

ПОПАРНО-ПЕРЕКРЕСТНОЕ СРАВНЕНИЕ ТРАНСКРИПТОМОВ МЛЕКОПИТАЮЩИХ В ИССЛЕДОВАНИИ ВЛИЯНИЯ ПОЛИПЛОИДИИ НА АКТИВНОСТЬ ЭКСПРЕССИИ ГЕННЫХ МОДУЛЕЙ РАЗВИТИЯ

© О. В. Анацкая,^{1,*} Е. А. Эренпрейса,² Н. Н. Никольский,¹ А. Е. Виноградов¹

¹Институт цитологии РАН, Санкт-Петербург, 194064,

и ²Латвийский биомедицинский учебно-исследовательский центр, Рига, LV-1067, Латвия;

* электронный адрес: olga.anatskaya@gmail.com

Разработка и развитие высокочувствительных методов биоинформатики связаны с необходимостью получения сведений о влиянии эпигенетических изменений на активность генов. Одним из факторов, вызывающих такие изменения, является полиплоидия, которая, сохраняя баланс дозы генов, может оказывать лишь довольно слабое влияние на их экспрессию. В настоящее время единой концепции относительно действия полиплоидии на транскриптом нет. Поэтому мы разработали интегративный биоинформационный метод определения слабых влияний полиплоидии на активность генов с помощью попарноперекрестного анализа транскриптомов тканей млекопитающих с разной степенью полиплоидизации. Преимуществом метода является его способность отделять видо- и тканеспецифические эффекты от эволюционно консервативных. Применение метода продемонстрировано на примере анализа генных модулей и сетей межбелковых взаимодействий, вовлеченных в координацию процессов развития. Исследования были проведены на полнотранскриптомных данных для сердца и печени человека и мыши. Показано, что активированные плоидностью гены обогащают (т. е. представлены выше случайного уровня) биологические процессы базы Gene Ontology (GO) и пути базы KEGG, относящиеся к развитию, морфогенезу и биологии стволовых клеток (включая сигнальные пути Hippo, Pi3K, WNT, Hedgehog и TGF- β) в большей степени, чем подавленные плоидностью гены. Структура и состав сетей межбелковых взаимодействий, построенных для регулируемых плоидностью генов, подтвердили результаты анализа генных модулей. Таким образом, наши данные впервые показали, что полиплоидия может регулировать модули развития, повышая их активность.

Ключевые слова: сравнительный анализ транскриптомов, сети межбелковых взаимодействий, генные модули, полиплоидия, развитие и дифференцировка, стволовая клетка, эмбриональный фенотип.

Принятые сокращения: GO-процессы — процессы из базы Gene Ontology (Gene Ontology Consortium, 2015).

Развитие геномных технологий, позволяющих получать данные по полному геному, транскриптому, протеому и интерактому, создало новые возможности для изучения клеточных процессов. Типичным результатом таких экспериментов являются первичные данные с многотысячными списками генов, нуждающимися в дальнейшей обработке. Таким образом, новые технологии предъявляют новые требования к методам анализа данных.

С появлением множества общих и специализированных баз, полученных с помощью постгеномных технологий, появились методы интегративного анализа комплементарных данных. Стало возможным, например, сопоставлять последовательности ДНК, РНК и белков, экспрессию генов на уровне мРНК и белка, а также регуляторные мотивы нуклеотидных последовательностей (<http://www.ncbi.nih.gov>).

Качественным скачком, позволившим многократно повысить чувствительность цитонного анализа, стало создание методов идентификации генных модулей и сетей взаимодействий между генами, белками и микро-РНК.

Эти методы позволили уйти от необходимости применения консервативного (и произвольного) двукратного порога различия экспрессии генов, пригодного для оценки воздействия лекарств, ядов и резких изменений внешней среды (El-Samad, Madhani, 2011; Ideker et al., 2011). Появилась возможность с высокой достоверностью улавливать разницу менее чем в 2 раза. Это снижение порога уже позволило оценить небольшие изменения в биологических процессах, вызванные относительно небольшими флуктуациями условий внутреннего микроокружения, например температуры, артериального давления и уровня гормонов, небольшие изменения иммунного и пролиферативного статуса (Arkin, Schaffer, 2011; Nurse, Hayles, 2011; Dow, Lowe, 2012; Tyagi, 2015). Но, несмотря на значительный прогресс, разработка методов анализа полногеномных данных остается по-прежнему актуальной.

Разработка и развитие высокочувствительных методов биоинформатики связаны, в частности, с необходимостью получения сведений о влиянии эпигенетических

изменений на транскриптом. Одной из частых и в то же время наименее изученных причин эпигенетических изменений являются геномные дупликации или полиплоидизация соматических клеток. Сохраняя баланс дозы генов, геномные дупликации могут оказывать лишь слабое влияние на работу отдельных генов. Однако модифицируя активность тысяч генов одновременно, полиплоидия участвует в регуляции процессов постнатального развития, дифференцировки, ответа на стресс и патогенеза многих заболеваний (Lee et al., 2009). Показано, что полиплоидизация является важным фактором начальных этапов трансформации клеток (Ganem et al., 2007). Фигурально можно сказать, что слабые, подобные бликам, сдвиги уровня экспрессии отдельных генов, суммируясь, могут превратиться в красочный салют.

В последнее время внимание к проблеме изменения активности генов под воздействием геномных дупликаций резко возросло. Влияние полиплоидии на транскриптом было исследовано на гепатоцитах, мегакариоцитах, клетках васкулярного эпителия и плаценты. Оценка различий экспрессии с помощью консервативного двукратного порога привело исследователей к выводу о том, что из нескольких тысяч генов на полиплоидию не реагирует ни один (как в случае гепатоцитов) или же реагируют всего несколько десятков генов (Pandit et al., 2013). При этом для каждого клеточного типа этот небольшой список генов оказался разным. Так, у клеток васкулярного эпителия подавляется несколько генов, относящихся к иммунной системе (Jones, Ravid, 2004), у мегакариоцитов зарегистрирована активация тканеспецифических генов, а у децидуальных клеток — активация ядерных генов, кодирующих белки митохондрий (Ma et al., 2011). Эти разногласия свидетельствуют о том, что полные эффекты соматической полиплоидии (если они существуют) пока еще не выявлены.

Это побудило нас к созданию методики идентификации тонких эффектов геномных дупликаций с помощью сравнения транскриптомов тканей разных видов млекопитающих с разной степенью полиплоидизации. В настоящей работе мы приводим пример использования разработанного нами метода для изучения влияния геномных дупликаций на активность эмбриональной программы, которая включает в себя модули развития и роста, морфогенеза и дифференцировки.

Материал и методика

Основой разработанного метода является одновременное попарноперекрестное межвидовое и межтканевое сравнение активности транскриптома в органах с разной степенью полиплоидии, например в сердце и печени человека и мыши. Известно, что человек имеет высокополиплоидное сердце (4—8 геномов на ядро кардиомиоцита) и преимущественно диплоидную печень (около 2 геномов на ядро гепатоцита). Мышь, напротив, имеет преимущественно диплоидное сердце (около 2 геномов на ядро кардиомиоцита) и высокополиплоидную печень (4—8 геномов на ядро гепатоцита) (Anatskaya, Vinogradov, 2004a, 2004b). Далее для упрощения мы будем пользоваться терминами «полиплоидные органы» (состоящие преимущественно из полиплоидных клеток, как сердце человека и печень мыши) и «диплоидные органы» (состоящие преимущественно из диплоидных клеток сердце мыши и печень человека).

Источником данных служила база данных BioGPS (Wu et al., 2013), содержащая полнотранскриптомные данные, полученные с помощью олигонуклеотидных микрочипов. Были отобраны только охарактеризованные гены, т. е. представленные в базе NCBI Gene (Brown et al., 2015). Кроме того, гены должны были быть гомологичны (ортологичны) для человека и мыши. Ортологичными называют гомологичные гены из геномов разных организмов, образовавшиеся из одного и того же исходного гена в результате процесса видообразования (в отличие от паралогичных гомологичных генов, которые образовались в результате дупликации исходного гена внутри одного и того же генома). Ортологичные гены наиболее близки по структуре кодируемых ими белков (среди всех генов сравниваемых геномов), и поэтому предполагается, что они сохранили исходную функцию (Waterhouse et al., 2011; Powell et al., 2012).

Полные протеомы человека и мыши были взяты из базы данных RefSeq (Pruitt et al., 2014). С помощью высокочувствительного алгоритма Смита—Ватермана, реализованного в программе «Search» из пакета Fasta (Pearson, 2004), мы провели тотальное сравнение каждого белка человека против всех белков мыши (и наоборот, каждого белка мыши против всех белков человека). После этого мы отобрали пары генов с наиболее высокой статистической значимостью обоих взаимных сравнений кодируемых ими белков (reciprocal best hit), которые можно считать ортологичными.

После отбора пар ортологичных генов мы проанализировали изменение экспрессии в каждой паре. Мы считали, что гены меняют активность в зависимости от плоидности, если их экспрессия менялась в одинаковом направлении в обоих реципрокных сравнениях в соответствии с различием в степени полиплоидизации тканей (сердце человека → сердце мыши и печень мыши → печень человека). Кроме того, по крайней мере один ген из ортологичной пары должен быть достоверно экспрессирован в каждой паре сравниваемых тканей. Для того чтобы провести дополнительную фильтрацию сигнала от видо- и тканеспецифического шума, мы нормировали (т. е. разделили) экспрессию каждого гена в исследуемой ткани (сердце и печени) к среднему уровню экспрессии этого гена в остальных тканях данного организма.

Гены, которые соответствовали всем перечисленным критериям, были разделены на активированные и ингибированные и проанализированы на обогащение ими биологических процессов базы Gene Ontology (GO) (Gene Ontology Consortium, 2015) по нашей методике (Vinogradov, Anatskaya, 2007; Vinogradov, 2015). Под обогащением имеется в виду статистически значимое превышение случайно ожидаемого уровня. В частности, мы использовали усредненное изменение уровней экспрессии для обоих перекрестных сравнений (сердце человека → сердце мыши и печень мыши → печень человека). Для этой цели мы готовили преддетерминированные (априорные) функциональные группы генов на основе базы GO, отдельно для каждой ее категории — Biological processes, Molecular functions и Cellular components. Для каждой категории GO мы собирали все ее субкатегории (используя направленный ациклический граф, описывающий структуру данной базы). Ген относили к какой-либо категории в том случае, если он входил в любую из субкатегорий данной категории.

Оценку статистической значимости различий в экспрессии производили с помощью рандомизации. С этой

целью мы делали для каждой тестируемой группы генов 20 000 случайных выборок из полного набора генов (с размером каждой выборки, равным размеру тестируемой группы генов). Затем мы определяли, в каком проценте случайных выборок абсолютная величина усредненного изменения экспрессии генов равна или превышает абсолютную величину усредненного изменения экспрессии в тестируемой группе генов.

Использование рандомизации для определения статистической значимости разницы экспрессии между средним для генного модуля и средним для всего набора генов мы считаем более эффективным, чем использование методов параметрической (например, *t*-тест, ANOVA) или непараметрической статистики (тесты Манна—Уитни—Вилкоксона, Крускал—Уоллиса и др.). Использование рандомизации предпочтительнее, потому что параметрическая статистика требует нормального распределения величин (чего обычно не наблюдается), а непараметрическая теряет часть информации и поэтому обладает меньшей мощностью (т. е. меньшей разрешающей способностью). Метод рандомизации не зависит от особенностей распределения и сохраняет всю информацию (хотя требует написания специальных программ и значительно большего компьютерного времени для анализа).

Независимо от того, каким способом было сделано определение генных модулей с измененной экспрессией, в статистическую значимость этого определения необходимо ввести поправку на множественность сравнений (эта поправка всегда приводит к снижению уровня значимости). Это связано с тем, что в данном случае делается анализ сразу множества модулей (тысяч или даже десятков тысяч). Поэтому обычный порог статистической значимости, принятый в научных исследованиях ($p < 0.05$), здесь не подходит. Такая значимость ($p < 0.05$) может наблюдаться случайным образом в 5 из 100 тестируемых модулей. Поправку для случая независимых множественных сравнений обычно делают по методу Бонферони. Он заключается в умножении уровня значимости на число протестированных модулей. Т. е. в случае тестирования 1000 генных модулей нужно, чтобы статистическая значимость для модуля с измененной экспрессией была не хуже, чем $p < 0.00005$ (для того чтобы после умножения на 1000 значимость оказалась еще приемлемой, т. е. $p < 0.05$). Однако применение метода Бонферони оправдано только для тестирования полностью независимых модулей.

В нашем случае генные модули не являются полностью независимыми (например, некоторые модули могут частично включать в себя одни и те же гены, некоторые категории GO могут быть полностью или частично включены в другие, более общие категории). В этом случае метод Бонферони является излишне консервативным. Для случаев неполной независимости множественных сравнений разработаны специальные методы поправки статистической значимости (в этом случае поправка получается меньше, чем по Бонферони). Мы использовали определение «false discovery rate» (Storey, Tibshirani, 2003), при котором одновременно анализируется распределение всего набора полученных значений статистической значимости. Полученный параметр обозначается «q-value» и может рассматриваться как «p-value», в который внесена поправка на множественность тестов.

Сети межбелковых взаимодействий для продуктов генов, входящих в биологические процессы GO, которые

регулируют биологию стволовых клеток, строили с помощью сервера базы STRING (Szklarczyk et al., 2015). Достоверность обогащения GO-процессов и путей базы KEGG (Kanehisa, Goto, 2000) белками полученной сети, функциональные характеристики белков сети (например, их принадлежность к транскрипционным факторам, модуляторам хроматина и сигнальным молекулам), а также достоверность обогащения сети межбелковыми взаимодействиями (по сравнению со случайной выборкой с тем же числом генов и с тем же распределением числа межбелковых взаимодействий) также определяли с помощью сервера базы STRING (Szklarczyk, et al., 2015). Число связей (ребер), отходящих от белков (узлов) сети, находили с помощью программы Cytoscape. Это необходимо для идентификации центральных регуляторов (хабовых узлов) сети. Достоверность различий в числе генных модулей определяли с помощью биномиального теста (сравнивали числа различных модулей как их доли от общего числа генных).

Результаты и обсуждение

В настоящей работе мы предлагаем интегративный сравнительно-биоинформационный метод для нахождения биологических модулей, характеризующих влияние полиплоидии на профиль экспрессии генов. Преимуществом метода является то, что он дает возможность выделить видо- и тканеспецифические эффекты и обнаружить эволюционно консервативные, а значит, функционально значимые изменения, вызванные геномными дупликациями. Фильтрацию видоспецифических шумов обеспечивает перекрестно-межвидовое сравнение, а фильтрацию тканеспецифических — межтканевое сравнение. Применение метода мы показываем на примере анализа процессов, вовлеченных в координацию эмбриональной программы. Активность генов считали зависимой от плоидности в том случае, если отношение количества их мРНК в двух попарных сравнениях для полиплоидных и диплоидных органов (сердце человека/сердце мыши и печень мыши/печень человека) превышало 1.3 раза. Такой уровень отреза используют в тех случаях, когда необходимо выявить тонкие различия в экспрессии на больших массивах данных (Ideker et al., 2011; Nurse, Hayles, 2011). Результатом отбора стали 875 и 952 гена с более высоким и более низким уровнями мРНК в полиплоидных органах по сравнению с диплоидными органами. Далее мы будем называть эти гены соответственно активированными и подавленными.

База со сведениями об аннотациях генов (<http://www.ncbi.nlm.nih.gov/pubmed>) и анализ функциональных модулей GO с помощью нашей программы (Vinogradov, Anatskaya, 2007; Vinogradov, 2015) дали нам возможность идентифицировать GO- процессы, достоверно обогащенные регулируемые плоидностью генов. Затем из общего списка процессов мы извлекли те, которые содержат в названии термины «development», «embryo», «morphogenesis», «differentiation» и «stem cell». В результате было получено 92 GO-процесса, естественным образом разделившихся на две функциональные группы. В первую группу, которую мы назвали «развитие и дифференцировка», вошли 82 GO-процесса, координирующие развитие органов и систем, регуляцию развития, эмбриональное развитие, а также морфогенез и дифференцировку (рис. 1). Вторая группа, названная нами «стволовая клетка», объединила

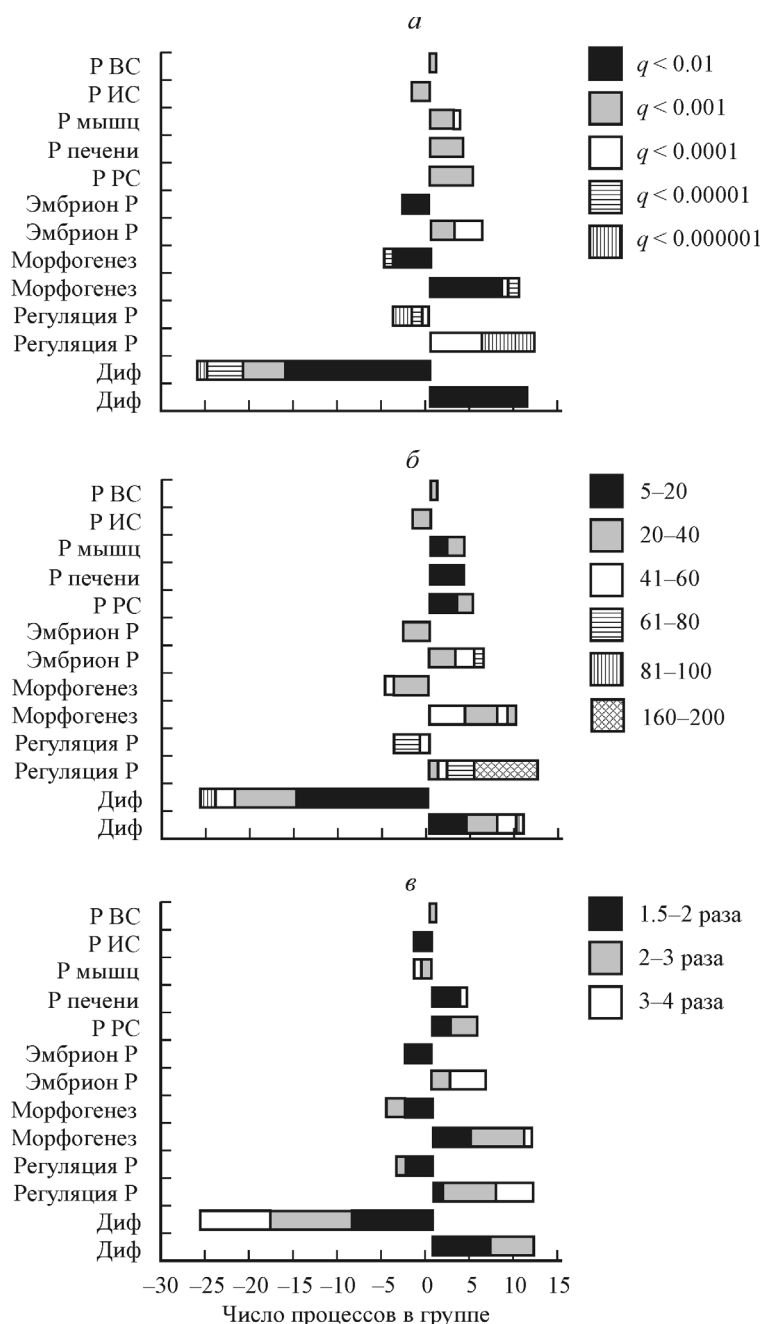


Рис. 1. Функциональные и количественные характеристики GO-процессов, относящихся к развитию и дифференцировке, которые обогащены регулируемыми плоидностью генами в сердце и печени человека и мыши.

Показано распределение GO-процессов по уровню достоверности q (а, где q — значение p , в которое внесена поправка на множественность тестов, см. раздел «Материал и методика»), числу генов (б) и среднему уровню различия уровней экспрессии между полиплоидными и диплоидными органами (в). По горизонтали — число GO-процессов, обогащенных активированными (положительное) и подавленными (отрицательное) генами; по вертикали — название групп процессов. P — развитие, BC — взрослое сердце, IC — иммунная система, PC — репродуктивная система, Эмбрион — эмбриональное, Диф — дифференцировка.

10 GO-процессов, вовлеченных в биологию стволовых клеток, включая развитие, поддержание, дифференцировку и пролиферацию (рис. 2).

Процессы развития и дифференцировки. Статистический анализ и функциональное распределение процессов для группы «развитие и дифференцировка» свидетельствуют о том, что полиплоидия ассоциирована с проявлениями эмбриональности в ключевых биологических путях (рис. 1, а–в). Наши результаты свидетельствуют о том, что активированные гены обогащают боль-

шое число процессов, относящихся к развитию сердца, печени, репродуктивной и эндокринной систем, эмбриональному развитию, морфогенезу и регуляции развития, чем подавленные гены (48 и 34 модуля, $p < 0.014$; биномиальный тест). Для активированных генов также характерны более высокая достоверность обогащения процессов (рис. 1, а) и большее число генов в процессах (рис. 1, б). Кроме того, оценка усредненной экспрессии генов, входящих в процессы развития и морфогенеза, показала большую степень индукции, чем подавления (рис. 1, в).

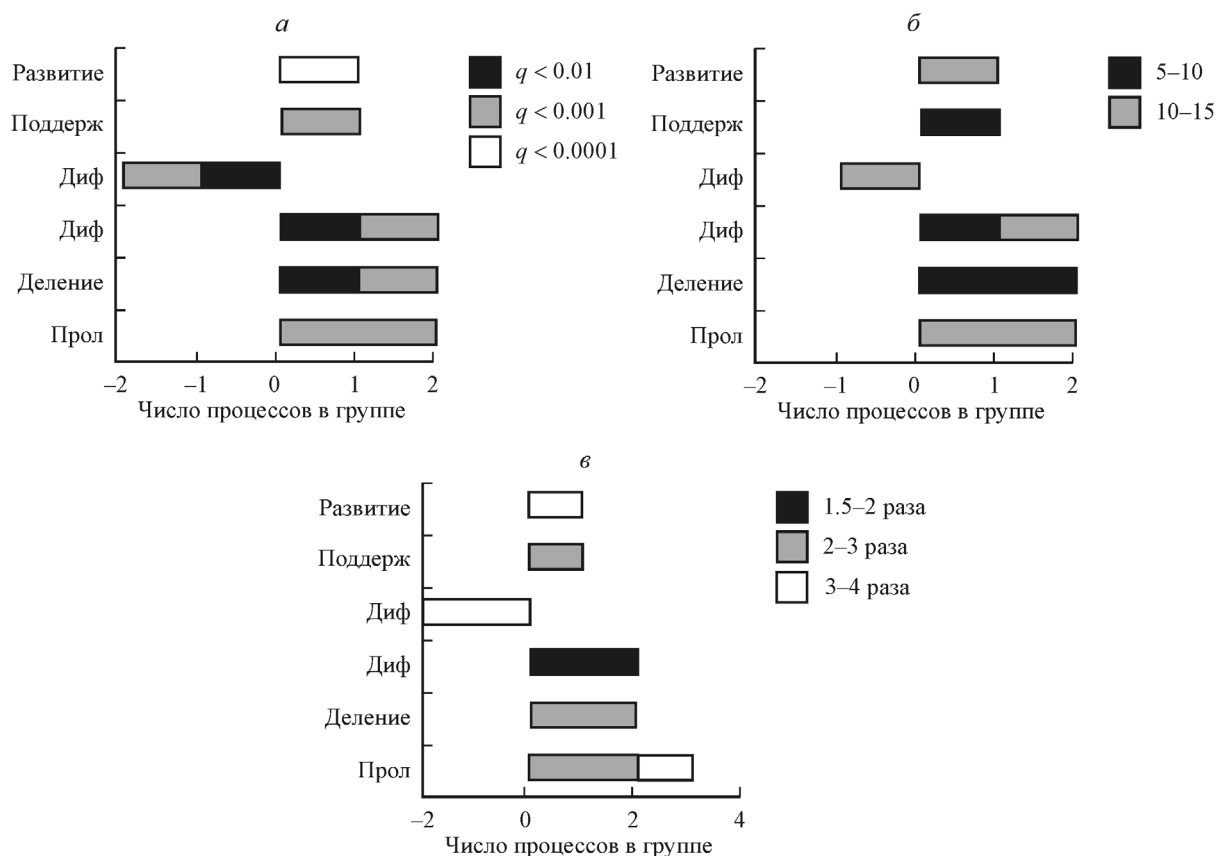


Рис. 2. Функциональные и количественные характеристики GO-процессов, относящихся к биологии стволовых клеток, которые обогащены регулируемыми плоидностью генами в сердце и печени человека и мышши.

Распределение GO-процессов: *а* — по уровню достоверности (q -value); *б* — по числу генов в процессах; *в* — по среднему уровню различий экспрессии между полиплоидными и диплоидными органами. По горизонтали — число GO-процессов, обогащенных активированными (положительное) и подавленными (отрицательное) генами; по вертикали — название групп процессов. Поддерж — поддержание, Диф — дифференцировка, Прол — пролиферация.

Самая высокая достоверность и максимальное число генов были зарегистрированы для процессов регуляции развития. Из 11 процессов 7 содержат более 160 генов, а 6 имеют $q < 10^{-6}$ (рис. 1, *а, б*). Максимальная индукция (более чем в 3.5 раза) была найдена для процессов эмбрионального развития (рис. 1, *в*).

Важно то, что для процессов дифференцировки отмечен противоположный характер изменений: ингибированные гены показывают большее число обогащенных процессов, чем индуцированные (рис. 1, *а, б*). Кроме того, для процессов, обогащенных подавленными генами, отмечены более значительные различия средней экспрессии генов между полиплоидными и диплоидными органами, чем для процессов, обогащенных активированными генами (рис. 1, *в*). Таким образом, данные о подавлении модулей дифференцировки дополняют и усиливают результаты об индукции антагонистичных модулей, включая модули развития и морфогенеза.

В настоящее время сведений, подтверждающих взаимосвязь между соматической полиплоидией и усилением эмбриональной программы, немного. Так, была установлена положительная корреляция между гиперполиплоидизацией кардиомиоцитов и проявлением эмбрионального фенотипа, видимого из индукции эмбриональных маркеров — тяжелой цепи бета-миозина, фактора HIF1- α (фактора гипоксии 1- α), увеличением числа и площади ядрышек, а также повышением ростового потенциала (Anatskaya, Vinogradov, 2002, 2004a, 2004b, 2007; Ана-

цкая и др., 2010, 2013a, 2013b; Anatskaya et al., 2010). Кроме того, есть и косвенные подтверждения проявлений эмбриональности при полиплоидизации кардиомиоцитов. С одной стороны, установлена взаимосвязь между заболеваниями сердца (гипертония, ишемическая болезнь сердца, врожденные пороки, кардиомиопатии) и избыточной полиплоидизацией кардиомиоцитов (Brodsky, Uryvaeva, 1977; Romyantsev et al., 1990; Meckert et al., 2005; Ahuja et al., 2007; Senyo et al., 2013). С другой стороны, существуют данные об эмбрионализации метаболизма, фенотипа и сократительной функции кардиомиоцитов при тех же самых заболеваниях (Razeghi et al., 2001; Reiser et al., 2001).

Модули биологии стволовых клеток (СК). Биологические процессы GO, координирующие развитие, пролиферацию и поддержание СК, обогащены только активированными плоидностью генами. В то же время процессы, связанные с дифференцировкой СК, обогащены как активированными, так и подавленными плоидностью генами примерно в равной степени (рис. 2). Самая высокая достоверность обогащения и максимальное число генов были зарегистрированы для процессов развития и пролиферации СК (рис. 2, *а, б*). Для этих процессов характерна также наиболее сильная индукция экспрессии (рис. 2, *в*).

Сеть межбелковых взаимодействий для генов модулей, относящихся к биологии СК. 1. Особенности архитектуры сети. Поскольку гены-ре-

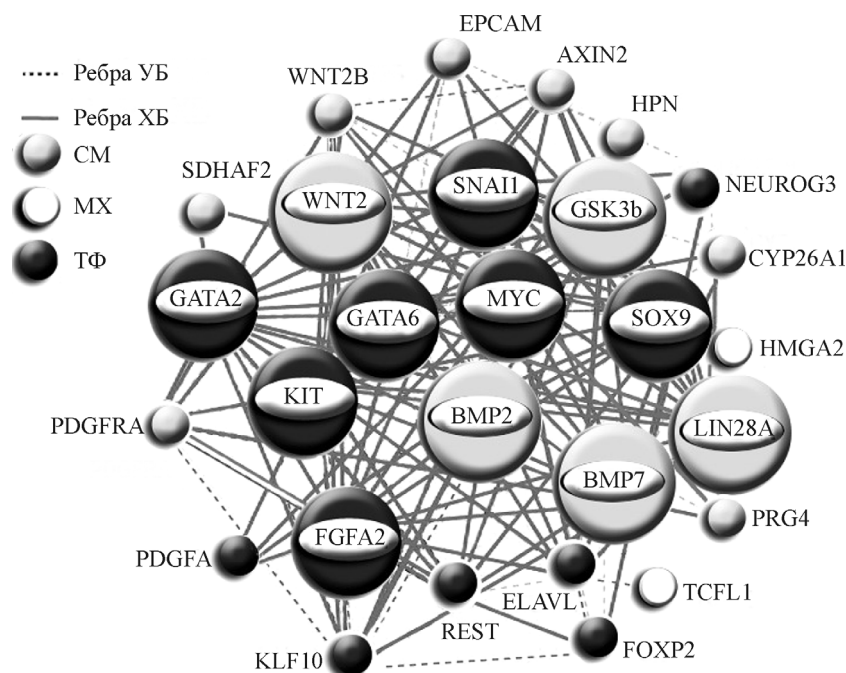


Рис. 3. Функциональная характеристика сети межбелковых взаимодействий для регулируемых плоидностью генов GO-процессов, относящихся к биологии стволовых клеток.

Крупные символы — хабовые белки (ХБ), мелкие символы — узловы белки (УБ). CM — сигнальные молекулы, MX — модуляторы хроматина, ТФ — транскрипционные факторы. Функциональные характеристики белков сети различаются интенсивностью цвета. Вероятность взаимодействия между белками $S < 0.50$ (здесь и далее S определена с помощью сервера String, см. раздел «Материал и методика»). Видно, что в сети присутствует много регуляторов, включая ТФ, CM и MX, что свидетельствует о влиянии полиплоидии на базовые механизмы регуляции программ развития.

гуляторы биологии СК играют центральную роль в координации программ развития, важно было исследовать особенности их функциональных взаимодействий. Для этого мы сконструировали сеть межбелковых взаимодействий для генов, входящих в GO-процессы регуляции биологии СК (рис. 3). Видно, что из 29 белков 28 являются узлами сети, соединенными связями. Достоверность насыщения сети связями, определенная с помощью сервера STRING (Szkarczyk et al., 2015), достигла $p < 10^{-31}$. Также видно, что сеть относительно гомогенна и не имеет скопления узлов в виде отдельных кластеров и иерархических структур. Имеется лишь тенденция к сгущению плотности связей к центру. Такой хаотический тип архитектуры характерен для сетей с динамическими свойствами, т. е. сетей, включающих в себя динамические регуляторы (Терентьев и др., 2009). Взаимодействуя с разными партнерами в разное время, такие регуляторы могут связывать между собой разные функциональные модули (Терентьев и др., 2004; Han et al., 2004; Lee et al., 2008; Vidal et al., 2011). Следует отметить, что проверка GO-модулей на наличие функциональных связей является одним из способов тестирования точности нормировки и адекватности метода отбора генов (Ideker et al., 2011). Высокая насыщенность сетей связями подтверждает, что набор тестируемых генов практически не зашумлен, а значит, метод нормировки данных и биоинформационный метод отбора генов были подобраны правильно (El-Samad, Madhani, 2011; Ideker et al., 2011; Vidal et al., 2011).

2. *Регуляторы (хабы) сети межбелковых взаимодействий.* Для того чтобы выяснить, содержит ли полученная 28-узловая сеть важные регуляторы (хабы), мы оценили число межбелковых взаимосвязей (ребер) для каждого белка (узла), а также число транскрипционных факторов и регуляторов хроматина. Наиболее высокую

плотность ребер (более 15) показали белки генов Мус, Sox9, Kit, SNAIL, FGFR2, BMP7, BMP2, WNT2 и GSK3b (рис. 3). Следует отметить, что белки, имеющие более 15 ребер, принято относить к хамам, которые связывают сеть в единое целое (Ideker et al., 2011). В связи с этим можно считать, что исследуемая сеть сильно насыщена регуляторами. Принципиальная роль хабовых генов была недавно подтверждена экспериментами по их удалению из метаболических сетей. В частности, было показано, что нокаут хабовых генов сопровождается повышением летальности (Vidal et al., 2011). В то же время исчезновение из сети обычных генов, которые имеют мало связей, не приводит к подобному эффекту (Rajarathinam, Lin, 2006; Терентьев и др., 2009). О высокой насыщенности полученной сети регуляторами экспрессии генов свидетельствуют также большое число транскрипционных факторов (12 из 28 генов сети) (*Snail*, *Gata2*, *Kit*, *Gata6*, *Klf10*, *Sox9*, *Foxp2*, *Myc*, *Pdgfa*, *Rest*, *Neurog3* и *Elavl1*) и присутствие двух модуляторов хроматина (*Hmga2* и *Tcf11*). Насыщенность сети межбелковых взаимодействий большим числом регуляторов подтверждает правильность нашей гипотезы о том, что геномные дупликации влияют на программы развития.

3. *Насыщенность сети GO-модулями.* Успешная проверка точности нормировки данных, адекватности метода и правильности гипотезы дает нам основания для детального исследования биологических функций 28-узловой сети взаимодействий белков — продуктов генов, представленных в процессах стволовости в базе GO. Для этого мы проанализировали насыщенность сети генами из биологических процессов базы GO и путями базы KEGG, которые относятся к стволовым клеткам. На рис. 3, а, б показаны модули с максимальной достоверностью обогащения генами сети. Видно, что среди GO-процессов (рис. 4,

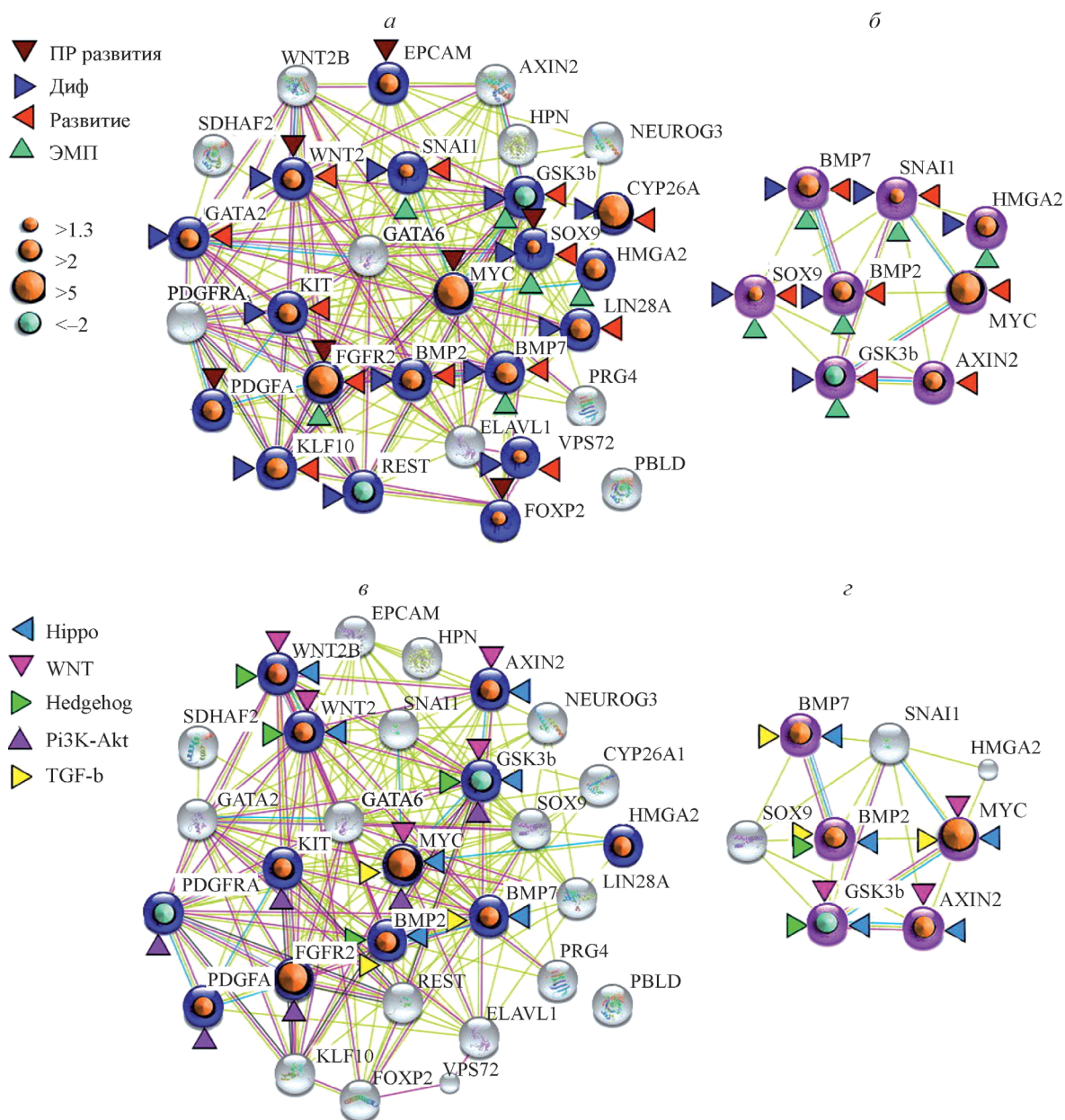


Рис. 4. Сети межбелковых взаимодействий для регулируемых плоидностью генов, входящих в модули биологии стволовых клеток.

a, в — сети межбелковых взаимодействий, построенные с достоверностью $S < 0.5$; *б, г* — функциональные ядра сетей межбелковых взаимодействий для крупных сетей (представленных на: *a* и *в* соответственно), построенные с достоверностью $S < 0.9$. Размер внутреннего кружка отражает различия экспрессии гена между полиплоидными и диплоидными органами. Показаны процессы GO и пути базы KEGG, обогащенные генами белковые сети с достоверностью q не менее 10^{-5} . Цветными треугольниками показаны модули стволовых клеток, в которые входит ген. Серые кружки — белки генов, не входящих в модули, символы с кружком внутри соответствуют белкам генов, входящих в GO-процессы (*a, б*) и пути базы KEGG (*в, г*). ПР развития — позитивная регуляция развития, Диф — дифференцировка, ЭМП — эпителиально-мезенхимный переход; Hippo, WNT, Hedgehog, Pi3-AKT и TGF-b — сигнальные пути базы KEGG с соответствующими названиями. Видно, что в функциональных ядрах сетей (*б* и *г*) сохранено обогащение GO-процессами и KEGG-путями, присутствующими в основных сетях (*a* и *в*).

a) наиболее высокую достоверность демонстрируют пути развития, дифференцировки и пролиферации стволовых клеток, а также путь эпителиально-мезенхимного перехода. Следует отметить, что практически все хабовые (узловые) белки являются общими для нескольких модулей, что свидетельствует об их сплетении и синергизме.

Максимальная активация экспрессии (в 4.2 ± 0.46 раза) была зарегистрирована для модуля пролиферации СК. Фрагмент этого пути включает в себя белки генов *Myc*, *Wnt2*, *Pdgfa*, *Fgfr2*, *Epcam*, *Fpxp2* и *Vps72*. Ин-

дукция модуля произошла главным образом из-за сильно-го повышения экспрессии гена *Myc* (более чем в 5 раз). Модули развития и дифференцировки индуцированы слабее (в 2.37 ± 0.31 и 2.41 ± 0.34 раза соответственно) и состоят практически полностью из генов, которые создают баланс между ранней специализацией клеток (белки генов *Gata2*, *Sox9*, *Bmp2*, *7*, *Gsk3b* и *Cyp26a1*) и их самообновлением (белки генов *Klf10*, *Wnt2*, *Kit*, *Lin28a*, *Fgfr2* и *Snai1*). Сходство генного состава для процессов пролиферации и дифференцировки можно объяснить тем, что для

стволовых клеток эти процессы являются разными этапами специализации. Развитие — более ранняя фаза, а дифференцировка — более поздняя (Hanna et al., 2010).

Средний уровень активации генов модуля эпителиально-мезенхимного перехода (ЭМП) составил 2.23 ± 0.28 раза (рис. 4, а). Известно, что многие регуляторы этого модуля координируют не только программы развития, но и трансформацию клеток (Thiery et al., 2009; De Craene, Berx, 2013). Во фрагменте ЭМП-модуля семь индуцированных генов (*Wnt2*, *Sox9*, *Fgfr2*, *Snail*, *Bmp2*, 7 и *Hmga2*). Известно, что гены *Wnt2*, *Sox9*, *Fgfr2*, *Snail*, *Bmp2* и 7 ассоциированы с плюрипотентностью (Puisieux et al., 2014). Ген *Snail* способствует повышению клеточной подвижности (De Craene, Berx, 2013). Ген *Hmga2* способствует активации хроматина и процессов трансформации в стволовых и нормальных клетках, а также в клетках опухолей (Wu et al., 2011; Polyanichko, Chikhirzhina, 2013; Chikhirzhina et al., 2014; Lee et al., 2015).

Среди подавленных генов сети плюрипотентности мы идентифицировали только ген *Gsk3b*, который участвует в энергетическом обмене и негативной регуляции пролиферации (Katoh et al., 2006). Ранее взаимосвязь между полиплоидными клетками и индукцией ЭМП была зарегистрирована в единственной работе, сделанной на гигантских трансформированных клетках карциномы яичников человека (Zhang et al., 2013). Взаимосвязь между полиплоидией и проявлениями признаков ЭМП может свидетельствовать о вовлеченности геномных дупликаций в начальные этапы трансформации клеток.

Тестирование сети межбелковых взаимодействий на обогащенность путями из базы KEGG (рис. 4, б) выявила достоверное насыщение сигнальными путями Hippo (в 3.8 ± 0.26 раза; гены *Bmp2*, 7, *Gsk3b*, *Axin2*, *Wnt2*, *Wnt2b* и *Myc*), Hedgehog (в 2.9 ± 0.21 раза; гены *Bmp2*, *Gsk3b*, *Wnt2* и *Wnt2b*), Wnt (в 2.4 ± 0.23 раза; гены *Gsk3b*, *Wnt2*, *Wnt2b* и *Axin2*), PI3-киназы (в 2.7 ± 0.23 раза; гены *Gsk3b*, *Myc*, *Kit*, *Pdgfra*, *Pdgfra* и *Fgfr2*) и фактора роста TGF- β (в 3.2 ± 0.26 раза; гены *Bmp2*, 7 и *Myc*). Данные из литературы свидетельствуют о том, что эти пути координируют регуляцию процессов развития и плюрипотентности (Yin, Zhang, 2011; Dalton et al., 2013; Atlasi et al., 2014).

Таким образом, преобладание активированных генов по сравнению с подавленными генами, а также преимущественная активация важных позитивных регуляторов (*Myc*, *Taz*, *Kit*, *Wnt2*, *Wnt2b* и *fgfr2*) и снижение активности негативного регулятора *Gsk3b* пути Wnt (Katoh, 2006; Atlasi et al., 2014) дополняют и подтверждают наши данные, полученные при исследовании ГО-процессов. Активацию программ эмбриональности при полиплоидизации ранее регистрировали только в трансформированных клетках и в культуре стареющих и трансформированных фибробластов (Salmina et al., 2010; Erenpreisa Cragg, 2010, 2013; Erenpreisa et al., 2014). В связи с этим представленные в настоящей работе результаты для нормальных клеток живых организмов можно считать приоритетными.

Для того чтобы выделить функциональное ядро сети, мы использовали повышенный уровень достоверности межбелковых взаимодействий, stringency — S, определенный с помощью базы String (рис. 4). В результате получилась 8-узловая сеть, представленная генами *Hmga2*, *Myc*, *Axin2*, *Gsk3b*, *Snail*, *Sox9*, *Bmp2* и 7). Тестирование субсети в отношении насыщенности ГО-процессами (рис. 4, а) и KEGG-путями (рис. 4, б) дало неожиданный результат: сохранилась обогащенность всеми важными модулями

основной сети, включая модули ЭМП, развития и дифференцировки СК, а также KEGG-путями Hippo, TGF- β и WNT. Существование плотного функционального ядра, включающего в себя общие регуляторы наиболее важных путей стволовости, свидетельствует о том, что полиплоидия может затрагивать основные механизмы регуляции программ развития, плюрипотентности и ЭМП.

З а к л ю ч е н и е. Соматическая полиплоидия ассоциирована со многими патологическими состояниями, включая канцерогенез и дегенеративные расстройства (Schoenfelder, Fox, 2015). В то же время физиологическая роль этого явления остается еще невыясненной. Не подтверждаются первоначальные предположения о том, что геномные дупликации связаны с терминальной дифференцировкой и старением (Pandit et al., 2013). Более того, данные, полученные при изучении трансформированных клеток, свидетельствуют о том, что полиплоидия способствует активации важных маркеров плюрипотентности (Erenpreisa et al., 2014). В связи с этим сведения о влиянии соматической полиплоидии на транскрипционную активность модулей развития и структуру их регуляторных сетей, полученные с помощью разработанного нами сравнительно-биоинформационного метода, помогут улучшить понимание роли полиплоидии в нормальной физиологии, трансформации и других патологических состояниях.

Авторы выражают признательность за ценные консультации И. А. Гамалей, Е. Б. Буровой, Е. Н. Попову, Б. В. Толкуновой, А. Н. Томилину.

Работа выполнена при финансовой поддержке Российского научного фонда (проект 14-50-00068).

Список литературы

- Анацкая О. В., Матвеев И. В., Сидоренко Н. В., Харченко М. В., Кропотов А. В., Виноградов А. Е. 2013а. Изменения в сердце неонатальных крыс после криптоспориоза разной степени тяжести. Журн. эволюц. биох. физиол. 49 (5) : 357—365. (Anatskaya O. V., Matveev I. V., Sidorenko N. V., Kharchenko M. V., Kropotov A. V., Vinogradov A. E. 2013a. Changes in the heart of neonatal rats after cryptosporidial gastroenteritis of different degrees of severity. J. Evol. Biochem. Physiol. 49 (5) : 509—518.)
- Анацкая О. В., Сидоренко Н. В., Бейер Т. В., Виноградов А. Е. 2010. Неонатальный гастроэнтерит как причина долговременной атрофии, деформации и необратимой гиперполиплоидизации кардиомиоцитов. Кардиология. 50 (10) : 35—44. (Anatskaya O. V., Sidorenko N. V., Beyer T. V., Vinogradov A. E. 2010. Neonatal Gastroenteritis triggers long-term cardiomyocyte atrophy, remodeling and irreversible hyperpolyploidization. Kardiologiya. 50 (12) : 35—44.)
- Анацкая О. В., Сидоренко Н. В., Матвеев И. В., Кропотов А. В., Харченко М. В., Виноградов А. Е. 2013б. Пороговый ответ кардиомиоцитов неонатальных крыс на постепенное увеличение криптоспориозной инвазии. Цитология. 55 (8) : 527—538. (Anatskaya O. V., Sidorenko N. V., Matveev I. V., Kropotov A. V., Kharchenko M. V., Vinogradov A. E. 2013b. Threshold rat neonatal cardiomyocyte response to gradual cryptosporidial infection severity increase. Tsitologiya. 55 (8) : 527—538.)
- Терентьев А. А., Молдогязиева Н. Т., Шайтан К. В. 2009. Динамическая протеомика в моделировании живой клетки. Белок-белковые взаимодействия. Биохимия. 74 (13) : 1586—1607. (Terentiev A. A., Moldogazieva N. T., Shaitan K. V. 2009. Dynamic proteomics in modeling of the living cell. Protein-protein interactions. Biochemistry. 74 (13) : 1586—1607.)

- Ahuja P., Sdek P., Mac Lellan W. R. 2007. Cardiac myocyte cell cycle control in development, disease, and regeneration. *Physiol. Rev.* 87: 521—544.
- Anatskaya O. V., Sidorenko N. V., Beyer T. V., Vinogradov A. E. 2010. Neonatal cardiomyocyte ploidy reveals critical windows of heart development. *Int. J. Cardiol.* 141: 81—91.
- Anatskaya O. V., Sidorenko N. V., Vinogradov A. E., Beyer T. V. 2007. Impact of neonatal cryptosporidial gastroenteritis on epigenetic programming of rat hepatocytes. *Cell Biol. Int.* 31: 420—427.
- Anatskaya O. V., Vinogradov A. E. 2002. Myocyte ploidy in heart chambers of birds with different locomotor activity. *J. Exp. Zool.* 293: 427—441.
- Anatskaya O. V., Vinogradov A. E. 2004a. Heart and liver as developmental bottlenecks of mammal design: evidence from cell polyploidization. *Biol. J. Linn. Soc.* 83: 175—186.
- Anatskaya O. V., Vinogradov A. E. 2004b. Paradoxical relationship between protein content and nucleolar activity in mammalian cardiomyocytes. *Genome.* 47: 565—578.
- Arkin A. P., Schaffer D. V. 2011. Network news: innovations in 21st century systems biology. *Cell.* 144: 844—849.
- Atlasi Y., Looijenga L., Fodde R. 2014. Cancer stem cells, pluripotency, and cellular heterogeneity: a WNTer perspective. *Curr. Top. Develop. Biol.* 107: 373—404.
- Brodsky W. Y., Uryvaeva I. V. 1977. Cell polyploidy: its relation to tissue growth and function. *Int. Rev. Cytol.* 50: 275—332.
- Brown G. R., Hem V., Katz K. S., Ovetsky M., Wallin C., Ermolaeva O., Tolstoy I., Tatusova T., Pruitt K.D., Maglott D.R., Murphy T.D. 2015. Gene: a gene-centered information resource at NCBI. *Nucl. Acids Res.* 43: D36—D42.
- Chikhirzhina E. V., Starkova T. J., Polyanichko A. M. 2014. Interaction between chromosomal protein HMGB1 and DNA studied by DNA-melting analysis. *J. Spectroscopy.* 1044: 167—172.
- Dalton S. 2013. Signaling networks in human pluripotent stem cells. *Curr. Opin. Cell Biol.* 25: 241—246.
- De Craene B., Berx G. 2013. Regulatory networks defining EMT during cancer initiation and progression. *Nat. Rev. Cancer.* 13: 97—110.
- Dow L. E., Lowe S. W. 2012. Life in the fast lane: mammalian disease models in the genomics era. *Cell.* 148: 1099—1109.
- El-Samad H., Madhani H.D. 2011. Can a systems perspective help us appreciate the biological meaning of small effects? *Develop. Cell.* 19 (21): 11—13.
- Erenpreisa J., Cragg M. S. 2010. MOS, aneuploidy and the ploidy cycle of cancer cells. *Oncogene.* 29: 5447—5451.
- Erenpreisa J., Cragg M. S. 2013. Three steps to the immortality of cancer cells: senescence, polyploidy and self-renewal. *Cancer Cell Int.* 13: 92.
- Erenpreisa J., Salmina K., Huna A., Jackson T. R., Azquez-Martin A., Cragg M. S. 2014. The «virgin birth», polyploidy, and the origin of cancer. *Oncoscience.* 2: 3—14.
- Ganem N. J., Storchova Z., Pellman D. 2007. Tetraploidy, aneuploidy and cancer. *Curr. Opin. Genet. Develop.* 17: 157—162.
- Gene Ontology Consortium. 2015. Gene ontology consortium: going forward. *Nucl. Acids Res.* 43: D1049—D1056.
- Han J. D., Bertin N., Hao T., Goldberg D. S., Berriz G. F., Zhang L. V., Dupuy D., Walhout A. J., Cusick M. E., Roth F. P., Vidal M. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature.* 430: 88—93.
- Hanna J. H., Saha K., Jaenisch R. 2010. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell.* 143: 508—525.
- Ideker T., Dutkowski J., Hood L. 2011. Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell.* 144: 860—863.
- Jones M. R., Ravid K. 2004. Vascular smooth muscle polyploidization as a biomarker for aging and its impact on differential gene expression. *J. Biol. Chem.* 279: 5306—5313.
- Kanehisa M., Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* 28: 27—30.
- Katoh M. 2006. Cross-talk of WNT and FGF signaling pathways at GSK3b to regulate beta-catenin and SNAIL signaling cascades. *Cancer Biol. Ther.* 9: 1059—1064.
- Lee D. S., Park J., Kay K. A., Christakis N. A., Oltvai Z. N., Barabasi A. L. 2008. The implications of human metabolic network topology for disease comorbidity. *Proc. Nat. Acad. Sci. USA.* 105: 9880—9885.
- Lee H. O., Davidson J. M., Duronio R. J. 2009. Endoreplication: polyploidy with purpose. *Genes Develop.* 23: 2461—2477.
- Lee J., Ha S., Jung C. K., Lee H. H. 2015. High-mobility-group A2 overexpression provokes a poor prognosis of gastric cancer through the epithelial-mesenchymal transition. *Int. J. Oncol.* 46: 2431—2438.
- Ma X., Gao F., Rusie A., Hemingway J., Ostmann A. B., Sroga J. M., Jegga A. G., Das S. K. 2011. Decidual cell polyploidization necessitates mitochondrial activity. *PLoS ONE.* 6: e26774.
- Meckert P. C., Rivello H. G., Vigliano C., Gonzalez P., Favalloro R., Laguens R. 2005. Endomitosis and polyploidization of myocardial cells in the periphery of human acute myocardial infarction. *Cardiovasc. Res.* 67: 116—123.
- Nurse P., Hayles J. 2011. The cell in an era of systems biology. *Cell.* 144: 850—854.
- Pandit S. K., Westendorp B., de Bruin A. 2013. Physiological significance of polyploidization in mammalian cells. *Trends Cell Biol.* 23: 556—566.
- Pearson W. 2004. Finding protein and nucleotide similarities with FASTA. *Curr. Protoc. Bioinform.* Chapter 3: Unit 3.9.
- Polyanichko A., Chikhirzhina E. 2013. Interaction between DNA and chromosomal proteins HMGB1 and H1 studied by IR/VCD spectroscopy. *J. Mol. Struct.* 1044: 167—172.
- Powell S., Szklarczyk D., Trachana K., Roth A., Kuhn M., Muller J., Arnold R., Rattei T., Letunic I., Doerks T., Jensen L. J., von Mering C., Bork P. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucl. Acids Res.* 40: D284—D289.
- Pruitt K. D., Brown G. R., Hiatt S. M., Thibaud-Nissen F., Astashyn A., Ermolaeva O., Farrell C. M., Hart J., Landrum M. J., McGarvey K. M., Murphy M. R., O'Leary N. A., Pujar S., Rajput B., Rangwala S. H., Riddick L. D., Shkeda A., Sun H., Tamez P., Tully R. E., Wallin C., Webb D., Weber J., Wu W., DiCuccio M., Kitts P., Maglott D.R., Murphy T. D., Ostell J. M. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 42: D756—D763.
- Puisieux A., Brabletz T., Caramel J. 2014. Oncogenic roles of EMT-inducing transcription factors. *Nat. Cell Biol.* 16: 488—494.
- Rajarathinam T., Lin Y. H. 2006. Topological properties of protein-protein and metabolic interaction networks of *Drosophila melanogaster*. *Gen Proteom. Bioinform.* 4: 80—89.
- Razeghi P., Young M. E., Alcorn J. L., Moravec C. S., Frazier O. H., Taegtmeyer H. 2001. Metabolic gene expression in fetal and failing human heart. *Circulation.* 104: 2923—2931.
- Reiser P. J., Portman M. A., Ning X. H., Schomisch Moravec C. 2001. Human cardiac myosin heavy chain isoforms in fetal and failing adult atria and ventricles. *Amer. J. Physiol. Heart Circ. Physiol.* 280: H1814—H1820.
- Rumyantsev P. P., Erokhina I. L., Antipanova E. M., Martynova M. G. 1990. DNA and sex chromatin content in nuclei of conductive system and working myocytes of normal and hypertrophied human heart. *Acta Histochem. Suppl.* 39: 225—237.
- Salmina K., Jankevics E., Huna A., Perminov D., Radovica I., Klymenko T., Ivanov A., Jascenko E., Scherthan H., Cragg M., Erenpreisa J. 2010. Up-regulation of the embryonic self-renewal network through reversible polyploidy in irradiated p53-mutant tumour cells. *Exp. Cell Res.* 316: 2099—2112.
- Schoenfelder K. P., Fox D.T. 2015. The expanding implications of polyploidy. *J. Cell Biol.* 209: 485—491.
- Senyo S. E., Steinhilber M. L., Pizzimenti C. L., Yang V. K., Cai L., Wang M., Wu T. D., Guerquin-Kern J. L., Lechene C. P., Lee R. T. 2013. Mammalian heart renewal by pre-existing cardiomyocytes. *Nature.* 493: 433—436.
- Storey J. D., Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc. Nat. Acad. Sci. USA.* 100: 9440—9445.

- Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A., Santos A., Tsafou K. P., Kuhn M., Bork P., Jensen L. J., von Mering C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucl. Acids Res.* 43 : D447—D452.
- Thiery J. P., Acloque H., Huang R. Y., Nieto M. A. 2009. Epithelial-mesenchymal transitions in development and disease. *Cell.* 139 : 871—890.
- Tyagi S. 2015. Tuning noise in gene expression. *Mol. Syst. Biol.* 5 : 805.
- Vidal M., Cusick M. E., Barabasi A. L. 2011. Interactome networks and human disease. *Cell.* 144 : 986—998.
- Vinogradov A. E., Anatskaya O. V. 2007. Organismal complexity, cell differentiation and gene expression: human over mouse. *Nucl. Acids Res.* 35 : 6350—6356.
- Vinogradov A. E. 2015. Consolidation of slow or fast but not moderately evolving genes at the level of pathways and processes. *Gene.* 561 : 30—34.
- Waterhouse R. M., Zdobnov E. M., Tegenfeldt F., Li J., Kriventseva E. V., 2011. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. *Nucl. Acids Res.* 39 : D283—D288.
- Wu C., Macleod I., Su A. I. 2013. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucl. Acids Res.* 41 : D561—D565.
- Wu J., Liu Z., Shao C., Gong Y., Hernando E., Lee P., Narita M., Muller W., Liu J., Wei J. J. 2011. HMGA2 overexpression-induced ovarian surface epithelial transformation is mediated through regulation of EMT genes. *Cancer Res.* 71 : 349—359.
- Yin M., Zhang L. 2011. Hippo signaling: a hub of growth control, tumor suppression and pluripotency maintenance. *J. Genet. Genomics.* 38 : 471—781.
- Zhang S., Mercado-Urbe I., Hanash S., Liu J. 2013. iTRAQ-based proteomic analysis of polyploid giant cancer cells and budding progeny cells reveals several distinct pathways for ovarian cancer development. *PLoS ONE.* 8 : e80120.

Поступила 25 VI 2015

PAIRWISE CROSS-SPECIES TRANSCRIPTOME ANALYSIS OF POLYPLOIDY-ASSOCIATED EXPRESSION CHANGES OF DEVELOPMENTAL GENE MODULES

O. V. Anatskaya,^{1,*} E. A. Erenpreisa,² N. N. Nikolsky,¹ A. E. Vinogradov¹

¹ Institute of Cytology RAS, St. Petersburg, 194064,
and ² Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia;
* e-mail: olga.anatskaya@gmail.com

Design and development of highly sensitive method bioinformatics are important for investigation of casual relationships between epigenetic changes and gene activity. Cell polyploidy may trigger such changes. However, maintaining the balance of gene dosage, polyploidy may provide only a rather weak effect on their expression. Currently, there is no comprehensive and concordant data in regard to ploidy-associated transcriptomic changes. To find out how polyploidy affects gene activity, we have developed an integrative bioinformatic method of pairwise cross-species transcriptome analysis of mammalian tissues with various polyploidy degrees. The main benefit of this approach is its ability to separate species- and tissue-specific noises of evolutionary conserved effects. We demonstrate the application of the method for the analysis of gene modules and protein interactions networks coordinating programs of development, differentiation and pluripotency. The analysis was performed with transcriptomes of polyploid and diploid organs (human and mouse heart and liver). Our data indicate that ploidy-induced genes enrich Gene Ontology (GO) biological processes and KEGG pathways related to development, morphogenesis and stem cells biology (including Hippo, Pi3K, WNT, Hedgehog and TGF- β pathways) with higher degree than ploidy-inhibited genes. Thus, our data are the first to show that polyploidy may induce and coordinate developmental modules.

Key words: comparative transcriptome analysis, protein interaction networks, gene modules, polyploidy, development, differentiation, stem cell, fetal and embryonic phenotype.