

## ВЫВОД СХЕМЫ РЕГУЛЯЦИИ ФАКТОРОВ ТРАНСКРИПЦИИ ИЗ КОВАРИАЦИИ ГЕННОЙ ЭКСПРЕССИИ НА ПРИМЕРЕ ЕСТЕСТВЕННЫХ ПОПУЛЯЦИЙ *Drosophila melanogaster*

© 2018 г. Н.М. Осман\* \*\*, Т.Х. Китапчи\*, С. Влахо\*,  
З. Вундерлих\*\*\*, С.В. Нуждин\* \*\*\*\*

\*Университет Южной Калифорнии, Лос-Анжелес, Калифорния, США

\*\*Национальный исследовательский центр, 12622, Докки, Гиза, Египет

\*\*\*Университет Калифорнии, 92697, Ирвин, Калифорния, США

\*\*\*\*Санкт-Петербургский политехнический университет Петра Великого,  
195251, Санкт-Петербург, Политехническая ул., 29

E-mail: snuzhdin@usc.edu

Поступила в редакцию 02.12.16 г.

Генные регуляторные сети контролируют сложные программы, которые управляют развитием. Расшифровка связи между факторами транскрипции и генами-мишенями является сложной задачей отчасти потому, что факторы транскрипции связываются с тысячами мест в геноме, но контролируют экспрессию через подмножество этих случаев связывания. Мы предполагаем, что можем объединить естественное изменение уровня экспрессии и предсказание сайтов связывания факторов транскрипции, чтобы идентифицировать цели этих факторов. Мы собрали данные РНК-секвенирования из 71 генетически различных эмбрионов F1 *Drosophila melanogaster* и рассчитали корреляции между факторами транскрипции и уровнями экспрессии потенциальных генов-мишеней, которые мы назвали «регуляторная сила». Для того чтобы отделить прямые и косвенные мишени факторов транскрипции, мы предположили, что прямые мишени будут иметь перевес сайтов связывания в их восходящих регионах. Используя 14 факторов транскрипции, активных во время эмбриогенеза, мы обнаружили, что 12 факторов транскрипции показали существенную корреляцию между их силой связывания и регуляторной силой на нисходящих мишенях, и 10 факторов транскрипции показали существенную корреляцию между числом сайтов связывания и регуляторной силой на генах-мишенях. Основные функции, например работа гена *bicoid* в качестве активатора, и специфические взаимодействия, которые мы наблюдали между нашими факторами транскрипции, например работа гена *twist* как репрессора генов *sloppy paired* и *odd paired*, как правило, совпадают с литературными данными.

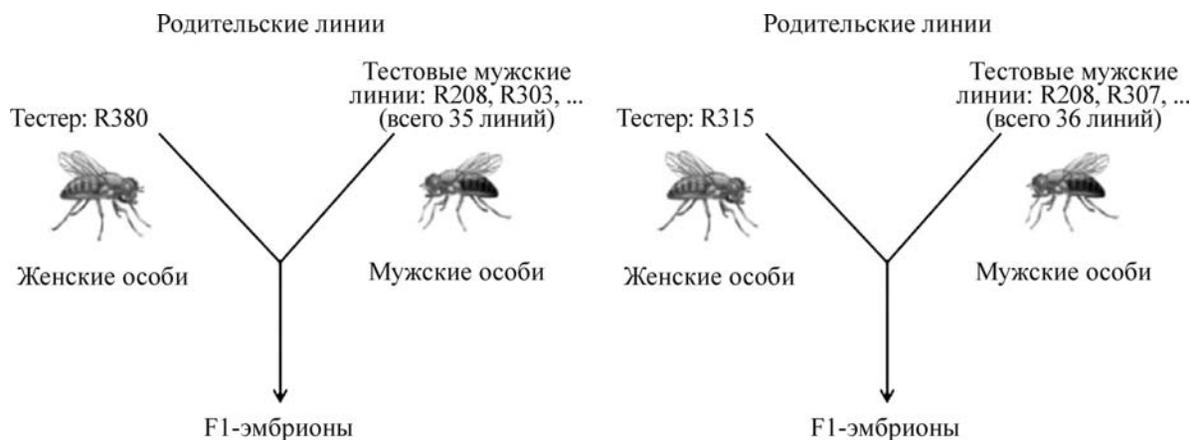
*Ключевые слова:* *Drosophila melanogaster*, генные регуляторные сети, факторы транскрипции.

Запутанные генные регуляторные сети отвечают за явления формирования структуры в процессе развития, которые создают сложное взрослое животное из одного оплодотворенного яйца. Эти сети состоят из факторов транскрипции (ТФ), комплексов ремоделирования хроматина, коактиваторов и компонентов сигнальных путей и в значительной степени закодированы в геноме кусками регуляторной ДНК, например энхансерами. Энхансеры расположены в некодирующих участках генома и состоят из сайтов связывания ТФ [1,2]. Число, сила и расположение сайтов связывания ТФ в пределах эн-

хансера помогают определить характер экспрессии, управляемой энхансером [3–7].

Сети, которые определяют передне-заднюю и дорзо-вентральную оси на ранних стадиях развития дрозофилы – признанные модельные системы для изучения генных регуляторных сетей. Передне-задняя структурирующая сеть является транскрипционным каскадом, в котором гены в восходящих звеньях каскада регулируют экспрессию генов в нисходящих звеньях. Гены материнского эффекта запускают экспрессию *gap*-генов. Эти *gap*-гены, которые представляют собой самое верхнее зиготическое звено каскада, управляют характером работы *pair-rule*-генов, которые затем определяют паттерны экспрессии генов полярности сегментов [8,9]. Дорзо-вентральная система формирования структу-

Сокращения: ТФ – факторы транскрипции, РПКМ – число ридов на килобазу, отнесенное на миллион ридов.



**Рис. 1.** Схемы скрещивания для образцов F1. Мы создали эмбрионы F1 с использованием двух тестерных женских штаммов, R380 и R315 из DGRP. В общей сложности было выполнено 71 различное скрещивание с использованием этих двух тестерных штаммов и различных штаммов самцов.

ры начинается с вентральной активации сигнального пути Toll, который активирует дорсальный, принадлежащий к семейству ТФ NF- $\kappa$ B, который активирует несколько других ТФ, важных для детализации дорзо-вентральной оси [9,10].

Одна из задач изучения генных регуляторных сетей – найти все соединения между ТФ, участвующими в сети. Традиционно это достигается за счет сочетания мутантных экспериментов и экспериментов по избыточной экспрессии ТФ, угнетения энхансеров, биохимического футпринтинга сайтов связывания ТФ и более новых подходов функциональной геномики, таких как ChIP-seq и RNA-seq, а также объединения со статистическим анализом для изучения генных регуляторных сетей [11–18].

В данной работе мы совмещаем полногеномное транскрипционное профилирование отдельных линий *D. melanogaster* с анализом сайтов связывания ТФ, чтобы найти связи в ранне-эмбриональных структурирующих сетях. Мы предполагаем, что звенья ТФ и его гены-мишени будут совместно меняться и используем биоинформационный анализ предсказанных сайтов связывания ТФ для разделения прямых и косвенных эффектов, аналогично более ранним исследованиям [19–21]. Для измерения ковариации между ТФ и их генами-мишенями мы использовали естественные вариации в ТФ и уровнях экспрессии гена-мишени, измеренных у эмбрионов в возрасте четырех–пяти часов [22], и ожидали значительное изменение экспрессии между генотипами, поскольку более ранние исследования показали значительные различия на уровне транскрипта среди генотипов *D. melanogaster* для примерно 10% от всего

транскриптома организма [23,24]. Мы скрестили мужских особей с различными генотипами с двумя тестовыми женскими линиями, в результате чего получили эмбрионы F1 с различными наследственностями, которые управляют изменением ТФ и уровней экспрессии генов-мишеней (рис. 1). Так как мы использовали линейные регрессии для анализа, мы ограничились рассмотрением преимущественно аддитивной генетической изменчивости, которая лучше всего проявляется в F1 гетерозиготных особях [25]. Наша цель состояла в том, чтобы определить, как управление ТФ генами-мишенями коррелирует с силой и количеством сайтов связывания для ТФ в предположительной регуляторной ДНК гена-мишени, и чтобы увидеть, будут ли наши результаты совпадать с литературными данными что указывает на сильную корреляцию между силой связывания ТФ или количеством сайтов связывания и его регуляторной силой действия на ген-мишень [3,5].

## МАТЕРИАЛЫ И МЕТОДЫ

**Линии *D. melanogaster* и наборы эмбрионов.** Эксперименты проводили с использованием генетической референсной панели дрозофилы (DGRP) [26] и линии Винтерс, которые были собраны в органическом саду, находящемся в Винтерсе (Калифорния), и были инбредны для достижения гомозиготности [27]. Все популяции содержались на обычной пище на основе кукурузной муки и выдерживались при приближительной температуре 25°C при 12-часовом световом цикле. Женские особи из двух референсных линий, Raleigh 315 и Raleigh 380, были скрещены с самцами из разных линий Raleigh и Winters (рис. 1). Скрещивания были выпол-

нены в 6-унциевых бутылках с квадратным дном, которые накрывали чашкой Петри, содержащей виноградный сусло-агар. На следующее утро в 8 ч утра бутылки накрывали чашками Петри, содержащими дрожжи для сбора эмбрионов, выведенных в течение ночи. В 9 ч утра эти чашки Петри удаляли и бутылки накрывали новыми чашками Петри ровно на один час. В 10 ч утра новые чашки Петри, содержащие эмбрионы, собирали и инкубировали при 25°C в течение 4 ч, получая эмбрионы в возрасте четырех–пяти часов. После инкубации эмбрионы были дехлорированы погружением их в 50% раствор хлорной извести в течение полутора минут. Эмбрионы затем промывали деионизованной водой и хранили в растворе Ambion TRIzol (# 15596-026, Life Technologies, США) при температуре –80°C. В данном исследовании было проанализировано 71 скрещивание. 25 скрещиваний были секвенированы как две биологические репликации, а остальные 46 скрещиваний были секвенированы без репликаций (табл. 1). В общей сложности были секвенированы 96 образцов.

**Построение библиотек транскриптома и РНК-секвенирование.** Для каждого образца РНК экстрагировали с помощью набора Direct-Zol RNA-prep Kit в соответствии с протоколом от Zymo Research (США). мРНК очищали с помощью набора для очистки мРНК Ambion Dynabeads mRNA Purification Kit (Product # 61006) и фрагментировали с помощью фрагментирующего набора (Product # AM8740), с последующим синтезом кДНК со случайными гексамерными праймерами. Тупые концы были получены с помощью набора для быстрого затупления Quick Blunting Kit (NEB Product # E1201L) с добавлением единичного аденина с помощью фрагмента Кленова 3'–5'-экзонуклеазы (NEB Product # M0212L). Адаптеры от Illumina (США) лигировали на кДНК-фрагменты с помощью набора Quick Ligation Kit (NEB Product # M2200L). Выбор размера фрагментов производился с помощью бусин Agencourt AMPure XP (Product # A63880, Beckman Coulter, США) с отношением объема шариков к общему объему 0,7. Наконец, 96 образцов были помечены двенадцатью индексами Illumina и восьмью изготовленными на заказ бар-кодами и обогащены, прежде чем их секвенировали на 96-луночном планшете [28] секвенатора Illumina HiSeq 2500 методом спаренных концов в режиме высокопроизводительного секвенирования (режим ридов по 100 пар оснований). Сырые риды Illumina будут депонированы на NCBI после принятия, в настоящее время они доступны по

адресу <http://rri-nuzhdin-2.cts.usc.edu/thkitapci/data/>.

**Анализ РНК, включая картирование и нормализацию ридов секвенирования.** Риды полногеномного РНК-секвенирования RNA-Seq были картированы на референсный геном *D. melanogaster* (dm3/BDGP5.75) с использованием STAR (2.4.0k) с параметрами по умолчанию [29]. Только однозначно и согласованно картированные риды были использованы для дальнейшего анализа. Подсчет сырых ридов был произведен с использованием HTSeq с параметрами по умолчанию [30] с помощью файла аннотаций (dm3/BDGP5.75.gtf). Подсчеты ридов от образцов, соответствующих одному и тому же генотипу были объединены вместе. Подсчет сырых ридов нормализовали с помощью показателя RPKM (число ридов на килобазу, отнесенное на миллион ридов) [31]. Скрипты анализа доступны по адресу [https://github.com/thkitapci/Inference\\_of\\_TF\\_regulatory\\_networks.git](https://github.com/thkitapci/Inference_of_TF_regulatory_networks.git).

**Ковариационный анализ экспрессии генов.** Мы были заинтересованы в генах сегментации (матерински-экспрессированные ТФ (*Bicoid* и *caudal*), *gap*-генах (*giant*, *kruppel*, *knirps*, *hunchback* и *tailless*), первичных генах *pair-rule* (*even-skipped*, *hairy*, *runt* и *fushi tarazu*) и генах сегментарной полярности (*engrailed*), в дополнение к генам в дорзо-вентрального структурирования (*snail* и *twist*), которые играют важную роль в формировании структуры во время эмбриогенеза дрозофилы [32,33]. Кроме того, гены сегментации показали обилие вариаций экспрессии во время эмбриогенеза дрозофилы [5].

Для того чтобы охарактеризовать регуляторную силу между исследуемыми ТФ и их генами-мишенями, мы рассчитали коэффициент корреляции Спирмена между уровнями экспрессии каждого ТФ и других картированных генов (7805 генов), которые имели высокие уровни экспрессии в 59 образцах. Положительными регуляторными силами считались те, у которых коэффициенты корреляции > 0, отрицательными считались те, у которых коэффициенты корреляции < 0.

Для того чтобы найти предполагаемую регуляторную ДНК для каждого гена-мишени, мы обнаружили ДНК-доступные регионы в пределах окна в 5000 п.о. выше каждого гена-мишени, с помощью более ранних измерений ДНК-доступных областей на стадии 10 (4 ч) и стадии 11 (5 ч и 40 мин) эмбрионов дрозофилы [34]. Чтобы найти сайты связывания в каждой доступной области, мы использовали программу PATSER [35], предполагая 47%-е содержание GC, и позиционные весовые матрицы для каждого исследуемого ТФ из базы

Таблица 1. Детали экспериментальных скрещиваний

Скрещивание	Тестер	Тестируемая линия	Наличие репликации	Скрещивание	Тестер	Тестируемая линия	Наличие репликации
1	R315	R208	Да	71	R380	R208	Нет
5	R315	R307	Нет	72	R380	R303	Нет
7	R315	R324	Нет	79	R380	R335	Нет
8	R315	R335	Нет	80	R380	R357	Нет
11	R315	R360	Нет	82	R380	R360	Нет
12	R315	R365	Нет	83	R380	R365	Да
13	R315	R375	Да	84	R380	R375	Нет
16	R315	R399	Да	86	R380	R399	Да
17	R315	R427	Да	87	R380	R427	Да
20	R315	R517	Да	88	R380	R437	Нет
21	R315	R555	Нет	90	R380	R517	Нет
22	R315	R639	Нет	91	R380	R555	Да
23	R315	R707	Нет	92	R380	R639	Нет
26	R315	R730	Нет	93	R380	R707	Нет
27	R315	R732	Да	96	R380	R730	Нет
28	R315	R765	Да	97	R380	R732	Нет
29	R315	R774	Нет	98	R380	R765	Нет
32	R315	R820	Нет	99	R380	R774	Нет
37	R315	W3-33	Да	102	R380	R820	Да
41	R315	W3-37	Да	103	R380	R852	Да
42	R315	W3-38	Да	105	R380	W3-23	Нет
45	R315	W3-47	Нет	112	R380	W3-38	Нет
49	R315	W3-54	Нет	113	R380	W3-40	Нет
50	R315	W3-55	Нет	115	R380	W3-47	Нет
52	R315	W3-59	Нет	119	R380	W3-54	Да
54	R315	W3-62	Да	122	R380	W3-59	Нет
55	R315	W3-63	Нет	123	R380	W3-60	Нет
57	R315	W3-66	Нет	124	R380	W3-62	Нет
58	R315	W3-67	Да	125	R380	W3-63	Да
59	R315	W3-68	Да	129	R380	W3-68	Нет
60	R315	W3-69	Нет	130	R380	W3-69	Нет
64	R315	W3-79	Да	132	R380	W3-74	Да
65	R315	W3-80	Нет	136	R380	W3-82	Да
67	R315	W3-84	Нет	139	R380	W3-87	Нет
69	R315	W3-87	Нет	140	R380	W3-114	Нет
70	R315	W3-114	Да				

данных Fly Factor Survey [36] (табл. 2). Мы использовали опцию PATSER -li для вычисления порога обрезания оценки связывания на основе информации, содержащейся в каждой позиционной весовой матрице. Кроме того, мы усовершенствовали сайты, выбрав самые сильные 5, 10 или 15% сайтов связывания на основе оценки связывания PATSER, которая коррели-

рует с силой связывания. Мы использовали порог 10% в основном тексте, так как порог 5% оказался слишком жестким и значительно сократил количество наших образцов, а порог 15% оказался слишком низким и ликвидировал корреляцию между регуляторной силой и содержанием сайта связывания (рис. 2).

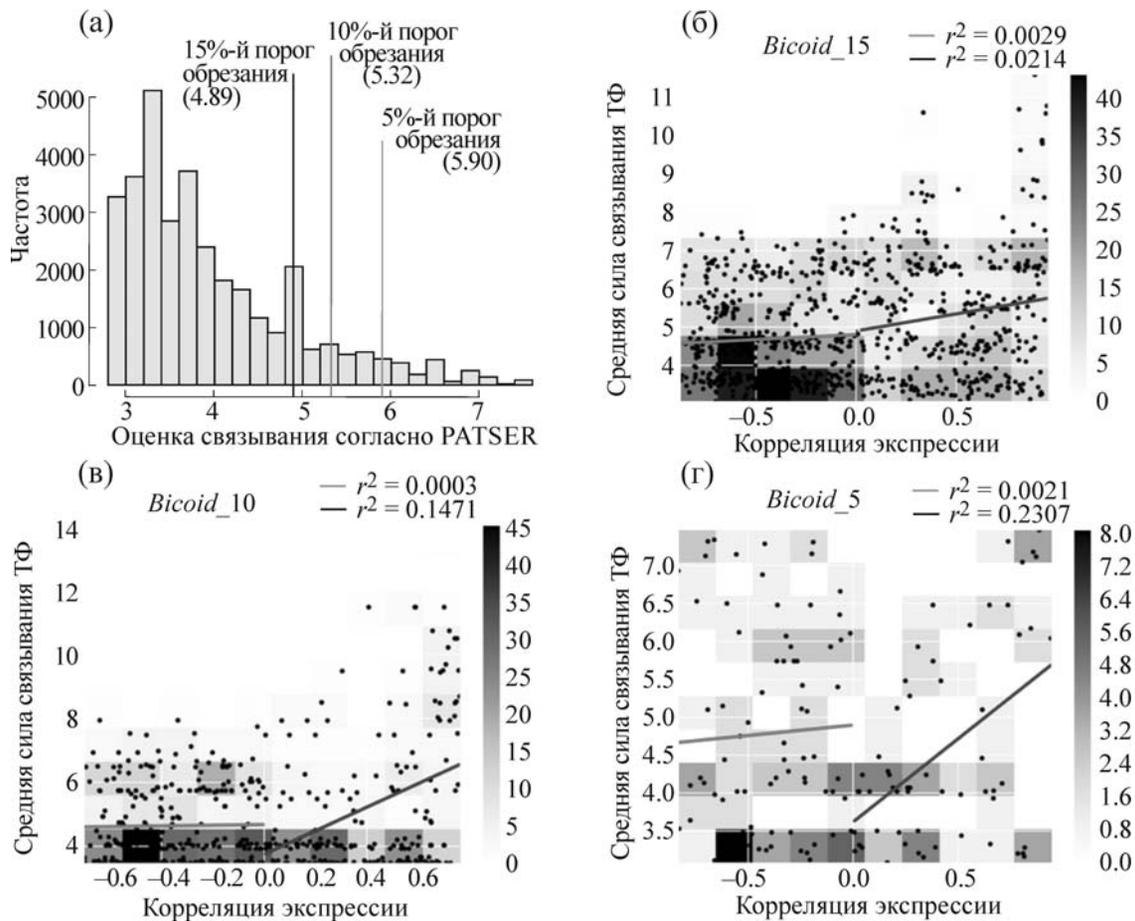
**Таблица 2.** Позиционные весовые матрицы из базы данных Fly Factor Survey для ТФ, использованных в данном исследовании

Название ТФ	Позиционные весовые матрицы												
	>bcd_FlyReg_FBgn0000166												
<i>Bicoid (bcd)</i>	A	9	8	45	47	1	1	2	5				
	C	18	3	1	0	0	44	26	12				
	G	3	1	1	1	16	0	3	18				
	T	18	36	1	0	31	3	17	13				
	>cad_FlyReg_FBgn0000251												
<i>Caudal (cad)</i>	A	0	2	0	1	6	0	3	7	0			
	C	5	0	0	0	2	4	1	3	4			
	G	3	3	0	0	1	3	9	2	0			
	T	5	8	13	12	4	6	0	1	9			
	>hb_FlyReg_FBgn0001180												
<i>Hunchback (hb)</i>	A	17	0	1	0	0	1	59	20				
	C	13	0	5	1	0	4	7	17				
	G	7	3	1	0	0	0	29	14				
	T	66	100	96	102	103	98	8	52				
	>gt_FlyReg_FBgn0001150												
<i>Giant (gt)</i>	A	3	0	5	0	0	6	0	0	0	5	5	2
	C	0	0	0	0	0	2	5	0	3	0	1	0
	G	0	1	2	0	5	0	2	4	0	0	0	1
	T	5	7	1	8	3	0	1	4	5	3	2	5
	>Kr_FlyReg_FBgn0001325												
<i>Kruppel (Kr)</i>	A	37	33	20	2	4	6	1	8	28	16		
	C	0	6	11	0	1	3	3	4	6	8		
	G	4	1	9	40	37	35	4	4	5	5		
	T	3	4	4	2	2	0	36	28	5	15		
	>kni_FlyReg_FBgn0001320												
<i>Knirps (kni)</i>	A	1	11	3	0	0	7	8	6	0	1		
	C	1	1	3	19	5	11	4	2	1	5		
	G	18	2	5	0	6	1	4	2	1	1		
	T	0	6	9	1	9	1	4	10	18	13		
	>tll_FlyReg_FBgn0003720												
<i>Tailless (tll)</i>	A	25	26	27	8	0	1	28					
	C	2	3	0	1	1	20	2					
	G	6	7	8	27	0	1	7					
	T	4	1	2	1	36	15	0					
	>eve_FlyReg_FBgn0000606												
<i>Even skipped (eve)</i>	A	6	0	2	3	16	19	0	3	7	3	1	0
	C	8	3	5	0	0	0	5	4	0	0	0	6
	G	0	10	3	0	2	0	0	5	6	2	7	1
	T	5	6	9	16	1	0	14	7	6	14	11	12

Название ТФ	Позиционные весовые матрицы												
<i>Hairy (h)</i>	>h_NAR_FBgn0001168												
	A	0	3	0	12	0	1	0	0	2	0		
	C	1	4	18	0	18	0	5	0	13	18		
	G	17	6	0	6	0	17	0	18	3	0		
	T	0	5	0	0	0	0	13	0	0	0		
<i>Runt (run)</i>	>run_Bgb_NBT_FBgn0013753												
	A	13	28	29	0	0	7	0	30	22			
	C	3	0	0	31	30	1	31	0	0			
	G	0	2	2	0	1	23	0	0	9			
	T	15	1	0	0	0	0	0	1	0			
<i>Fushi tarazu (ftz)</i>	>ftz_FlyReg_FBgn0001077												
	A	19	16	7	39	56	6	11	54				
	C	4	14	40	25	9	0	7	3				
	G	31	22	6	0	0	0	2	3				
	T	12	14	13	2	1	60	46	6				
<i>Engrailed (en)</i>	>en_FlyReg_FBgn0000577												
	A	20	5	7	38	52	0	0	27				
	C	3	15	7	0	0	2	6	0				
	G	18	12	0	4	0	2	14	26				
	T	12	21	39	11	1	49	33	0				
<i>Snail (sna)</i>	>sna_FlyReg_FBgn0003448												
	A	0	4	10	1	1	0	0	1	0	3		
	C	7	5	1	10	5	0	0	6	2	2		
	G	3	1	0	0	0	0	10	0	2	3		
	T	1	1	0	0	5	11	1	4	7	3		
<i>Twist (twi)</i>	>twi_FlyReg_FBgn0003900												
	A	1	3	1	0	11	0	7	0	0	1	1	2
	C	4	7	3	14	3	4	1	0	0	0	2	3
	G	3	3	10	1	1	1	5	0	12	2	2	8
	T	7	2	1	0	0	10	2	15	3	12	10	2

Используя эти сайты, мы рассчитали среднее значение силы или количество сайтов связывания в восходящей ДНК каждого целевого гена, отбрасывая любые гены-мишени с менее чем тремя сайтами связывания. Если ограничить анализ генами-мишенями с более чем двумя сайтами связывания, это увеличивало корреляцию между содержанием сайта связывания и регуляторной силой. Рассчитанные оценки PATSER, которые примерно соответствуют логарифмическому подобию  $k$ -меров, являющихся сайтами связывания ТФ, были использованы в качестве представителя силы сайта связывания в этом анализе.

Для того чтобы проверить, существует ли значимая корреляция между содержанием сайта связывания ТФ и его регуляторной силой, мы рассчитали коэффициент корреляции Спирмена между положительными и отрицательными регулятивными силами и усредненными силами на выборке сайтов связывания, используя программу R.  $P$ -значения были исправлены с помощью метода Бенджамини–Хохберга [37]. Все цифры были получены с использованием статистической программы R, и все сценарии были написаны на языке PERL. Сырые и нормированные файлы и скрипты подсчета, использованные для анализа, могут быть получены по



**Рис. 2.** Соотношение между регуляторной силой и средней силой сайта связывания *bicoid* для генов-мишеней с использованием верхних 15%-го, 10%-го и 5%-го порогов обрезания для *bicoid*. (а) – Пороговые значения, соответствующие верхним 15%-му, 10%-му и 5%-му порогам обрезания для *bicoid*. (б)–(г) – Соотношения между регуляторной силой и средней силой связывания *bicoid*. Каждая черная точка представляет собой мишень для *bicoid* по меньшей мере с тремя мотивами связывания, соответствующую верхним 15%-му (б), 10%-му (в) и 5%-му (г) порогам обрезания. По оси абсцисс – корреляция между уровнем экспрессии *bicoid* и его генов-мишеней во всех наших образцах. По оси ординат – среднее значение силы связывания для сайтов связывания *bicoid*, расположенных в предполагаемой области регуляторной ДНК для каждого гена-мишени. Темные и серые линии показывают линейную регрессию соответственно для положительно и отрицательно коррелирующих генов-мишеней, значения  $r^2$  отображены для этих наилучших эмпирических линий.

адресу [https://github.com/thkitapci/Inference\\_of\\_TF\\_regulatory\\_networks.git](https://github.com/thkitapci/Inference_of_TF_regulatory_networks.git).

**Анализ регуляторного взаимодействия между ТФ.** Для этого анализа мы использовали потенциальные гены-мишени для каждого исследуемого ТФ, чтобы найти подмножество потенциальных целей, которые были бы ТФ сами. Были произведены расчеты коэффициентов корреляции Спирмена между этими ТФ и проверки на значимость ( $p < 0,001$ , после многократной тестовой коррекции с методом Бенджамини–Хохберга) с использованием программы R.

**Анализ репликации.** Для того чтобы подтвердить, что наши результаты о регуляторных взаимодействиях между ТФ являются воспроизводимыми и имеют биологическое значение,

мы разделили наш набор данных на два набора на основе имеющегося тестерного генотипа, а именно скрещивания с R315 и скрещивания с R380. Коэффициенты корреляции Спирмена этих ТФ были вычислены независимо друг от друга в этих выборках и протестированы на значимость ( $p < 0,05$  после многократной тестовой коррекции с методом Бенджамини–Хохберга) с использованием программы R.

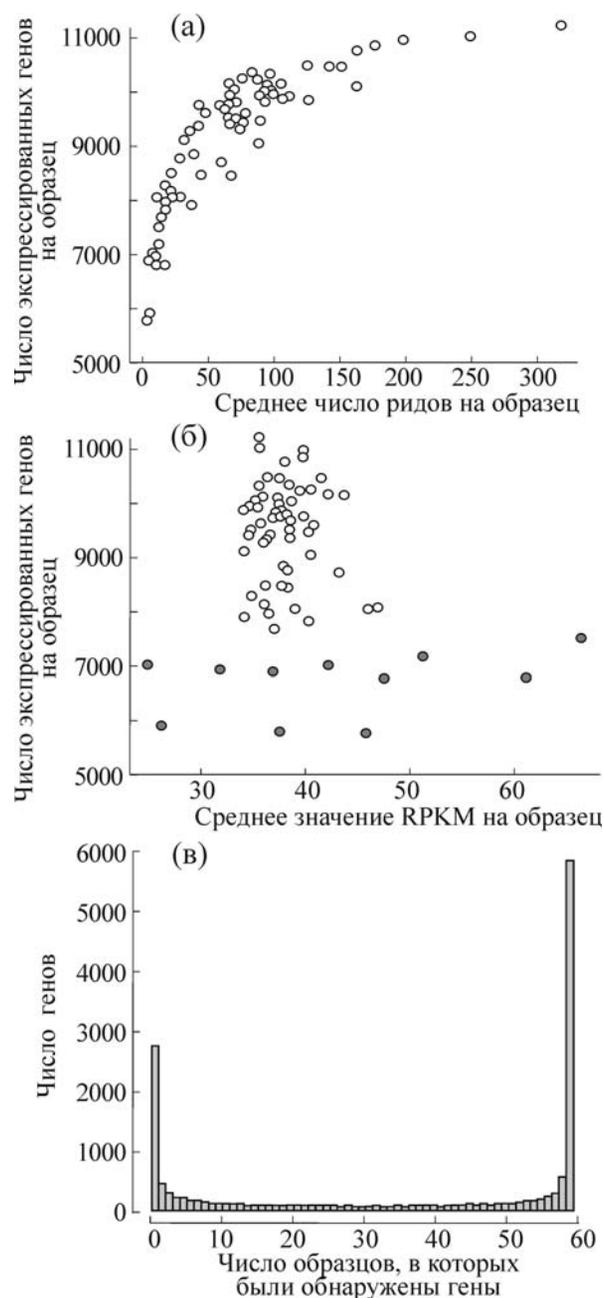
## РЕЗУЛЬТАТЫ

**Измерение транскриптома линий *D. melanogaster*.** Для измерения ковариации между ТФ и предполагаемыми генами-мишенями мы охарактеризовали транскриптом эмбрионов F1 отдельных линий дрозофилы (рис. 1) возрастом

четыре–пять часов после оплодотворения. Транскриптомы из 96 образцов *D. melanogaster* секвенировали и картировали риды на 15682 генов, которые были аннотированы в эталонном геноме *D. melanogaster*. Для того чтобы охарактеризовать качество секвенирования данных, число картированных генов и среднее число ридов в каждом образце были проанализированы. Определим, что ген картирован, если он имеет ненулевое значение рида. В каждом образце был разброс по среднему количеству ридов (рис. 3а). Более высокое среднее количество ридов позволяло определить большее число генов, как и ожидалось, наблюдалась положительная зависимость между средним количеством ридов и количеством картированных генов в образце. Однако эта зависимость не является линейной, так как увеличение глубины секвенирования будет определять большее количество генов только до некоторого предела. Мы видим, что при увеличении среднего значения количества ридов число секвенированных генов приближается к 11000, предполагаемому верхнему пределу количества генов, детектируемых в каждом образце (рис. 3а).

Для того чтобы учесть различия в глубине секвенирования из-за плохой подготовки проб, более низких концентраций РНК или смещения последовательности, которое является результатом фрагментации мРНК [38], мы нормализовали данные по длинам каждого отображенного гена и миллионам ридов для расчета показателя RPKM, количественной меры экспрессии генов [31]. Среднее значение RPKM для всех картированных генов было одинаковым во всех образцах, использованных в нисходящем анализе, со средним значением 37,9 RPKM и стандартным отклонением 2,8 RPKM. Двенадцать образцов из менее чем 7500 картированных генов показали широкий диапазон значений среднего и стандартного отклонения RPKM и были удалены из дальнейшего анализа (рис. 3б). Среди оставшихся 59 образцов 7805 генов были обнаружены, по крайней мере, в 50 образцах (рис. 3в). Мы рассмотрели эти 7805 генов в 59 образцах при нисходящем анализе. Для измерения «регуляторной силы» мы рассчитали коэффициент корреляции Спирмена между уровнями экспрессии 14 ТФ, активных в передне-задней и дорзо-вентральной структурирующих сетях с их потенциальными генами-мишенями.

**Ковариация экспрессии генов связана с содержанием сайта связывания ТФ.** Чтобы проверить гипотезу о том, что ковариация экспрессии генов (регуляторная сила) между ТФ и потенциальными генами-мишенями связана с



**Рис. 3.** Анализ данных транскриптома линии дрозофилы. (а) – Среднее количество ридов в зависимости от числа составляющих генов, выраженных до RPKM-нормализации. При увеличении числа ридов увеличивается также и число экспрессированных (картированных) генов, пока не достигнет плато ~11000 экспрессированных генов. (б) – Среднее значение RPKM в зависимости от числа экспрессированных генов. Для образцов с более чем 7500 экспрессированных (картированных) генов, среднее значение RPKM стабильно. Для образцов с менее 7500 экспрессированных генов (серые кружки) средние значения RPKM сильно различаются. Эти образцы удалялись из нисходящего анализа. (в) – Число генов, картированных по образцам. Только гены, которые экспрессируются по крайней мере в 50 образцах, были использованы в нисходящем анализе.

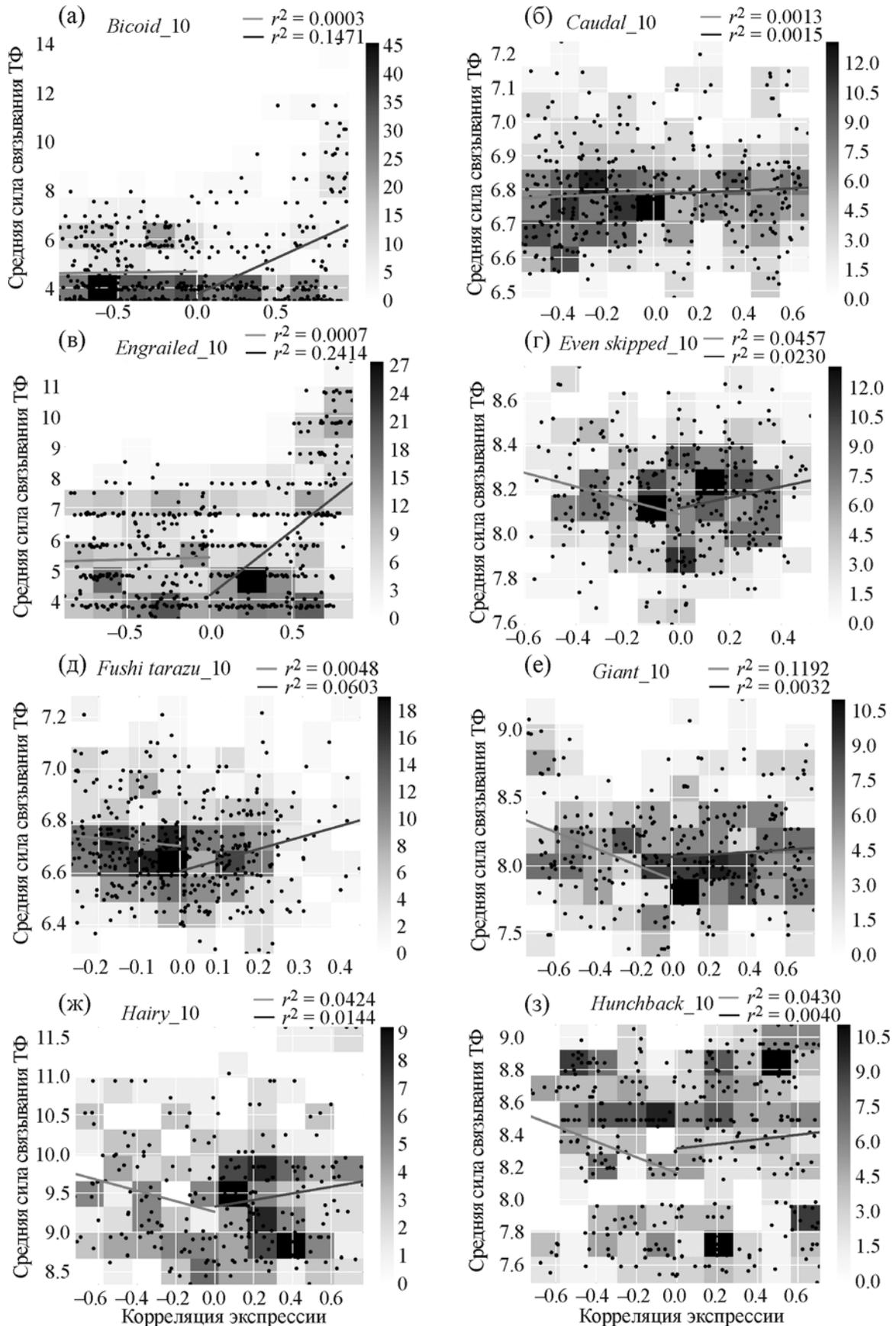
содержимым сайта связывания ТФ в регуляторной ДНК гена-мишени, мы создали два набора данных, которые описывают либо среднюю силу, либо количество сайтов связывания ТФ в предположительной регуляторной области ДНК каждого целевого гена. Для начала мы предположили, что регуляторная ДНК гена была расположена в 5 килобазах вверх по течению области его сайта инициации транскрипции и далее сузили эту область, чтобы включить только регионы, определенные как ДНКазы, гиперчувствительные на стадии 10 и стадии 11 развития, что примерно соответствует времени нашего окна сбора эмбрионов. Поскольку мы не имеем исчерпывающие аннотации энхансеров в геноме, мы предполагаем, что эти ближние, доступные области генома будут содержать энхансеры потенциальных генов-мишеней и рассмотрим последствия этого предположения в разделе «Обсуждение результатов». Сайты связывания для каждого исследуемого ТФ были обнаружены в этих областях с использованием позиционных весовых матриц каждого ТФ и программы PATSER (см. раздел «Материалы и методы»). Для того чтобы определить, какие сайты связывания рассматривать, для каждого ТФ мы использовали программу PATSER для вычисления порога обрезания оценки связывания на основе информации, содержащейся в каждой позиционной весовой матрице, а затем рассматривали сайты, попавшие в первые 5 и 10% от распределения оценок сайтов связывания, в качестве потенциального сайта связывания для каждого ТФ. Затем мы рассчитали среднюю силу или количество сайтов связывания, расположенных в предположительной регуляторной ДНК каждого потенциального целевого гена для каждого ТФ. На рис. 4 и 5 мы показываем результаты для 10%-го порога обрезания, а результаты для 5%-го порога обрезания – на рис. 6 и 7. Использование 15%-го порога ликвидировало корреляции между регуляторной силой и содержимым сайта связывания и, следовательно, считалось слишком слабым в качестве порогового значения (рис. 2).

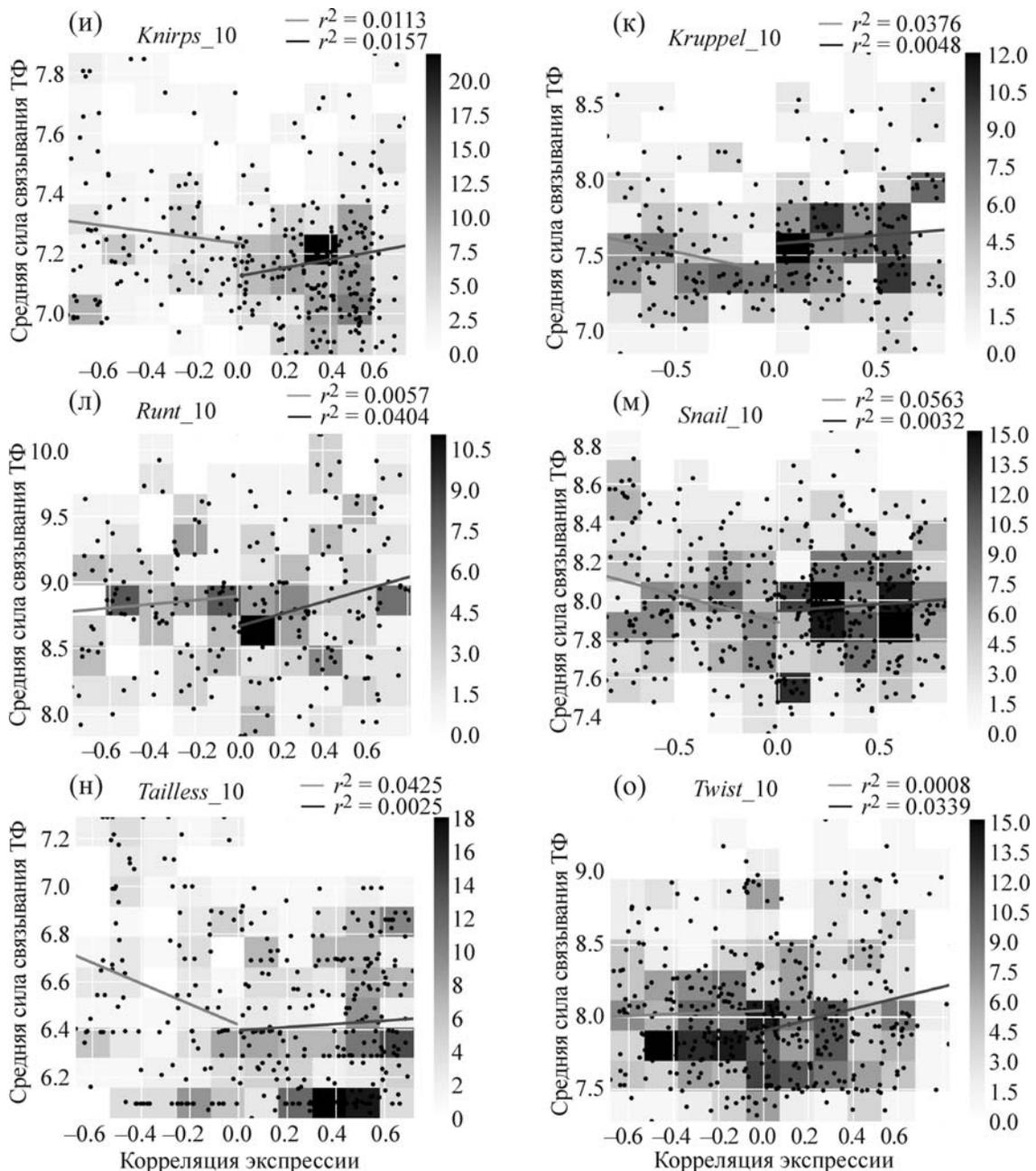
Чтобы сравнить регуляторные силы каждого ТФ с содержимым сайта связывания ТФ каждой потенциальной цели, мы разделили данные по регуляторной силе на положительные и отрицательные корреляции. Мы обнаружили сильную связь между положительными регуляторными силами и средними силами сайтов связывания из *bicoid* и *engrailed*, в то время как *giant* и *hunchback* имеют сильную взаимосвязь между отрицательными регуляторными силами и средними силами сайтов связывания на этой стадии развития дрозофилы (рис. 8). Мы также

обнаружили существенную корреляцию между регуляторной силой и средней силой сайтов связывания ТФ для *even skipped*, *fushi tarazu*, *hairy*, *kruppel*, *runt*, *snail*, *tailless* и *twist* (рис. 4 и 6, табл. 3 и 4). *Bicoid*, *engrailed*, *fushi tarazu*, *runt* и *twist* показывают значимые корреляции между положительной регуляторной силой и средней силой сайтов связывания ТФ, предположительно эти ТФ действуют как активаторы на данном этапе развития. *Even-skipped*, *giant*, *hairy*, *hunchback*, *kruppel*, *snail* и *tailless* имели существенную корреляцию между отрицательной регуляторной силой и средней силой сайтов связывания ТФ, предположительно действуя как репрессоры на этой стадии развития (рис. 4 и 6, табл. 3 и 4). Мы также обнаружили значительную корреляцию между числом сайтов связывания и активирующей силой *bicoid*, *engrailed*, *fushi tarazu*, *kruppel*, *runt* и *snail*. *Engrailed*, *even skipped*, *giant*, *hunchback*, *kruppel* и *tailless* имели значительную корреляцию между числом сайтов связывания и их подавляющей силой (рис. 5 и 7, табл. 3 и 4).

Эти наблюдения некоторым образом согласуются с известной биологией данных ТФ. *Bicoid* и *twist* [39–46], как известно, действуют в качестве активаторов транскрипции. *Giant*, *tailless* [43,47–50], *even-skipped* и *hairy* [51–56] известны как транскрипционные репрессоры. *Runt* и *fushi tarazu* могут выступать в качестве активаторов или репрессоров, но мы только обнаружили их в качестве активаторов в нашем исследовании [57–61]. *Hunchback* может выступать в качестве транскрипционного активатора или репрессора, хотя мы обнаружили только его репрессорную роль в нашем исследовании [43,62,63]. Существует ряд доказательств, что *kruppel* [62,64–66], *engrailed* [67–69] и *snail* [70] могут выступать в качестве бифункциональных ТФ.

**Выявление регуляторного взаимодействия между ТФ.** Поскольку передне-задняя и дорзо-вентральная структурирующие сети подразумевают взаимодействие между ТФ, мы проанализировали наши данные, чтобы увидеть, могли ли мы раскрыть эти взаимодействия. Мы обнаружили, что материнский ген *bicoid* имел положительную корреляцию с большинством ТФ, что согласуется с ролью *bicoid* в качестве активатора. В то же время *tailless* показал репрессорное взаимодействие с *giant* и *knirps*, что опять же согласуется с известной биологией [71–75]. *Twist*, предположительно, репрессорует *sloppy paired* и *odd paired*, как было предложено в предыдущем исследовании [32] (табл. 5, рис. 9).

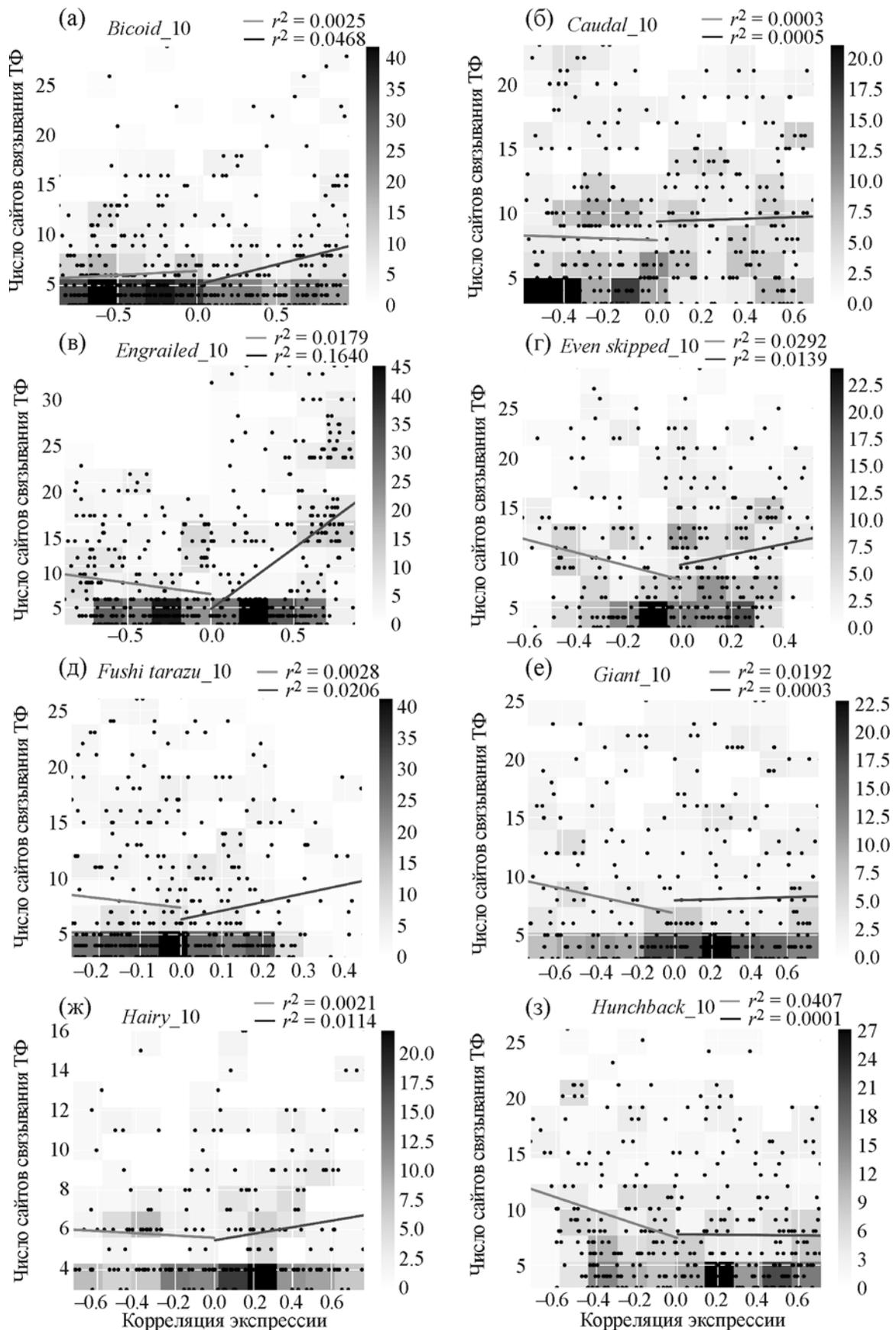


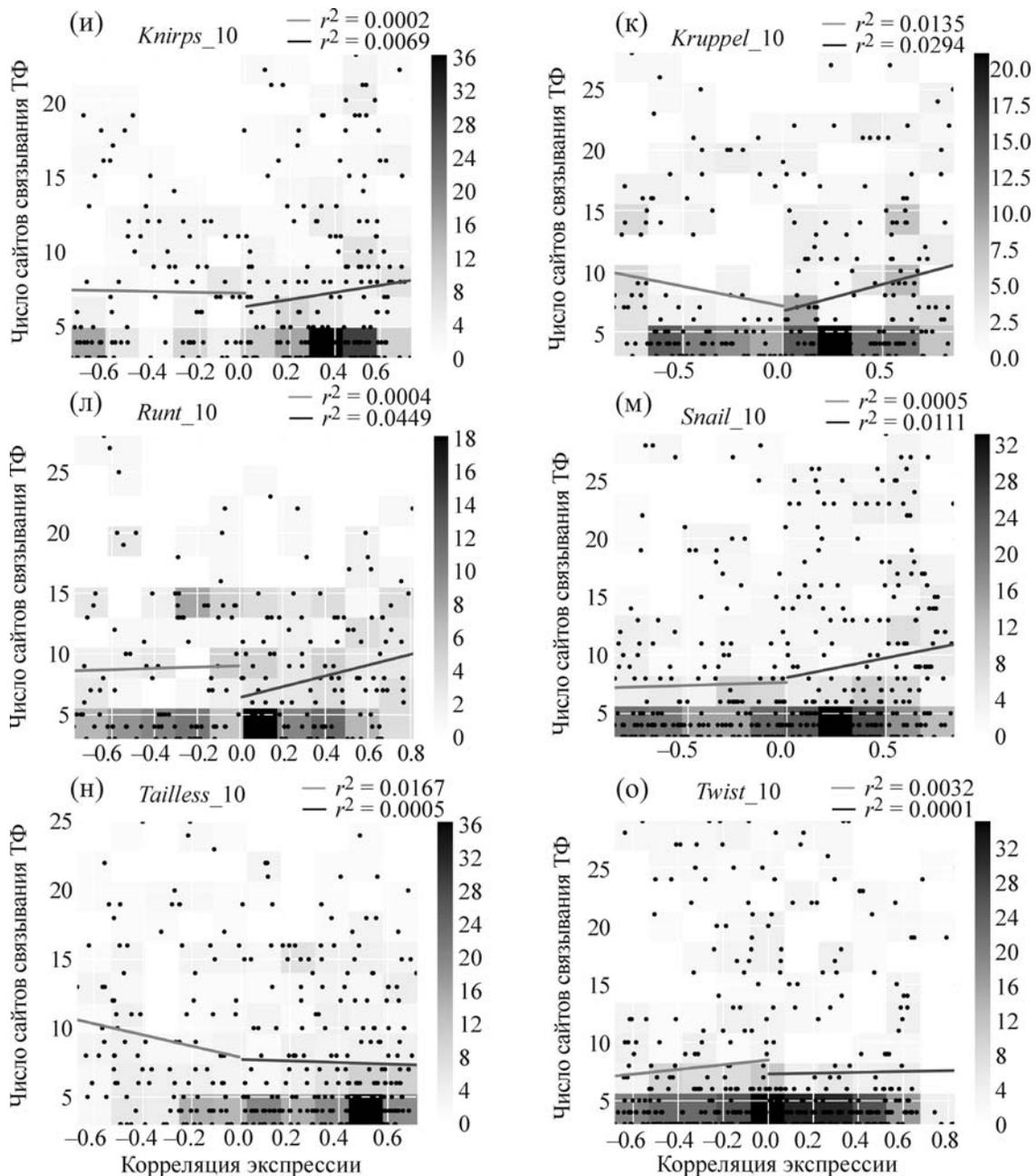


**Рис. 4.** Соотношение между регуляторной силой и средней силой сайта связывания ТФ для генов-мишеней с использованием верхнего 10%-го порога обрезания для каждого ТФ. Показана взаимосвязь между регуляторной силой и средней силой связывания для всех интересующих нас ТФ. Каждая черная точка представляет собой мишень для ТФ по меньшей мере с тремя мотивами связывания. По оси абсцисс – корреляция между уровнем экспрессии каждого ТФ и его генов-мишеней во всех наших образцах. По оси ординат – среднее значение силы связывания ТФ для сайтов связывания, расположенных в предполагаемой области регуляторной ДНК для каждого гена-мишени. Темные и серые линии показывают линейную регрессию соответственно для положительно и отрицательно коррелирующих генов-мишеней, значения  $r^2$  отображены для этих наилучших эмпирических линий.

Чтобы убедиться, что наши результаты не зависят от генетического фона, мы использовали два тестера женских штаммов (R315 и R380), аналогичные использованным в работе [5]. Когда мы проанализировали данные по

каждому из этих тестерных штаммов по отдельности, оценки регуляторной силы оказались примерно идентичны между двумя тестерами (рис. 10, табл. 6). Такая высокая степень репликации устанавливает надежность техники.



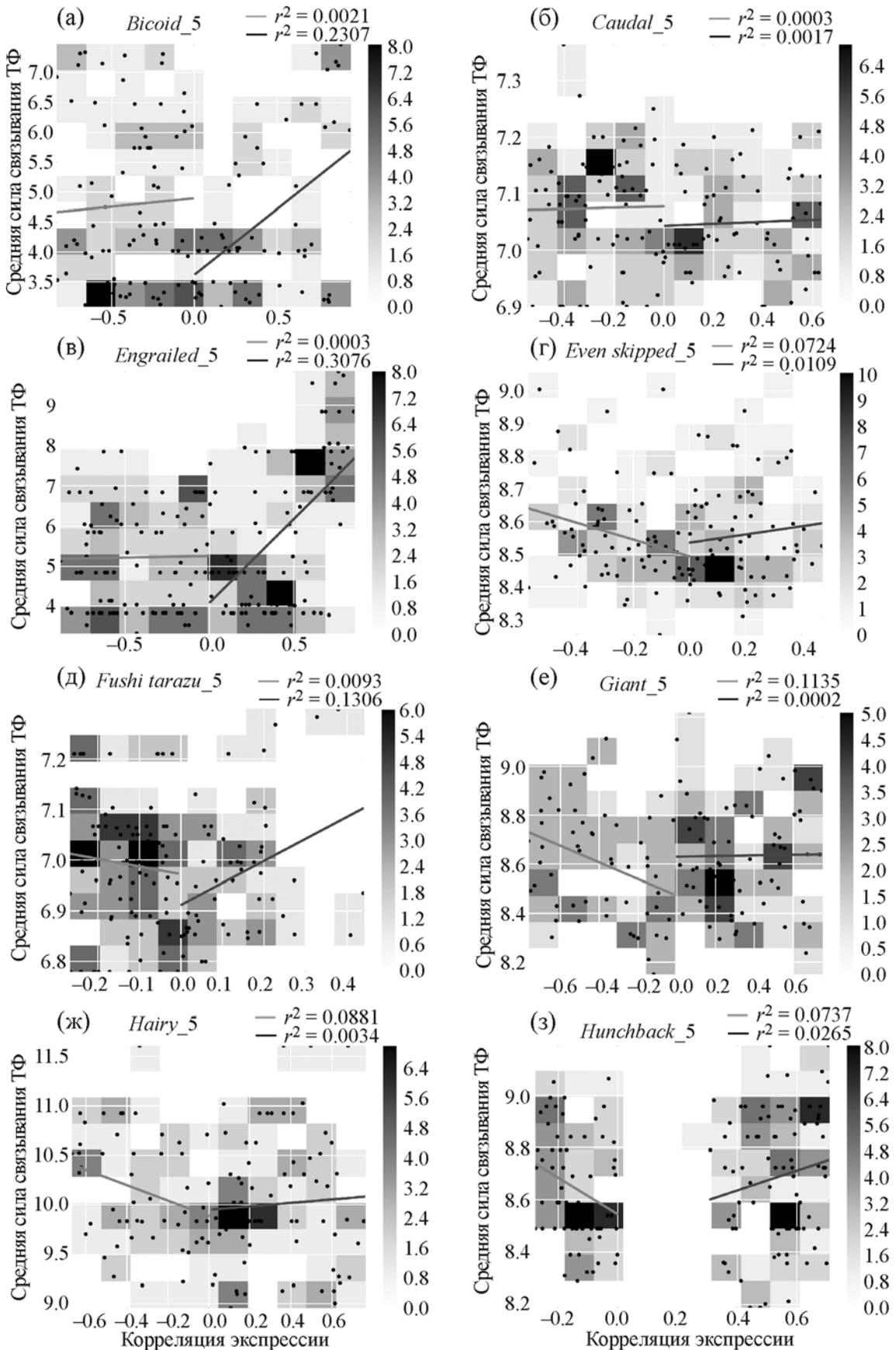


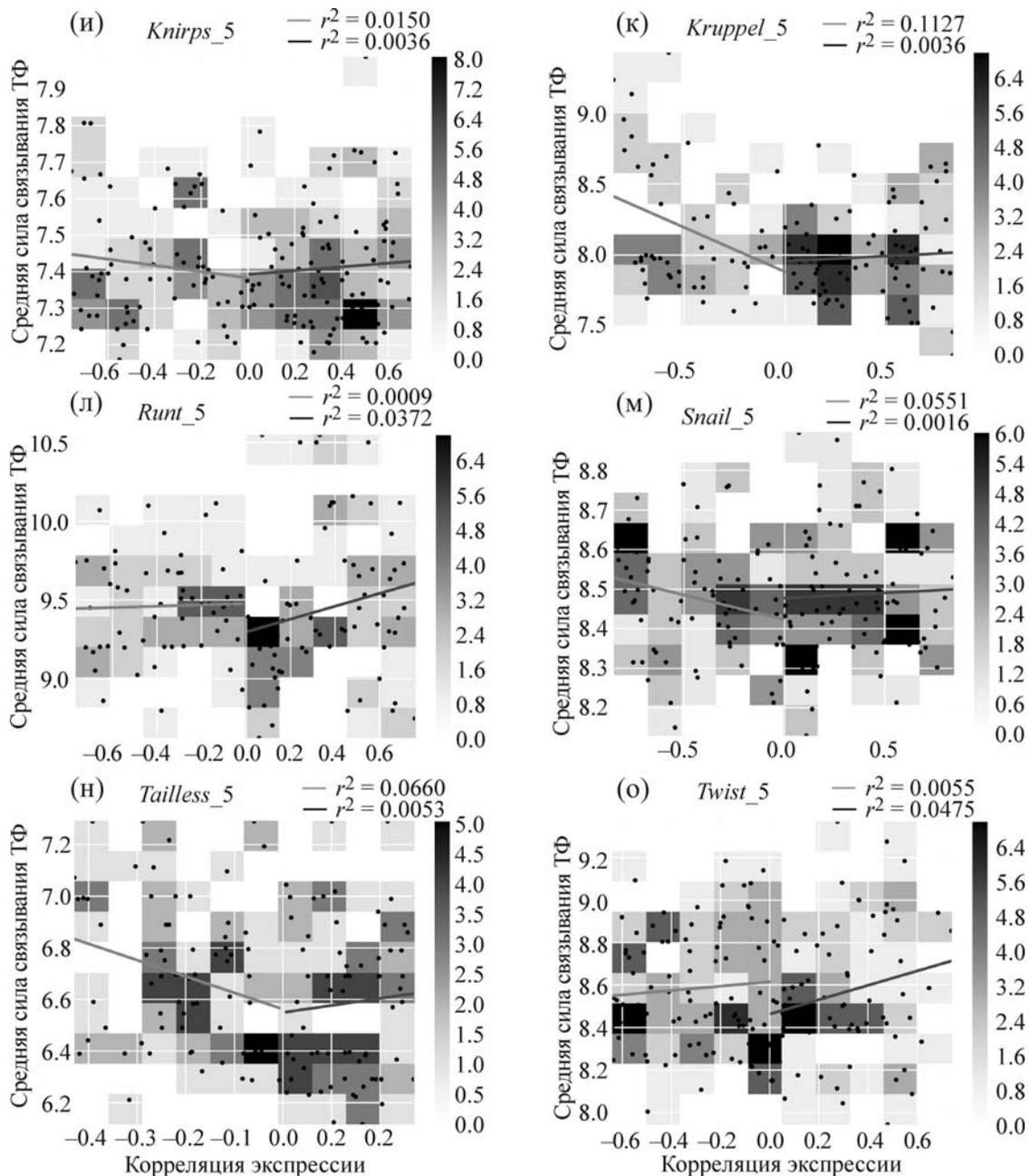
**Рис. 5.** Соотношение между регуляторной силой и числом сайтов связывания ТФ для генов-мишеней с использованием верхнего 10%-го порога обрезания для каждого ТФ. Показана взаимосвязь между регуляторной силой и числом сайтов связывания для всех интересующих нас ТФ. Каждая черная точка представляет собой мишень для ТФ по меньшей мере с тремя мотивами связывания. По оси абсцисс – корреляция между уровнем экспрессии каждого ТФ и его генов-мишеней во всех наших образцах. По оси ординат – число сайтов связывания ТФ, расположенных в предполагаемой области регуляторной ДНК для каждого гена-мишени. Темные и серые линии показывают линейную регрессию соответственно для положительно и отрицательно коррелирующих генов-мишеней, значения  $r^2$  отображены для этих наилучших эмпирических линий.

## ОБСУЖДЕНИЕ

В данном исследовании мы использовали транскрипционное профилирование линий *D. melanogaster* для анализа взаимосвязи между сайтами связывания ТФ и регуляторным кон-

тролем их генов-мишеней с использованием двух метрик связывания содержимого сайта: средняя численность и количество сайтов связывания. Мы обнаружили значимые корреляции для нескольких представляющих интерес

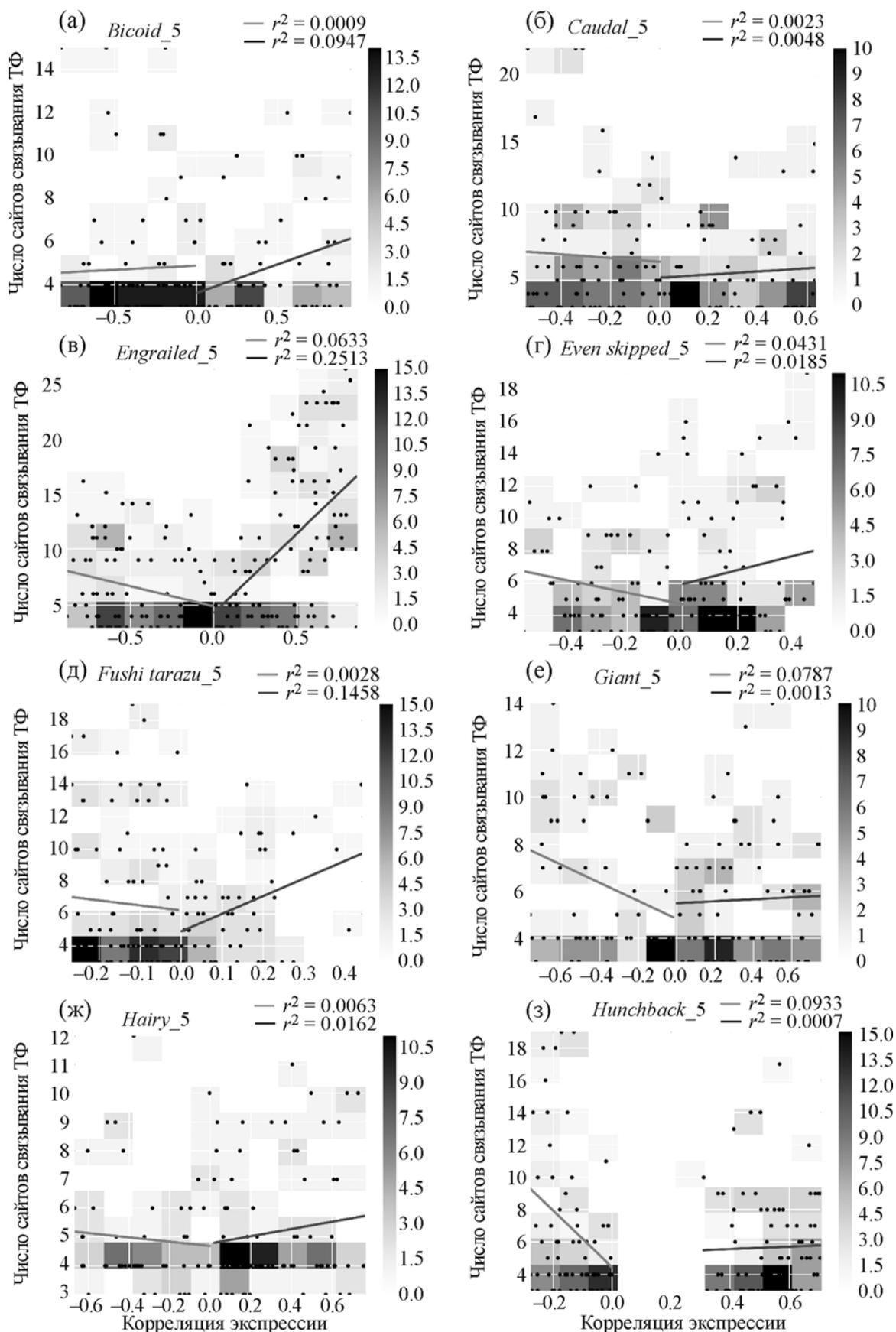


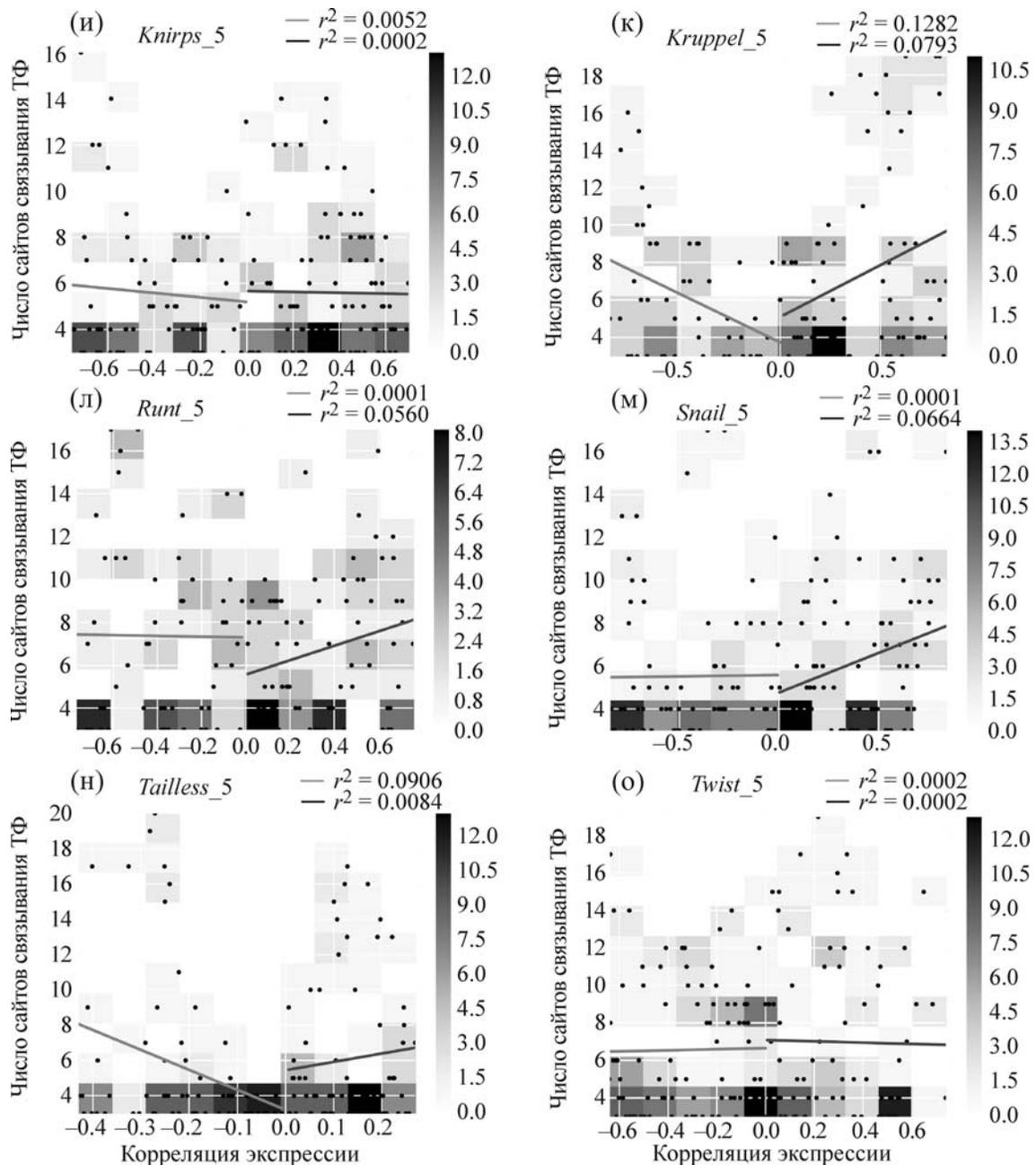


**Рис. 6.** Соотношение между регуляторной силой и средней силой сайта связывания ТФ для генов-мишеней с использованием верхнего 5%-го порога обрезания для каждого ТФ. Показана взаимосвязь между регуляторной силой и средней силой связывания для всех интересующих нас ТФ. Каждая черная точка представляет собой мишень для ТФ по меньшей мере с тремя мотивами связывания. По оси абсцисс – корреляция между уровнем экспрессии каждого ТФ и его генов-мишеней во всех наших образцах. По оси ординат – среднее значение силы связывания ТФ для сайтов связывания, расположенных в предполагаемой области регуляторной ДНК для каждого гена-мишени. Темные и серые линии показывают линейную регрессию соответственно для положительно и отрицательно коррелирующих генов-мишеней, значения  $r^2$  отображены для этих наилучших эмпирических линий.

ТФ. Это говорит о том, что сила и количество сайтов связывания для конкретного ТФ в регуляторных областях ДНК коррелируют с регуляторным контролем их генов-мишеней. Наши результаты согласуются с предыдущими ис-

следованиями, которые показали, что количество и сила сайтов связывания ТФ коррелируют, хотя и несовершенно, с регуляторной силой конкретного ТФ, действующей на его гены-мишени [3,76,77]. Предыдущее исследование пяти

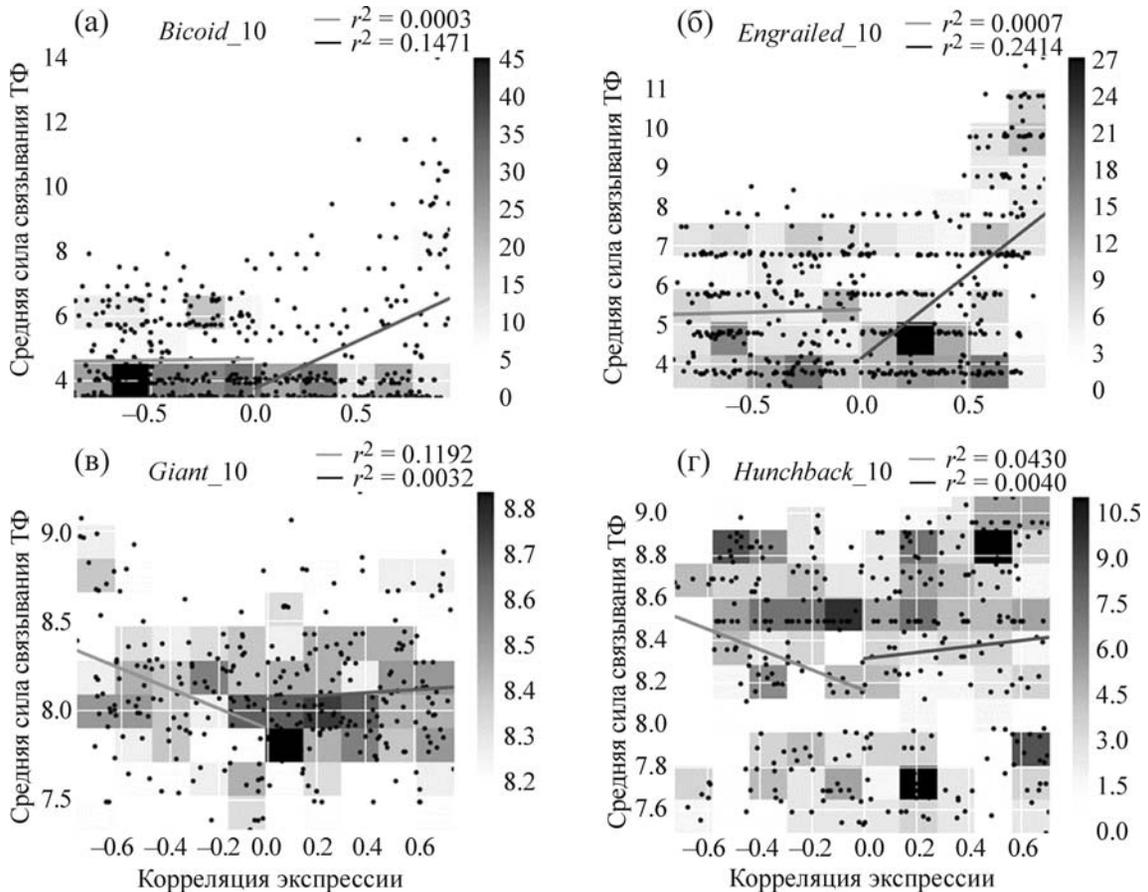




**Рис. 7.** Соотношение между регуляторной силой и числом сайтов связывания ТФ для генов-мишеней с использованием верхнего 5%-го порога обрезания для каждого ТФ. Показана взаимосвязь между регуляторной силой и числом сайтов связывания для всех интересующих нас ТФ. Каждая черная точка представляет собой мишень для ТФ по меньшей мере с тремя мотивами связывания. По оси абсцисс – корреляция между уровнем экспрессии каждого ТФ и его генов-мишеней во всех наших образцах. По оси ординат – число сайтов связывания ТФ, расположенных в предполагаемой области регуляторной ДНК для каждого гена-мишени. Темные и серые линии показывают линейную регрессию соответственно для положительно и отрицательно коррелирующих генов-мишеней, значения  $r^2$  отображены для этих наилучших эмпирических линий.

ТФ (*bicoid*, *caudal*, *giant*, *hunchback*, и *kruppel*) в течение первых пяти–восьми часов развития дрозофилы показали, что существует связь между силой связывания ТФ и регуляторным контролем их генов-мишеней [5].

*Engrailed* демонстрировал самую высокую корреляцию между регуляторным контролем его генов-мишеней и силой и количеством сайтов связывания (табл. 3 и 4). Мы предполагаем, что эта сильная корреляция обусловлена тем, что *engrailed* в общем случае имеет высокий уровень



**Рис. 8.** Соотношение между регуляторной силой и средней силой сайта связывания ТФ. Показана взаимосвязь между регуляторной силой и средней силой связывания ТФ для *bicoid*, *engrailed*, *giant* и *hunchback* с их генами-мишенями с использованием верхнего 10%-го порога обрезания для каждого ТФ. По оси абсцисс – корреляция между уровнем экспрессии каждого ТФ и его генов-мишеней во всех наших образцах. По оси ординат – среднее значение силы связывания ТФ для сайтов связывания, расположенных в предполагаемой области регуляторной ДНК для каждого гена-мишени. Каждая черная точка представляет собой мишень для ТФ по меньшей мере с тремя мотивами связывания. Темные и серые линии показывают линейную регрессию соответственно для положительно и отрицательно коррелирующих генов-мишеней, значения  $r^2$  отображены для этих наилучших эмпирических линий. (а, б) – Активирующие силы (ковариация положительной экспрессии) для *bicoid* и *engrailed* коррелируют со средней силой сайта связывания, показывая, что *bicoid* и *engrailed* выступают в качестве активаторов на данном этапе развития. (в, г) – Подавляющие силы (ковариация отрицательной экспрессии) *giant* и *hunchback* коррелируют со средней силой сайта связывания.

экспрессии в этот момент времени в процессе развития, особенно по сравнению с предыдущими стадиями развития, в то время как другие ТФ, рассмотренные нами, не показывают столь отчетливое различие в экспрессии [78,79].

В целом наши результаты показывают, что количество и сила значимых ассоциаций ковариации в экспрессии с силой связывания и количеством сайтов связывания невелики. Есть много причин для этого. Каждый из этих ТФ экспрессируется сложным пространственным и временным образом, так что данные по экспрессии целого эмбриона перекрывают регуляторные события в небольшом количестве клеток. Прогнозируемые сайты связывания ТФ мо-

гут не соответствовать связыванию в естественных условиях [80,81]. Наше предположение о том, что ДНК-доступные регионы в 5 кб выше гена-мишени будут регулировать его экспрессию, заставит нас пропустить энхансеры, расположенные в других частях генома, и может включать в себя регионы, которые не действуют в качестве энхансеров. Таким образом, наше представление о связывании ТФ лишь приблизительно. ТФ может взаимодействовать с кофакторами, которые изменяют способность ТФ активировать или репрессировать транскрипцию их генов-мишеней [82,83]. Кроме того, некоторые случаи, когда мы обнаружили высокие ковариации между ТФ и генами с низким средством

**Таблица 3.** Коэффициенты корреляции Спирмана между средней силой сайта связывания ТФ и числом сайтов связывания с ковариантностью генов-мишеней при использовании верхнего 10%-го порога обрезания для каждого ТФ

ТФ	Функция ТФ	Количество генов-мишеней с положительной ковариантностью	Коэффициенты корреляции Спирмана для средней положительной силы связывания (скорректированные значения $p$ )	Коэффициенты корреляции Спирмана для числа сайтов связывания с положительной ковариантностью (скорректированные значения $p$ )	Количество генов-мишеней с отрицательной ковариантностью	Коэффициенты корреляции Спирмана для средней отрицательной силы связывания (скорректированные значения $p$ )	Коэффициенты корреляции Спирмана для числа сайтов связывания с отрицательной ковариантностью (скорректированные значения $p$ )	Стандартное отклонение / среднее значение величины RPKM для ТФ	10%-е пороговые значения	Расчетные пороговые значения для числа сайтов связывания на основании данных RPKM для каждого ТФ	Максимальные значения позиционной весовой матрицы ТФ с референсным геномом
<i>Bicoid</i>	Maternal	196	<b>0,268</b> (0,0004)*	<b>0,209</b> (0,007)*	281	<b>0,0018</b> (0,488)	<b>0,024</b> (0,471)	37,13/ 58,86	<b>5,32</b>	2,891	7,54
<i>Caudal</i>	Maternal	150	<b>0,047</b> (0,293)	<b>0,027</b> (0,371)	195	<b>0,072</b> (0,247)	<b>0,0513</b> (0,376)	4,41/ 14,89	<b>4,89</b>	1,916	7,403
<i>Engrailed</i>	Segment-Polarity	280	<b>0,441</b> (1,38502e-13)*	<b>0,424</b> (1,659e-12)*	235	<b>0,023</b> (0,422)	<b>-0,137</b> (0,039)*	10,05/ 28,18	<b>5,31</b>	2,601	6,853
<i>Even skipped</i>	Pair-rule	160	<b>0,136</b> (0,103)	<b>0,123</b> (0,122)	145	<b>-0,179</b> (0,039)*	<b>-0,172</b> (0,041)*	3,25/ 5,17	<b>6,32</b>	3,521	9,635
<i>Fushi tarazu</i>	Pair-rule	187	<b>0,227</b> (0,0037)*	<b>0,179</b> (0,032)*	205	<b>-0,084</b> (0,201)	<b>-0,049</b> (0,441)	5,22/ 7,04	<b>4,45</b>	2,001	7,212
<i>Giant</i>	Gap	166	<b>0,058</b> (0,274)	<b>0,045</b> (0,319)	124	<b>-0,269</b> (0,014)*	<b>-0,191</b> (0,038)*	8,75/ 23,84	<b>6,17</b>	3,553	10,581
<i>Hairy</i>	Pair-rule	136	<b>0,052</b> (0,286)	<b>0,096</b> (0,204)	90	<b>-0,217</b> (0,037)*	<b>-0,040</b> (0,403)	6,33/ 13,18	<b>9,34</b>	7,273	11,609
<i>Hunchback</i>	Gap	185	<b>0,089</b> (0,194)	<b>0,077</b> (0,207)	151	<b>-0,224</b> (0,017)*	<b>-0,193</b> (0,030)*	8,00/ 30,37	<b>7,55</b>	4,431	8,729
<i>Knirps</i>	Gap	207	<b>0,091</b> (0,192)	<b>0,079</b> (0,206)	94	<b>-0,052</b> (0,400)	<b>0,065</b> (0,234)	5,93/ 17,92	<b>4,7</b>	2,042	8,280
<i>Kruppel</i>	Gap	150	<b>0,052</b> (0,283)	<b>0,170</b> (0,042)*	98	<b>-0,214</b> (0,038)*	<b>-0,226</b> (0,034)*	11,76/ 31,47	<b>5,42</b>	3,058	9,643
<i>Runt</i>	Pair-rule	126	<b>0,214</b> (0,028)*	<b>0,194</b> (0,038)*	91	<b>0,051</b> (0,401)	<b>0,045</b> (0,445)	6,21/ 16,97	<b>9,35</b>	6,761	10,636
<i>Snail</i>	DV	236	<b>0,057</b> (0,276)	<b>0,146</b> (0,039)*	170	<b>-0,196</b> (0,023)*	<b>0,034</b> (0,464)	16,78/ 50,16	<b>5,92</b>	3,139	8,951
<i>Tailless</i>	Gap	228	<b>0,076</b> (0,196)	<b>0,046</b> (0,312)	105	<b>-0,218</b> (0,037)*	<b>-0,201</b> (0,043)*	2,90/ 7,91	<b>4,74</b>	2,261	7,121
<i>Twist</i>	DV	199	<b>0,152</b> (0,044)*	<b>0,036</b> (0,329)	197	<b>0,012</b> (0,467)	<b>0,023</b> (0,479)	17,96/ 40,84	<b>6,03</b>	3,440	11,117

Примечание. Для генов-мишеней были рассчитаны отдельные коэффициенты корреляции, которые положительно и отрицательно изменяются в зависимости от конкретного ТФ. \* – Скорректированные значения  $p$ , которые существенны при  $\alpha = 0,05$ .

**Таблица 4.** Коэффициенты корреляции Спирмана между средней силой сайта связывания ТФ и числом сайтов связывания с ковариантностью генов-мишеней при использовании верхнего 5%-го порога обрезания для каждого ТФ

ТФ	Функция ТФ	Количество генов-мишеней с положительной ковариантностью	Коэффициенты корреляции Спирмана для средней положительной силы связывания (скорректированные значения $p$ )	Коэффициенты корреляции Спирмана для числа сайтов связывания с положительной ковариантностью (скорректированные значения $p$ )	Количество генов-мишеней с отрицательной ковариантностью	Коэффициенты корреляции Спирмана для средней отрицательной силы связывания (скорректированные значения $p$ )	Коэффициенты корреляции Спирмана для числа сайтов связывания с отрицательной ковариантностью (скорректированные значения $p$ )	Стандартное отклонение / среднее значение величины RPKM для ТФ	10%-е пороговые значения	Расчетные пороговые значения для числа сайтов связывания на основании данных RPKM для каждого ТФ	Максимальные значения позиционной весовой матрицы ТФ с референсным геномом
<i>Bicoid</i>	Maternal	58	0,376 (0,007)*	0,335 (0,015)*	79	0,069 (0,381)	0,087 (0,369)	37,13/ 58,86	5,9	2,891	7,54
<i>Caudal</i>	Maternal	62	0,102 (0,376)	0,052 (0,454)	76	0,047 (0,436)	0,061 (0,383)	4,41/ 14,89	5,32	1,916	7,403
<i>Engrailed</i>	Segment-Polarity	102	0,545 (4,3512e-08)*	0,497 (1,512e-06)*	90	0,018 (0,445)	-0,227 (0,041)*	10,05/ 28,18	5,9	2,601	6,853
<i>Even skipped</i>	Pair-rule	73	0,092 (0,381)	0,076 (0,371)	66	-0,291 (0,027)*	-0,261 (0,031)*	3,25/ 5,17	6,93	3,521	9,635
<i>Fushi tarazu</i>	Pair-rule	52	0,365 (0,014)*	0,287 (0,023)*	88	-0,091 (0,348)	-0,068 (0,372)	5,22/ 7,04	4,95	2,001	7,212
<i>Giant</i>	Gap	71	0,043 (0,395)	0,017 (0,557)	55	-0,332 (0,024)*	-0,299 (0,026)*	8,75/ 23,84	6,81	3,553	10,581
<i>Hairy</i>	Pair-rule	76	0,028 (0,405)	0,041 (0,462)	50	-0,299 (0,039)*	-0,135 (0,350)	6,33/ 13,18	10,2	7,273	11,609
<i>Hunchback</i>	Gap	85	0,139 (0,238)	0,131 (0,232)	65	-0,303 (0,026)*	-0,254 (0,033)*	8,00/ 30,37	7,78	4,431	8,729
<i>Knirps</i>	Gap	95	0,072 (0,384)	0,059 (0,443)	65	-0,092 (0,362)	0,043 (0,391)	5,93/ 17,92	5,42	2,042	8,280
<i>Kruppel</i>	Gap	77	0,067 (0,386)	0,229 (0,030)*	49	-0,287 (0,042)*	-0,293 (0,029)*	11,76/ 31,47	5,96	3,058	9,643
<i>Runt</i>	Pair-rule	71	0,288 (0,024)*	0,251 (0,027)*	55	0,034 (0,441)	0,030 (0,413)	6,21/ 16,97	9,88	6,761	10,636
<i>Snail</i>	DV	87	0,045 (0,391)	0,228 (0,029)*	74	-0,284 (0,027)*	0,077 (0,371)	16,78/ 50,16	6,7	3,139	8,951
<i>Tailless</i>	Gap	68	0,058 (0,388)	0,033 (0,493)	70	-0,265 (0,036)*	-0,253 (0,032)*	2,90/ 7,91	5,48	2,261	7,121
<i>Twist</i>	DV	77	0,239 (0,036)*	0,013 (0,559)	87	0,016 (0,448)	0,041 (0,395)	17,96/ 40,84	6,68	3,440	11,117

Примечание. Для генов-мишеней были рассчитаны отдельные коэффициенты корреляции, которые положительно и отрицательно изменяются в зависимости от конкретного ТФ. \* – Скорректированные значения  $p$ , которые существенны при  $\alpha = 0,05$ .

**Таблица 5.** Регуляторные взаимодействия между ТФ

ТФ	Ген-мишень	Коэффициент корреляции Спирмана	Скорректированная величина $p$
<i>Bicoid</i>	Caudal	0,62	1,56E-05
<i>Bicoid</i>	Knirps	0,66	3,00E-06
<i>Bicoid</i>	Kruppel	0,69	5,33E-07
<i>Bicoid</i>	Giant	0,67	1,72E-06
<i>Bicoid</i>	Hunchback	0,61	1,56E-05
<i>Bicoid</i>	Tailless	0,64	7,46E-06
<i>Bicoid</i>	Runt	0,68	1,43E-06
<i>Bicoid</i>	Hairy	0,60	4,10E-05
<i>Tailless</i>	Giant	-0,53	6,54E-05
<i>Tailless</i>	Knirps	-0,59	4,25E-05
<i>Twist</i>	Sloppy paired	-0,62	7,46E-06
<i>Twist</i>	Odd paired	-0,68	5,33E-07

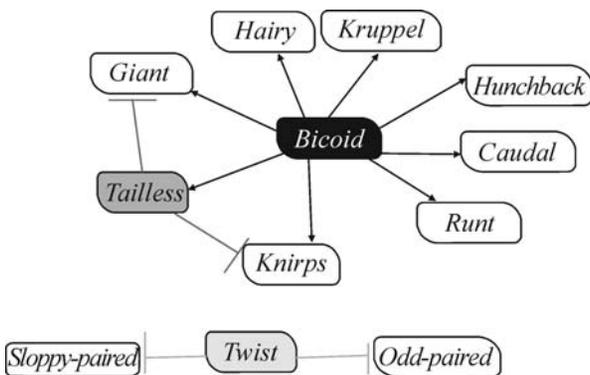
Примечание. Авторы считали значимыми скорректированные значения  $p < 0,001$ .

сайтов, могут быть связаны с наличием промежуточных белков, которые связываются с данным сайтом и увеличивают связывающий потенциал ТФ [82,84]. Несмотря на все эти оговорки, мы можем использовать предсказанные сайты связывания ТФ и ковариацию экспрессии для подтверждения некоторых известных регуляторных взаимодействий в этой генной сети [3,5,76].

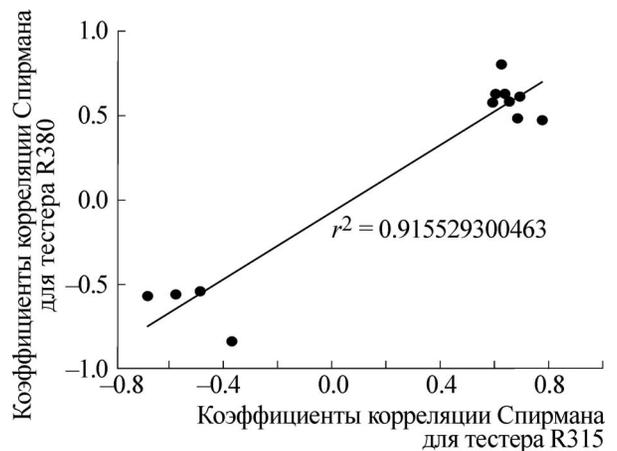
Мы можем представить себе, как экспрессия ТФ и регулируемых генов могут совместно меняться из-за косвенного регулирования. Тем не менее наблюдаемая взаимосвязь между силой регуляции и силой сайта связывания указывает на прямую составляющую в этой ковариации. Дальнейшее понимание происходит из сравнения активации и репрессии. Рассмотрим, на-

пример, активатор *bicoid*. Если ковариации происходят из-за косвенных последствий, их признаки могут быть как положительными, так и отрицательными. Можно было бы предположить, что тогда отрицательная регуляция должна быть косвенной, тогда как положительная регуляция может быть прямой или косвенной. Затем сила связывания не должна быть связана с величиной отрицательных ковариаций, в то время как положительная регуляция может, точно так, как отмечено на рис. 8. Сила такой ковариации, вероятно, отражает долю прямых влияний в нашем наборе данных.

Вообще говоря, число значимых ассоциаций ковариации в экспрессии с силой связывания



**Рис. 9.** Регуляторные взаимодействия между ТФ. Черными стрелками обозначена положительная корреляция между *bicoid* и другими ТФ. Темно-серыми линиями обозначена отрицательная корреляция между *tailless* и двумя *gap*-генами (*giant* and *knirps*), а светло-серыми линиями – отрицательная корреляция между *twist* и *odd-paired* и *sloppy-paired*.



**Рис. 10.** Демонстрация надежности результатов для генотипов штаммов тестеров. Скрещивание с использованием двух линий тестеров (R380 и R315) дало близкие значения коэффициентов корреляции для исследованных ТФ. Оси абсцисс и ординат соответствуют столбцам 3 и 4 в табл. 6.

**Таблица 6.** Регуляторные взаимодействия между ТФ для двух наборов данных, основанных на генотипе использованных тестеров

ТФ	Ген-мишень	Коэффициенты корреляции Спирмана (скорректированные значения $p$ ) для R315	Коэффициенты корреляции Спирмана (скорректированные значения $p$ ) для R380
<i>Bicoid</i>	<i>Caudal</i>	0,69 (3,7E-05)	0,61(0,004)
<i>Bicoid</i>	<i>Knirps</i>	0,59 (0,00036)	0,57 (0,00459)
<i>Bicoid</i>	<i>Kruppel</i>	0,77 (8,1E-06)	0,47 (0,019)
<i>Bicoid</i>	<i>Giant</i>	0,68 (3,7E-05)	0,48(0,0166)
<i>Bicoid</i>	<i>Hunchback</i>	0,63 (0,00018)	0,62(0,00331)
<i>Bicoid</i>	<i>Tailless</i>	0,65 (8,6E-05)	0,58(0,00448)
<i>Bicoid</i>	<i>Runt</i>	0,62(0,00018)	0,80(2,1E-05)
<i>Bicoid</i>	<i>Hairy</i>	0,60(0,00027)	0,62(0,00331)
<i>Tailless</i>	<i>Giant</i>	-0,49(0,0038)	-0,54(0,0071)
<i>Tailless</i>	<i>Knirps</i>	-0,37 (0,0311)	-0,84(2,1E-05)
<i>Twist</i>	<i>Sloppy paired</i>	-0,58 (0,00036)	-0,56(0,00459)
<i>Twist</i>	<i>Odd paired</i>	-0,68(3,7E-05)	-0,57(0,00448)

Примечание. Авторы считали значимыми скорректированные значения  $p < 0,05$ .

и числом сайтов связывания невелико. Принимая во внимание, что транскрипционный процесс является сложным и возможно потенциальное влияние промежуточных белков, мы не обеспокоены тем, что величина значимых ассоциаций ковариации низка. Корреляции, которые мы наблюдали с несколькими ТФ между силой связывания с конкретным геном-мишенью и регуляторной силой, поддерживают идею о том, что сила связывания ТФ является, в механистическом представлении, предсказателем силы регуляторного эффекта. Более того, наши результаты, говорящие о регуляторных взаимодействиях между ТФ, подтверждают результаты предыдущих исследований, таких как роль *bicoid* в качестве активатора.

Авторы благодарят П. Чанга и Е. Курмангалиева за их помощь и содержательное обсуждение.

Исследование, описанное в данной статье, было поддержано Национальным институтом здоровья (грант R00HD073191 – для З. Вундерлиха, гранты U01GM103804 и R01GM102227 – для С.В. Нужиной). Работа С.В. Нужиной также была поддержана грантами Российского научного фонда № 16-16-00007 (ковариационный анализ экспрессии генов и анализ регуляторных взаимодействий) и № 14-14-00302 (создание библиотек транскриптома и РНК-секвенирование).

#### СПИСОК ЛИТЕРАТУРЫ

1. E. H. Davidson, *Nature* **468**, 911 (2010).
2. I. S. Peter and E. H. Davidson, *Cell* **144**, 970 (2011).
3. X. Li, S. MacArthur, R. Bourgon, et al., *PLoS Biol.* **6**, e27 (2008).
4. J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, *Nat. Rev. Genet.* **10**, 252 (2009).
5. S. V. Nuzhdin, A. Rychkova, and M. W. Hahn, *Trends Genet.* **26**, 51 (2010).
6. Yang, S., et al., *Bioinformatics* **27**, 2972 (2011).
7. M. Levo and E. Segal, *Nat. Rev. Genet.* **15**, 453 (2014).
8. A. Nasiadka, B. H. Dietrich, and H. M. Krause, *Adv. Dev. Biol. Biochem.* **12**, 155 (2002).
9. S. Bonn and E. E. Furlong, *Curr. Opin. Genet. Dev.* **18**, 513 (2008).
10. M. Levine and E. H. Davidson, *Proc. Natl. Acad. Sci. USA* **102**, 4936 (2005).
11. K. Howard and P. Ingham, *Cell* **44**, 949 (1986).
12. C. M. Bergman, J. W. Carlson, and S. E. Celniker, *Bioinformatics* **21**, 1747 (2005).
13. R. Bonneau, *Nat. Chem. Biol.* **4**, 658 (2008).
14. B. W. Busser, M. L. Bulyk, and A. M. Michelson, *Curr. Opin. Genet. Dev.* **18**, 521 (2008).
15. R. E. Bumgarner and K. Y. Yeung, *Comput. Syst. Biol.* **541**, 225 (2009).
16. P. J. Park, *Nat. Rev. Genet.* **10**, 669 (2009).
17. S. Pepke, B. Wold, and A. Mortazavi, *Nat. Meth.* **6**, S22 (2009).
18. Y. R. Wang and H. Huang, *J. Theor. Biol.* **362**, 53 (2014).
19. D. J. Kliebenstein, *Plant Syst. Biol.* **553**, 227 (2009).
20. J. A. Lewis, I. M. Elkon, M. A. McGee, et al., *Genetics* **186**, 1197 (2010).
21. S. Mostafavi, et al., *J. Immunol.* **193**, 4485 (2014).
22. S. V. Nuzhdin, D. M. Tufts, and M.W. Hahn, *Evol. Dev.* **10**, 683 (2008).

23. W. Jin, R. M. Riley, R. D. Wolfinger, et al., *Nat. Genet.* **29**, 389 (2001).
24. S. V. Nuzhdin, M. L. Wayne, K. L. Harmon, and L. M. McIntyre, *Mol. Biol. Evol.* **21**, 1308 (2004).
25. S. V. Nuzhdin, M. L. Friesen, and L. M. McIntyre, *Trends Genet.* **28**, 421 (2012).
26. T. F. Mackay, et al., *Nature* **482**, 173 (2012).
27. D. Campo, K. Lehmann, C. Fjeldsted, et al., *Mol. Ecol.* **22**, 5084 (2013).
28. J. P. Dunham and M. L. Friesen, *Cold Spring Harb. Protoc.* **9**, 820 (2013).
29. A. Dobin, et al., *Bioinformatics* **29**, 15 (2013).
30. S. Anders, P. T. Pyl, and W. Huber, *Bioinformatics* **31** (2), 166 (2015).
31. A. Mortazavi, B. A. Williams, K. McCue, et al., *Nat. Meth.* **5**, 621 (2008).
32. T. Sandmann, C. Girardot, M. Brehme, et al., *Genes Dev.* **21**, 436 (2007).
33. J. A. Campos-Ortega and V. Hartenstein, *The embryonic development of Drosophila melanogaster* (Springer Sci. Business Media, 2013).
34. S. Thomas, et al., *Genome Biol.* **12**, R43 (2011).
35. G. Z. Hertz and G. D. Stormo, *Bioinformatics* **15**, 563 (1999).
36. L. J. Zhu, et al., *Nucleic Acids Res.* **39**, D111 (2011).
37. Y. Benjamini and Y. Hochberg, *J. Royal Stat. Soc. Series B* **57** (1), 289 (1995).
38. Z. Wang, M. Gerstein, and M. Snyder, *Nat. Rev. Genet.* **10**, 57 (2009).
39. G. Struhl, K. Struhl, and P. M. Macdonald, *Cell* **57**, 1259 (1989).
40. M. Leptin, *Genes Dev.* **5**, 1568 (1991).
41. R. M. Cripps, B. L. Black, B. Zhao, et al., *Genes Dev.* **12**, 422 (1998).
42. A. Stathopoulos, M. Van Drenth, A. Erives, et al., *Cell* **111**, 687 (2002).
43. M. D. Schroeder, et al., *PLoS Biol.* **2**, e271. (2004).
44. J. Zeitlinger, et al., *Genes Dev.* **21**, 385 (2007).
45. A. Ochoa-Espinosa, D. Yu, A. Tsigos, et al., *Proc. Natl. Acad. Sci. USA* **106**, 3823 (2009).
46. A. Porcher and N. Dostatni, *Curr. Biol.* **20**, R249 (2010).
47. X. Wu, R. Vakani, and S. Small, *Development* **125**, 3765 (1998).
48. G. F. Hewitt, et al., *Development* **126**, 1201 (1999).
49. É. Morán and G. Jiménez, *Mol. Cell. Biol.* **26**, 3446 (2006).
50. J. O. Yáñez-Cuna, E. Z. Kvon, and A. Stark, *Trends Genet.* **29**, 11 (2013).
51. A. S. Manoukian and H. M. Krause, *Genes Dev.* **6**, 1740 (1992).
52. S. Barolo and M. Levine, *EMBO J.* **16**, 2883 (1997).
53. G. Jiménez, Z. E. Paroush, and D. Ish-Horowicz, *Genes Dev.* **11**, 3072 (1997).
54. M. Kobayashi, R. E. Goldstein, M. Fujioka et al., *Development* **128**, 1805 (2001).
55. M. Fujioka, G. L. Yusibova, N. H. Patel, et al., *Development* **129**, 4411 (2002).
56. D. Bianchi-Frias, A. Orian, J. J. Delrow, et al., *PLoS Biol.* **2**, e178. (2004).
57. Y. Hiromi and W. J. Gehring, *Cell* **50**, 963 (1987).
58. S. G. Kramer, T. M. Jinks, P. Schedl, and J. P. Gergen, *Development* **126**, 191 (1999).
59. Y. Yu, M. Yussa, J. Song, et al., *Mech. Dev.* **83**, 95 (1999).
60. A. Nasiadka, A. Grill, and H. M. Krause, *Development* **127**, 2965 (2000).
61. J. C. Wheeler, K. Shigesada, J. P. Gergen, and Y. Ito, *Semin. Cell Dev. Biol.* **11**, 369 (2000).
62. P. I. Zuo, D. Stanojević, J. Colgan, et al., *Genes Dev.* **5**, 254 (1991).
63. M. V. Staller, et al., *Proc. Natl. Acad. Sci. USA* **112**, 785 (2015).
64. F. Sauer and H. Jäckle, *Nature* **353**, 563 (1991).
65. F. Sauer, J. D. Fondell, Y. Ohkuma, et al., *Nature* **375**, 162 (1995).
66. A. La Rosée-Borggreve, T. Häder, D. Wainwright, et al., *Mech. Dev.* **89**, 133 (1999).
67. J. Heemskerk, S. DiNardo, R. Kostriken, and P. H. O'Farrell, *Nature* **352**, 404 (1991).
68. T. Tabata, S. Eaton, and T. B. Kornberg, *Genes Dev.* **6**, 2635 (1992).
69. C. Alexandre and J. P. Vincent, *Development* **130**, 729 (2003).
70. M. Rembold, et al., *Genes Dev.* **28**, 167 (2014).
71. L. Sánchez and D. Thieffry, *J. Theor. Biol.* **211**, 115 (2001).
72. J. Jaeger, et al., *Genetics* **167**, 1721 (2004).
73. J. Jaeger, *Cell. Mol. Life Sci.* **68**, 243 (2011).
74. F. Liu, A. H. Morrison, and T. Gregor, *Proc. Natl. Acad. Sci. USA* **110**, 6724 (2013).
75. I. A. Gula and A. M. Samsonov, *Bioinformatics* **31**, 714 (2015).
76. S. MacArthur, et al., *Genome Biol.* **10**, R80 (2009).
77. J. M. Franco-Zorrilla, I. López-Vidriero, J. L. et al., *Proc. Natl. Acad. Sci. USA* **111**, 2367 (2014).
78. B. R. Graveley, et al., *Nature* **471**, 473 (2011).
79. A. S. Hammonds, et al., *Genome Biol.* **14**, R140 (2013).
80. M. D. Biggin and R. Tjian, *Funct. Integr. Genomics* **1**, 223 (2001).
81. M. Levine and R. Tjian, *Nature* **424**, 147 (2003).
82. S. Björklund, G. Almouzni, I. Davidson, et al., *Cell* **96**, 759 (1999).
83. A. Tanay, *Genome Res.* **16**, 962 (2006).
84. M. Mannervik, Y. Nibu, H. Zhang, and M. Levine, *Science* **284**, 606 (1999).

# Inference of Transcription Factor Regulation Patterns Using Gene Expression Covariation in Natural Populations of *Drosophila melanogaster*

N.M. Osman\* \*\*, T.H. Kitapci\*, S. Vlaho\*, Z. Wunderlich\*\*, and S.V. Nuzhdin\* \*\*\*\*

\*University of Southern California, Los Angeles, California 90007 USA

\*\*National Research Centre, Dokki, Giza, 12622 Egypt

\*\*\*University of California, Irvine, California 92697 USA

\*\*\*\*Peter the Great Saint Petersburg Polytechnical University,  
Polytekhnicheskaya ul. 29, St.Petersburg, 195251 Russia

Gene regulatory networks control the complex programs that drive development. Deciphering the connections between transcription factors and target genes is challenging, in part because transcription factors bind to thousands of places in the genome but control expression through a subset of these binding events. We hypothesize that we can combine natural variation of expression levels and predictions of transcription factors binding sites to identify transcription factors targets. We gathered RNA-seq data from 71 genetically distinct F1 *Drosophila melanogaster* embryos and calculated the correlations between transcription factors and potential target genes' expression levels, which we called "regulatory strength". To separate direct and indirect transcription factors targets, we hypothesized that direct transcription factors targets will have the preponderance of binding sites in their upstream regions. Using 14 transcription factors active during embryogenesis, we have found that 12 transcription factors showed a significant correlation between their binding strength and regulatory strength on downstream targets, and 10 transcription factors showed a significant correlation between the number of binding sites and the regulatory effect on target genes. The main functions such as the activator function of bicoid and the particular interactions we observed between our transcription factors, as the repressor function of Twist in regard to sloppy paired and odd paired loci, coincide with those described in literature.

*Key words: Drosophila melanogaster, gene regulatory networks, transcription factors*