

МЕТОДЫ ПОЛУЧЕНИЯ КОЛИЧЕСТВЕННЫХ ДАННЫХ ИЗ КОНФОКАЛЬНЫХ ИЗОБРАЖЕНИЙ КАРТИН ЭКСПРЕССИИ ГЕНОВ IN SITU

© С. Ю. Суркова,^{1,*} Е. М. Мясникова,¹ К. Н. Козлов,¹ А. А. Самсонова,²
Дж. Райниц,³ М. Г. Самсонова¹

¹ Отдел компьютерной биологии, Центр перспективных исследований
С.-Петербургского государственного политехнического университета,

² Кафедра генетики Медицинской школы Гарвардского университета, Бостон, США и

³ Кафедра прикладной математики и статистики и Центр генетики развития
Университета Стони Брука, Стони Брук, Нью-Йорк, США;

* электронный адрес: sestr_sve@mail.ru

В обзоре обобщены ранее разработанные нами методы получения количественной информации по экспрессии генов на основе конфокальных изображений. Эти методы включают в себя сегментацию изображений, получение количественных данных в виде текстовых таблиц, удаление неспецифического фонового сигнала и пространственную регистрацию картин экспрессии генов. Последовательное применение перечисленных процедур в конечном итоге позволяет построить атлас экспрессии генов во времени и пространстве путем усреднения данных на каждой из исследуемых стадий развития. Методы были изначально разработаны для извлечения количественных данных из изображений картин экспрессии генов сегментации у *Drosophila melanogaster*. Использование этих методов позволило авторам выявить новые важные закономерности формирования доменов экспрессии этих генов, а также построить пространственно-временной атлас экспрессии генов сегментации. Большая часть процедур достаточно универсальна и может с небольшими модификациями применяться для обработки широкого спектра биологических изображений.

Ключевые слова: методы обработки изображений, конфокальная микроскопия, количественная экспрессия генов, сегментация изображений, пространственная регистрация, удаление фонового сигнала.

Успехи молекулярной генетики эукариотических организмов привели к расшифровке структуры кодирующих областей генов, идентификации их контролирующих районов, содержащих промоторы, энхансеры и сайты присоединения транскрипционных факторов, а также к выявлению транскрипционных факторов, присоединяющихся к этим районам. Вместе с мРНК и белковыми продуктами все вышеперечисленные генетические элементы образуют генные регуляторные сети. Эти сети осуществляют координацию экспрессии генов, отвечающих за спецификацию клеток, дифференциацию тканей и в конечном итоге за онтогенез организма.

Понимание принципов организации, функционирования и эволюции генных регуляторных сетей требует детального количественного описания динамики каждого из их компонентов. Быстрое развитие экспериментальной базы в области молекулярной биологии позволяет успешно решать эту задачу. Например, для количественной оценки уровня экспрессии генов широко используется технология чипов ДНК (Banerjee et al., 2002), а для оценки скорости диффузии или транспорта молекул внутри клетки — метод восстановления флуоресценции после фото-выжигания (FRAP) (Shav-Tal, 2006).

Следует отметить, что, несмотря на высокую разрешающую способность, метод ДНК-чипов, так же как и

многие другие методы количественной оценки экспрессии генов (количественный PCR, CAT assays), имеет ограниченное применение при исследовании процессов развития. Дело в том, что все эти методы используют гомогенаты клеток и, таким образом, теряют информацию об экспрессии генов в пространстве.

В то же время на ранних стадиях эмбриогенеза детерминация клеток и формирование паттернов происходят в относительно небольших морфогенетических полях, в которых появление незначительных различий в пространственной экспрессии малого количества генов предшествует образованию клеток разного типа (Gilbert, 2003). В частности, у дрозофилы детерминация сегментов происходит в морфогенетическом поле, имеющем длину менее 0.3 мм вдоль переднезадней оси эмбриона и представляющем собой ту часть эмбриона на стадии синцитиальной бластодермы, из которой в будущем разовьется зародышевая полоска. Поэтому изучение механизмов детерминации требует получения детальной информации о пространственно-временной экспрессии генов in situ.

К настоящему времени разработано очень много методов для выявления экспрессии генов в живых и фиксированных биологических объектах как на уровне мРНК, так и на уровне белка. Одним из основных методов выявления биологических макромолекул является их имму-

нофлуоресцентное маркирование. Этот метод высокоспецифичен, поскольку основан на взаимодействии антител, меченных флуоресцентными красителями, с соответствующими им антигенами и часто используется для обнаружения и локализации антигенов в фиксированных клетках и тканях. Для регистрации сигнала в клетках и тканях используется флуоресцентная микроскопия, современные технологии позволяют применять для этого лазерный сканирующий конфокальный микроскоп. Преимущества конфокального микроскопа прежде всего состоят в том, что на детектор поступает сигнал только из фокальной плоскости и флуоресценция из близлежащих слоев исследуемого объекта не фиксируется. Поэтому результирующие изображения получаются более контрастными и менее размытыми. Кроме того, конфокальный микроскоп позволяет получать качественные цифровые изображения, готовые для дальнейшей компьютерной обработки.

В настоящее время конфокальная микроскопия получает все большее распространение, все более востребованными становятся различные и коммерческие, и некоммерческие средства обработки конфокальных изображений. Мы представляем обзор разработанных нами методов обработки конфокальных изображений картин (паттернов) экспрессии генов сегментации у плодовой мушки дрозофилы (Kozlov et al., 2000, 2002; Мясникова и др., 2001; Myasnikova et al., 2001, 2005; Janssens et al., 2005). Эти методы нацелены на получение количественных данных по экспрессии этих генов и создание единого атласа экспрессии генов. В конечном итоге это позволяет реконструировать с точностью до отдельной клетки экспрессию всех генов данной генной сети во времени и пространстве. Большая часть процедур достаточно универсальна и может быть с небольшими модификациями применена к извлечению количественных данных из изображений картин экспрессии генов у других организмов.

Степень детализации описания методов в обзоре достаточна для того, чтобы читатель мог сделать выводы о возможности их адаптации для других объектов, в то же время с детальной информацией об алгоритмах обработки можно ознакомиться в оригинальных публикациях.

Метод получения конфокальных изображений

Экспериментальные данные. Для получения информации об уровнях экспрессии генов применяли метод непрямого иммунофлуоресцентного маркирования (Dequin et al., 1984; Frasch et al., 1987), т. е. были последовательно использованы первые и вторые антитела. Примерно 1600 эмбрионов *Drosophila melanogaster* Oregon-R были зафиксированы и инкубированы с первыми антителами (Kosman et al., 1998). Использовались антитела против белковых продуктов следующих генов, контролирующих сегментацию у дрозофилы: материнских генов *bicoid* (*bcd*) и *caudal* (*cad*), ген генов *Kruppel* (*Kr*), *knirps* (*kni*), *giant* (*gt*), *hunchback* (*hb*) и *tailless* (*tll*), а также pair-rule генов *even-skipped* (*eve*), *fushi tarazu* (*ftz*), *hairy* (*h*), *runt* (*run*), *odd-skipped* (*odd*), *paired* (*prd*) и *sloppy-paired* (*slp*). Эти антитела были получены нами (Kosman et al., 1998), за исключением кроличьих антител против белка Eve (Azriaizu, Frasch, 1993). Для части эмбрионов были использованы вторые антитела, конъюгированные с FITC, Texas

Red или Cy5 (Jackson Labs., США), для остальных — конъюгированные с Alexa Fluor 488, 555, 647 или 700 (Molecular Probes, США) (Janssens et al., 2005).

Каждый эмбрион метили на экспрессию *eve* и двух других генов сегментации. Примерно половину эмбрионов дополнительно метили антителами, специфичными к гистонам H1-H4 (Janssens et al., 2005). Были использованы эмбрионы, начиная с цикла дробления 10 и заканчивая началом гастрюляции. Для сканирования брали только латерально ориентированные эмбрионы.

Конфокальную микроскопию осуществляли, как описано ранее (Kosman et al., 1997; Janssens et al., 2005; Myasnikova et al., 2005). Эмбрионы сканировали с помощью конфокального микроскопа Leica TCS4D, используя иммерсионный объектив $15 \times / 0.50$ (Kosman et al., 1997) или 20×0.70 Plan Apo системы Leica TCS SP2 (Janssens et al., 2005). Для каждого эмбриона в каждом из четырех каналов микроскопа сканировали два оптических среза, отстоящих друг от друга на 1 мкм. Для уменьшения уровня шума в регистрируемом сигнале каждый из двух оптических срезов сканировали 16 раз, после чего полученные изображения усредняли. Таким образом, каждый пиксель результирующего изображения эмбриона, получаемого в одном канале, представлял собой среднее из 32 измерений. Всего для каждого эмбриона были получены 3 результирующих изображения (по 1 в каждом канале). Эти изображения имеют размер 1024×1024 пикселя, а для записи сигнала использовали формат 8 бит. Параметры конфокального микроскопа «gain» и «offset» настраивали таким образом, чтобы пиксели, лежащие вне эмбриона, имели бы значение, равное 0, а пиксели, имеющие наибольшую яркость, равнялись бы 255. Пиксели с максимальной яркостью выбирали следующим образом: моменту развития, на котором данный ген сегментации имеет максимальную интенсивность, соответствует характерная картина экспрессии этого гена. Таким образом, опытный экспериментатор может найти на слайде для каждого гена эмбрион с картиной экспрессии, соответствующей максимуму интенсивности, и использовать небольшое количество наиболее ярких пикселей этого паттерна для настройки фотоумножителя. Очевидно, что такая настройка позволяет достаточно точно оценить количественно (в отн. ед.) уровни экспрессии одного гена, однако она не позволяет сравнивать значения уровней экспрессии разных генов.

Распределение эмбрионов по времени развития. Для реконструкции временной динамики экспрессии генов на основе индивидуальных фиксированных эмбрионов было необходимо их классифицировать по времени развития. Для циклов дробления 10—13 использовали простой подсчет ядер в текстовых файлах, содержащих количественные данные по экспрессии, полученных путем сегментации изображений (см. следующий раздел). Так как интерфазы этих циклов составляют только 6—14 мин, этого было достаточно, чтобы получить временную шкалу для этого периода. Цикл 14A намного дольше и составляет примерно 50 мин, поэтому эмбрионы внутри этого цикла были разделены на 8 эквивалентных по возрасту временных классов на основе динамики экспрессии гена *eve* (рис. 1).

Практическим критерием эквивалентности временного класса по возрасту является неспособность опытного наблюдателя различить картины экспрессии генов у эмбрионов, принадлежащих к одному классу. Поскольку эмбрионы были сканированы без учета их возраста и все 8

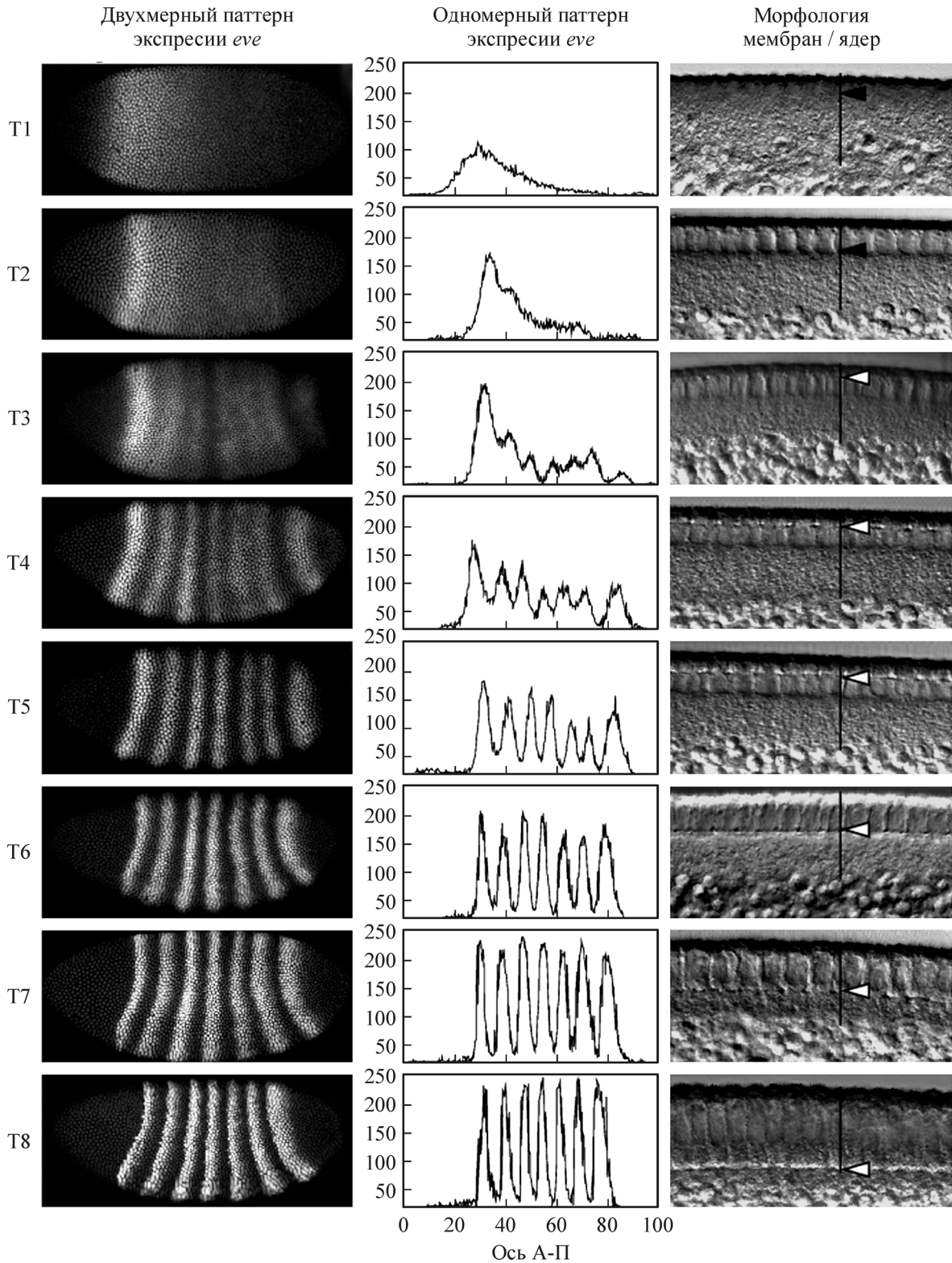
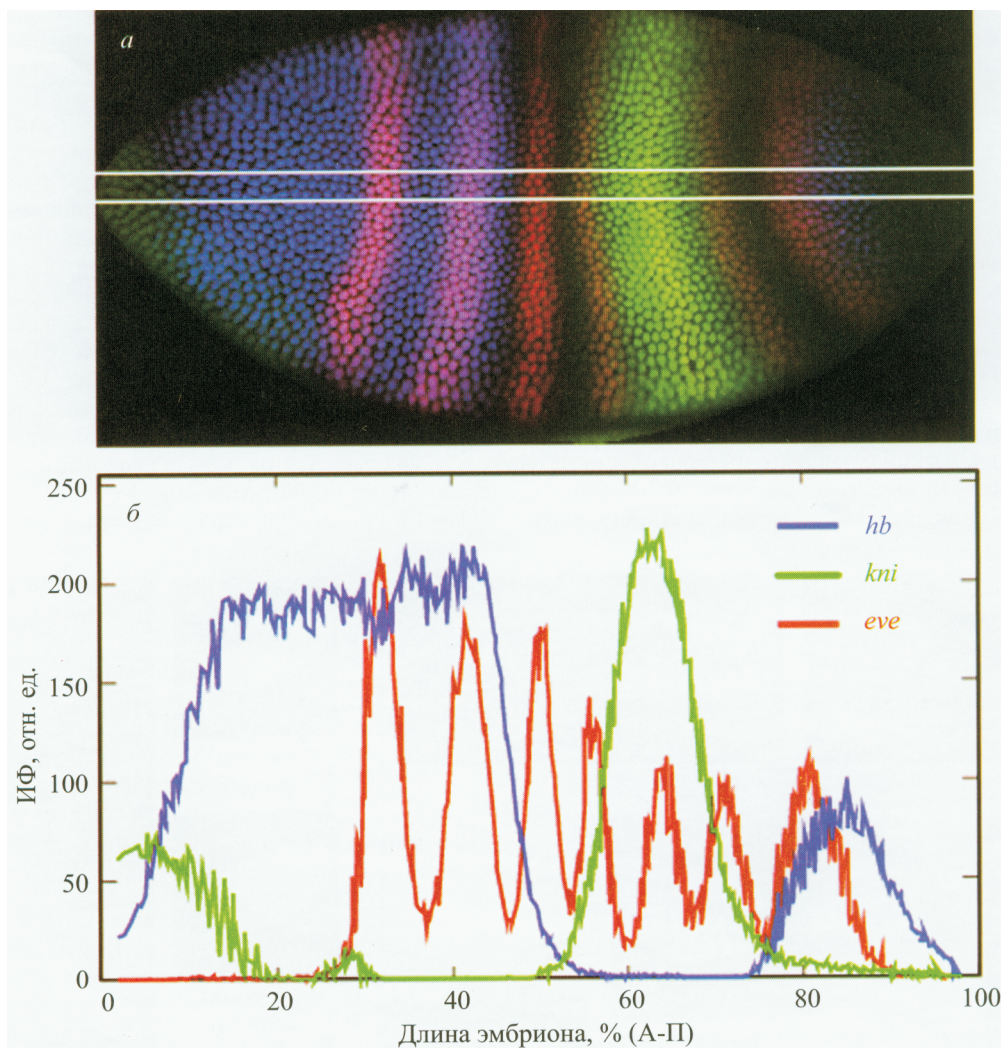


Рис. 1. Разделение эмбрионов дрозофилы цикла 14А на 8 эквивалентных по возрасту временных классов.

Для типичных эмбрионов каждого из классов 1—8 представлены изображения картины экспрессии *eve* (левый ряд), одномерные картины экспрессии этого гена (центральный ряд) и изображения морфологии бластодермы, полученные с дорсальной стороны тех же эмбрионов в режиме ДИК (дифференциально-интерференционного контраста; правый ряд); черной стрелкой показано изменение формы ядер (круглой на удлиненную), белой — фронт растущей мембраны, вертикальной черной линией — размер кортекса. Все изображения, представленные на рисунке, имеют одинаковую ориентацию относительно anteriорно-постериорной оси (А-П).



временных классов содержали приблизительно одинако-

Ориентация эмбрионов, одномерные кар-

Рис. 2. Извлечение одномерных количественных данных из изображений картин экспрессии генов сегментации у дрозофилы. *а* — эмбрион дрозофилы цикла 14А, меченный на экспрессию генов *hb*, *kni* и *eve*, обозначенных соответственно синим, зеленым и красным цветами; белыми линиями выделена центральная 10%-ная полоса. *б* — одномерные картины экспрессии трех генов, показанных на рис. 2, *а*. ИФ интенсивность флуоресценции.

вое количество эмбрионов, разумно предположить равномерное распределение возраста в выборке эмбрионов, принадлежащих к циклу 14А, а также считать длительность каждого временного класса примерно равной 6.5 мин развития (Myasnikova et al., 2001; Surkova et al., 2008).

Как видно на рис. 1, классификация эмбрионов, основанная на анализе динамики картин экспрессии гена *eve*, хорошо согласуется со степенью инвагинации мембран — морфологическим маркером, используемым для определения возраста эмбрионов. Во временном классе 1 ядра эмбриона имеют круглую форму, во временном классе 2 форма ядер становится удлинённой, а инвагинация мембран начинается во временном классе 3 и заканчивается к концу цикла 14А. Данные по морфологии blastoderмы были получены в режиме дифференциально-интерференционного контраста (ДИК) для 120 эмбрионов в системе Leica TCS4D и для всех эмбрионов в системе Leica TCS SP2 (рис. 1).

тины экспрессии. Изображения картин экспрессии *eve* на рис. 1 (левый ряд) были стандартно ориентированы относительно антериорно-постериорной (А-П) и дорсально-вентральной (Д-В) осей так, что передний (антериорный) конец эмбриона находился слева, а спинная (дорсальная) часть — сверху. Для получения одномерных данных по экспрессии генов (рис. 1, центральный ряд; рис. 2) брали только координаты ядер из центральной 10%-ной полосы (45—55 %) по оси Д-В (оси *Y*), вырезанной в направлении А-П (оси *X*) (рис. 2, б). Таким образом, у каждого ядра сохраняется лишь его *X*-координата, в то время как *Y*-координата отбрасывается и паттерны представляются в виде графиков, отражающих изменение экспрессии генов вдоль оси *X*.

Доступ к данным. Все изображения и количественные данные содержатся в базе данных FlyEx (<http://urchin.spbcas.ru/flyex/>; <http://flyex.ams.sunysb.edu/FlyEx/>), доступной по сети Интернет (Poustelnikova et al., 2004).

Методы обработки изображений

1. *Сегментация изображений.* Для изучения динамики процесса формирования паттерна необходима детальная информация о характере экспрессии во времени и пространстве всех регуляторов морфогенетического поля, в котором происходит этот процесс. Доступные в настоящее время крупные коммерческие пакеты (например, VisiQuest компании Accusoft Corp., США, и MatLab компании MathWorks, США) и их свободные аналоги (SCIRun, и TiViPe, Япония) не всегда представляют все методы обработки изображений, необходимые для количественной оценки пространственной экспрессии генов. Приведенный оригинальный метод сегментации был разработан на основе конфокальных изображений картин экспрессии генов в эмбрионе дрозофилы (Janssens et al., 2005). Этот метод позволяет не только выделить анализируемые объекты (отдельные ядра) одновременно у нескольких изображений, полученных в разных каналах микроскопа, но и в результате извлекать количественную информацию об экспрессии генов в виде текстовых таблиц. Для получения

количественных данных об активности генов в каждом ядре эмбриона дрозофилы необходимо решить следующие задачи: 1) привести экспериментальные изображения в стандартную ориентацию (см. предыдущий раздел); 2) выделить область, занимаемую объектом на изображении, и отсечь пустые области по краям; 3) построить «ядерную маску», т. е. бинарное изображение, в котором «включены» только пиксели, соответствующие ядрам в эмбрионе; 4) вычислить координаты центроидов ядер и среднюю интенсивность флуоресценции в каждом ядре для каждого канала микроскопа, т. е. для каждого сканированного белка (Kosman et al., 1997; Janssens et al., 2005). Большая часть из приведенных ниже операций является стандартной и может быть использована для обработки изображений других биологических объектов.

Приведение изображений к стандартной ориентации. Вначале для каждого из трех генных продуктов, регистрируемых у одного эмбриона, строится одно изображение на основе усреднения двух оптических срезов (рис. 3). Это делается для того, чтобы обеспечить высокий уровень сигнала в максимально возможном чис-

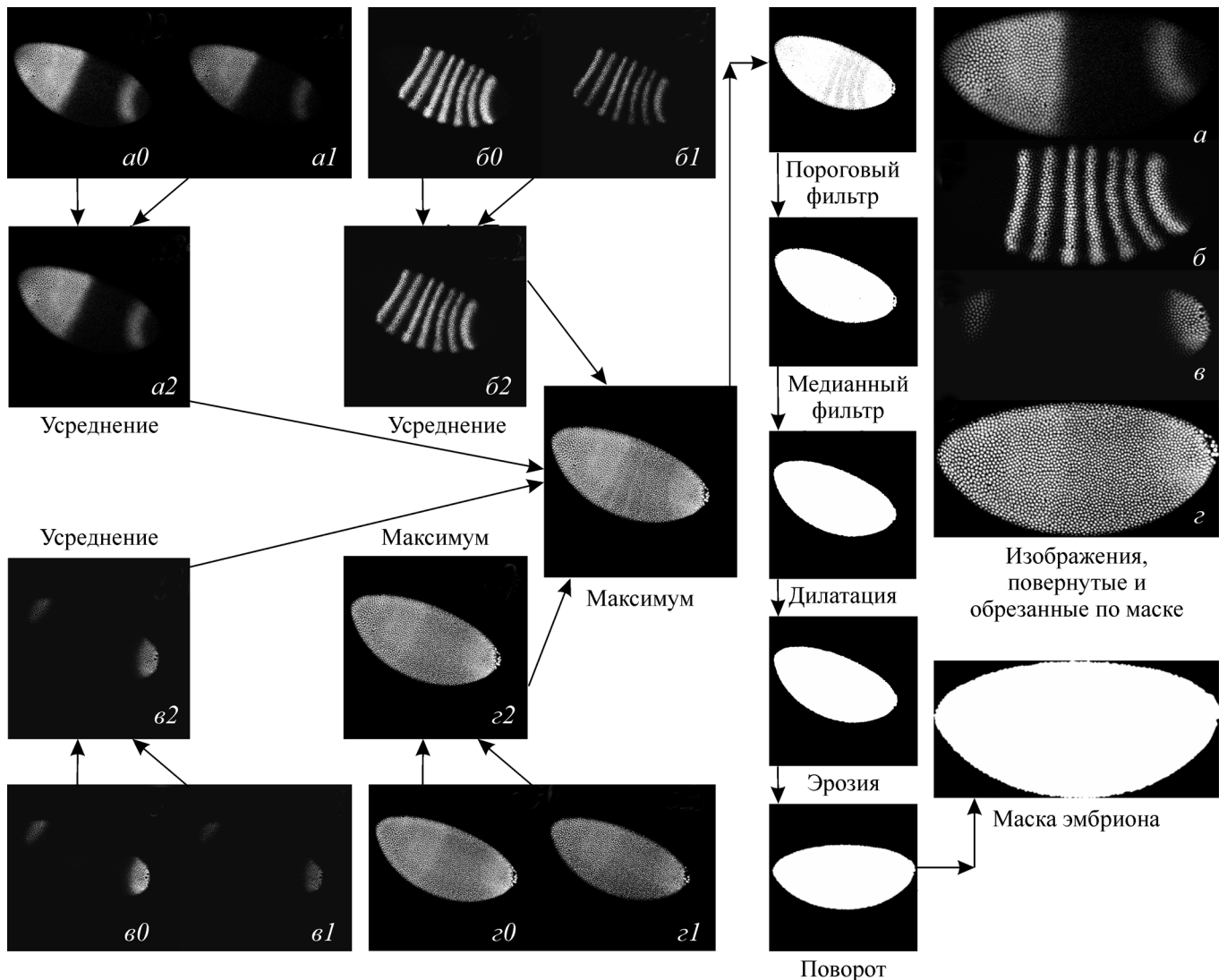


Рис. 3. Предварительная обработка и приведение конфокальных изображений эмбрионов дрозофилы к стандартной ориентации.

$a_0, a_1, b_0, b_1, v_0, v_1$ — отдельные оптические z-срезы для картин экспрессии генов a – v ; a_2 – v_2 — изображения, полученные усреднением двух срезов для каждого гена; z_0, z_1 — отдельные z-срезы для картин экспрессии ядерного маркера (гистона); z_2 — изображение пиксельного максимума для этих срезов.

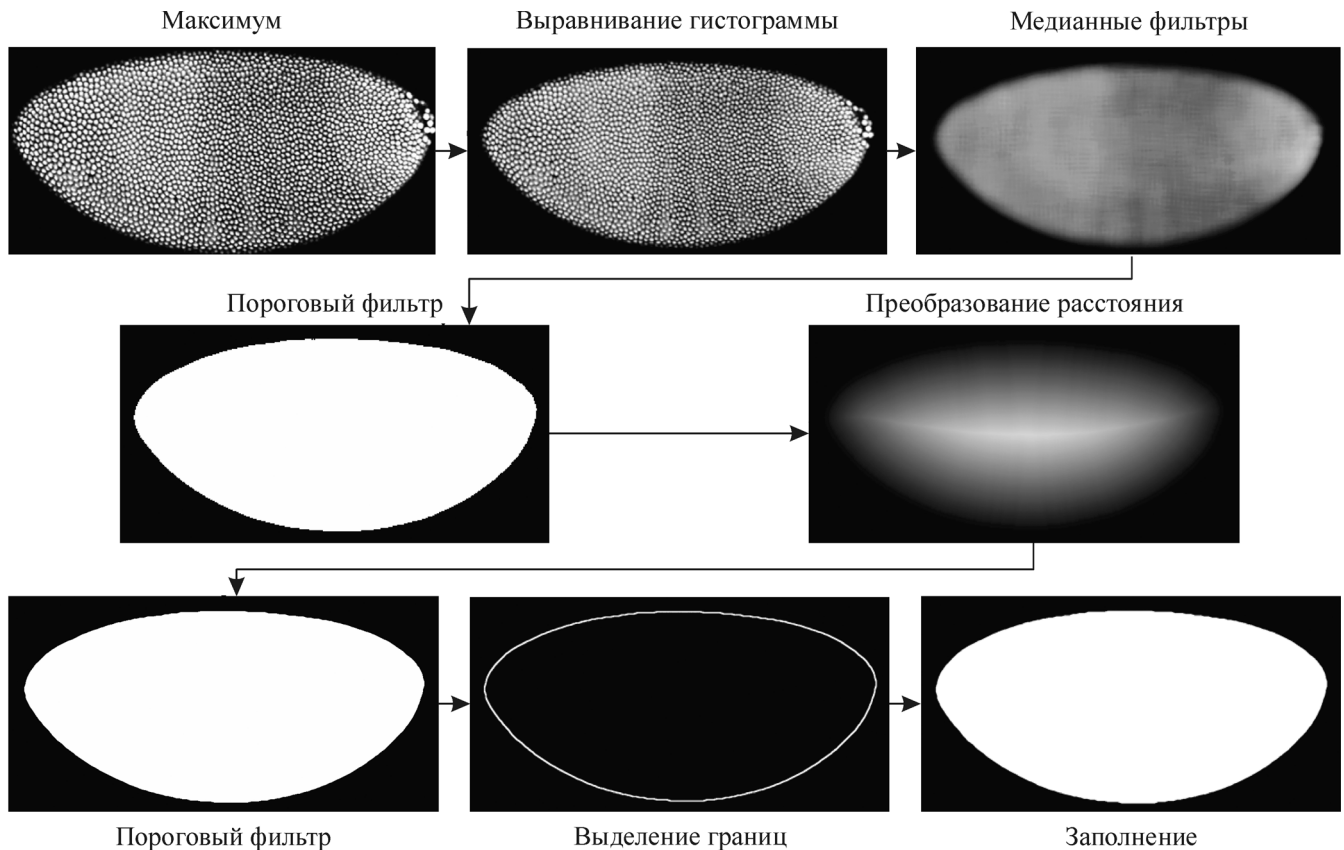


Рис. 4. Построение гладкой маски эмбриона.

ле ядер, включая ядра, расположенные на краях эмбриона. Для изображений, прокрашенных на гистоновые белки, используемые в качестве ядерных маркеров, на основе этих двух срезов строится пиксельный максимум (рис. 3), и подобное усиление сигнала на последующих этапах обработки способствует лучшему качеству построения ядерной маски (рис. 5, 6).

С целью вычисления угла поворота эмбриона для приведения его в стандартную ориентацию (см. выше), а также для удаления ненулевых пикселей, находящихся вне территории эмбриона, требуется построение первичной (грубой) маски всего эмбриона. Для этого создается пиксельный максимум из четырех изображений, полученных в разных каналах микроскопа. Далее к этому изображению применяют пороговый и медианный фильтры, а также несколько циклов эрозии и дилатации (Gonzales, Woods, 2002; Гонсалес, Вудс, 2005) (рис. 3). Маска всего эмбриона представляет собой бинарное изображение, у которого только пиксели, находящиеся в области, занимаемой эмбрионом, имеют значение 1. Угол поворота эмбриона вычисляется с помощью инвариантов моментов маски (Hu, 1962). Маска всего эмбриона и изображения картин экспрессии генов в четырех каналах поворачивается на этот угол, а неинформативные области по краям отбрасываются (рис. 3). В результате изображения картин экспрессии оказываются обрезанными по размерам маски (Kosman et al., 1997; Janssens et al., 2005).

Построение гладкой маски. Завершающим подготовительным этапом перед сегментацией изображений является создание гладкой маски всего эмбриона, которая в точности повторяет форму эмбриона (рис. 4). Эта процедура осуществляется после поворота и обрезания

изображений, так как структурные элементы (Gonzales, Woods, 2002), используемые при этом, чувствительны к ориентации изображений. Вначале строится новое изображение пиксельного максимума на основе предварительно повернутых и обрезанных изображений. Затем к полученному изображению применяются выравнивание гистограммы для максимального усиления контраста, а также медианный фильтр, осуществляющий сглаживание и уменьшение шума. В конечном итоге с помощью порогового фильтра получается бинарное изображение, и к нему применяется трансформация эвклидова расстояния. Эта процедура создает изображение в градации серого, где каждый пиксель маски имеет значение, соответствующее его эвклидову расстоянию от границы. При последующем применении медианных и порогового фильтров это приводит к созданию маски с очертаниями, более соответствующими натуральным контурам эмбриона. С помощью алгоритма Шена—Кастана (Shen, Castan, 1986) выделяется край полученной маски, из которого после заполнения и эрозии получается новая гладкая маска всего эмбриона (рис. 4). В конечном итоге изображения, полученные во всех каналах микроскопа, обрезают по размеру новой гладкой маски (Janssens et al., 2005).

Сегментация изображений и получение количественных данных. Для извлечения информации об экспрессии генов в каждом ядре эмбриона строится так называемая ядерная маска на основе изображения, прокрашенного на гистоновые белки, если оно имеется (см. выше), или же на основе пиксельного максимума экспрессии трех остальных генов (рис. 5). В данном случае для усиления контраста и уточнения границ ядер применяется локальное выравнивание гистограммы (Gonzales,

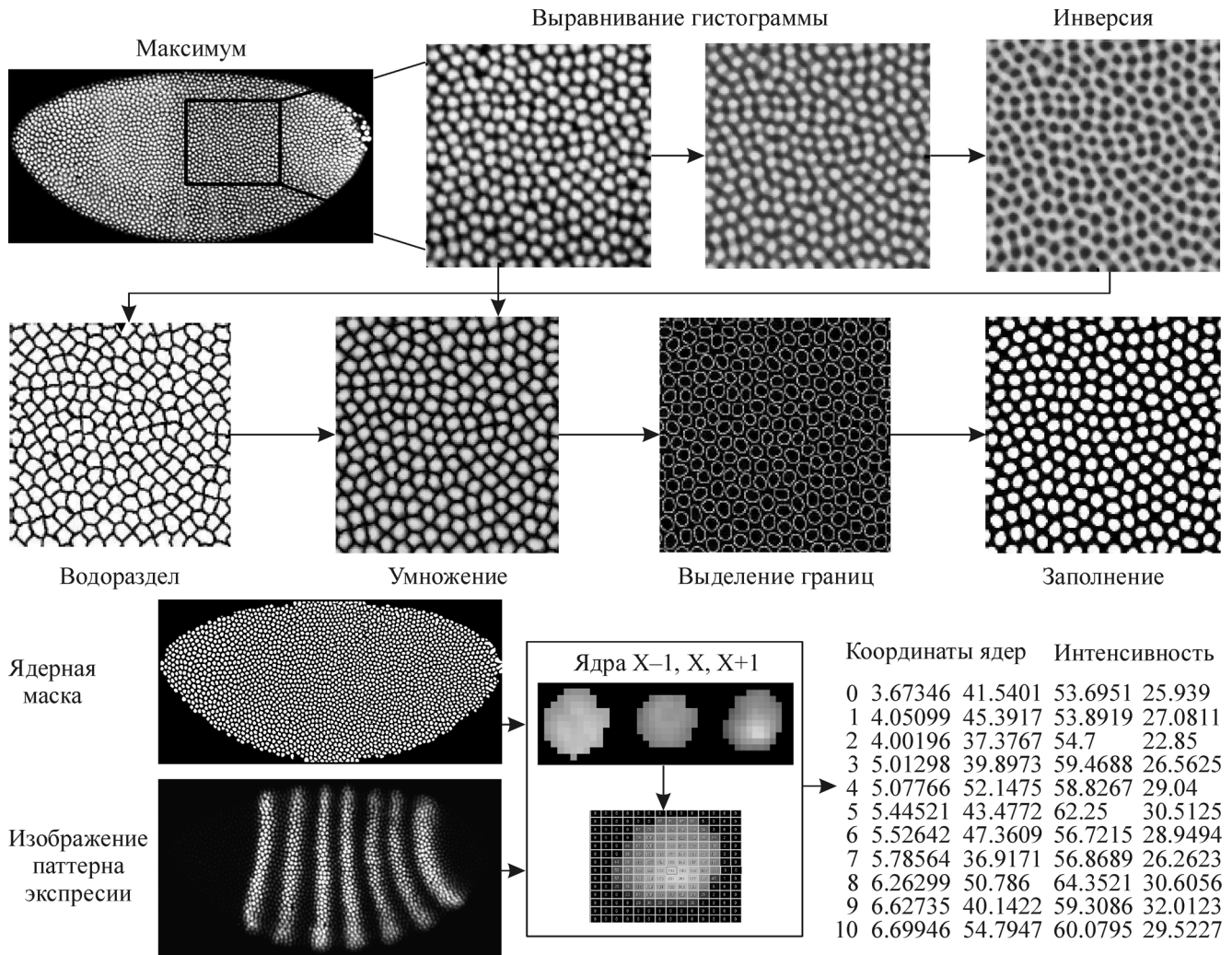


Рис. 5. Получение количественных данных.

Woods, 2002). Гранулированный шум (speckle noise) удаляется алгоритмом Кримминса (Crimmins, 1985), несколько циклов медианной фильтрации производят дальнейшее удаление шума из изображения. Далее, после инвертирования значений всех пикселей, к изображению применяется преобразование водораздела (watershed) (Гонсалес, Вудс, 2005). Область водораздела определяется как территория, занятая одним ядром, ограниченная линией шириной в 1 пиксель. Каждая область водораздела характеризуется уникальным значением градации серого цвета (Janssens et al., 2005). В результате применения эрозии и порогового фильтра это изображение преобразуется в бинарное, где границы водораздела имеют значение 0 (рис. 5).

После умножения этого бинарного изображения на изображение пиксельного максимума или изображение гистоновых белков каждое ядро отделяется от соседних ядер границей из нулевых пикселей. Эта операция позволяет разделить некоторые ядра, слившиеся на изображении между собой. В завершение эрозия с последующим преобразованием расстояния (Vincent, 1993) и пороговым фильтром позволяет удалить из маски посторонние пятна, не являющиеся ядрами эмбриона. После операций выделения границ (edge detection) и заполнения результирующая маска представляет собой бинарное изображение, в котором области, соответствующие ядрам, имеют значе-

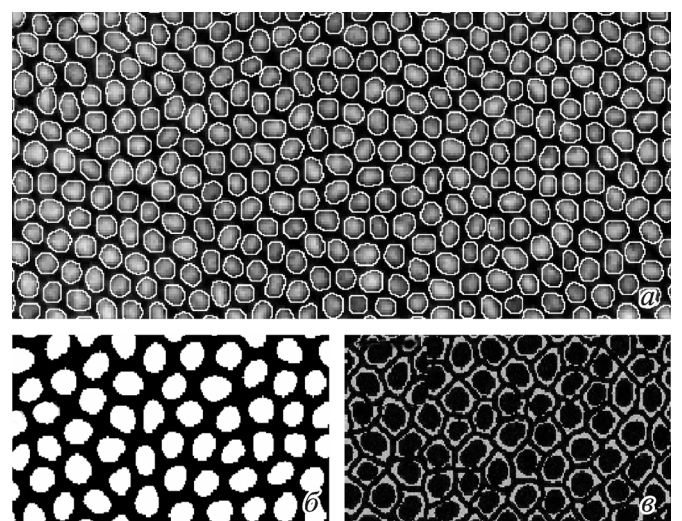


Рис. 6. Контроль качества извлечения количественных данных из изображений.

a — фрагмент изображения, полученного при наложении ядерной маски на изображение, полученное в гистоновом канале; *б, в* — 2 класса пикселей: соответствующие ядрам (*б*) и соответствующие областям водораздела, но не ядрам (*в*).

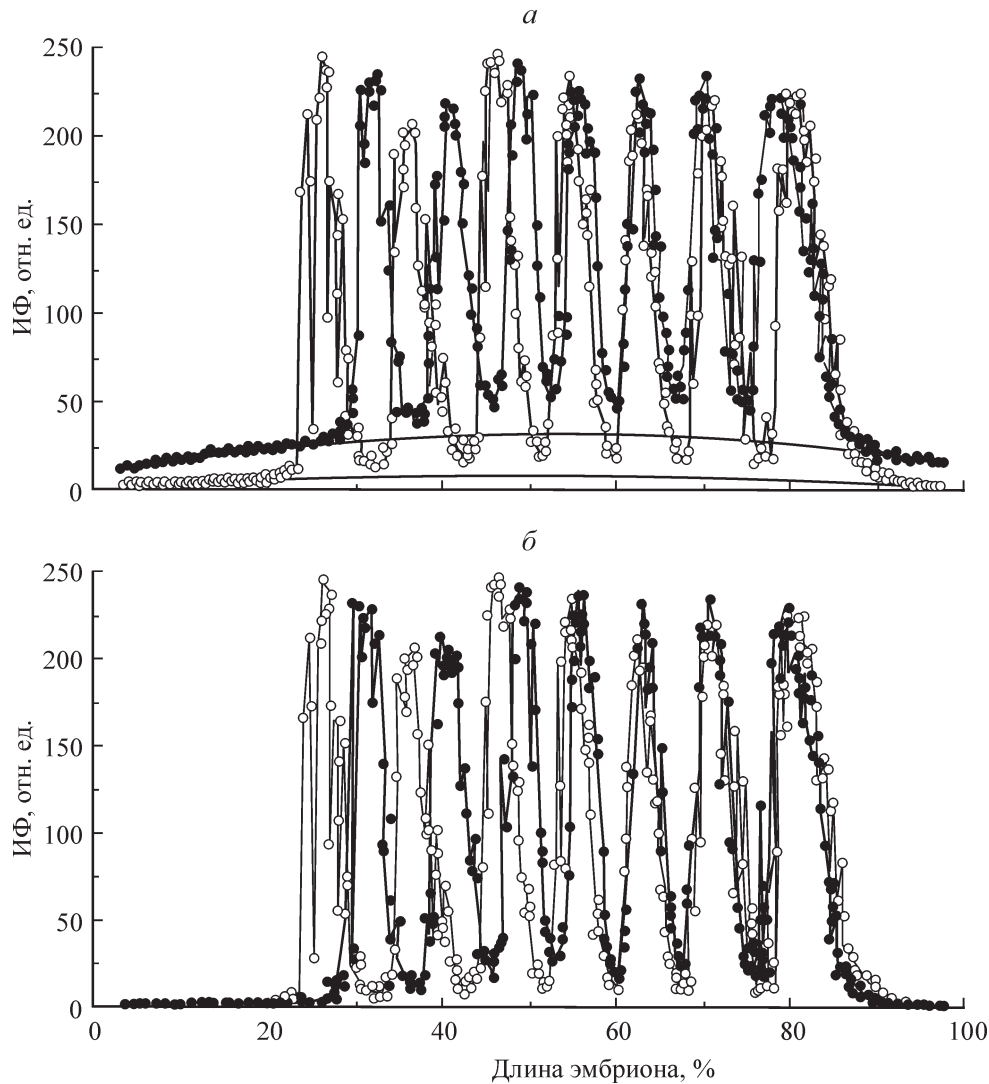


Рис. 7. Две картины (*a*, *б*) экспрессии гена *eve*, меченные одинаковыми первыми кроличьими антителами и разными специфичными к белкам кролика вторыми антителами, сопряженными с Cy5 (белый) или Texas Red (черный).

a — представлены одномерные картины, полученные путем вырезания 10%-ной полосы из центральной части эмбрионов вдоль оси А-П; *б* — те же картины экспрессии после удаления фона.

ние 1. Качество построения ядерной маски обычно контролируется визуально путем ее наложения на изображение, полученное при сканировании гистонов или на пиксельный максимум изображений, полученных в других каналах микроскопа (рис. 6, *a*). Помимо этого, нами был разработан численный метод контроля качества ядерной маски, основанный на том, что разброс интенсивностей для пикселей внутри ядра должен быть меньше, чем между пикселями внутри и вне ядра. Множество пикселей внутри и вне ядер рассматриваются как два класса (рис. 6, *б*, *в*).

Полученная бинарная маска используется для извлечения количественных данных об экспрессии генов. Координаты центра каждого ядра вычисляются с помощью инвариантов моментов (Hu, 1962). Наложение маски на изображения, полученные в каждом канале микроскопа, позволяет вычислить среднюю концентрацию продуктов всех сканированных генов в каждом ядре (Janssens et al., 2005). Конечный результат представляет собой таблицу, содержащую координаты *X* и *Y* каждого ядра в процентах соответственно от длины и ширины эмбриона, а

также усредненную интенсивность флуоресценции (относительный уровень экспрессии) для каждого из сканированных у данного эмбриона генов (рис. 5). Разработанная процедура извлечения количественной информации из биологических изображений реализована в пакете ProStack (Processing of Stacks), позволяющем осуществлять визуальное построение сложных процедур обработки данных и изображений (Matveeva et al., 2006).

2. *Удаление фонового сигнала.* Хорошо известно, что наряду со специфическим окрашиванием методы иммунофлуоресцентного маркирования биологических образцов *in situ* приводят к появлению неспецифического сигнала, называемого фоном. Возникновение этого сигнала вызвано многими факторами, из которых важная роль принадлежит первым и вторым антителам (рис. 7). Грубое удаление неспецифического фона может производиться вручную непосредственно на конфокальном микроскопе для каждого конкретного изображения путем установки значения «offset». Однако на практике это весьма затруднительно, во-первых, из-за необходимости настраивать режим сканирования для каждого отдельного изоб-

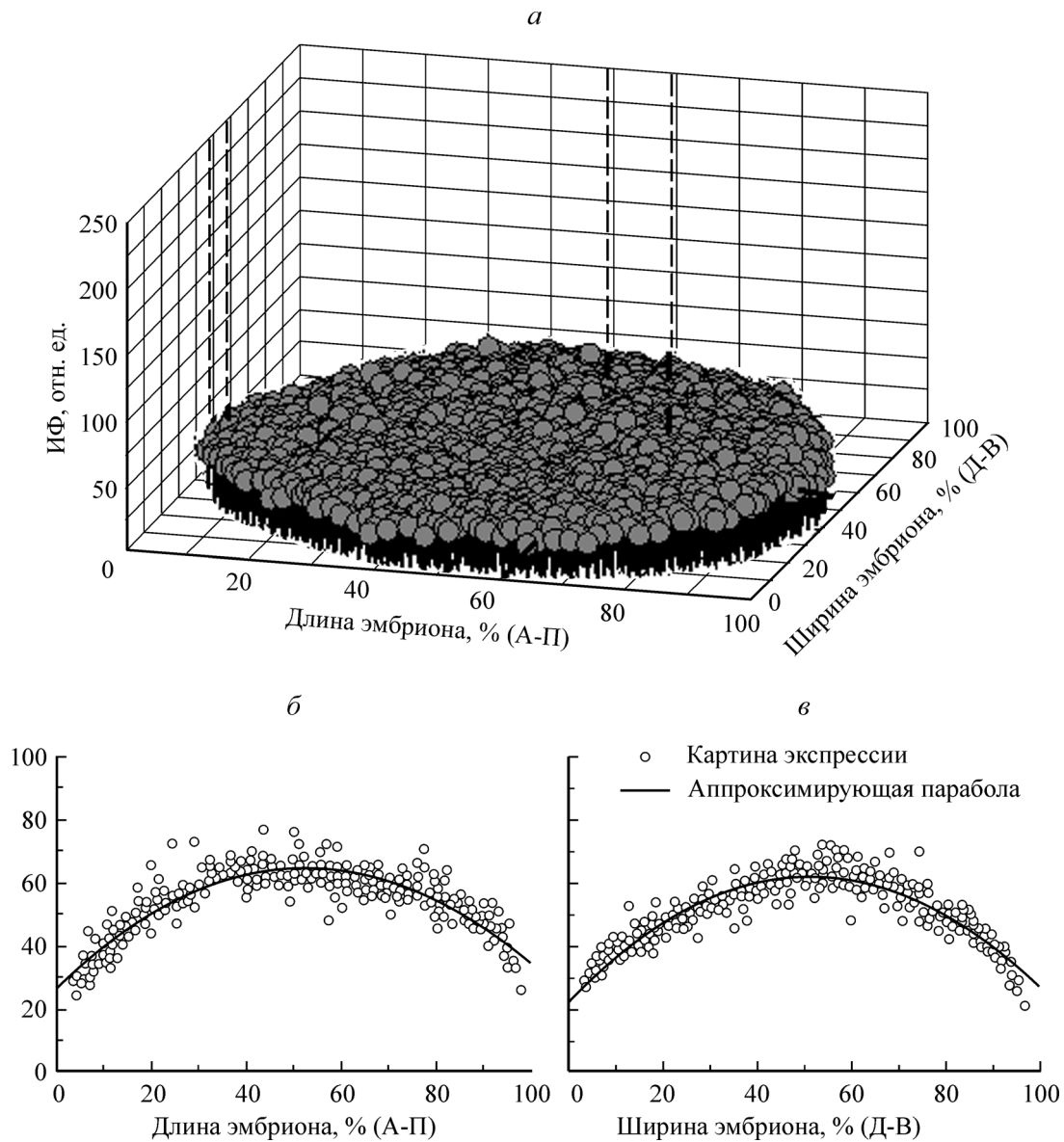


Рис. 8. Картина экспрессии *eve* у эмбриона, гомозиготного по $Df(3R)eve^{1.27}$, у которого не экспрессируется белок Eve.

a — двумерный паттерн, в котором индивидуальные ядра изображены в виде серых кружков, а черные отрезки, соединяющие ядра с плоскостью XY, характеризуют уровень интенсивности флуоресценции (ИФ) в ядрах; *б, в* — экспрессия в 10%-ных полосах, вырезанных из центральной части двумерной картины экспрессии по осям А-П (анteriorно-постeriorной) и Д-В (дорсально-вентральной); на двумерном паттерне эти полосы показаны штриховой линией; серые точки — одномерная картина экспрессии, сплошная линия — проекция аппроксимирующего параболоида.

ражения и, во-вторых, из-за того, что интенсивность фонового сигнала может распределяться неравномерно внутри биологического объекта.

Нами был разработан метод удаления неспецифического сигнала (Myasnikova et al., 2005), который может применяться для изображений, получаемых для разных объектов при фиксированных настройках конфокального микроскопа (см. выше). Очевидно, что даже незначительный уровень фона исказит численное значение уровня экспрессии гена. Более того, уровень фона варьирует от эмбриона к эмбриону и от эксперимента к эксперименту, что не позволяет провести сравнение данных, полученных в разных экспериментах. Помимо этого, уровень неспецифического окрашивания варьирует даже среди картин экспрессии одного гена, полученных с использованием разных вторых антител, сопряженных с разными флуорофорами (рис. 7).

Разработанный метод удаления фонового сигнала основан на наблюдении, что уровень флуоресценции в нуль-мутантах, окрашенных на продукт мутантного гена, хорошо аппроксимируется двумерным параболоидом или, в более общем случае, выпуклой поверхностью второго порядка (рис. 8). Этот параболоид очень близок к симметричному по осям *X* (А-П) и *Y* (Д-В). Параболическое распределение уровня фона объясняется, по всей видимости, свойствами конфокального микроскопа, что следует из экспериментов по изменению установок микроскопа. Было установлено, что увеличение значения «offset» приводит к уменьшению уровня фонового сигнала в изображении, в то время как уменьшение — к появлению параболического фона с возрастающей кривизной. Основная идея метода удаления фона состоит в определении неэкспрессирующих областей эмбриона, которые затем используются для аппроксимации фонового сигнала, с тем

чтобы затем удалить его масштабированием картины экспрессии. Эта процедура выполняется за несколько этапов.

Определение неэкспрессирующих областей. Неэкспрессирующими областями считаются участки, в которых данный ген не экспрессируется в большинстве ядер. Для каждого гена эти участки первоначально определяются на основе тщательного визуального изучения картин экспрессии у всех эмбрионов. Неэкспрессирующие области, идентифицированные при таком анализе, затем уточняются на двумерной картине экспрессии у каждого эмбриона (рис. 9).

Поскольку экспрессия генов сегментации является в основном функцией позиции относительно оси А-П эмбриона (см. выше), она достаточно полно может быть представлена в виде одномерного сигнала. Следовательно, в большинстве случаев достаточно задать границы неэкспрессирующих областей только вдоль оси X . Например, у гена *eve* эти области лежат в пределах 0—25 и 90—100 % длины эмбриона (рис. 9). Тем не менее в двумерном представлении имеется некоторое искривление полос экспрессии вдоль оси Y . В эмбрионах из нашего набора данных такое искривление усиливается еще больше механической деформацией, вызванной давлением покровного стекла. Таким образом, для определения двумерных неэкспрессирующих областей необходимо предварительно выпрямить искривленные полосы экспрессии (Myasnikova et al., 2005).

Аппроксимация и удаление фона. Фоновый сигнал аппроксимируется квадратичным параболом по опорным точкам, выделенным из неэкспрессирующих областей эмбриона (рис. 9). Аппроксимирующий парабола строится при помощи итеративной оптимизационной процедуры. Окончательно фон удаляется из всего эмбриона линейным отображением интенсивности, которое преобразует значения флуоресценции, равные или меньше уровня фона в ноль, а максимально возможную флуоресценцию (255 ед.) — саму в себя (Myasnikova et al., 2005). Примеры нормализации паттернов разных генов сегментации у эмбрионов разного возраста приводятся на рис. 10.

Оценка точности метода удаления фона. Метод удаления фонового сигнала был тщательно протестирован на эмбрионах, гомозиготных по нуль-мутации генов *bcd*, *eve*, *gt*, *kni* и *Kr* и окрашенных для выявления продукта мутантного гена. Картины экспрессии у таких мутантов после удаления фона трансформируются в нулевой уровень экспрессии во всем эмбрионе (рис. 8). Визуальная оценка картин экспрессии показывает, что метод удаления фона дает хорошие результаты для эмбрионов, относящихся к циклу 14А. Типичные результаты удаления фона из картин экспрессии *bcd*, *hb*, *Kr*, *eve* и *run* на этой стадии приведены на рис. 10. Для картин экспрессии генов *gar* и *raig-gule* на более ранних стадиях развития (циклы 10—12) метод работает несколько хуже из-за низкой интенсивности сигнала и отсутствия четких границ областей экспрессии (Myasnikova et al., 2005).

3. *Регистрация картин экспрессии генов.* Конфокальный микроскоп позволяет получать одновременно картины экспрессии лишь нескольких генов. Однако в конечном итоге интерес представляет пространственно-временная динамика экспрессии всех генов исследуемой генной сети. В силу индивидуальной вариативности размеров эмбрионов информацию об относительном расположении картин экспрессии разных генов в пространстве невозможно получить простым совмещением картин экс-

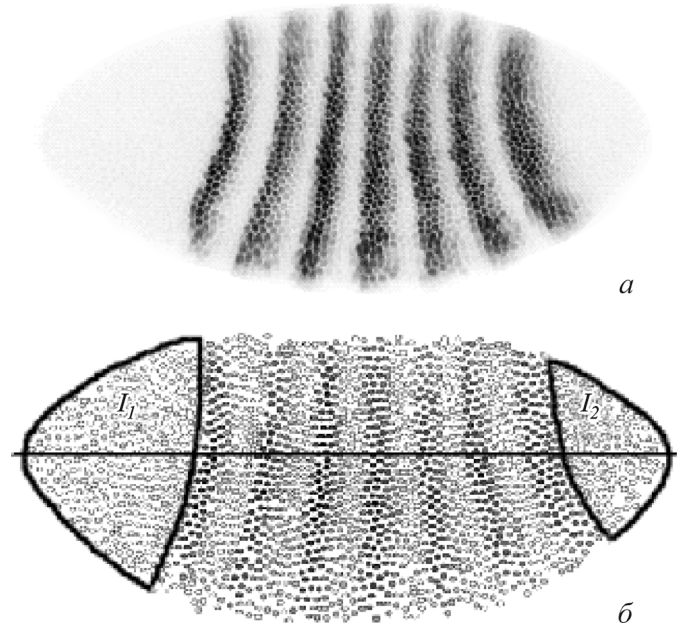


Рис. 9. Выделение неэкспрессирующих областей.

а — конфокальное изображение эмбриона на стадии поздней бластодермы, меченного на экспрессию гена *eve*; для наглядности изображение инвертировано, белые пиксели соответствуют отсутствию сигнала, черные — максимальному уровню сигнала. *б* — картина экспрессии для эмбриона, изображенного на рис. 9, *а*, реконструированная из количественных данных; индивидуальные ядра изображены кружками, градации серого пропорциональны интенсивности флуоресценции в ядрах; I_1 и I_2 — неэкспрессирующие области.

прессии у индивидуальных эмбрионов, окрашенных антителами к разным белкам. Чтобы решить эту задачу, данные, полученные в индивидуальных эмбрионах, должны быть приведены к общей системе координат с помощью процедуры регистрации. В данных работах (Мясникова и др., 2001; Myasnikova et al., 2001; Kozlov et al., 2002) нами был использован метод регистрации, основанный на выделении контрольных точек в изображении (Brown, 1992) и на преобразовании координат для максимально полного совмещения этих точек в разных изображениях. В качестве контрольных точек обычно используют какие-либо характерные признаки изображений. В описываемом наборе данных все эмбрионы были сканированы для выявления экспрессии гена *eve*, и картина экспрессии этого гена высокодинамична во времени. Поэтому в качестве контрольных точек использовали характерные признаки этого паттерна — экстремумы одномерной картины экспрессии, полученной путем экстракции данных из центральной 10%-ной полосы эмбриона (рис. 2). После выделения контрольных точек для каждого изображения находили такое преобразование координат, которое минимизировало расстояние между соответствующими точками в разных изображениях, а затем трансформации координат подвергали все регистрируемые изображения.

Выбор контрольных точек для регистрации. 1. *Аппроксимация сплайнами.* На графике, представляющем собой одномерную экспрессию гена, полосы и межполосные пространства имеют вид соответственно пиков и впадин (рис. 1, *центральный ряд*). Наиболее естественным представляется использовать в качестве контрольных точек для последующей регистрации X -координаты экстремумов на кривой, однако определить положение экстремумов непосредственно по исходным данным

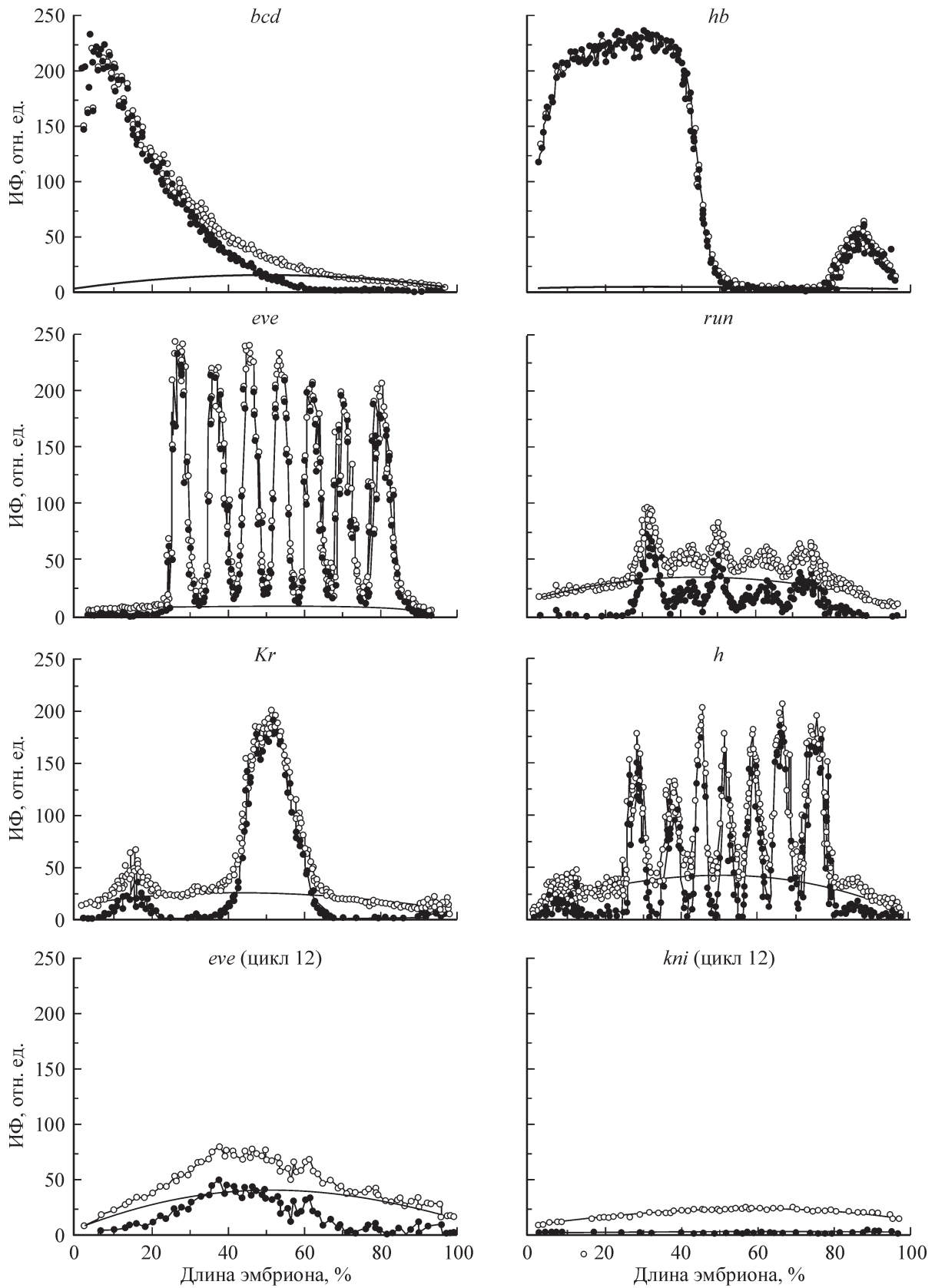


Рис. 10. Удаление фона из индивидуальных картин экспрессии ряда генов.

За исключением случаев, когда цикл деления ядер указан, все остальные картины экспрессии получены в эмбрионах, принадлежащих циклу 14А. Ненормализованные картины экспрессии изображены белыми кружками, картины экспрессии после удаления фонового сигнала — черными кружками; параболоид, аппроксимирующий фон, изображен сплошной линией.

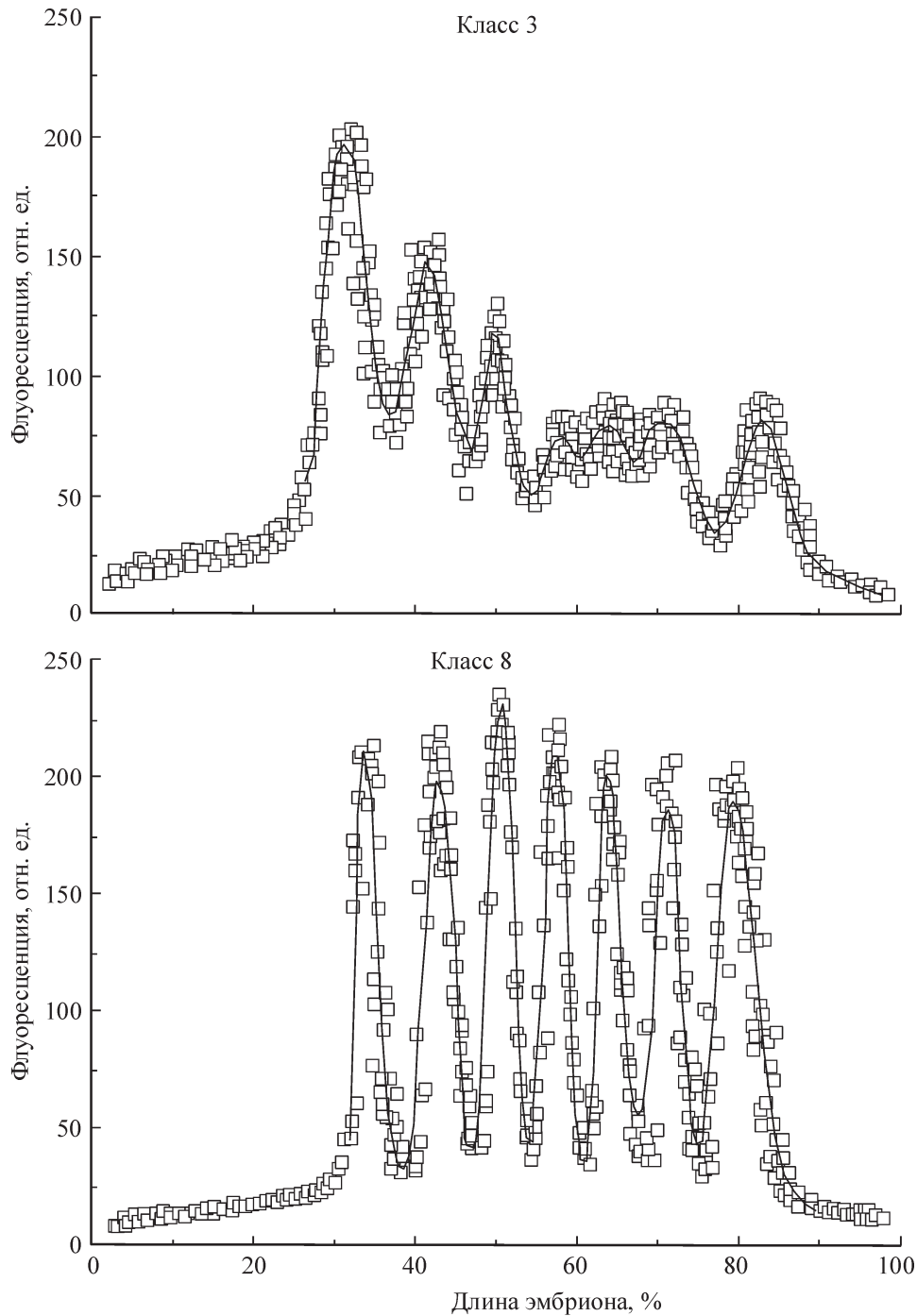


Рис. 11. Аппроксимация сплайнами.

Экспериментальная кривая концентраций белка Eve (точки) и ее аппроксимация сплайнами (сплошная линия); X -координаты экстремумов сплайна принимаются за координаты экстремумов сигнала.

трудно ввиду наличия значительного шума. Чтобы преодолеть эту трудность и локализовать экстремумы, был применен метод аппроксимации картин экспрессии квадратичными сплайнами (Мясникова и др., 2001; Myasnikova et al., 2001). Этот метод обеспечивает непрерывную аппроксимацию исходной кривой и характеризует каждый паттерн конечным множеством параметров. Простейшая аппроксимация достигается квадратичным сплайном с нефиксированными узлами, для которого выполняется требование непрерывности первой производной в узловых точках. Система узлов выбирается как множество различ-

ных точек на оси X , ограничивающих область каждого пика. Положение узла определяется как переходная точка между пиком и соседней впадиной. В результате аппроксимации каждая кривая характеризуется множеством узлов и параметров сплайна. По этим параметрам вычисляется X -координата каждого экстремума. На примере, представленном на рис. 11, продемонстрировано, что после аппроксимации положение пиков на кривой можно определить с хорошей точностью (Myasnikova et al., 2001).

2. Быстрое избыточное двоичное вейвлет-преобразование (Fast Redundant Dyadic Wavelet transform — FRDWT).

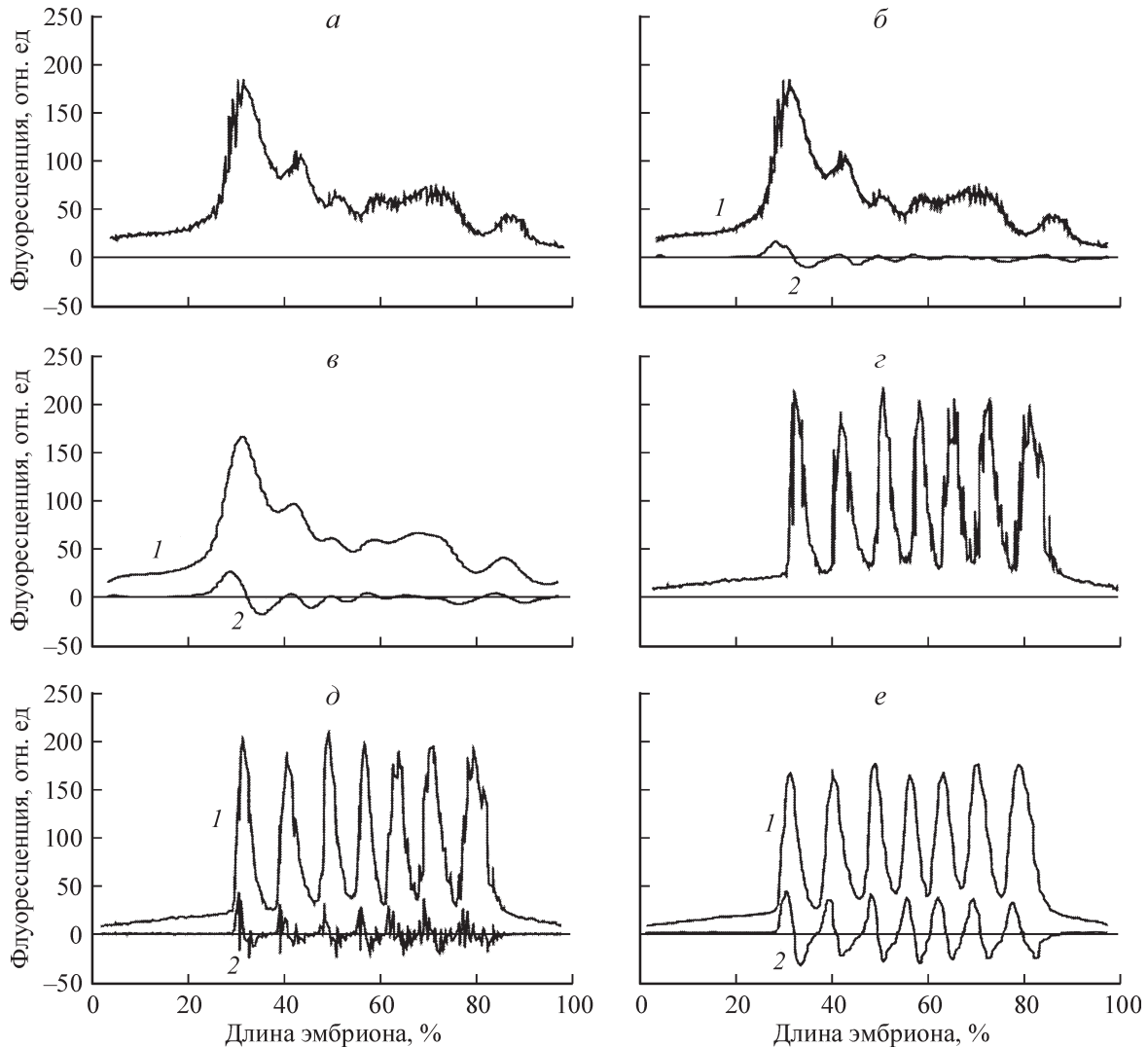


Рис. 12. Пример вейвлетного изображения.

Картини экспрессии гена *eve* во временных классах 3 (*a–в*) и 8 (*г–е*); *a, г* — исходный сигнал; *б, д* — аппроксимирующая (1) и детализирующая (2) последовательности для первого (*б*) и второго (*д*) уровней разложения; *в, е* — третий уровень разложения сигнала, используемый для извлечения контрольных точек.

Второй способ выделения контрольных точек основан на итеративном применении вейвлетного преобразования (Kozlov et al., 2000, 2002). Вейвлетное преобразование (Unser, 1996; Малоземов и др., 1998) позволяет получать локальную высокочастотную и крупномасштабную информацию об объекте. Его применение обеспечивает возможность одновременного исследования данных в физическом (время и координата) и частотном пространствах. Нам необходимо выбрать тип и базис преобразования таким образом, чтобы выделить из сигнала информацию о первой производной исходного сигнала. Такую возможность предоставляет FRDWT (Unser et al., 1994).

Основными свойствами FRDWT являются хорошее подавление шума и точное выделение пространственно локализованных признаков. Сигнал раскладывается на две последовательности — аппроксимирующую (low pass) и детализирующую (high pass) (соответственно кривые 1 и 2 на рис. 12). Ввиду избыточности преобразования число элементов каждой последовательности совпадает с числом наблюдений в исходной выборке, что позволяет сохранить полную информацию о локализации

экстремумов, содержащуюся в исходном сигнале. На каждом уровне разложения исходный сигнал сглаживается путем удаления шума вплоть до определенной частоты и представляется в виде аппроксимирующей последовательности. Детализирующая составляющая содержит информацию о тех признаках исходного сигнала, которые определяются выбором соответствующего вейвлетного базиса (Unser, 1996). Для определения положения экстремумов, т. е. нулей первой производной, применяем базисные функции, которые включают в себя характеристики первой производной в детализирующую последовательность. Разложение повторяется итеративно, причем на каждом последующем этапе вместо исходного сигнала раскладывается соответствующая аппроксимирующая составляющая (рис. 12).

3. Регистрация изображений. Регистрация изображений производится посредством масштабирования двумерных картин экспрессии вдоль оси абсцисс посредством аффинного преобразования, минимизирующего суммарное расстояние между X -координатами всех контрольных точек на разных паттернах и средним по-

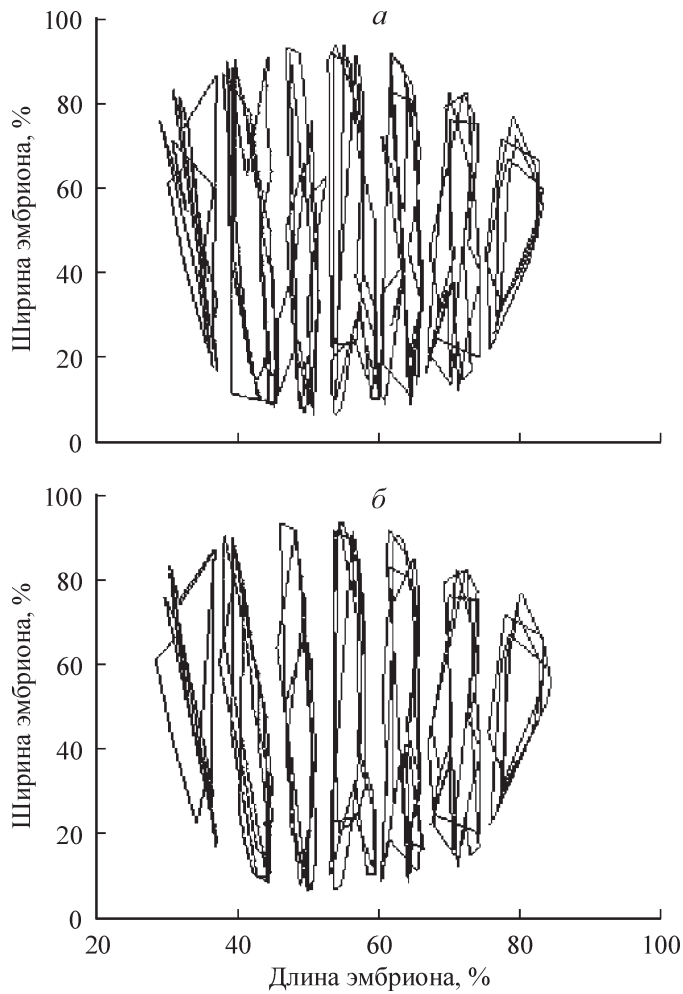


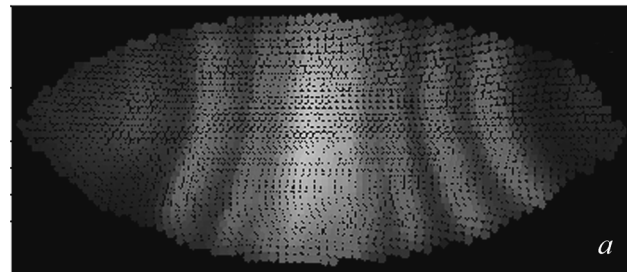
Рис. 13. Пространственная регистрация.

а, б — двумерные картины экспрессии гена *eve* у четырех эмбрионов класса 8 соответственно до и после регистрации; *линии* соответствуют уровню интенсивности флуоресценции, составляющему примерно 50% от максимального уровня экспрессии.

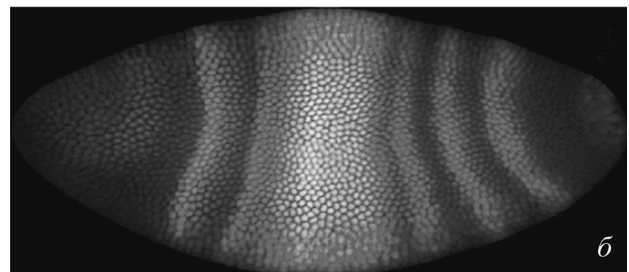
ложением соответствующих контрольных точек по всем регистрируемым изображениям (Kozlov et al., 2002) (рис. 13). Разработанный метод был использован для регистрации картин экспрессии генов сегментации у эмбрионов, относящихся к одному временному классу. Пример результата регистрации картин экспрессии гена *eve* у четырех эмбрионов, отнесенных к временному классу 8, показан на рис. 13. Очевидно, что регистрация уменьшает пространственную вариабельность расположения паттернов. Оценка точности метода регистрации изображений производилась по величинам стандартных отклонений *X*-координат экстремумов паттерна *eve* для каждого временного класса. Во временном классе 3, когда паттерн экспрессии еще не сформировался и можно выделить только несколько контрольных точек, эти значения составляют примерно 3.5% длины эмбриона. Точность выделения экстремумов, а следовательно, и регистрации выше в более поздних временных классах. В этот период значения стандартных отклонений позиций экстремумов становятся меньше 1% длины эмбриона, это примерно соответствует диаметру одного ядра в цикле 14А и говорит о хорошем качестве регистрации (Kozlov et al., 2002).

4. Построение единого набора данных. Главной целью пространственной регистрации и удаления фона являлось построение карты взаимного расположения доменов экспрессии всех генов исследуемой генной сети на каждом отрезке времени. Таким образом, области экспрессии каждого гена на карте представлены своим «эталонным» или «интегрированным» паттерном, который характеризуется средними для каждого периода времени значениями интенсивности флуоресценции. Для генной сети, контролирующей сегментацию у дрозофилы, целью является построение интегрированных паттернов для всех эмбрионов, принадлежащих к каждому из временных классов (рис. 1). Идеальный интегрированный паттерн данного гена мог бы быть построен в той гипотетической ситуации, если все эмбрионы имели бы одинаковую ядерную структуру, т. е. состояли бы из множества ядер с одинаковыми для всех эмбрионов координатами. В таком случае средние интенсивности флуоресценции

Интегрированный паттерн

*gt, Kr, eve*

Индивидуальный эмбрион

*gt, Kr, eve*

Интегрированный паттерн

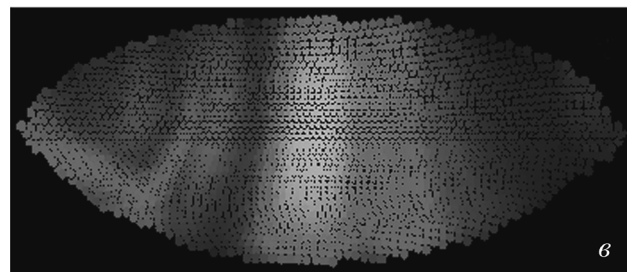
*gt, Kr, kni*

Рис. 14. Двухмерный атлас экспрессии генов сегментации во временном классе 8.

а — интегрированные картины экспрессии *gt, Kr* и *eve*; *б* — картина экспрессии тех же генов у индивидуального эмбриона того же класса; *в* — «виртуальный» эмбрион, содержащий картины экспрессии *Kr, kni* и *gt*.

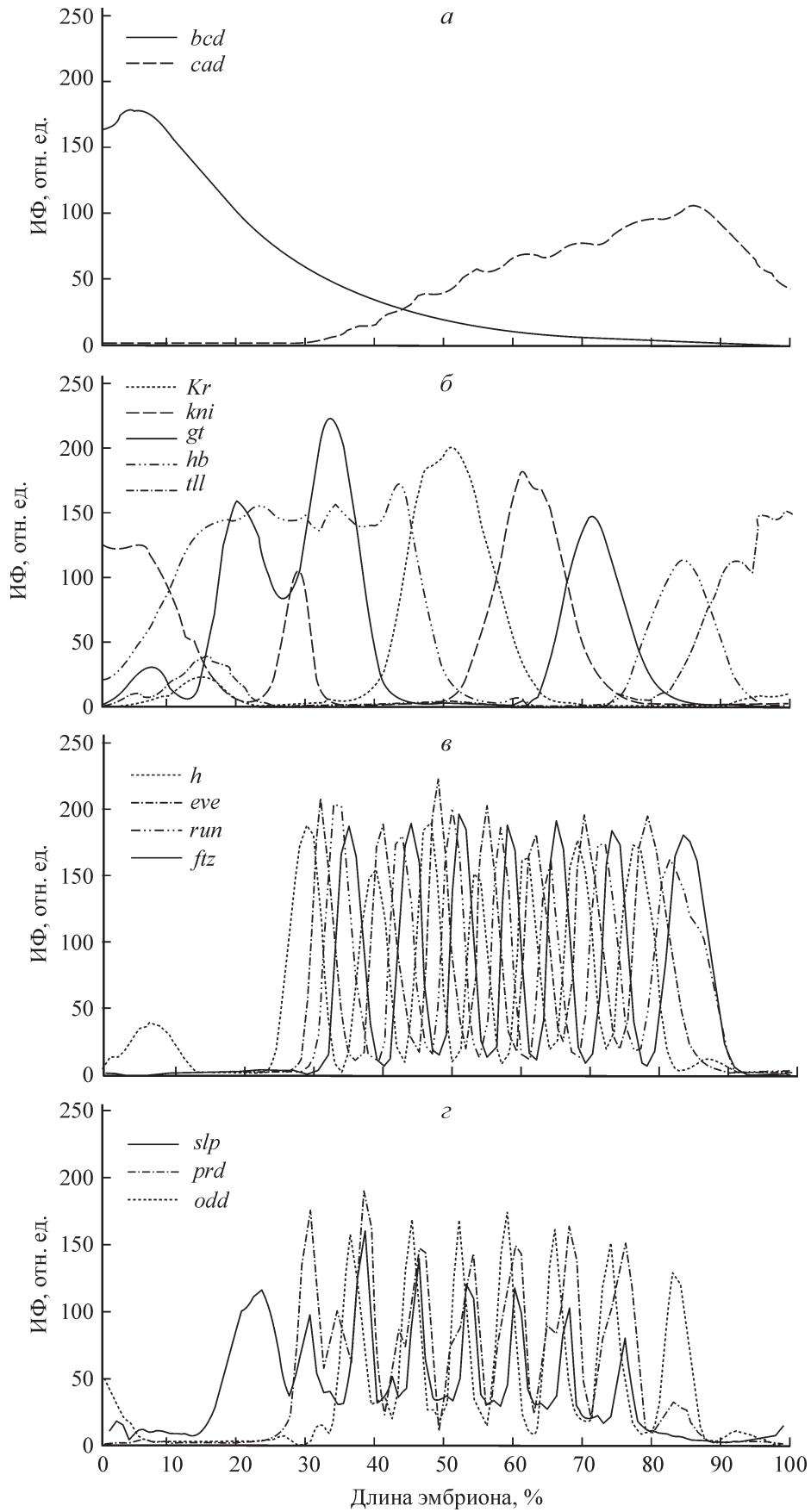


Рис. 15. Одномерный атлас экспрессии генов сегментации у дрозофилы.

a — материнские градиенты во временном классе 6; *б* — гены gap во временном классе 8; *в*, *г* — гены pair-rule во временном классе 8.

могли бы быть вычислены независимо в каждом ядре. Однако на самом деле количество и положение ядер в разных эмбрионах, а также плотность распределения ядер внутри каждого индивидуального эмбриона непостоянны. Поэтому должна быть вычислена усредненная ядерная структура паттерна, учитывающая эти особенности. Эта задача была решена путем определения пространственной плотности распределения ядер, что позволило вычислить среднее значение диаметра ядра в каждой точке усредненного эмбриона (Kozlov et al., 2002).

Конструирование двухмерных интегрированных картин экспрессии осуществили, используя зарегистрированные данные, и для эмбрионов каждого временного класса. Для этого каждому ядру каждого индивидуального эмбриона ставили в соответствие ядро усредненной ядерной структуры паттерна, имеющее наиболее близкие координаты, после чего вычисляли среднюю интенсивность флуоресценции по всем индивидуальным ядрам, отнесенным к данному среднему ядру. Пример интегрированной двухмерной картины экспрессии генов *gt*, *Kr* и *eve* у эмбрионов временного класса 8 представлены на рис. 14, а. Сравнение этой картины экспрессии с картиной экспрессии тех же генов у индивидуального эмбриона класса 8 (рис. 14, б) показывает, что интегрированная картина в целом правильно передает основные особенности экспрессии индивидуальных генов, включая форму и размер доменов. Представленная на рис. 14, в картина экспрессии генов *Kr*, *kni* и *gt* интересна тем, что в использованном в работе наборе данных не было эмбрионов, одновременно окрашенных для выявления экспрессии этих генов. Отображая интегрированные картины экспрессии на усредненную ядерную структуру паттерна, можно создавать так называемые виртуальные эмбрионы с новыми комбинациями генов для визуализации взаимного расположения их доменов экспрессии (Kozlov et al., 2002).

Для конструирования одномерных интегрированных картин экспрессии генов сегментации координаты ядер каждого одномерного паттерна из набора данных, прошедшего регистрацию, группируются вдоль оси *X* по *R*-интервалам (Myasnikova et al., 2001). Затем внутри каждого интервала вычисляется среднее значение интенсивности флуоресценции по всем эмбрионам данного временного класса. Число *R* выбирается из следующих соображений: например, в цикле 14А в центральной части эмбриона диаметр одного ядра составляет примерно 1 % от его длины, и тогда *R* должно быть равно 100, для того чтобы правильно смоделировать ядерную структуру паттерна. Одномерные интегрированные картины экспрессии материнских генов, генов *gap* и генов *pair-rule* показаны на рис. 15.

Заключение

В данной статье приведен обзор ранее разработанных нами методов получения количественной информации по экспрессии генов на основе конфокальных изображений. Описанные методы обработки включают в себя сегментацию изображений, удаление фонового сигнала, пространственную регистрацию картин экспрессии и интеграцию данных (Kozlov et al., 2000, 2002; Мясникова и др., 2001; Myasnikova et al., 2001, 2005; Janssens et al., 2005). Эти процедуры являются базовыми для получения и обработки количественных данных по экспрессии генов и могут использоваться как последовательно, так и в различных

сочетаниях. Широкий спектр процедур обработки, а также легкость адаптации к другим биологическим объектам являются основными преимуществами представленных методов.

Большинство из разработанных в настоящее время методов обработки изображений чаще всего направлены на решение какой-либо одной проблемы (например, регистрацию изображений) и поэтому включают в себя только одну или две процедуры. Наиболее распространенной и востребованной из них является сегментация изображений клеток и тканей (Ortiz de Solorzano et al., 1999, 2001; Umesh Adiga, Chaudhuri, 1999; Chawla et al., 2004). Существенным недостатком большинства алгоритмов сегментации и реализованных на основе их программных систем является необходимость участия пользователя для установки различных параметров обработки. Предложенный нами метод сегментации полностью автоматический. Единственным этапом, требующим визуального контроля, является определение ориентации эмбриона (рис. 3). Другие коллективы исследователей также проводили сегментацию конфокальных изображений бластодермы дрозофилы, используя собственные оригинальные методы. Например, перемещая квадратное окно, размер которого примерно совпадал со средним размером ядра, вдоль дорсальной стороны конфокального изображения эмбриона, были получены количественные данные по экспрессии *bcd* и *hb* (Houchmandzadeh et al., 2002). При этом в каждом квадрате авторы вычисляли среднее значение уровня флуоресценции. Позже тот же метод был применен и другими исследователями (Gregor et al., 2005). В обоих случаях реализация метода осуществлялась в коммерческом пакете MatLab, Mathworks (США). Недостатком этого метода является то, что он оперирует понятием среднего размера ядра и не выделяет на изображении реальные ядра, что затрудняет возможность расширения этого метода для сегментации изображений других биологических объектов.

Одной из последних разработок является метод сегментации трехмерных стеков конфокальных изображений бластодермы дрозофилы (Luegno et al., 2006). Этот метод основан на определении локальных максимумов флуоресценции ядер и на извлечении так называемых зерен (*seeds*) для каждого ядра с целью идентификации позиций ядер, окрашенных Sytox Green Dye для выявления ДНК. «Зерна» в дальнейшем наращиваются на основе изображения, полученного из исходного путем применения порогового фильтра. Этот метод, так же как и предыдущие, реализован в пакете MatLab, его преимущество состоит в том, что он изначально разрабатывался для обработки трехмерных данных. Недостатком является проблема с выделением «зерен», так как большинство ядер имеет несколько локальных максимумов и требуется сложная процедура корректировки, к тому же наращивание «зерен» осуществляется на основе масок, полученных с помощью порогового фильтра, которые не всегда правильно воспроизводят форму и размер ядер.

В настоящее время известно мало методов удаления неспецифического сигнала из конфокальных изображений, полученных с применением методов иммунохимии, а также методов регистрации конфокальных изображений. Многие авторы удаляют неспецифический фон простым вычитанием определенной величины интенсивности из общего сигнала. Было показано (Gregor et al., 2007), что в полученных с помощью двухфотонного конфокального микроскопа изображениях живых эмбрионов дрозофилы, у которых эндогенный ген *bcd* был замещен

на конструкт, экспрессирующий eGFP-Vcd, уровень неспецифического свечения примерно однороден внутри одного эмбриона. Однако этот случай является исключением, и, несомненно, должны существовать алгоритмы, позволяющие учитывать различия в уровне фонового сигнала в разных частях конфокального изображения.

Недавно была опубликована процедура регистрации изображений, отличная от разработанной нами (Sorzano et al., 2006). Метод, названный эластичной регистрацией, эластично деформирует изображения для совмещения одних и тех же черт у картин экспрессии разных эмбрионов. После трансформации все регистрируемые изображения приводятся к одному размеру по осям X и Y . Данный метод крайне эффективен, но не может быть использован для оценки позиций доменов экспрессии, так как трансформирует координаты исходных изображений.

По сравнению с вышеописанными методами основными преимуществами предложенной в настоящей работе обработки конфокальных изображений бластодермы дрозофилы являются широкий спектр процедур, возможность поточной обработки данных, а также легкость адаптации к другим биологическим объектам. Метод был применен нами для обработки около 5000 изображений экспрессии генов сегментации у дрозофилы на уровне белка. Полученные количественные данные были успешно использованы для расшифровки регуляторных механизмов, обеспечивающих сдвиги границ экспрессии доменов генов *gap* в ходе раннего развития эмбриона (Jaeger et al., 2004b), а также для выявления регуляторных механизмов, контролирующих экспрессию генов *gap* (Jaeger et al., 2004a). Также были охарактеризованы процессы динамической фильтрации варибельности картин экспрессии зиготических генов сегментации во времени.

Основным недостатком большинства проблемно-ориентированных пакетов обработки данных и изображений является трудность их адаптации для решения сходных задач на других объектах. Описанные в данной статье процедуры сегментации, удаления фона и построения интегрированных паттернов были успешно адаптированы для обработки данных по экспрессии генов сегментации на уровне мРНК. У дрозофилы основная разница в экспрессии генов, кодирующих транскрипционные факторы на уровне белка и мРНК, заключается в том, что РНК локализуется не только в ядре, но и в цитоплазме. Для сегментации подобных изображений была применена модифицированная процедура водораздела (рис. 5), конструирующая маску для ядер и окружающей цитоплазмы. Кроме того, были внесены небольшие изменения в процедуру поиска опорных точек при удалении неспецифического фона. В результате были созданы усредненные картины экспрессии на уровне мРНК для разных репортерных конструкторов гена *eve* и на их основе предложена новая модель транскрипционной регуляции (Janssens et al., 2006). Помимо этого, метод также частично адаптирован для обработки трехмерных данных, полученных с помощью конфокального микроскопа. К настоящему моменту обработано несколько стеков изображений, полученных у эмбрионов, окрашенных для выявления активности гена *rhomboid* в ядрах. В настоящее время процедура построения ядерной маски в процессе сегментации (рис. 5) успешно модифицирована также для маскирования экспрессирующих областей и получения количественных данных по экспрессии генов в раннем развитии коралла *Acropora millepora* и морского анемона *Nematostella vectensis* (Kozlov et al., 2007).

Авторы выражают благодарность д-ру Д. Косману и д-ру К. Алонсо за помощь в получении экспериментальных данных.

Работа выполнена при финансовой поддержке грантов NIH RR07801, АФГиР (GAP) RUB1-1578 и NWO-РФФИ 047.011.2004.013.

Список литературы

- Goncalves P., Vudc P. 2005. Цифровая обработка изображений. М.: Техносфера. 1072 с.
- Закс Л. 1976. Статистическое оценивание. М.: Статистик. 457 с.
- Малоземов В. Н., Певный А. Б., Третьяков А. А. 1998. Быстрое вейвлетное преобразование дискретных периодических сигналов и изображений. Проблемы передачи инф. 34 (2) : 77—85.
- Мясникова Е., Самсонова А., Самсонова М., Рейнитуц Д. 2001. Пространственная регистрация данных об экспрессии генов *in situ*. Молекуляр. биол. 35 (6) : 1110—1115.
- Самсонова М. Г., Гурский В. В., Козлов К. Н., Самсонов А. М. 2006. Системный подход к исследованию развития организмов. Научно-технические ведомости СПбГПУ. 2 : 222—234.
- Azpiazu N., Frasch M. 1993. *tinman* and *bagpipe*: two homeobox genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Develop.* 7 : 1325—1340.
- Banerjee N., Zhang M. Q. 2002. Functional genomics as applied to mapping transcription regulatory networks. *Curr. Opin. Microbiol.* 5 : 313—317.
- Brown L. G. 1992. A survey of image registration techniques. *ACM Computing Surveys.* 24 (4) : 325—376.
- Chawla M. K., Lin G., Olson K., Vazdarjanova A., Burke S. N., McNaughton B. L., Worley P. F., Guzowski J. F., Roysam B., Barnes C. A. 2004. 3D-catFISH: a system for automated quantitative three-dimensional compartmental analysis of temporal gene transcription activity imaged by fluorescence *in situ* hybridization. *J. Neurosci. Methods.* 139 : 13—24.
- Crimmins T. R. 1985. Geometric filter for speckle reduction. *Appl. Opt.* 24 : 1438—1443.
- Dequin R., Saumweber H., Sedat J. 1984. Proteins shifting from the cytoplasm into the nuclei during early embryogenesis of *Drosophila melanogaster*. *Develop. Biol.* 104 : 37—48.
- Frasch M., Hoey T., Rushlow C. 1987. Characterization and localization of the *even-skipped* protein of *Drosophila*. *EMBO J.* 6 : 749—759.
- Gilbert S. F. 2003. *Developmental biology*. Sunderland, Massachusetts: Sinauer Associates. 839 p.
- Gonzalez R. C., Woods R. E. 2002. *Digital image processing*. Upper Saddle River, N. J.: Prentice Hall. 730 p.
- Gregor T., Bialek W., de Ruyter van Steveninck R. R., Tank D. W., Wieschaus E. F. 2005. Diffusion and scaling during early embryonic pattern formation. *Proc. Nat. Acad. Sci. USA.* 102 : 18 403—18 407.
- Gregor T., Tank D. W., Wieschaus E. F., Bialek W. 2007. Probing limits of positional information. *Develop. Cell.* 130.4 : 153—164.
- Hu M.-K. 1962. Visual pattern recognition by moment invariants. *IRE transactions of information theory*. *IRE Transactions on Information Theory.* 8 : 179—187.
- Houchmandzadeh B., Wieschaus E., Leibler S. 2002. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature.* 415 : 748—749.
- Jaeger J., Blagov M., Kosman D., Kozlov K. N., Manu, Myasnikova E., Surkova S., Vanario-Alonso C. E., Samsonova M., Sharp D. H., Reinitz J. 2004a. Dynamical analysis of regulatory interactions in the *gap* gene system of *Drosophila melanogaster*. *Genetics.* 167 : 1721—1737.
- Jaeger J., Surkova S., Blagov M., Janssens H., Kosman D., Kozlov K., Manu, Myasnikova E., Vanario-Alonso C.-E., Samsonova

- va M., Sharp D., Reinitz J. 2004b. Dynamic control of positional information in the early *Drosophila* embryo. *Nature*. 430 : 368—371.
- Janssens H., Hou S., Jaeger J., Kim A. R., Myasnikova E., Sharp D., Reinitz J. 2006. Quantitative and predictive model of transcription control of the *Drosophila melanogaster even-skipped* gene. *Nature Genetics*. 38 : 1159—1165.
- Janssens H., Kosman D., Vanario-Alonso C.-E., Jaeger J., Samsonova M., Reinitz J. 2005. A high-throughput method for quantifying gene expression data from early *Drosophila* embryos. *Develop. Gen. Evol.* 215 : 374—381.
- Kosman D., Reinitz J., Sharp D. H. 1997. Automated assay of gene expression at cellular resolution. In: Proc. of the 1998 Pacific Symposium on Biocomputing. Singapore: World Scientific Press. 6—17.
- Kosman D., Small S., Reinitz J. 1998. Rapid preparation of a panel of polyclonal antibodies to *Drosophila* segmentation proteins. *Develop. Gen. Evol.* 208 : 290—294.
- Kozlov K., Myasnikova E., Pisarev A., Samsonova M., Reinitz J. 2002. A method for two-dimensional registration and construction of the two-dimensional atlas of gene expression patterns *in situ*. In *Silico Biology*. 2 : 125—141.
- Kozlov K., Myasnikova E., Samsonova M., Reinitz J., Kosman D. 2000. Method for spatial registration of the expression patterns of *Drosophila* segmentation genes using wavelets. *Computational Technologies*. 5 : 112—119.
- Kozlov K., Pisarev A., Matveeva A., Kaandorp J., Samsonova M. 2007. Image processing package ProStack for quantification of biological images. In: Proc. of the 4th International Symposium on Networks in Bioinformatics (ISNB). Amsterdam, The Netherlands. 204.
- Luengo-Hendriks C. L., Keränen S. V., Fowlkes C. C., Simirnenko L., Weber G. H., Henriquez C., Kaszuba D., Hamann B., Eisen M., Malik J., Sudar D., Biggin M. D., Knowles D. W. 2006. 3D measurement of morphology and gene expression in the *Drosophila* blastoderm at cellular resolution. *Genome Biology*. 7 : R123.
- Matveeva A., Kozlov K., Samsonova M. 2006. Extraction of quantitative gene expression data from the images of gene expression patterns with ProStack and iSIMBioS. In: Proc. of the 4rd TICSP Workshop on Computational Systems Biology (WCSB 2006). Tampere, Finland. 65—68.
- Myasnikova E., Samsonova A., Kozlov K., Samsonova M., Reinitz J. 2001. Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics*. 17 : 3—12.
- Myasnikova E., Samsonova A., Samsonova M., Reinitz J. 2002. Support vector regression applied to the determination of the developmental age of a *Drosophila* embryo from its segmentation gene expression patterns. *Bioinformatics*. 18 : S87—S95.
- Myasnikova E., Samsonova M., Kosman D., Reinitz J. 2005. Removal of background signal from *in situ* data on the expression of segmentation genes in *Drosophila*. *Develop. Gen. Evol.* 215 : 320—326.
- Ortiz de Solórzano C., Garcia Rodriguez E., Jones A., Pinkel D., Gray J. W., Sudar D., Lockett S. J. 1999. Segmentation of confocal microscope images of cell nuclei in thick tissue sections. *J. Microsc.* 193 : 212—226.
- Poustelnikova E., Pisarev A., Blagov M., Samsonova M., Reinitz J. 2004. A database for management of gene expression data *in situ*. *Bioinformatics*. 20 : 2212—2221.
- Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. V. 1988. *Numerical Recipes in C: the art of scientific computing*. Cambridge: Cambridge Univ. Press. 208 p.
- Shav-Tal Y. 2006. The living test-tube: imaging of real-time gene expression. *Soft Matter*. 2 : 361—370.
- Shen J., Castan S. 1986. An optimal linear operator for edge detection. In: Proc. of the IEEE Conference on computer vision and pattern recognition. Miami, FL. 109—114.
- Sorzano C. O. S., Blagov M., Thevenaz P., Myasnikova E., Samsonova M., Unser M. 2006. Algorithm for spline-based elastic registration in application to confocal images of gene expression. *Pattern Recognition and Image Analysis*. 16 : 93—96.
- Umesh Adiga P. S., Chaudhuri B. B. 1999. Efficient cell segmentation tool for confocal microscopy tissue images and quantitative evaluation of FISH signal. *Microsc. Res. Tech.* 44 : 49—68.
- Unser M. 1996. A practical guide to the implementation of the wavelet transform. In: *Wavelets in medicine and biology*. A. Aldroubi M. A. Unser (Eds). Boca Raton, FL, USA: CRC Press. 37—73.
- Unser M., Aldroubi A., Schiff S. 1994. Fast implementation of the continuous wavelet transform with integral scales. *IEEE Trans. Signal Process.* 42 : 3519—3523.
- Vincent L. 1993. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Transactions Image on Processing*. 2 : 176—201.

Поступила 23 IV 2007

METHODS FOR ACQUISITION OF QUANTITATIVE FROM CONFOCAL IMAGES OF GENE EXPRESSION *IN SITU*

S. Yu. Surkova,^{1,*} E. M. Myasnikova,¹ K. N. Kozlov,¹ A. A. Samsonova,² J. Reinitz,³ M. G. Samsonova¹

¹ Department of Computational Biology, Center for Advanced Studies, St. Petersburg State Polytechnical University, Russia, ² Department of Genetics, Harvard Medical School, Boston, USA, and ³ Department of Applied Mathematics and Statistics and Center for Developmental Genetics, Stony Brook University, Stony Brook, New York, USA; * e-mail: sestr sve@mail.ru

In this review we summarize original methods for the extraction quantitative information from the confocal images of gene expression patterns. These methods include image segmentation, extraction of quantitative numerical data on gene expression, removal of background signal and spatial registration. Finally it is possible to construct a spatiotemporal atlas of gene expression from individual images obtained at each developmental stage. Initially all methods were developed to extract quantitative numerical information from confocal images of segmentation gene expression in *Drosophila melanogaster*. Application of these methods to *Drosophila* images makes it possible to reveal new mechanisms of formation of segmentation gene expression domains, as well as to construct the quantitative atlas of segmentation gene expression. Most image processing procedures can be easily adapted to process a wide range of biological images.

Key words: methods of image processing, confocal microscopy, quantitative gene expression, image segmentation, spatial registration, background removal.