печени, скелетной мышце, бурой жировой ткани или головном мозге во время спячки [132]. Была также показана роль AMPK в регуляции остановки энергетического метаболизма у цыплят в различных типах скелетных мышц [133]. В нашем исследовании данный ген не меняет значимо уровня экспрессии и таким образом подчеркивает разницу между состоянием гибернации в ходе эмбрионального развития и спячкой взрослой особи.

Путь передачи сигналов PPAR также регулирует липидный обмен. Многочисленные исследования показали увеличение количества PPAR в различных органах на уровне мРНК и белка во время оцепенения у нескольких видов, впадающих в спячку [128, 134]. Ген PPARD статистически значимо активируется только на ранних стадиях развития – 6 часов, а далее сохраняется только тренд на увеличение экспрессии в точке В. Промоторы PPARA значимо понижают экспрессию при охлаждении эмбрионов в точке HH16 (2.5 дня) и имеют тренд на инактивацию в остальных экспериментах. HMGCS2 – ключевой ген кетогенеза, регулируемый PPARA, был активен только в эмбрионах HH16 и инактивировался в ответ на охлаждение. Такие результаты согласуются с данными по ко-экспрессии, где пурпурный модуль с большим количеством

генов каскада PPAR (рок 40) отвечает на охлаждение в точке HH16 (Рисунок 36Б). Из пути PPAR ген ACADL, ответственный за бета окисление жирных кислот понижал экспрессию при охлаждении во всех экспериментах, тогда как UBC имел обратный профиль экспрессии. UBC является важным источником убиквитина во время пролиферации и клеточного стресса [135].

В последнее время предполагается, что активация PGC-1 α отвечает за защиту скелетной мышцы от атрофии в течение длительных периодов оцепенения при спячке [136]. Однако у эмбрионов курицы активность этого гена находится на уровне шума.

Предыдущие исследования показали, что сигнализация инсулина ингибируется в скелетной мышце серых мышшиных лемуров [137] и что активность киназы Akt подавляется во время оцепенения в нескольких тканях сусликов [138]. Подавленная активность Akt сопровождается уменьшением активации mTOR, приводящим к состоянию ингибирования синтеза белка во время оцепенения у животных, впадающих в спячку [138, 139]. В случае с эмбрионами, ни Akt, ни mTOR гены значимо не меняют экспрессию в ответ на охлаждение, что еще раз подчеркивает разницу между гибернацией животных, впадающих в спячку и эмбриональной гипотермальной диапаузой.

Анализ состояния эмбрионального ареста выявил маркерных гены, такие как NANOG, OCT4, REX1, которые также характерны для эмбриональных стволовых клеток (ЭСК) человека. Среди маркеров ЭСК также отмечены гены Sox2, TERT, щелочная фосфатаза, Cvh, Brachyury, Gata6 [140, 141]. NANOG активен на ранних этапах развития (Рисунок 13) и в эксперименте с охлаждением статистически значимо повышает уровень экспрессии в ответ на стресс в точках HH1 и HH4. OCT4 гомолог PouV (Pou5f3) также активен в эмбрионах HH1 и HH4, но имеет противоположный тренд в отличие от NANOG. Sox2 и Cvh имеют профиль экспрессии специфичный для поздних стадий и вместе с TERT, ALPL, ALPI, Gata6 значимо не меняют активность при

охлаждении, хоть в некоторых случаях и просматривается тренд на обратимую инактивацию. *Bachyugy* в эксперименте с охлаждением активен только в точке НН4, где незначимо повышает экспрессию в ответ на стресс.

ЗАКЛЮЧЕНИЕ

В нашем исследовании были обнаружены промоторы и связанные с ними гены, реагирующие на охлаждение на всех этапах развития эмбриона. Эта группа не характеризуется особыми метаболическими или сигнальными путями, но связана в общем с митохондриями, рибосомами и убиквитинилированием. Для более полной картины регуляции индуцированного гипометаболизма, мы построили сеть генов, на основе ко-экспрессии, внутри которой определили группы генов (модули), с высокой корреляцией профилей экспрессии. Для этих модулей мы определили их функциональную роль – эмбриогенез, стрессовый ответ и другие, транскрипционные факторы, связывающие промоторы внутри этих модулей, характерные метаболические пути, а также нашли модули, в которых наиболее представлены гены универсального ответа на охлаждение. На данном этапе не удалось связать регуляцию гипотермальной диапаузы с каким-то известным метаболическим каскадом, кроме PPAR, который принимает участие на этапе HH16. Однако созданная ко-экспрессионная сеть с обозначенными генными модулями представляет основной набор генов, необходимых для успешного обратимого перехода эмбриона курицы в состояние гипометаболизма при охлаждении на различных этапах развития.

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 промоторы 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

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Wispe J. Effect of environmental hypothermia on dorsal aortic blood flow in the chick embryo, stages 18 to 24 / Wispe J., Hu N., Clark E.B. // *Pediatric Research* – 1983.
2. Nichelmann M. Ontogeny of thermoregulation in precocial birds / Nichelmann M., Tzschentke B. // *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* – 2002.
3. Haussler D. Genome 10K: A proposal to obtain whole-genome sequence for 10000 vertebrate species / Haussler D., O'Brien S.J., Ryder O.A., Turner S. // *Journal of Heredity* – 2009. – Т. 100 – № 6 – С.659–674.
4. Koepfli K. The Genome 10K Project: a way forward. / Koepfli K., Paten B., Genome 10K Community of Scientists, O'Brien S.J. // *Annual review of animal biosciences* – 2015. – Т. 3 – С.57–111.
5. Wu J.Q. Systematic analysis of transcribed loci in {ENCODE} regions using {RACE} sequencing reveals extensive transcription in the human genome / Wu J.Q., Du J., Rozowsky J., Zhang Z., Urban A.E., Euskirchen G., Weissman S., Gerstein M., Snyder M. // *Genome Biology* – 2008. – Т. 9 – № 1 – C.R3.
6. Oszolak F. RNA sequencing: advances, challenges and opportunities / Oszolak F., Milos P.M. // *Nature Reviews Genetics* – 2011. – Т. 12 – № 2 – С.87–98.
7. Forrest A.R.R. A promoter-level mammalian expression atlas / Forrest A.R.R., Kawaji H., Rehli M., Baillie J.K. // *Nature* – 2014. – Т. 507 – № 7493 – С.462–470.
8. Morrissy S. Digital gene expression by tag sequencing on the illumina genome analyzer / S. Morrissy, Y. Zhao, A. Delaney, J. Asano, N. Dhalla, I. Li, H. McDonald, P. Pandoh, A. L. Prabhu, A. Tam, M. Hirst, M. Marra – , 2010.– 1-36с.

9. Saha S. Using the transcriptome to annotate the genome / Saha S., Sparks A.B., Rago C., Akmaev V., Wang C.J., Vogelstein B., Kinzler K.W., Velculescu V.E. // Nature Biotechnology – 2002.
10. Parkinson J. Expressed sequence tags: An overview / Parkinson J., Blaxter M. // Methods in Molecular Biology – 2009.
11. Langfelder P. WGCNA: An R package for weighted correlation network analysis / Langfelder P., Horvath S. // BMC Bioinformatics – 2008.
12. Zhang G. Comparative genomic data of the Avian Phylogenomics Project. / Zhang G., Li B., Li C., Gilbert M.T.P., Jarvis E.D., Wang J. // GigaScience – 2014. – T. 3 – № 1 – C.26.
13. Feingold E. The ENCODE (ENCyclopedia Of DNA Elements) Project. / Feingold E., Good P., Guyer M., Consortium E.P. // Science – 2004. – T. 306 – № 5696 – C.636–40.
14. Andersson R. An atlas of active enhancers across human cell types and tissues. / Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., Zhao X., Schmidl C., Suzuki T., Ntini E., Arner E., Valen E., Li K., Schwarzfischer L., Glatz D., Raithel J., Lilje B., Rapin N., Bagger F.O., Jørgensen M., Andersen P.R., Bertin N., Rackham O., Burroughs a M., Baillie J.K., Ishizu Y., Shimizu Y., Furuhashi E., Maeda S., Negishi Y., Mungall C.J., Meehan T.F., Lassmann T., Itoh M., Kawaji H., Kondo N., Kawai J., Lennartsson A., Daub C.O., Heutink P., Hume D. a, Jensen T.H., Suzuki H., Hayashizaki Y., Müller F., Forrest A.R.R., Carninci P., Rehli M., Sandelin A. // Nature – 2014. – T. 507 – № 7493 – C.455–61.
15. Arner E. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells / Arner E., Daub C.O., Vitting-Seerup K., Hayashizaki Y. // Science – 2015.

16. Hoskins R.A. Genome-wide analysis of promoter architecture in *Drosophila melanogaster* / Hoskins R.A., Landolin J.M., Brown J.B., Sandler J.E., Takahashi H., Lassmann T., Yu C., Booth B.W., Zhang D., Wan K.H., Yang L., Boley N., Andrews J., Kaufman T.C., Graveley B.R., Bickel P.J., Carninci P., Carlson J.W., Celniker S.E. // *Genome Res* – 2011. – T. 21 – № 2 – C.182–192.
17. Nepal Chirag, Hadzhiev Y. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis / Nepal Chirag, Hadzhiev Y., Previti C., Haberle V., Abdelhamid R.F., Anand S., Gehrig J., Akalin A., Kockx C.E.M., Sloot A.A.J. Van Der, Ijcken W.F.J. Van, Stupka E. – 2013. – C.1938–1950.
18. Haberle V. Two independent transcription initiation codes overlap on vertebrate core promoters / Haberle V., Li N., Hadzhiev Y., Plessy C., Previti C., Nepal C., Gehrig J., Dong X., Akalin A., Suzuki A.M., Ijcken W.F.J. Van, Armant O., Ferg M., Strähle U., Carninci P., Müller F., Lenhard B. // *Nature* – 2014.
19. Lizio M. Systematic analysis of transcription start sites in avian development / Lizio M., Deviatiiarov R., Nagai H., Galan L., Arner E., Itoh M., Lassmann T., Kasukawa T., Hasegawa A., Ros M.A., Hayashizaki Y., Carninci P., Forrest A.R.R., Kawaji H., Gusev O., Sheng G. // *PLoS Biology* – 2017. – T. 15 – № 9 – C.e2002887.
20. Takahashi H. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing / Takahashi H., Lassmann T., Murata M., Carninci P. // *Nature Protocols* – 2012.
21. Itoh M. Automated workflow for preparation of cDNA for cap analysis of gene expression on a single molecule sequencer / Itoh M., Kojima M., Nagao-Sato S., Saijo E., Lassmann T., Kanamori-Katayama M., Kaiho A., Lizio M., Kawaji H., Carninci P., Forrest A.R.R., Hayashizaki Y. // *PLoS ONE* – 2012. – T. 7 – № 1.

22. Plessy C. Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. / Plessy C., Bertin N., Takahashi H., Simone R., Salimullah M., Lassmann T., Vitezic M., Severin J., Olivarius S., Lazarevic D., Hornig N., Orlando V., Bell I., Gao H., Dumais J., Kapranov P., Wang H., Davis C.A., Gingeras T.R., Kawai J., Daub C.O., Hayashizaki Y., Gustincich S., Carninci P. // *Nature methods* – 2010. – T. 7 – № 7 – C.528–34.
23. Salimullah M. NanoCAGE: A high-resolution technique to discover and interrogate cell transcriptomes / Salimullah M., Mizuho S., Plessy C., Carninci P. // *Cold Spring Harbor Protocols* – 2011.
24. Carninci P. High-efficiency full-length cDNA cloning by biotinylated CAP trapper / Carninci P., Kvam C., Kitamura A., Ohsumi T., Okazaki Y., Itoh M., Kamiya M., Shibata K., Sasaki N., Izawa M., Muramatsu M., Hayashizaki Y., Schneider C. // *Genomics* – 1996.
25. Kanamori-Katayama M. Unamplified cap analysis of gene expression on a single-molecule sequencer / Kanamori-Katayama M., Itoh M., Kawaji H., Lassmann T., Katayama S., Kojima M., Bertin N., Kaiho A., Ninomiya N., Daub C.O., Carninci P., Forrest A.R.R., Hayashizaki Y. // *Genome Research* – 2011. – T. 21 – № 7 – C.1150–1159.
26. Murata M. Detecting expressed genes using CAGE / Murata M., Nishiyori-Sueki H., Kojima-Ishiyama M., Carninci P., Hayashizaki Y., Itoh M. // *Methods in Molecular Biology* – 2014.
27. Kodzius R. CAGE: cap analysis of gene expression / Kodzius R., Kojima M., Nishiyori H., Nakamura M., Fukuda S., Tagami M., Sasaki D., Imamura K., Kai C., Harbers M., Hayashizaki Y., Carninci P. // *Nature Methods* – 2006. – T. 3 – № 3 – C.211–222.
28. Dreos R. The eukaryotic promoter database in its 30th year: Focus on non-

- vertebrate organisms / Dreos R., Ambrosini G., Groux R., Perier R.C., Bucher P. // Nucleic Acids Research – 2017.
29. Rie D. De An integrated expression atlas of miRNAs and their promoters in human and mouse / Rie D. De, Abugessaisa I., Alam T., Hoon M.J.L. De // Nature Biotechnology – 2017. – T. 35 – № 9.
30. Warren W.C. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure / Warren W.C., Hillier L.W., Tomlinson C., Minx P., Kremitzki M., Graves T., Markovic C., Bouk N., Pruitt K.D., Thibaud-Nissen F., Schneider V., Mansour T.A., Brown C.T., Zimin A., Hawken R., Abrahamsen M., Pyrkosz A.B., Morisson M., Fillon V., Vignal A., Chow W., Howe K., Fulton J.E., Miller M.M., Lovell P., Mello C. V., Wirthlin M., Mason A.S., Kuo R., Burt D.W., Dodgson J.B., Cheng H.H. // *G3* Genes|Genomes|Genetics – 2017.
31. Stern C.D. The chick: A great model system becomes even greater / Stern C.D. // Developmental Cell – 2005.
32. Hamburger V. A series of normal stages in the development of the chick embryo. 1951. / Hamburger V., Hamilton H.L. // Developmental dynamics: an official publication of the American Association of Anatomists – 1992.
33. Hillier L.W. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution / Hillier L.W., Miller W., Birney E., Wilson R.K. // Nature – 2014.
34. Schmid M. Third Report on Chicken Genes and Chromosomes 2015: / Schmid M., Smith J., Burt D.W., Zhou H. // Cytogenetic and Genome Research – 2015.
35. LOMBARD R.E. Evolution of the tetrapod ear: an analysis and reinterpretation / LOMBARD R.E., BOLT J.R. // Biological Journal of the Linnean Society – 1979.
36. Schreiber A.M. Tadpole skin dies autonomously in response to thyroid hormone

at metamorphosis / Schreiber A.M., Brown D.D. // Proceedings of the National Academy of Sciences – 2003.

37. Downs K.M. The enigmatic primitive streak: Prevailing notions and challenges concerning the body axis of mammals / Downs K.M. // BioEssays – 2009.

38. Bertocchini F. The hypoblast of the chick embryo positions the primitive streak by antagonizing nodal signaling / Bertocchini F., Stern C.D. // Developmental Cell – 2002.

39. Zagorski M. Decoding of position in the developing neural tube from antiparallel morphogen gradients / Zagorski M., Tabata Y., Brandenberg N., Lutolf M.P., Tkačik G., Bollenbach T., Briscoe J., Kicheva A. // Science – 2017.

40. И. Г. Акмаев, М. А. Александрова, Ю. И. Афанасьев, В. П. Бабминдра, Л. П. Бобова, Т. Г. Боровая, В. Ш. Вагапова, Е. В. Виноградова, В. Г. Гололобов, В. Л. Горячкина, Б. А. Григорян, Р. К. Данилов, Н. В. Дедух, Р. В. Деев, Н. А. Дмитриева, И. В. Дюйзен, В. В. В. Ю. Руководство по гистологии / В. В. Ю. И. Г. Акмаев, М. А. Александрова, Ю. И. Афанасьев, В. П. Бабминдра, Л. П. Бобова, Т. Г. Боровая, В. Ш. Вагапова, Е. В. Виноградова, В. Г. Гололобов, В. Л. Горячкина, Б. А. Григорян, Р. К. Данилов, Н. В. Дедух, Р. В. Деев, Н. А. Дмитриева, И. В. Дюйзен, В. / под ред. Данилов Р. К. – Т.1 – СПб: «СпецЛит», 2010. Вып. 2– 831с.

41. Rafferty A.R. Arrested embryonic development: A review of strategies to delay hatching in egg-laying reptiles // Proc. R. Soc. B Biol. Sci. – 2012.

42. Lopes F.L. Embryonic diapause and its regulation // Reproduction. – 2004.

43. Shanbhag B.A. Lowering body temperature induces embryonic diapause during prolonged egg retention in the lizard, *Calotes versicolor* / Shanbhag B.A., Saidapur S.K., Radder R.S. // Naturwissenschaften – 2003.

44. Geiser F. Hibernation // *Curr. Biol.* – 2013.
45. Whittow G.G.C. The early development of thermoregulation in birds / Whittow G.G.C., Tazawa H. // *Physiological zoology* – 1991.
46. Tazawa H. Response of egg temperature, heart rate and blood pressure in the chick embryo to hypothermal stress / Tazawa H., Nakagawa S. // *Journal of Comparative Physiology B* – 1985.
47. Janke O. Metabolic responses of chicken and muscovy duck embryos to high incubation temperatures / Janke O., Tzschentke B., Höchel J., Nichelmann M. // *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* – 2002.
48. Tazawa H. Hypothermal effect on the gas exchange in chicken embryo / Tazawa H. // *Respiration Physiology* – 1973.
49. Hiroshi T. Tolerance of chick embryos to low temperatures in reference to the heart rate / Hiroshi T., Hermann R. // *Comparative Biochemistry and Physiology -- Part A: Physiology* – 1986.
50. Fasenko G.M. Egg storage and the embryo , 2007.
51. Wong F. eChickAtlas: An introduction to the database / Wong F., Welten M.C.M., Anderson C., Bain A.A., Liu J., Wicks M.N., Pavlovska G., Davey M.G., Murphy P., Davidson D., Tickle C.A., Stern C.D., Baldock R.A., Burt D.W. // *Genesis* – 2013.
52. Bloom G.A. The importance of mentoring in the development of coaches and athletes / Bloom G.A., Durand-Bush N., Schinke R.J., Salmela J.H. // *International Journal of Sport Psychology* – 1998.
53. Hwang G.W. Whi2 enhances methylmercury toxicity in yeast via inhibition of Akr1 palmitoyltransferase activity / Hwang G.W., Fukumitsu T., Ogiwara Y., Takahashi T., Miura N., Kuge S., Naganuma A. // *Biochimica et Biophysica Acta* -

General Subjects – 2016.

54. Alev C. Manipulating the avian epiblast and epiblast-derived stem cells / Alev C., Nakano M., Wu Y., Horiuchi H., Sheng G. // *Methods in Molecular Biology* – 2013.
55. Head R. Unamplified Cap Analysis of Gene Expression on a single molecule sequencer / Head R., Type M. – 2011. – C.1150–1159.
56. Robinson M.D. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. / Robinson M.D., McCarthy D.J., Smyth G.K. // *Bioinformatics (Oxford, England)* – 2010.
57. Hoskins R. a Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. / Hoskins R. a, Landolin J.M., Brown J.B., Sandler J.E., Takahashi H., Lassmann T., Yu C., Booth B.W., Zhang D., Wan K.H., Yang L., Boley N., Andrews J., Kaufman T.C., Graveley B.R., Bickel P.J., Carninci P., Carlson J.W., Celniker S.E. // *Genome Res* – 2011. – T. 21 – № 2 – C.182–192.
58. McLeay R.C. Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data / McLeay R.C., Bailey T.L. // *BMC Bioinformatics* – 2010.
59. Deviatiiarov R. Application of a CAGE method to an avian development study / R. Deviatiiarov, M. Lizio, O. Gusev – , 2017.
60. Lizio M. Systematic analysis of transcription start sites in avian development / Lizio M., Deviatiiarov R., Nagai H., Galan L., Arner E., Itoh M., Lassmann T., Kasukawa T., Hasegawa A., Ros M.A., Hayashizaki Y., Carninci P., Forrest A.R.R., Kawaji H., Gusev O., Sheng G. // *PLoS Biology* – 2017. – T. 15 – № 9.
61. Andrews S. FastQC: A quality control tool for high throughput sequence data. / Andrews S. // *Babraham Bioinformatics* – 2010.
62. Bolger A.M. Trimmomatic: A flexible trimmer for Illumina sequence data / Bolger A.M., Lohse M., Usadel B. // *Bioinformatics* – 2014.

63. Li H. Fast and accurate long-read alignment with Burrows-Wheeler transform / Li H., Durbin R. // *Bioinformatics* – 2010.
64. Bailey T.L. MEME Suite: Tools for motif discovery and searching / Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. // *Nucleic Acids Research* – 2009.
65. Haberle V. CAGEr: Precise TSS data retrieval and high-resolution promoterome mining for integrative analyses / Haberle V., Forrest A.R.R., Hayashizaki Y., Carninci P., Lenhard B. // *Nucleic Acids Research* – 2015.
66. Muret K. Long noncoding RNA repertoire in chicken liver and adipose tissue / Muret K., Klopp C., Wucher V., Esquerré D., Legeai F., Lecerf F., Désert C., Boutin M., Jehl F., Acloque H., Giuffra E., Djebali S., Foissac S., Derrien T., Lagarrigue S. // *Genetics Selection Evolution* – 2017.
67. Li A. ALDB: A Domestic-Animal Long Noncoding RNA Database / Li A., Zhang J., Zhou Z., Wang L., Liu Y., Liu Y. // *PLOS ONE* – 2015.
68. Zhao Y. NONCODE 2016: An informative and valuable data source of long non-coding RNAs / Zhao Y., Li H., Fang S., Kang Y., Wu W., Hao Y., Li Z., Bu D., Sun N., Zhang M.Q., Chen R. // *Nucleic Acids Research* – 2016.
69. Zhang G. Comparative genomics reveals insights into avian genome evolution and adaptation / Zhang G., Li C., Li Q., Froman D.P. // *Science* – 2014.
70. Seki R. Functional roles of Aves class-specific cis-regulatory elements on macroevolution of bird-specific features / Seki R., Li C., Fang Q., Hayashi S., Egawa S., Hu J., Xu L., Pan H., Kondo M., Sato T., Matsubara H., Kamiyama N., Kitajima K., Saito D., Liu Y., Gilbert M.T.P., Zhou Q., Xu X., Shiroishi T., Irie N., Tamura K., Zhang G. // *Nature Communications* – 2017.
71. Kornblihtt A.R. Promoter usage and alternative splicing // *Curr. Opin. Cell Biol.* –

2005.

72. Rach E.A. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the *Drosophila* genome / Rach E.A., Yuan H.Y., Majoros W.H., Tomancak P., Ohler U. // *Genome Biology* – 2009.

73. Laval F. The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells / Laval F., Acloque H., Bertocchini F., MacLeod D.J., Boast S., Bachelard E., Montillet G., Thenot S., Sang H.M., Stern C.D., Samarut J., Pain B. // *Development* – 2007.

74. Kim J.Y. The Nucleolar Protein GLTSCR2 Is an Upstream Negative Regulator of the Oncogenic Nucleophosmin-MYC Axis / Kim J.Y., Cho Y.E., Park J.H. // *American Journal of Pathology* – 2015.

75. Baala L. Homozygous silencing of T-box transcription factor EOMES leads to microcephaly with polymicrogyria and corpus callosum agenesis / Baala L., Briault S., Etchevers H.C., Laumonier F., Natiq A., Amiel J., Boddaert N., Picard C., Sbiti A., Asermouh A., Attié-Bitach T., Encha-Razavi F., Munnich A., Sefiani A., Lyonnet S. // *Nature Genetics* – 2007.

76. Zhang B. A General Framework for Weighted Gene Co-Expression Network Analysis / Zhang B., Horvath S. // *Statistical Applications in Genetics and Molecular Biology* – 2005.

77. Eisenberg E. Human housekeeping genes, revisited // *Trends Genet.* – 2013.

78. Burt D.W. Chicken genome: Current status and future opportunities // *Genome Res.* – 2005.

79. Abe H. Abundance, arrangement, and function of sequence motifs in the chicken promoters / Abe H., Gemmell N.J. // *BMC Genomics* – 2014.

80. Chen J. A systems-level approach reveals new gene regulatory modules in the

developing ear / Chen J., Tambalo M., Barembaum M., Ranganathan R., Simões-Costa M., Bronner M.E., Streit A. // *Development* – 2017.

81. Khan M.A.F. Computational tools and resources for prediction and analysis of gene regulatory regions in the chick genome. / Khan M.A.F., Soto-Jimenez L.M., Howe T., Streit A., Sosinsky A., Stern C.D. // *Genesis* (New York, N.Y. : 2000) – 2013.

82. Roellig D. Dynamic transcriptional signature and cell fate analysis reveals plasticity of individual neural plate border cells / Roellig D., Tan-Cabugao J., Esaian S., Bronner M.E. // *eLife* – 2017.

83. Uchikawa M. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals // *Dev. Cell.* – 2003.

84. Alev C. Transcriptomic landscape of the primitive streak / Alev C., Wu Y., Kasukawa T., Jakt L.M., Ueda H.R., Sheng G. // *Development* – 2010.

85. Lowe C.B. Feather Development Genes and Associated Regulatory Innovation Predate the Origin of Dinosauria / Lowe C.B., Clarke J.A., Baker A.J., Haussler D., Edwards S. V. // *Molecular Biology and Evolution* – 2015.

86. Abu-Bonsrah K.D. CRISPR/Cas9 Targets Chicken Embryonic Somatic Cells in Vitro and in Vivo and generates Phenotypic Abnormalities / Abu-Bonsrah K.D., Zhang D., Newgreen D.F. // *Scientific Reports* – 2016.

87. Véron N. CRISPR mediated somatic cell genome engineering in the chicken / Véron N., Qu Z., Kipen P.A.S., Hirst C.E., Marcelle C. // *Developmental Biology* – 2015.

88. Cheng A.W. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system / Cheng A.W., Wang H., Yang H., Shi L., Katz Y., Theunissen T.W., Rangarajan S., Shivalila C.S., Dadon D.B., Jaenisch R.

// Cell Research – 2013.

89. Hsu P.D. DNA targeting specificity of RNA-guided Cas9 nucleases / Hsu P.D., Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Agarwala V., Li Y., Fine E.J., Wu X., Shalem O., Cradick T.J., Marraffini L.A., Bao G., Zhang F. // Nature Biotechnology – 2013.

90. Dubrulle J. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation / Dubrulle J., McGrew M.J., Pourquié O. // Cell – 2001.

91. Morales A. V. Snail genes at the crossroads of symmetric and asymmetric processes in the developing mesoderm / Morales A. V., Acloque H., Ocaña O.H., Frutos C.A. de, Gold V., Nieto M.A. // EMBO Reports – 2007.

92. Dady A. Timing and kinetics of E- to N-cadherin switch during neurulation in the avian embryo / Dady A., Blavet C., Duband J.L. // Developmental Dynamics – 2012.

93. Leenen F.A.D. Where does transcription start? 5'-RACE adapted to next-generation sequencing / Leenen F.A.D., Vernocchi S., Hunewald O.E., Schmitz S., Molitor A.M., Muller C.P., Turner J.D. // Nucleic Acids Research – 2015.

94. Sandelin A. Mammalian RNA polymerase II core promoters: Insights from genome-wide studies // Nat. Rev. Genet. – 2007.

95. Carey M.F. The primer extension assay / Carey M.F., Peterson C.L., Smale S.T. // Cold Spring Harbor Protocols – 2013.

96. Kawaji H. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing / Kawaji H., Lizio M., Itoh M., Kanamori-Katayama M., Kaiho A., Nishiyori-Sueki H., Shin J.W., Kojima-Ishiyama M., Kawano M., Murata M., Ninomiya-Fukuda N., Ishikawa-Kato S., Nagao-Sato S., Noma S., Hayashizaki Y., Forrest A.R.R., Carninci P. // Genome

Research – 2014.

97. Sharma C.M. Differential RNA-seq: The approach behind and the biological insight gained // *Curr. Opin. Microbiol.* – 2014.

98. Batut P. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression / Batut P., Dobin A., Plessy C., Carninci P., Gingeras T.R. // *Genome Research* – 2013.

99. Innocenti N. Whole-genome mapping of 5' RNA ends in bacteria by tagged sequencing: A comprehensive view in *Enterococcus faecalis* / Innocenti N., Golumbeanu M., D'Hérouël A.F., Lacoux C., Bonnin R.A., Kennedy S.P., Wessner F., Serron P., Bouloc P., Repoila F., Aurell E. // *RNA* – 2015.

100. Eid J. Real-time DNA sequencing from single polymerase molecules / Eid J., Fehr A., Gray J., Turner S. // *Science* – 2009.

101. Gilbert L.A. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation / Gilbert L.A., Horlbeck M.A., Adamson B., Villalta J.E., Chen Y., Whitehead E.H., Guimaraes C., Panning B., Ploegh H.L., Bassik M.C., Qi L.S., Kampmann M., Weissman J.S. // *Cell* – 2014.

102. Newman A.M. Robust enumeration of cell subsets from tissue expression profiles / Newman A.M., Liu C.L., Green M.R., Gentles A.J., Feng W., Xu Y., Hoang C.D., Diehn M., Alizadeh A.A. // *Nature Methods* – 2015.

103. Gong T. DeconRNASeq: A statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data / Gong T., Szustakowski J.D. // *Bioinformatics* – 2013.

104. Li Y. A mixture model for expression deconvolution from RNA-seq in heterogeneous tissues / Li Y., Xie X. // *BMC Bioinformatics* – 2013.

105. Scialdone A. Resolving early mesoderm diversification through single-cell

expression profiling / Scialdone A., Tanaka Y., Jawaid W., Moignard V., Wilson N.K., Macaulay I.C., Marioni J.C., Göttgens B. // Nature – 2016.

106. Patel A.P. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma / Patel A.P., Tirosh I., Trombetta J.J., Shalek A.K., Gillespie S.M., Wakimoto H., Cahill D.P., Nahed B. V., Curry W.T., Martuza R.L., Louis D.N., Rozenblatt-Rosen O., Suvà M.L., Regev A., Bernstein B.E. // Science – 2014.

107. Chang I.K. Simple method for isolation of primordial germ cells from chick embryos / Chang I.K., Tajima A., Yasuda Y., Chikamune T., Ohno T. // Cell Biology International Reports – 1992.

108. Renfree M.B. Diapause / Renfree M.B., Shaw G. // Annual Review of Physiology – 2000.

109. Bailey T.L. Inferring direct DNA binding from ChIP-seq / Bailey T.L., MacHanick P. // Nucleic Acids Research – 2012.

110. Opitz O.G. Interaction between Sp1 and cell cycle regulatory proteins is important in transactivation of a differentiation-related gene / Opitz O.G., Rustgi A.K. // Cancer Research – 2000.

111. Iwahori S. Identification of phosphorylation sites on transcription factor Sp1 in response to DNA damage and its accumulation at damaged sites / Iwahori S., Yasui Y., Kudoh A., Sato Y., Nakayama S., Murata T., Isomura H., Tsurumi T. // Cellular Signalling – 2008.

112. Olofsson B.A. Phosphorylation of Sp1 in Response to DNA Damage by Ataxia Telangiectasia-Mutated Kinase / Olofsson B.A., Kelly C.M., Kim J., Hornsby S.M., Azizkhan-Clifford J. // Molecular Cancer Research – 2007.

113. Dong S. Crucial roles of MZF1 and Sp1 in the transcriptional regulation of the peptidylarginine deiminase type I gene (PADI1) in human keratinocytes / Dong S.,

Ying S., Kojima T., Shiraiwa M., Kawada A., Méchin M.C., Adoue V., Chavanas S., Serre G., Simon M., Takahara H. // *Journal of Investigative Dermatology* – 2008.

114. Bouwman P. Transcription factor Sp3 is essential for post-natal survival and late tooth development / Bouwman P., Göllner H., Elsässer H.P., Eckhoff G., Karis A., Grosveld F., Philipsen S., Suske G. // *EMBO Journal* – 2000.

115. Marin M. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. / Marin M., Karis a, Visser P., Grosveld F., Philipsen S. // *Cell* – 1997.

116. Safe S. Sp transcription factor family and its role in cancer // *Eur. J. Cancer.* – 2005.

117. Ganss B. Zinc finger transcription factors in skeletal development / Ganss B., Jheon A. // *Crit Rev Oral Biol Med* – 2004. – T. 15 – № 5 – C.282–297.

118. Chamboredon S. v-Jun downregulates the SPARC target gene by binding to the proximal promoter indirectly through Sp1/3 / Chamboredon S., Briggs J., Vial E., Hurault J., Galvagni F., Oliviero S., Bos T., Castellazzi M. // *Oncogene* – 2003.

119. Chen C. Ubiquitin-proteasome degradation of KLF5 transcription factor in cancer and untransformed epithelial cells / Chen C., Sun X., Ran Q., Wilkinson K.D., Murphy T.J., Simons J.W., Dong J.T. // *Oncogene* – 2005.

120. Galvagni F. Sp1 and Sp3 physically interact and co-operate with GABP for the activation of the utrophin promoter / Galvagni F., Capo S., Oliviero S. // *Journal of Molecular Biology* – 2001.

121. Gabriele M. YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction / Gabriele M., Vulto-van Silfhout A.T., Germain P.L., Vries B.B.A. de // *American Journal of Human Genetics* – 2017.

122. Shi Y. Everything you have ever wanted to know about Yin Yang 1... // *Biochim. Biophys. Acta - Rev. Cancer.* – 1997.
123. Crosby S.D. The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGGGGGCG (GSG) element-binding protein family. / Crosby S.D., Puetz J.J., Simburger K.S., Fahrner T.J., Milbrandt J. // *Molecular and cellular biology* – 1991.
124. Okazawa M. The Etv1 transcription factor activity-dependently downregulates a set of genes controlling cell growth and differentiation in maturing cerebellar granule cells / Okazawa M., Abe H., Nakanishi S. // *Biochemical and Biophysical Research Communications* – 2016.
125. Besnard A. Elk-1 a transcription factor with multiple facets in the brain // *Front. Neurosci.* – 2011.
126. Khan A. JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework / Khan A., Fornes O., Stigliani A., Gheorghe M., Castro-Mondragon J.A., Lee R. Van Der, Bessy A., Chèneby J., Kulkarni S.R., Tan G., Baranasic D., Arenillas D.J., Sandelin A., Vandepoele K., Lenhard B., Ballester B., Wasserman W.W., Parcy F., Mathelier A. // *Nucleic Acids Research* – 2018. – T. 46 – № D1 – C.D260–D266.
127. Ding M. Secreted IGFBP5 mediates mTORC1-dependent feedback inhibition of IGF-1 signalling / Ding M., Bruick R.K., Yu Y. // *Nature Cell Biology* – 2016.
128. Xu Y. Molecular signatures of mammalian hibernation: Comparisons with alternative phenotypes / Xu Y., Shao C., Fedorov V.B., Goropashnaya A. V., Barnes B.M., Yan J. // *BMC Genomics* – 2013.
129. Melvin R.G. Torpor induction in mammals: recent discoveries fueling new ideas // *Trends Endocrinol. Metab.* – 2009.

130. Zhang J. Regulation of Torpor in the Gray Mouse Lemur: Transcriptional and Translational Controls and Role of AMPK Signaling / Zhang J., Tessier S.N., Biggar K.K., Wu C.W., Pifferi F., Perret M., Storey K.B. // *Genomics, Proteomics and Bioinformatics* – 2015.
131. Lanaspá M.A. Opposing activity changes in AMP deaminase and AMP-activated protein kinase in the hibernating ground squirrel / Lanaspá M.A., Epperson L.E., Li N., Cicerchi C., Garcia G.E., Roncal-Jimenez C.A., Trostel J., Jain S., Mant C.T., Rivard C.J., Ishimoto T., Shimada M., Sanchez-Lozada L.G., Nakagawa T., Jani A., Stenvinkel P., Martin S.L., Johnson R.J. // *PLoS ONE* – 2015.
132. Horman S. Evaluation of the role of AMP-activated protein kinase and its downstream targets in mammalian hibernation / Horman S., Hussain N., Dilworth S.M., Storey K.B., Rider M.H. // *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology* – 2005.
133. Hu X. Effects of feed deprivation on the AMPK signaling pathway in skeletal muscle of broiler chickens / Hu X., Liu L., Song Z., Sheikahmadi A., Wang Y., Buyse J. // *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology* – 2016.
134. Han Y. Adaptation of peroxisome proliferator-activated receptor alpha to hibernation in bats / Han Y., Zheng G., Yang T., Zhang S., Dong D., Pan Y.-H. // *BMC Evolutionary Biology* – 2015.
135. Ryu K.Y. The mouse polyubiquitin gene UbC is essential for fetal liver development, cell-cycle progression and stress tolerance / Ryu K.Y., Maehr R., Gilchrist C.A., Long M.A., Bouley D.M., Mueller B., Ploegh H.L., Kopito R.R. // *EMBO Journal* – 2007.
136. Xu R. Hibernating squirrel muscle activates the endurance exercise pathway despite prolonged immobilization / Xu R., Andres-Mateos E., Mejias R., MacDonald

E.M., Leinwand L.A., Merriman D.K., Fink R.H.A., Cohn R.D. // *Experimental Neurology* – 2013.

137. Tessier S.N. Regulation of the PI3K/AKT Pathway and Fuel Utilization During Primate Torpor in the Gray Mouse Lemur, *Microcebus murinus* / Tessier S.N., Zhang J., Biggar K.K., Wu C.W., Pifferi F., Perret M., Storey K.B. // *Genomics, Proteomics and Bioinformatics* – 2015.

138. Wu C.-W. Regulation of the mTOR signaling network in hibernating thirteen-lined ground squirrels / Wu C.-W., Storey K.B. // *Journal of Experimental Biology* – 2012.

139. Lee K. Molecular mechanism underlying muscle mass retention in hibernating bats: Role of periodic arousal / Lee K., So H., Gwag T., Ju H., Lee J.W., Yamashita M., Choi I. // *Journal of Cellular Physiology* – 2010.

140. Zhang X. Derivation of human embryonic stem cells from developing and arrested embryos. // *Stem Cells*. – 2006.

141. Jahanpanah M. Pluripotency properties of embryonic stem cells isolated from stage X blastoderm of Mazandaran native chicken / Jahanpanah M., Pourasgari F., Mohammadi-Sangcheshmeh A., Ardeshirylajimi A., Azarnia M., Sharifi S.D., Soleimani M. // *Cellular and Molecular Biology* – 2014.