

Анализ геномного распределения сайтов связывания транскрипционных факторов GAGA и CNC в развитии *Drosophila melanogaster*

И.И. Брусенцов , Д.А. Карагодин, Э.М. Баричева, Т.И. Меркулова

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

На основании доступных данных ChIP-seq и ChIP-chip экспериментов, выполненных с использованием антител к транскрипционным факторам GAGA и CNC, проведен анализ распределения по геному участков связывания этих факторов в эмбриогенезе (возраст эмбрионов 0–12 и 16–24 ч), а также на стадии белой предкуколки дрозофилы. Показано, что основная часть мест связывания GAGA и CNC попадает в промоторные районы и интроны генов, при этом максимальная плотность пиков связывания обоих факторов приходится на область старта транскрипции. Показано также, что у эмбрионов в возрасте как 0–12, так и 16–24 ч развития наблюдается неслучайная колокализация GAGA и CNC, в то время как на стадии белой предкуколки колокализации этих транскрипционных факторов в масштабе генома выявить не удается. Для того чтобы очертить круг генов, для которых возможна совместная регуляция GAGA и CNC, было осуществлено исследование их совместного распределения в аннотированных регуляторных районах (промоторная область и участки, соответствующие 5'-UTR и 3'-UTR мРНК). Оказалось, что совокупности генов, в регуляторных районах которых обнаружено связывание обоих факторов, сильно различаются на разных стадиях. Если у эмбрионов в возрасте 0–12 ч 353 гена характеризуются пересечением пиков ChIP-seq GAGA и CNC, а в возрасте 16–24 ч число таких генов составляет 611, то всего лишь 61 ген является «общим» для обеих стадий. Предполагается, что различные подгруппы генов-мишеней этих факторов регулируются разными сочетаниями изоформ GAGA и CNC, паттерны экспрессии которых изменяются в ходе эмбриогенеза дрозофилы. Функциональный анализ генов, в регуляторных районах которых найдена колокализация GAGA и CNC на всех исследованных стадиях развития, показывает обогащение генами, контролирующими эмбриогенез, развитие нервной системы и крыла дрозофилы.

Ключевые слова: транскрипционные факторы; GAGA; CNC; развитие; *D. melanogaster*.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Брусенцов И.И., Карагодин Д.А., Баричева Э.М., Меркулова Т.И. Анализ геномного распределения сайтов связывания транскрипционных факторов GAGA и CNC в развитии *Drosophila melanogaster*. Вавиловский журнал генетики и селекции. 2016;20(6):762-769. DOI 10.18699/VJ16.193

HOW TO CITE THIS ARTICLE:

Brusentsov I.I., Karagodin D.A., Baricheva E.M., Merkulova T.I. Analysis of whole-genome binding patterns of GAGA and CNC transcription factors during *Drosophila melanogaster* development. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):762-769. DOI 10.18699/VJ16.193

Analysis of whole-genome binding patterns of GAGA and CNC transcription factors during *Drosophila melanogaster* development

I.I. Brusentsov , D.A. Karagodin, E.M. Baricheva, T.I. Merkulova

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

On the basis of available data of ChIP-seq and ChIP-chip experiments performed using antibodies against GAGA and CNC transcription factors, genome-wide binding mapping of these factors at hours 0–12 and 16–24 of *Drosophila* embryogenesis, as well as on white pre-pupae stage, was conducted. It was shown that the bulk of GAGA and CNC binding falls into promoter regions and introns, with the maximal density of peaks in the vicinity of the transcription start site. Moreover, in both 0–12 and 16–24 hour old embryos GAGA and CNC are frequently co-localized, while on white pre-pupae stage there is no co-localization of these factors on a genome-wide scale. In order to select a set of genes potentially co-regulated by GAGA and CNC, the study of their co-binding in annotated regulatory regions (promoter areas and segments corresponding to the 5'-UTR and 3'-UTR of mRNA) was performed. The results obtained clearly demonstrated that the sets of genes characterized by co-binding of both factors vary greatly at different stages. Thus from 353 genes with overlapped GAGA and CNC binding loci on the 0–12 hour old embryos and 611 genes on the 16–24 hour old embryos only 61 genes “belong” to both stages. For an explanation it is proposed that different sets of target genes are regulated by combinations of various GAGA and CNC isoforms, which are characterized by distinct expression patterns during *Drosophila* embryogenesis. Functional annotation analysis of genes, in whose regulatory regions both GAGA and CNC were found at all investigated stages, demonstrates enrichment by genes controlling embryogenesis, neurogenesis and wing development. The data obtained suggest the interaction of GAGA and CNC during *D. melanogaster* embryogenesis.

Key words: transcription factors; GAGA; CNC; development; *D. melanogaster*.

Развитие и жизнедеятельность многоклеточного организма осуществляются за счет дифференциальной экспрессии многих тысяч генов, определенные наборы которых транскрибируются в различных типах клеток на разных этапах развития и отвечают на разнообразные внешние сигналы (Tomancak, Ohler, 2010; Меркулова и др., 2013). Известно, что регуляция экспрессии большинства генов происходит на транскрипционном уровне, на котором центральную роль играют транскрипционные факторы (ТФ) – регуляторные белки, опознающие определенные последовательности ДНК в регуляторных районах (Frith et al., 2008; Charoensawan et al., 2010). В зависимости от состава используемых сайтов связывания ТФ в регуляторных районах генов формируются различные ансамбли этих белков, и такая комбинаторика сайтов и взаимодействующих с ними ТФ имеет ключевое значение для определения уровня экспрессии различных генов, обусловленного стадией развития организма, типом клетки и внешними воздействиями (Harbison et al., 2004; Ochoa-Espinosa, Small, 2006; Bulger, Groudine, 2011; Меркулова и др., 2013).

D. melanogaster на протяжении многих лет является одной из наиболее популярных моделей для изучения сети генных взаимодействий и роли различных ТФ на разных стадиях онтогенеза, включая эмбриональное развитие (Adryan, Teichmann, 2010; Tomancak, Ohler, 2010). В ходе изучения эмбриогенеза *D. melanogaster* было выявлено более 10 ТФ (Bicoid, Nanos, Kruppel, Knirps, Giant, Hunchback, Caudal и др.) координированное действие которых определяет ранние стадии формирования эмбриона (Ingham, 1988; Rivera-Pomar, Jäckle, 1996; Segal et al., 2008). Проведенные позднее полногеномные исследования показали, что в контроле онтогенеза дрозофилы участвуют сотни ТФ, хотя конкретная роль большинства из них еще не определена (Adryan, Teichmann, 2010; Hammonds et al., 2013). До сих пор очень мало известно также о взаимодействии различных ТФ друг с другом в ходе развития дрозофилы. Поэтому представляется очевидным, что наряду с общим анализом полногеномных данных необходимы более конкретизированные исследования накопленного экспериментального материала, в частности изучение колокализации отдельных ТФ в масштабе генома и выявление круга генов, совместно регулируемых этими факторами.

Транскрипционный фактор GAGA, кодируемый геном *Trithorax-like (Trl)*, необходим для нормального развития *D. melanogaster* (Farkas et al., 1994). Он экспрессируется на всех стадиях развития, но особенно активно во время эмбриогенеза (<http://flybase.org/reports/FBgn0013263.html>). В ходе анализа полногеномных профилей связывания ТФ было установлено, что GAGA специфично связывается с так называемыми HOT-регионами (или Hotspot), местами с повышенной плотностью сайтов связывания разных ТФ по сравнению с остальным геномом (Kvon et al., 2012; Slattery et al., 2014). Поскольку GAGA присутствует в большинстве тканей и на всех стадиях онтогенеза, должен существовать механизм, обеспечивающий изменение наборов генов, регулируемых им в разных условиях. ТФ GAGA представлен несколькими изоформами, и одним из возможных механизмов специфичности его действия

является связывание с разными сайтами различных его гомо- и гетеродимеров (Espinosa et al., 1999). Другим возможным механизмом может быть взаимодействие GAGA с другими ТФ и белками хроматина для обеспечения изменения специфичности связывания с разными сайтами (Pagans et al., 2002). В этом случае изменение перекрывания профилей связывания ТФ на разных стадиях развития будут обеспечивать изменение набора регулируемых генов. В рамках проекта modENCODE (www.modencode.org; Roy et al., 2010) были получены профили связывания для более чем 80 ТФ на разных стадиях развития дрозофилы. Однако в настоящее время только для очень небольшого количества этих факторов получены профили связывания для нескольких следующих друг за другом стадий развития, что дает возможность проследить динамику колокализации районов связывания разных ТФ и сделать предположения об их возможном взаимодействии в течение развития. Для GAGA профили были получены у эмбрионов нескольких возрастов, а также на стадии белой предкуколки (БП). Только для трех ТФ (SuHW, STAT92E, CNC) также получены профили для этих же периодов развития. SuHW является важным компонентом инсуляторных комплексов, и взаимное расположение его сайтов и сайтов связывания GAGA исследовалось в том числе и на геномном уровне (Nègre et al., 2010). Для STAT92E и CNC имеются литературные данные о перекрывании их сайтов связывания с сайтами GAGA, полученные при анализе профилей только для 0–12-часовых эмбрионов (Nègre et al., 2011). На наш взгляд, наибольший интерес для изучения возможного взаимодействия с GAGA представляет CNC, кодируемый геном *cap-n-collar (cnc)*. CNC является важным регулятором эмбриогенеза дрозофилы, в частности он необходим для формирования головного сегмента эмбриона (Mohler et al., 1995).

Целью настоящей работы было изучение распределения сайтов связывания для двух важных регуляторов эмбрионального развития *D. melanogaster* – ТФ GAGA и CNC – в течение эмбрионального развития и на стадии БП. Мы определили также круг генов, которые, возможно, совместно регулируются этими факторами.

Материалы и методы

В анализ были взяты данные полногеномных экспериментов по иммунопреципитации хроматина из проекта modENCODE (www.modencode.org; Roy et al., 2010). Эксперименты были выполнены с использованием антител к константным областям транскрипционных факторов GAGA и CNC на линии *D. melanogaster iso1 (y, bw cn sp)* для эмбрионов в возрасте 0–12 и 16–24 ч и на стадии БП. Экспериментальные данные для эмбрионов 0–12 ч были получены с применением технологии ChIP-chip, остальные – с помощью ChIP-seq. Исходные данные были извлечены из архивов NCBI (<http://www.ncbi.nlm.nih.gov/geo/> для ChIP-chip и <http://www.ncbi.nlm.nih.gov/sra/> для ChIP-seq) (таблица). В работе использовали версию 6 аннотации генома дрозофилы (Hoskins et al., 2015; ftp://ftp.ensembl.org/pub/release-83/gtf/drosophila_melanogaster/Drosophila_melanogaster.BDGP6.83.gtf.gz).

Поскольку ChIP-chip для 0–12-часовых эмбрионов проводили в рамках одного эксперимента, данные о про-

Raw ChIP-seq and ChIP-on-chip data and peak detection results

Transcription factor	Stage	ID GEO	Number of peaks	Mean peak size (bp)
GAGA	Embryo 0–12 h	GSM409071	3 902	1 826
	Embryo 16–24 h	SRR063878, SRR063879, SRR063880*	6 682	658
	White prepupa	SRR099124, SRR099125, SRR099126*	18 668	346
CNC	Embryo 0–12 h	GSM399627	697	1 037
	Embryo 16–24 h	SRR390242, SRR390243, SRR390244*	823	618
	White prepupa	SRR350870, SRR350871, SRR350872*	415	568

* input control.

филе распределения пиков взяты без предварительной обработки.

Предпроцессинг данных ChIP-seq проводили с использованием программы Trimmomatic-3.2.2 (Bolger et al., 2014). Средняя длина прочтения составила 36 п. н. Выравнивание проводили на последовательность генома дрозофилы версии 6 с помощью пакета программ Bowtie2 (Langmead, Salzberg, 2012). Для удаления эффекта дубликации прочтений использовали пакет программ SAMtools 1.2 (Li et al., 2009). Поиск пиков связывания проводили с помощью программы MACs2 2.0.9 (qvalue = 0.01) (Zhang et al., 2008). В качестве контроля для бутстреп-анализа были сформированы искусственные данные по каждому транскрипционному фактору, содержащие то же количество пиков для каждой хромосомы и размеры пиков, соответствующие средней протяженности пиков в экспериментальных данных, но со случайным распределением по геному. Относительно этой же выборки оценивали обогащение различных районов генома пиками связывания CNC и GAGA, полученными в эксперименте. Плотность распределения пиков (ρ) в различных районах генома (промоторы, 5'- и 3'-UTRs, интроны, кодирующие районы и межгенные промежутки) рассчитывали по формуле:

$$\rho = \frac{Nr}{N + Sr},$$

где Nr – количество пиков данного ТФ, попадающих в определенный район генома, N — общее количество пиков данного ТФ, Sr — суммарная площадь данного района генома в мегабазах.

Координаты HOT-районов и STARR-Seq энхансеров взяты из ранее опубликованных работ (Roy et al., 2010; Arnold et al., 2013). Преобразование ординат к версии 6 аннотации генома дрозофилы производили с использованием Drosophila Sequence Coordinates Converter (http://flybase.org/static_pages/downloads/COORD.html).

Анализ обогащения генов-мишеней выполняли с использованием пакета программ WebGestalt (<http://www.webgestalt.org/>).

Результаты

При сравнении результатов экспериментов по иммунопреципитации хроматина было выявлено значительно большее количество мест связывания GAGA фактора с ДНК по сравнению с CNC (см. таблицу). Причем если в случае с CNC это количество остается относительно ста-

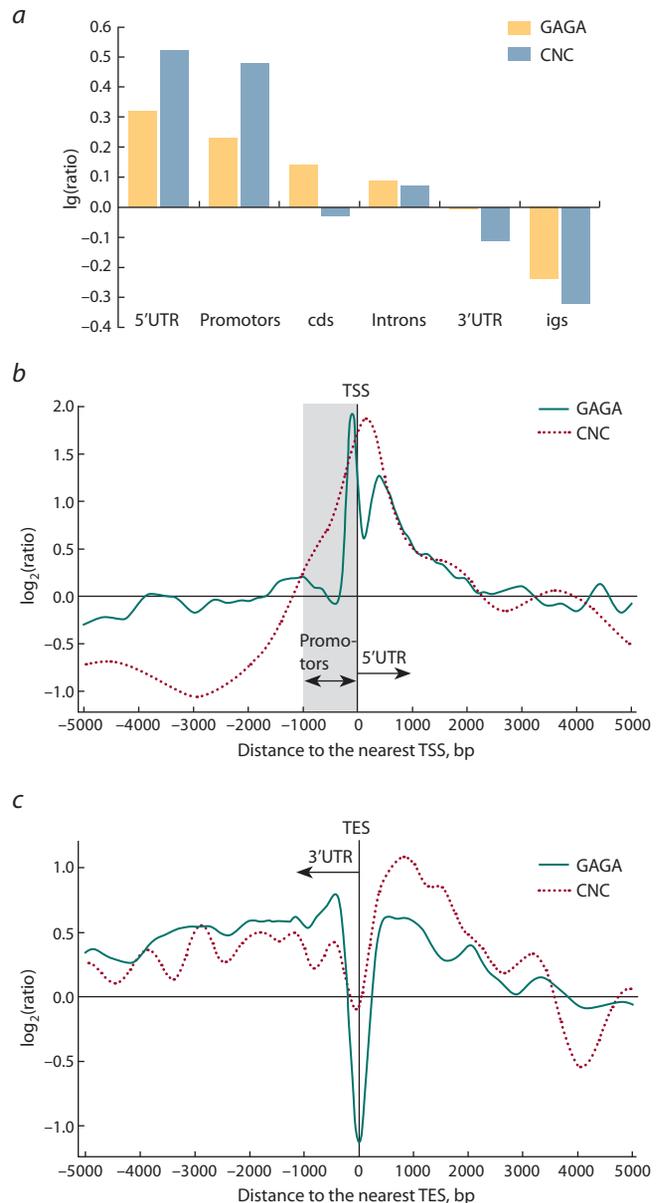


Fig. 1. (a) Distribution of CNC and GAGA-binding peaks in the *Drosophila* genome. Designations: cds, coding sequences; igs, intergenic spacers. (b) Regions enriched and depleted in TFBSs around TSS. (c) Same around TES. Y-axis: logarithmic ratios between the densities of GAGA and CNC peaks around (b) TSS and (c) TES and the densities of the peaks in an artificial sample in the same region (see Materials and methods).

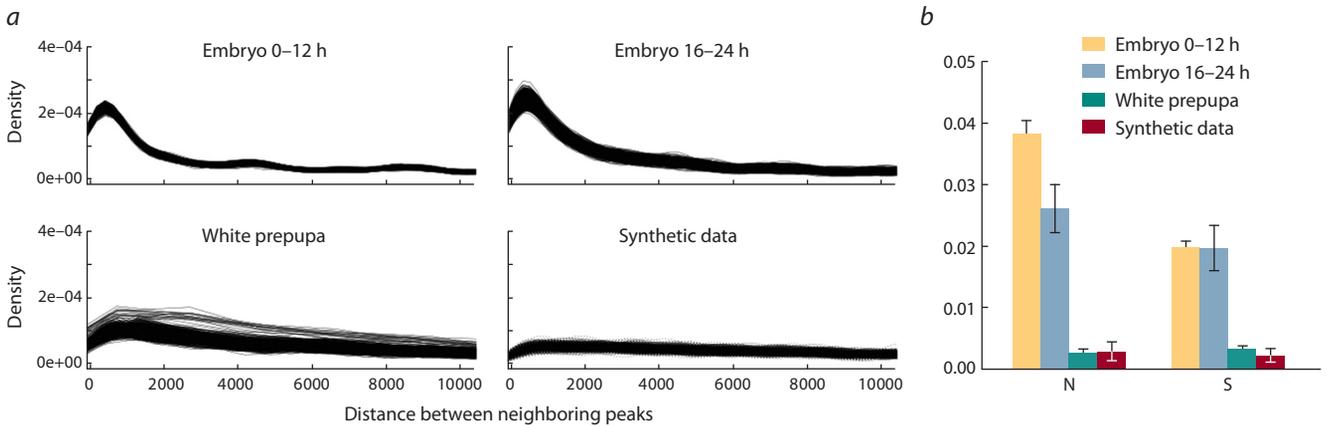


Fig. 2. Arrangement of GAGA- and CNC-binding peaks in the *D. melanogaster* genome.

(a) Distance between the tops of neighboring CNC and GAGA peaks, bp; (b) shares of the numbers (N) and coverage areas (S) of overlapping peaks in the total number of peaks $N_{\text{over}} / (N_{\text{CNC}} + N_{\text{GAGA}})$ and the total coverage area $S_{\text{over}} / (S_{\text{CNC}} + S_{\text{GAGA}} - S_{\text{over}})$. Error bars indicate 95% Gaussian confidence ranges ($\pm 1.96 \sigma$).

бильным во всех стадиях, то количество пиков связывания GAGA растет по мере развития мух.

При оценке распределения пиков связывания GAGA и CNC по геному для всех исследованных стадий было выявлено, что районы начала генов, включающие промоторную область (интервал от старта транскрипции до -1000 п. н.) и участок, соответствующий 5'-UTR мРНК, существенно обогащены пиками обоих факторов (рис. 1, а). Следует отметить достаточно высокую плотность пиков GAGA в кодирующих областях генов (см. рис. 1, а). Возможно, это связано с участием GAGA в процессе элонгации транскрипции (O'Brien et al., 1995). Максимальная плотность пиков и GAGA и CNC приходится на область старта транскрипции (TSS) (см. рис. 1, б). Такой характер распределения мест связывания в геноме зафиксирован и для других транскрипционных факторов (Farnham, 2009) и подтверждает принятую точку зрения о важной роли промоторных районов генов в регуляции транскрипции (Меркулова и др., 2013). Минимальная плотность пиков обоих факторов наблюдается в районе сайта терминации транскрипции (TES), что ранее было зафиксировано для GAGA (Nègre et al., 2010), а для CNC наблюдается впервые.

Для исследования возможного взаимодействия транскрипционных факторов была проведена оценка минимального расстояния между пиками GAGA и CNC на разных стадиях развития мух. Чтобы нивелировать влияние разницы в количестве пиков, полученных в разных экспериментах, на результат вычисления, для каждой стадии были случайным образом сформированы подвыборки, содержащие 3902 пиков GAGA и 415 CNC, и на этих подвыборках оценивали расстояние между вершинами соседних пиков ТФ. Достоверность определяли с использованием бутстреп-анализа (999 реплик). Результаты представлены на рис. 2. Видно, что на эмбриональной стадии развития пики исследуемых ТФ тяготеют друг к другу. На стадии БП колокализации сайтов GAGA и CNC на геномном уровне практически не наблюдается: из рис. 2, б видно, что

расстояние между пиками на этой стадии достоверно не отличается от расстояний между пиками в искусственно созданной нами выборке. При оценке количества и площади перекрытия пиков мы также видим колокализацию CNC и GAGA на эмбриональных стадиях и ее отсутствие на стадии БП в масштабе генома (см. рис. 2, б).

Для исследования функциональной роли участков возможного совместного связывания CNC и GAGA мы проанализировали перекрытие всех таких участков с HОТ-регионами и STARR-seq энхансерами. HОТ-регионы – это участки ДНК, в которых плотность сайтов связывания ТФ повышена (Moogman et al., 2006). Было продемонстрировано, что HОТ-регионы часто обладают энхансерной активностью (Kvon et al., 2012). Также ранее был предложен экспериментальный метод для полногеномного выявления энхансеров (STARR-seq), с помощью которого авторы выявили более 10000 участков, обладающих энхансерной активностью в S2 и OSC клетках (Arnold et al., 2013). Было установлено, что районы колокации GAGA и CNC с высокой частотой перекрываются и с HОТ-регионами, и с энхансерами (рис. 3). С HОТ-регионами перекрывается 90 %, а с энхансерами – 42 % районов колокации CNC и GAGA, что достоверно больше, чем перекрытие каждого из ТФ по отдельности. Таким образом, значительная часть районов колокации может обладать энхансерной активностью и участвовать в активации экспрессии рядом расположенных генов.

Особый интерес представляет возможная колокализация изучаемых ТФ в регуляторных районах генов. В качестве таких районов были выбраны промоторная область и прилегающий к ней участок, соответствующий 5'-UTR мРНК, а также участок, соответствующий 3'-UTR мРНК. Оказалось, что у эмбрионов в возрасте 0–12 ч 353 гена характеризуются пересечением пиков связывания GAGA и CNC в этих районах, в 12–24 ч – число таких генов составляет 611, а на стадии БП – 180 (рис. 4). Из приведенных на рис. 4 данных видно также, что число ге-

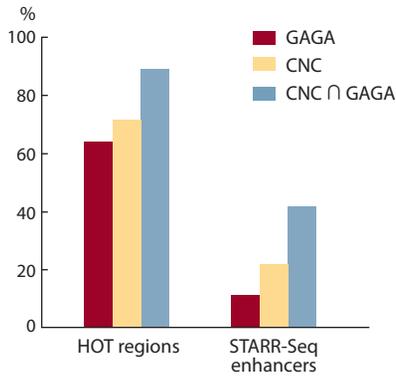


Fig. 3. Shares of GAGA- and CNC-binding regions and their putative linkage (CNC ∩ GAGA) that overlap HOT regions and S2 enhancers.

Data are integrated over all *Drosophila* developmental stages investigated. The share of regions with CNC and GAGA colocalization is significantly ($p < 0.01$ according to Fisher's exact test) greater than the share of each factor in both cases.

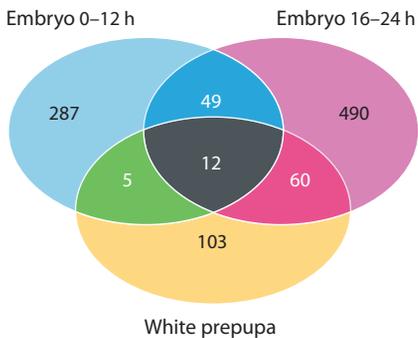


Fig. 4. Numbers of genes with peaks of GAGA and CNC-binding sites at the 0–12 h and 16–24 h embryonic stages and at the white prepupa stage.

нов, в регуляторных районах которых обнаружено совместное связывание GAGA и CNC на более одной стадии, весьма невелико. Так, только в 12 из них колокализация ТФ наблюдалась на всех стадиях. Для эмбрионов в возрасте 0–12 и 16–24 ч количество таких «общих» районов равно 61, в 0–12 ч и БП – 17, а в 16–24 ч и БП – 72. Наиболее вероятным объяснением этому явлению служит изменение паттернов экспрессии изоформ GAGA и CNC в ходе эмбриогенеза дрозофилы. Это позволяет предполагать, что разные подгруппы генов-мишеней GAGA и CNC регулируются различными сочетаниями изоформ GAGA и CNC.

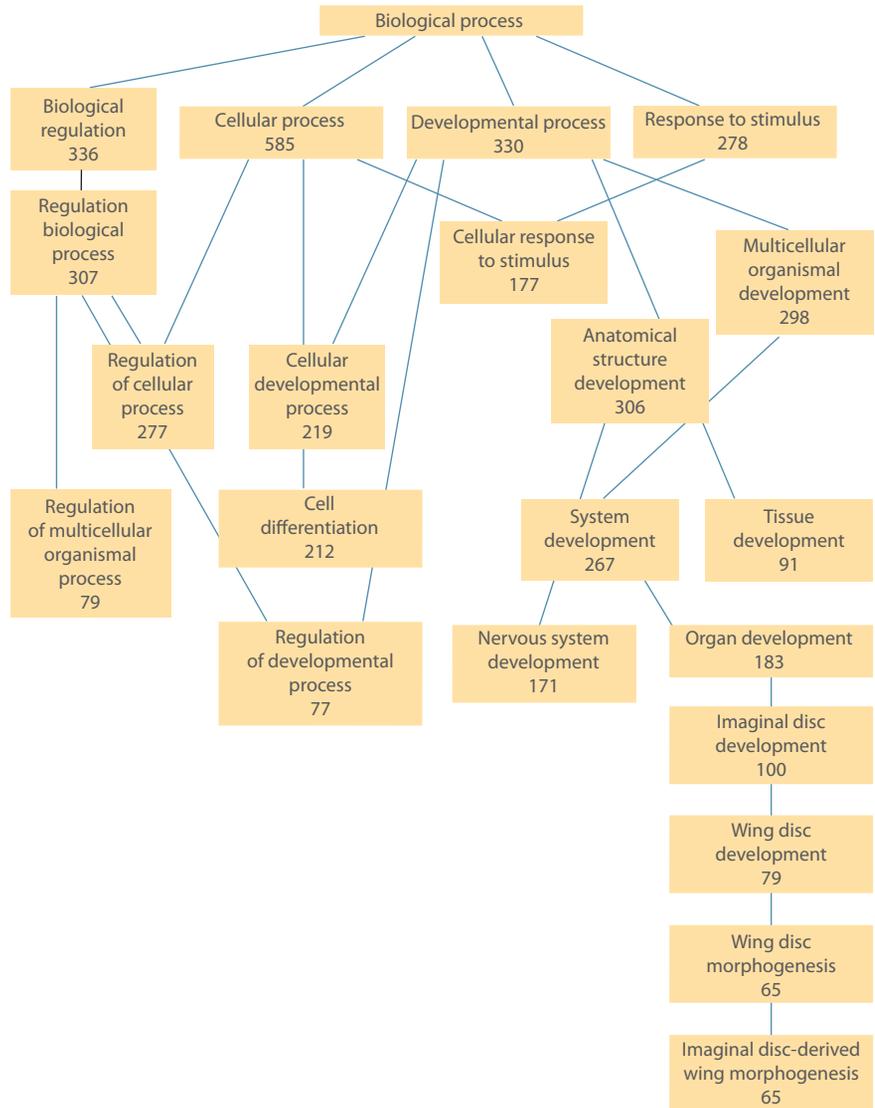


Fig. 5. Involvement of genes targeted by CNC and GAGA TFs in various biologic processes according to analysis with WebGestalt software ($p < 0.02$).

Мы проанализировали функции всех выявленных нами предполагаемых генов-мишеней обоих ТФ (1006 генов). В результате выяснилось, что совокупность этих генов обогащена определенными группами. Так, по термину «молекулярные функции» выявлено повышенное содержание генов, кодирующих транскрипционные факторы (60 генов). По термину «биологические функции» среди генов-мишеней обоих ТФ также наблюдалось существенное обогащение генами-регуляторами различных процессов (336 генов). Это, вероятно, связано с тем, что *cnc* и *Trl* находятся на вершине регуляторных каскадов, а их мишени-регуляторы осуществляют регуляцию второго порядка. Более детально процессы, в которые вовлечены предполагаемые гены-мишени, представлены на рис. 5. Среди них интересны процессы развития нервной системы и морфогенеза крыла. Представленность генов, связанных с этими процессами, не одинакова на разных стадиях развития. Так, если обогащение по генам развития нервной системы наблюдается на всех изучаемых стадиях (эмбрионы 0–12 ч, эмбрионы 16–24 ч, БП), то обогащение по генам, вовлеченным в морфогенез крыльев, – только на стадии эмбриогенеза 16–24 ч, что коррелирует со временем инвагинации имагинальных дисков (Campos-Ortega, Hartenstein, 1997).

Обсуждение

Разработка полногеномных технологий изучения экспрессии генов, функциональной организации хроматина и локализации различных ТФ в масштабе генома (RNA-seq, ChIP-chip, ChIP-seq, DamID) привела к существенному расширению знаний о характерных для разных типов клеток профилях транскрипции генов, модификаций хроматина и распределения ТФ, а также об изменениях этих профилей в ходе развития многоклеточного организма (Filion et al., 2010; Roy et al., 2010; Lelli et al., 2012; Diehl, Boyle, 2016). Однако в настоящее время темпы накопления экспериментальных данных, получаемых с помощью указанных технологий, значительно превышают темпы извлечения из них биологической информации. Так, в базах проекта modENCODE содержится уже довольно много данных ChIP-seq экспериментов по определению полногеномного распределения ряда ТФ на разных стадиях развития *D. melanogaster*, однако их анализ далек от завершения. В частности, для весьма немногих ТФ выполнены работы по выявлению совместной локализации этих факторов (Slattery et al., 2014), которая может служить весомым указанием на их взаимодействие в процессе регуляции тех или иных наборов генов. Среди факторов, представляющих интерес для подобного изучения, мы выделили два важных регулятора развития дрозофилы, ТФ GAGA и CNC, для которых в modENCODE содержатся данные ChIP-chip и ChIP-seq экспериментов, выполненных во время 0–12 ч и 16–24 ч развития эмбриона, а также на стадии белой предкуколки.

Известно, что гены *Trl* и *cnc* экспрессируются на всех стадиях онтогенеза дрозофилы (Перельгина и др., 1993; Soeller et al., 1993; <http://www.flyexpress.net/>). Экспрессия *Trl* достигает максимума у эмбрионов 4–12 ч развития, а затем постепенно снижается (Soeller et al., 1993; Bhat et al., 1996; Karagodin et al., 2013). В раннем эмбриогенезе транскрипты *Trl* равномерно распределены по всему эмбриону, а в дальнейшем транскрипция наблюдается преимущественно в формирующейся нервной системе (Перельгина и др., 1993). Основными транскриптами гена *Trl* являются два: длиной 2.5 и 3.0 т.н. На стадии 0–4 ч развития эмбрионов преобладает 2.5 т.н. транскрипт, а начиная с 4 ч увеличивается количество транскрипта длиной 3.0 т.н., и с 8–12 ч он становится преобладающим и остается таковым до конца эмбриогенеза (Soeller et al., 1993; Karagodin et al., 2013). Известно, что транскрипт длиной 2,5 т.н. соответствует изоформе белка GAGA-519, а транскрипт длиной 3.0 т.н. – изоформе GAGA-581 (Karagodin et al., 2013). Несмотря на то что соотношение основных транскриптов гена *Trl* меняется в ходе онтогенеза, установлено, что только в первые 6 ч развития эмбрионов в них доминирует белок GAGA-519, а в дальнейшем обе изоформы выявляются примерно в равных количествах (Benyajati et al., 1997).

Транскрипты гена *cnc* на первых стадиях эмбриогенеза так же, как и транскрипты гена *Trl*, равномерно распределены во всем эмбрионе, а в дальнейшем транскрипция *cnc* в основном ограничена областью формирующейся головы (<http://www.flyexpress.net/>). Выявлены три главных транскрипта гена *cnc* длиной 6.6; 5.4; 3.3 т.н. (McGinnis et al., 1998). У 2–8-часовых эмбрионов преобладающими

являются транскрипты 5.4 и 4.0 т.н., у 8–12-часовых эмбрионов основной транскрипт – 5.4 т.н., а у 12–24-часовых присутствуют все три транскрипта. Выявлены соответствующие этим транскриптам изоформы белка CNC длиной 533 а.о. (изоформа А, 3.3 т.н. транскрипт), 805 а.о. (изоформа В, 5.4 т.н. транскрипт) и 1383 а.о. (изоформа С, 6.6 т.н. транскрипт). Транскрипты, дающие начало изоформам А и С, и сами белки нарабатываются еще во время оогенеза, поэтому они выявляются в раннем эмбриогенезе еще до начала зиготической транскрипции. В дальнейшем эти изоформы появляются вновь уже после 11 ч эмбриогенеза и присутствуют до конца этого процесса. Значительное количество изоформы В выявляется в головном отделе эмбриона на стадиях 6–14 (3–12 ч) эмбриогенеза, тогда как в другое время этот белок обнаруживается в эмбрионе в очень малом количестве (McGinnis et al., 1998). Таким образом, если разные изоформы белка GAGA выявляются в течение большей части эмбриогенеза приблизительно в равных количествах, то разные изоформы белка CNC распределены по разным периодам эмбриогенеза достаточно специфично.

Анализ взаимного расположения пиков связывания CNC и GAGA позволил выявить районы, в которых пики перекрываются на определенной стадии развития. В этих районах сайты связывания CNC и GAGA располагаются достаточно близко друг к другу, для того чтобы было возможно влияние связывания одного из ТФ на связывание второго и даже физическое взаимодействие между ними для совместного связывания. Нужно отметить, что данные ChIP-seq и ChIP-chip относятся к целым эмбрионам и БП, поэтому нельзя исключить наложение пиков ТФ от разных типов клеток. Однако следует учитывать высокую степень перекрывания районов экспрессии генов *cnc* и *Trl*. Так, в раннем эмбриогенезе GAGA и CNC присутствуют во всех клетках эмбриона, а на более поздних стадиях их экспрессия перекрывается в головном сегменте эмбриона. Поэтому кажется маловероятным, что большая часть наблюдаемого перекрытия возникает за счет наложения пиков от разных типов клеток. Как следствие, поиск мест совместной локализации CNC и GAGA служит первым этапом для выявления возможной совместной регуляции генов этими ТФ в течение развития дрозофилы.

Проведенный нами анализ высветил новые аспекты возможного совместного действия этих факторов в ходе эмбрионального развития *D. melanogaster*. Мы показали, что в периоды 0–12 и 16–24 ч развития эмбриона наблюдается неслучайная колокализация GAGA и CNC, в то время как на стадии белой предкуколки колокализации этих факторов в масштабе генома выявить не удастся. Данные указывают на возможность совместной регуляции GAGA и CNC определенного круга генов во время эмбриогенеза. Для того чтобы выявить эти гены, было осуществлено исследование колокализации изучаемых ТФ в аннотированных регуляторных районах (промоторная область и участки, соответствующие 5'-UTR и 3'-UTR мРНК). Оказалось, что совокупности генов, в регуляторных районах которых обнаружено связывание обоих ТФ, сильно различаются на разных стадиях. Так, если в период 0–12 ч эмбриогенеза 353 гена характеризуются пересечением пиков ChIP-seq GAGA и CNC, а в

период 12–24 ч число таких генов составляет 611, то всего лишь 61 ген является «общим» для обоих периодов. Мы предполагаем, что такое варьирование числа генов, характеризующихся совместным связыванием обоих белков, определяется тем, что на разных стадиях развития преобладают различные изоформы CNC и GAGA, которые, по-видимому, обладают определенной избирательностью по отношению к сайтам связывания этих белков на ДНК и/или в составе хроматина.

Функциональный анализ генов, в регуляторных районах которых найдена колокализация ТФ GAGA и CNC, выделяет группы генов, контролирующих эмбриогенез, развитие нервной системы и развитие крыла дрозофилы. Следует отметить, что ТФ GAGA так же, как и ТФ CNC, влияют на ход этих процессов (Iyer et al., 2013; Schertel et al., 2015). Так, хорошо известно, что GAGA регулирует активность многих генов, контролирующих эмбриогенез дрозофилы (Granok et al., 1995). Ген *snr* также принимает участие в этом процессе, влияя на работу генов развития головы (Veraksa et al., 2000). Однако до настоящего времени не было данных о совместном участии GAGA и CNC в регуляции транскрипции конкретных генов. Полученные нами результаты указывают на существование обширной и динамично меняющейся сети генов, совместно регулируемых этими ТФ в ходе эмбрионального развития дрозофилы, и открывают новые перспективы в изучении этого процесса.

Acknowledgments

This work was supported by State Budgeted Project 0324-2015-0003 and the Russian Foundation for Basic Research, project 16-04-01018-a.

Conflict of interest

The authors declare no conflict of interest.

References

Adryan B., Teichmann S.A. The developmental expression dynamics of *Drosophila melanogaster* transcription factors. *Genome Biol.* 2010; 11(4):R40.

Arnold C.D., Gerlach D., Stelzer C., Boryn L.M., Rath M., Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science.* 2013;339(6123):1074-1077.

Benyajati C., Mueller L., Xu N., Pappano M., Gao J., Mosammaparast M., Conklin D., Granok H., Craig C., Elgin S. Multiple isoforms of GAGA factor, a critical component of chromatin structure. *Nucl. Acids Res.* 1997;25(16):3345-3353.

Bhat K.M., Farkas G., Karch F., Gyurkovics H., Gausz J., Schedl P. The GAGA factor is required in the early *Drosophila* embryo not only for transcriptional regulation but also for nuclear division. *Development (Cambridge, England).* 1996;122(4):1113-1124.

Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England).* 2014; 30(15):2114-2120.

Bulger M., Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 2011;144(3):327-339.

Campos-Ortega J.A., Hartenstein V. *The Embryonic Development of Drosophila Melanogaster.* Berlin; Heidelberg: Springer, 1997.

Charoensawan V., Wilson D., Teichmann S.A. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucl. Acids Res.* 2010;38(21):7364-7377.

Diehl A.G., Boyle A.P. Deciphering ENCODE. *Trends Genet.* 2016; 32(4):238-249.

Espinosa M.L., Jimenez-Garcia E., Vaquero A., Canudas S., Bernues J., Azorin F. The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. *J. Biol. Chemistry.* 1999;274:16461-16469.

Farkas G., Gausz J., Galloni M., Reuter G., Gyurkovics H., Karch F. The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature.* 1994;371(6500):806-808.

Farnham P.J. Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 2009;10(9):605-616.

Filion G.J., van Bommel J.G., Braunschweig U., Talhout W., Kind J., Ward L.D., Brugman W., de Castro I.J., Kerkhoven R.M., Bussemaker H.J., van Steensel B. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell.* 2010; 143(2):212-224.

Frith M.C., Valen E., Krogh A., Hayashizaki Y., Carninci P., Sandelin A. A code for transcription initiation in mammalian genomes. *Genome Res.* 2008;18(1):1-12.

Granok H., Leibovitch B.A., Shaffer C.D., Elgin S.C. Chromatin. GAGA over GAGA factor. *Curr. Biol.* 1995;5(3):238-241.

Hammonds A.S., Bristow C.A., Fisher W.W., Weiszmann R., Wu S., Hartenstein V., Kellis M., Yu B., Frise E., Celniker S.E. Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biol.* 2013;14(12):R140.

Harbison C.T., Gordon D.B., Lee T.I., Rinaldi N.J., Macisaac K.D., Danford T.W., Hannett N.M., Tagne J.-B., Reynolds D.B., Yoo J., Jennings E.G., Zeitlinger J., Pokholok D.K., Kellis M., Rolfe P.A., Takusagawa K.T., Lander E.S., Gifford D.K., Fraenkel E., Young R.A. Transcriptional regulatory code of a eukaryotic genome. *Nature.* 2004;431(7004):99-104.

Hoskins R.A., Carlson J.W., Wan K.H., Park S., Mendez I., Galle S.E., Booth B.W., Pfeiffer B.D., George R.A., Svirskas R., Krzywinski M., Schein J., Accardo M.C., Damia E., Messina G., Méndez-Lago M., de Pablos B., Demakova O.V., Andreyeva E.N., Boldyreva L.V., Marra M., Carvalho A.B., Dimitri P., Villasante A., Zhimulev I.F., Rubin G.M., Karpen G.H., Celniker S.E. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* 2015;25(3):445-458.

Ingham P.W. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature.* 1988;335(6185):25-34.

Iyer E.P.R., Iyer S.C., Sullivan L., Wang D., Meduri R., Graybeal L.L., Cox D.N. Functional genomic analyses of two morphologically distinct classes of *Drosophila* sensory neurons: post-mitotic roles of transcription factors in dendritic patterning. *PLoS ONE.* 2013;8(8):e72434.

Karagodin D.A., Omelina E.S., Fedorova E.V., Baricheva E.M. Identification of functionally significant elements in the second intron of the *Drosophila melanogaster* Trithorax-like gene. *Gene.* 2013;520(2): 178-184.

Kvon E.Z., Stampfel G., Yanez-Cuna J.O., Dickson B.J., Stark A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Develop.* 2012;26(9): 908-913.

Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 2012;9(4):357-359.

Lelli K.M., Slattery M., Mann R.S. Disentangling the many layers of eukaryotic transcriptional regulation. *Ann. Rev. Genet.* 2012;46(1):43-68.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. & 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England).* 2009;25(16):2078-2079.

McGinnis N., Ragnhildstveit E., Veraksa A., McGinnis W. A cap “n” collar protein isoform contains a selective Hox repressor function. *Development (Cambridge, England).* 1998;125(22):4553-4564.

Merkulova T.I., Ananko E.A., Ignat'eva E.V., Kolchanov N.A. Regulatory transcription codes in eukaryotic genomes. *Genetika = Genetics (Moscow).* 2013;49(1):37-54. (in Russian)

Mohler J., Mahaffey J.W., Deutsch E., Vani K. Control of *Drosophila* head segment identity by the bZIP homeotic gene *cnc*. *Development (Cambridge, England).* 1995;121(1):237-247.

- Moorman C., Sun L.V., Wang J., de Wit E., Talhout W., Ward L.D., Greil F., Lu X.-J., White K.P., Bussemaker H.J., van Steensel B. Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 2006;103:12027-12032.
- Nègre N., Brown C.D., Ma L., Bristow C.A., Miller S.W., Wagner U., ... White K.P. A cis-regulatory map of the *Drosophila* genome. *Nature*. 2011;471(7339):527-531. <http://doi.org/10.1038/nature09990>.
- Nègre N., Brown C.D., Shah P.K., Kheradpour P., Morrison C.A., Henikoff J.G., ... White K.P. A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genetics*. 2010;6(1):e1000814. <http://doi.org/10.1371/journal.pgen.1000814>.
- O'Brien T., Wilkins R.C., Giardina C., Lis J.T. Distribution of GAGA protein on *Drosophila* genes in vivo. *Genes Development*. 1995;9(9):1098-1110.
- Ochoa-Espinosa A., Small S. Developmental mechanisms and cis-regulatory codes. *Curr. Opin. Gen. Dev.* 2006;16(2):165-170.
- Pagans S., Ortiz-Lombardia M., Espinas M.L., Bernues J., Azorin F. The *Drosophila* transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. *Nucl. Acids Res.* 2002;30(20):4406-4413.
- Perelygina L.M., Baricheva E.M., Sebeleva T.E., Kokoza V.A. The evolutionarily conserved gene Nc70F expressed in nerve tissue of *Drosophila melanogaster* encodes a protein homologous to the mouse delta transcription factor. *Genetika = Genetics (Moscow)*. 1993;29(10):1597-1607.
- Rivera-Pomar R., Jäckle H. From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet.* 1996;12(11):478-483.
- Roy S., Ernst J., Kharchenko P.V., Kheradpour P., Negre N., Eaton M.L., Landolin J.M., Bristow C.A., Ma L.J., Lin M.F., ... Kellis M. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*. 2010;330(6012):1787-1797.
- Schertel C., Albarca M., Rockel-Bauer C., Kelley N.W., Bischof J., Hens K., van Nimwegen E., Basler K., Deplancke B. A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating *Drosophila* wing development. *Genome Res.* 2015;25(4):514-523.
- Segal E., Raveh-Sadka T., Schroeder M., Unnerstall U., Gaul U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature*. 2008;451(7178):535-540.
- Slattery M., Ma L., Spokony R.F., Arthur R.K., Kheradpour P., Kundaje A., Nègre N., Crofts A., Ptashkin R., Zieba J., Ostapenko A., Suchy S., Victorsen A., Jameel N., Grundstad A.J., Gao W., Moran J.R., Rehm E.J., Grossman R.L., Kellis M., White K.P. Diverse patterns of genomic targeting by transcriptional regulators in *Drosophila melanogaster*. *Genome Res.* 2014;24(7):1224-1235.
- Soeller W.C., Oh C.E., Kornberg T.B. Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* 1993;13(12):7961-7970.
- Tomancak P., Ohler U. Mapping the complexity of transcription control in higher eukaryotes. *Genome Biol.* 2010;11(4):115.
- Veraksa A., McGinnis N., Li X., Mohler J., McGinnis W. Cap "n" collar B cooperates with a small Maf subunit to specify pharyngeal development and suppress deformed homeotic function in the *Drosophila* head. *Development (Cambridge, England)*. 2000;127(18):4023-4037.
- Zhang Y., Liu T., Meyer C.A., Eeckhoutte J., Johnson D.S., Bernstein B.E., Nusbaum C., Myers R.M., Brown M., Li W., Liu X.S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.