

С.А. Бабичев, П.И. Бидюк, А.И. Корнелюк, В.И. Литвиненко

## Анализ современного состояния проблемы предобработки данных при оценке уровня экспрессии генов

Рассмотрены методы создания и обработки микромассивов ДНК для оценки уровня экспрессии генов. Для каждого этапа эксперимента проведен анализ существующих методов и средств осуществления текущей операции, сформулированы основные преимущества и недостатки данного метода; выделены задачи, решение которых способствует повышению эффективности процесса оценки уровня экспрессии генов.

The methods of creating and processing DNA microarray for estimating the level of gene expression are studied. The existing methods and means of realizing the current operation are analyzed for each phase of the experiment. The advantages and disadvantages of the given method are formulated. Based on the done analysis, the tasks are allocated, their solutions promote the effectiveness of the process of the level of gene expression evaluation.

Розглянуто методи створення та обробки мікромасивів ДНК для оцінки рівня експресії генів. Для кожного етапу експерименту проведено аналіз існуючих методів і засобів здійснення поточної операції, сформульовано основні переваги та недоліки даного методу; виділено задачі, розв'язання яких сприяє підвищенню ефективності процесу оцінки рівня експресії генів.

**Введение.** Согласно современным толкованиям биологический организм представляет собой открытую термодинамическую систему, в которой происходят непрерывные преобразования энергии, определяющие характер функционирования данной системы [1]. При этом характер развития биологической системы в основном определяется генетической информацией, реализующейся через генетический код. В процессе экспрессии генов происходит ее декодирование – передача информации от генов к молекулярным структурам организма. Экспрессия генов осуществляется в клетке биологического организма как наименьшая часть организованной живой материи. Клетка представляет собой элементарную ячейку жизнедеятельности биологических организмов [2, 3], а ДНК содержит генетическую информацию о функционировании клетки, этапах ее дальнейшего развития и размножения. Таким образом, расшифровка информации, заложенной в ДНК, позволяет предсказать развитие организма, а также спрогнозировать возможные варианты развития генетических заболеваний с целью принятия мер по профилактике и лечению.

Данным вопросам в настоящее время посвящено достаточно большое количество работ. В [2–5] авторами рассматриваются и анализируются основные принципы применения совре-

менных информационных технологий для организации хранения, обработки и анализа биологических данных, приводится информация о структуре клетки и методам оценки экспрессии генов и белков. В [6] дан обзор основных направлений использования и развития технологий микромассивов ДНК. Под термином «микромассив» авторы понимают стеклянный или полимерный слайд, на который каким-либо способом наносятся молекулы мишени исследуемого вещества. Поскольку на современном этапе в большинстве случаев исследуемыми молекулами есть молекулы ДНК, то в дальнейшем будем считать, что микромассив представляет собой совокупность большого количества признаков, каждый из которых представляет собой уникальную молекулу ДНК.

Впервые комплексный подход для анализа микромассивов применил в 1995 г. *M. Schena* [7]. Он разработал технологию создания микромассивов, которая лежит в основе современных промышленных методов их производства [8, 9]. Область применения технологий микромассивов на современном этапе затрагивает следующие сферы деятельности: криминалистика [10, 11], биология [12–14], медицина [15–17], фармакология [18, 19], археология и антропология [20].

Технология микромассивов позволяет решить следующие задачи [21]:

**Ключевые слова:** микромассив ДНК, экспрессия генов, предобработка данных.

- определение суммарной экспрессии генов, соответствующих отдельной субпопуляции (суммарной экспрессии группы генов);

- идентификация генов, уровень экспрессии которых позволяет различить биологические объекты, неразличимые другими методами анализа (раковые опухоли);

- изучение моделей изменения активности генов при различных стрессовых условиях (уровень химической обработки).

В общем случае задачи анализа данных микромассивов можно разбить на два вида: задачи прогнозирования и задачи определения скрытых закономерностей в данных. Задача прогнозирования позволяет по имеющимся в данных закономерностям предсказать с определенной вероятностью характер поведения биологической системы на следующих этапах ее развития. Второй тип задач заключается в том, чтобы выявить в имеющихся данных скрытые закономерности, что позволяет на более глубоком уровне понять характер поведения данной биологической системы.

Анализ литературных источников по подготовке, проведению, анализу и обработке результатов эксперимента позволяет выделить следующие этапы [1, 2, 6, 9, 12, 17, 21]:

- формулировка целей и задач исследования, включающая в себя постановку проблемы, формирование задач исследования и формулировку возможных гипотез;

- проектирование эксперимента. На этом этапе осуществляется отбор исследуемых факторов, определение условий проведения эксперимента и определение критериев качества эксперимента;

- подготовка проб микромассивов в соответствии с поставленными задачами исследований. Описание приготовленных проб. Подготовка образцов с мРНК исследуемых генов.

- проведение процесса гибридизации, промывка полученного микромассива;

- сканирование изображения, в процессе чего осуществляется преобразование матрицы интенсивностей изображения в цифровую матри-

цу, значение каждой ячейки в которой пропорционально интенсивности света в ней;

- предварительная обработка полученного цифрового изображения, включающая в себя процесс нормализации данных микромассива, фоновая обработка изображения, фильтрация, т.е. минимизация шумовой компоненты сигнала;

- статистический анализ полученных данных, включающий в себя визуализацию полученной информации, кластеризацию или классификацию объектов исследования, нейросетевой анализ и т.д.;

- интерпретация полученных результатов, включающая в себя процесс кросс-валидации, биологическую валидацию, проведение статистических тестов, планирование дальнейших экспериментов, выдвижение новых гипотез и т.д.

Несмотря на то что микромассивы в настоящее время используются в различных сферах человеческой деятельности, задача получения полезной информации не тривиальна. Сложность заключается, прежде всего, в многомерности и уровне зашумленности исходных данных, что определяется существующими методами создания микромассивов и считывания с них информации. Поэтому разработка новых технологий и методик обработки данных ДНК микромассивов с целью повышения точности прогноза поведения исследуемой биологической системы одна из актуальных задач современной биоинформатики.

### **Биологические основы процесса оценки экспрессии генов**

Макромолекула ДНК, образованная двумя комплементарными полинуклеотидными цепями, несет генетическую информацию, закодированную в линейной последовательности нуклеотидов в каждой цепи ДНК [1–3], (рис. 1). Генетическая информация реализуется согласно правилам генетического кода – соответствия комбинаций нуклеотидных триплетов определенным аминокислотам в белках.

Процесс считывания информации с ДНК заключается в следующем: спираль ДНК разворачивается, каждая из спиралей копируется в матричную РНК (мРНК) в соответствии с

принципом комплементарности нуклеотидов. Этот процесс называется транскрипцией гена.



Рис. 1

В результате транскрипции образуется матричная РНК (мРНК), в которой содержится генетическая информация, закодированная в молекуле ДНК. Далее в процессе трансляции информация, содержащаяся в мРНК, считывается, в результате чего синтезируются определенные молекулы белков. Таким образом, гены содержат в себе информацию о том, какой тип белка, где и в каком количестве будет синтезирован. Вид синтезируемых белков и характер их взаимодействия определяют фенотип клетки организма. Блок-схема процесса экспрессии генов представлена на рис. 2.

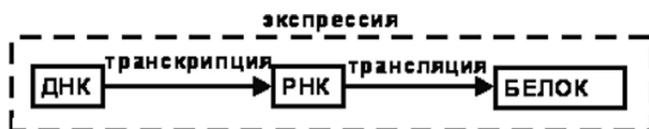


Рис. 2

Решение данной задачи позволяет сделать оценку количества вырабатываемых в организме белков, что дает возможность спрогнозировать характер дальнейшего развития организма в целом или его составляющей. Существуют методы измерения экспрессии генов на уровне образовавшихся белков. Однако методически проще сделать оценку экспрессии генов на уровне матричных РНК. Измеряя концентрацию образовавшейся мРНК, можно судить об уровне экспрессии данного гена.

Существует несколько способов оценки экспрессии генов при помощи определения концентрации мРНК.

*Первый* из них – секвенирование кДНК, синтезированной по мРНК, выделенной из определенной ткани организма [22–24]. К главным преимуществам метода секвенирования

относится низкий уровень шума на исследуемых образцах, что способствует высокой точности измерений. Основной недостаток – это сложность в подготовке опытных образцов. Высокие технические требования к опытным образцам и условиям проведения эксперимента резко повышают стоимость данного метода в сравнении с другими методами оценки экспрессии генов. Кроме того, следует отметить, что метод секвенирования эффективен для определения того или иного гена, оценка концентрации определенного типа генов, т.е. уровня экспрессии, связана с определенными сложностями.

Ко *второму* методу оценки уровня экспрессии генов отнесем метод полимеразной цепной реакции (ПЦР), комбинированной с обратной транскрипцией [25]. Суть данного метода заключается в многократном копировании в искусственных условиях определенного фрагмента молекулы ДНК, синтезированного по мРНК. Применение данного метода эффективно в случае, когда необходимо выявить наличие или отсутствие того или иного гена посредством выявления молекулы ДНК, соответствующего данному гену. Именно этим качеством определяются области применения данного метода, такие как криминалистика, медицинская диагностика, установление отцовства и т.д. К основным недостаткам метода относится то, что он позволяет определить наличие только одного специализированного гена. Поэтому его используют в основном для проверки результатов работы других методов по оценке концентрации мРНК.

*Третьим* методом оценки уровня экспрессии генов есть метод анализа микромассивов ДНК [6, 7]. Основным преимуществом данного метода считается то, что его использование позволяет одновременно оценить уровень экспрессии совокупности исследуемых генов, провести их сравнительный анализ на различных стадиях развития исследуемого организма. Кроме того, затраты на проведение эксперимента существенно ниже, чем при использовании других методов. Основной недостаток метода – высокая зашумленность экспериментальных

данных, возникающая как на этапе подготовки микромассива, так и на этапе считывания с него информации.

### Модификации микромассивов и методы синтеза ДНК-мишеней

Существующие в настоящее время микромассивы различаются по типу используемой основы и модификацией, нанесенной на основу поверхности [6, 17]. В качестве основы используется стекло, пластик, фольга, металл и другие твердые материалы. На основу наносятся различные биологические макромолекулы, такие как молекулы ДНК, РНК, белки, полисахариды. Иммуобилизуемые молекулы называются зондами, а молекулы, находящиеся в исследуемом образце – мишенями. Зонды наносят на поверхность, как правило, в определенной априори известной последовательности. ДНК-мишень может быть нанесена на поверхность двумя способами: доставкой или методом прямого синтеза (*in situ*) [26]. В первом случае мРНК синтезируются на мишенях при помощи полимеразной цепной реакции с использованием универсальных или геноспецифических праймеров или посредством олигонуклеотидного фосфоромидитного синтеза [6]. Преимущества данных методов заключаются в доступности, высокой эффективности синтеза мишеней, относительно низкой стоимости используемых микромассивов. Доставка мишеней осуществляется при этом при помощи контактного или неконтактного принтингов [27]. При контактном принтинге ДНК-мишени непосредственно контактируют с зондами на поверхности микромассива. При неконтактном принтинге капли с высококонцентрированными ДНК-мишенями распыляются на поверхность микромассива [28]. В результате применения метода доставки на поверхности микромассива образуются множество пятен диаметром 15–230 мкм, каждое из которых представляет уникальную ДНК-молекулу.

В основе применения метода прямого синтеза или синтеза *in situ* лежат полупроводниковые технологии, используемые в компьютерной индустрии [29]. Поверхность микромассива создается с использованием метода фотоли-

тографии, запатентованного для создания ДНК-микромассива компанией «Affymetrix». Данный метод предполагает использование большого количества фотомасок, представляющих собой фотозащитные группы, селективно возбуждаемые ультрафиолетовым светом. Фотомаски используются для направленного упорядоченного синтеза и позволяют производить микромассивы с плотностью ДНК-мишеней приблизительно 500 тыс. элементов на 1 см<sup>2</sup>.

### Процесс гибридизации и создание ДНК-микромассивов

ДНК-микромассив представляет собой организованное упорядоченное размещение олигонуклеотидных последовательностей на твердом плоском носителе [7]. Поверхность микромассива делится на сотни тысяч областей, в каждой из которых находятся пробы с известной последовательностью нуклеотидов, способных связать определенный участок мРНК. В исследуемом веществе находятся одинарные цепочки мРНК с нанесенными флуоресцентными метками, соответствующие тому или иному гену биологического организма. В соответствии с принципом комплементарности молекулы мРНК исследуемого вещества вступают в реакцию с соответствующими пробами на микрочипе. Чем больше молекул мРНК прореагирует с данной пробой, тем выше освещенность флуоресцирующего вещества в данном месте. Таким образом, по интенсивности света можно судить о концентрации прореагировавших молекул мРНК, т.е. об уровне экспрессии генов. Реакция объединения фрагментов мРНК исследуемого образца с комплементарной пробой на микромассиве называется процессом гибридизации (рис. 3).

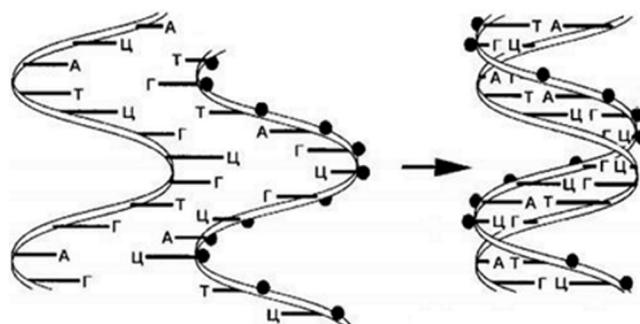


Рис. 3

В результате гибридизации происходит соединение комплементарных одноцепочных нуклеиновых кислот в одну молекулу, при этом между комплементарными основаниями образуются водородные связи.

Процесс создания ДНК-микромассива предполагает наличие следующих этапов:

- приготовление нужного количества исследуемых образцов, каждый из которых содержит идентичные последовательности мРНК, соответствующие заданному гену;

- нанесение флуоресцентной метки на исследуемые образцы, при этом каждому типу образца соответствует свой цвет;

- перемешивание полученных смесей и приведение их во взаимодействие с приготовленной ранее заготовкой микромассива с нанесенными зондами, роль которых играют априори известные последовательности нуклеотидов. Процесс гибридизации сопровождается медленным перемешиванием полученной смеси в течение 5–10 минут;

- промывка полученной смеси, в результате чего удаляются негибридизированные зонды, т.е. в процессе фильтрации на микромассиве остаются двухцепочные гибридизированные молекулы ДНК, соответствующие исследуемым генам.

Блок-схема процесса создания ДНК-микромассива приведена на рис. 4.

К основным преимуществам метода оценки уровня экспрессии генов при помощи ДНК-микромассива есть то, что он позволяет за короткое время получить оценку экспрессии десятков тысяч генов исследуемого образца, при

минимальных финансовых затратах на подготовку и проведение эксперимента. Однако данный метод имеет ряд недостатков, а именно:

- в процессе гибридизации между зондами на микромассиве и мишенями в образце могут возникать побочные реакции, вызванные тем, что некоторые гены в образце могут быть частично комплементарными зонду на микромассиве, но присутствующих связей может быть достаточно для вступления в реакцию. Данное явление получило название кросс-гибридизации [30]. Следствие кросс-гибридизации – возникновение биологического шума;

- с помощью микромассивов можно оценивать уровень экспрессии только тех генов, для которых предусмотрены пробы.

Отметим, что качественно проведенная гибридизация способствует уменьшению ошибки в процессе обработки результатов эксперимента по оценке уровня экспрессии генов. Следовательно, качество анализируемого микромассива определяется прежде всего качеством проведения процесса гибридизации.

### Методы получения изображения микромассива

Считывание информации с микромассива проводится методом сканирования или фотографическим способом. Получение изображения при помощи сканера описано в [31]. В процессе перемещения головки сканера происходит поэтапное считывание изображения с микромассива, при этом яркие оптические характеристики отдельных пятен на микромассиве преобразуются в цифровые параметры, пропорциональные цветовым характеристикам

соответствующего пятна. Основные характеристики оптических сканеров: чувствительность, разрешающая способность, скорость сканирования, динамический диапазон считываемой информации, число каналов передачи считываемой информации,

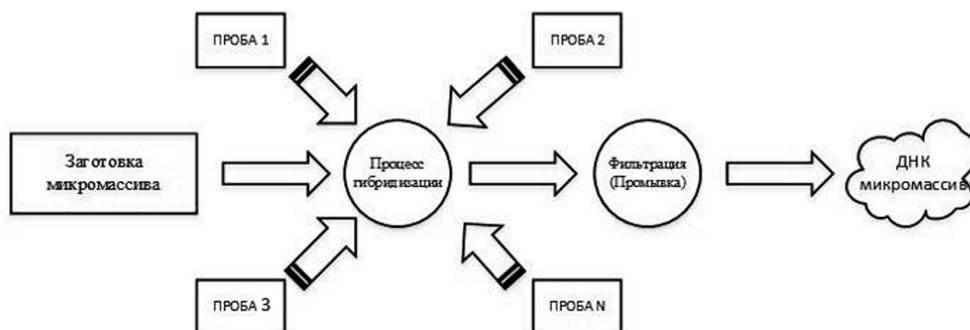


Рис. 4

детективность, размер кадра и формат представления считываемой информации [32]. Скорость перемещения головки оптических сканеров составляет приблизительно  $500 \text{ мм}^2/\text{мин}$  с разрешающей способностью пять мкм. Динамический диапазон определяется отношением максимального и минимального уровней сигнала, идентифицируемых сканером. Чувствительность определяется минимальной разницей сигналов, распознаваемых сканером. Детективность определяется минимальным уровнем сигнала, распознаваемым сканером. Число каналов определяется количеством длин волн, различаемых сканером при считывании информации. В процессе подготовки мишеней используются различные красители для изготовления меток соответствующих молекул мРНК. Очевидно, что цвет красителя различных молекул мРНК должен быть различим сканером, в противном случае цвета, соответствующие различным молекулам, не будут распознаны и погрешность эксперимента резко возрастет. Данный факт необходимо учитывать в процессе создания ДНК-мишеней. Механизм работы и устройство оптических сканеров, а также их сравнительная характеристика приводится в [33, 34].

При использовании фотографического метода считывания информации получается изображение микромассива на одном снимке. Устройство и принцип действия фотографического устройства приведен в [35]. Отметим, что фотографический метод обладает большей погрешностью считывания информации, вследствие чего он не получил широкого распространения.

Основное преимущество микромассивов ДНК заключается в том, что они позволяют одновременно анализировать огромное количество исследуемых генов. Вследствие этого очевидно, что микромассив ДНК должен обладать несколькими свойствами, которые противоречат одно другому. С одной стороны, увеличение количества ячеек микромассива способствует увеличению информации об исследуемом объекте. С другой – объем доступной для исследования информации ограничен, что позволяет уменьшить размеры микромассива.

Современные методы считывания информации такие, как атомно-силовая микроскопия [36, 37], позволяют детектировать сигнал в ячейках микромассивов размерами несколько десятков нанометров. Методы производства таких микромассивов основаны на нанолитографии [38]. Основным недостатком данного метода считается то, что производство таких микромассивов – процесс достаточно трудоемкий, что сказывается на их стоимости, поэтому в настоящее время наиболее распространен оптический метод считывания информации с микромассива.

Анализ методов производства микромассивов и считывания с них информации позволяет выделить следующие факторы, вызывающие вариацию признаков, соответствующих одному гену:

- биологические факторы, обусловленные характером протекания биологических процессов в результате проведения процесса гибридизации;
- технические факторы, обусловленные погрешностью процесса обработки микромассива, окрашивания образцов и гибридизации;
- технологические факторы, обусловленные погрешностью сканирования и обработки полученных изображений и включающие в себя фоновый шум в фотоусилительной трубе и оптический шум, связанный с рассеянием света при его отражении и прохождении через фильтры и линзы.

Анализ факторов, обуславливающих погрешность проведения эксперимента и обработки его результатов, позволяет сделать вывод, что реальными путями повышения точности проведения эксперимента есть уменьшение допусков на технические и технологические факторы. В первом случае требуемая точность может быть достигнута созданием более высоких технологий изготовления и подготовки микромассива. Во втором – при условии хорошего качества сканирования изображений, требуемой точности эксперимента можно достигнуть за счет создания эффективных путей обработки полученного изображения.

## Компьютерные методы предобработки изображения ДНК-микромассива

В системах компьютерной обработки визуальной информации анализируемый экспериментально полученный сигнал представляется следующим образом:

$$I = S + Ns + Nk, \quad (1)$$

где  $I$  – матрица значений, полученных в результате проведения эксперимента;  $S$  – полезная составляющая сигнала, характеризующего исследуемый образец;  $Ns$  – шумовая компонента исследуемого сигнала;  $Nk$  – некорректные локальные области на исходном микромассиве данных, возникающие по причине неспецифической гибридизации.

Ключевой проблемой при решении задач классификации или кластеризации объектов посредством анализа микромассивов ДНК есть выделение из исходной матрицы значений интенсивностей света, соответствующих уровню экспрессии таких генов, т.е. полезной составляющей, определяющей реальный уровень экспрессии того или иного гена. Успешное решение данной проблемы повышает точность прогноза состояния биологического объекта в процессе его диагностирования. Решению данной проблемы посвящены работы [21, 39–41]. Основные этапы анализа и обработки изображения ДНК-микромассива представлены на рис. 5.

### Методы учета фоновой поправки

На первом этапе обработки изображения проводится фоновая поправка, необходимость которой вызвана несовершенством системы сканирования изображения. Одним из первых методов коррекции фона изображения компанией *Affymetrix* предложен разностный метод *MM-PM-проб* (*Ideal Mismatch*). Для этого на микромассивах ДНК-проб помимо нуклеотидных зондов, в точности соответству-

ющих исследуемым генам (*Perfect Match probes*), размещались зонды, в которых средний нуклеотид замещался ему комплементарным (*Mismatch probes*).

По интенсивности свечения *MM-проб* судят о величине светового фона микромассива. Использование данного метода предполагает следующее: для каждого набора проб, соответствующих одному гену, вычислялось значение фоновой интенсивности света  $SB$ , представляющее собой одношаговое среднее, взвешенное по множеству логарифмов отношений *PM-интенсивностей* к *MM-интенсивностям* для каждой пары проб.

Пусть  $u_k = \{u_{1k}, u_{2k}, \dots, u_{ik}\}$  – множество логарифмов отношения интенсивностей *PM* и *MM*  $k$ -го подмножества проб, где  $i = 1, \dots, n_k$  – количество проб в  $k$ -м подмножестве. Для  $i$ -й пробы  $k$ -го подмножества логарифм отношения интенсивностей вычисляется следующим образом:

$$u_{ik} = \log_2 \frac{PM_{ik}}{MM_{ik}}. \quad (2)$$

Для определения среднего взвешенного Тьюки вычисляется для каждого элемента выборки отклонение от среднего по формуле:

$$\Delta u_{ik} = \frac{u_{ik} - M_k}{cS + \varepsilon}, \quad (3)$$

где  $M_k$  – медиана  $k$ -го подмножества проб;  $c$  – параметр, определяющий чувствительность системы к отклонению от среднего;  $\varepsilon$  – малый параметр, исключающий возможность деления на ноль. Параметры  $c$  и  $\varepsilon$  подбираются эмпи-

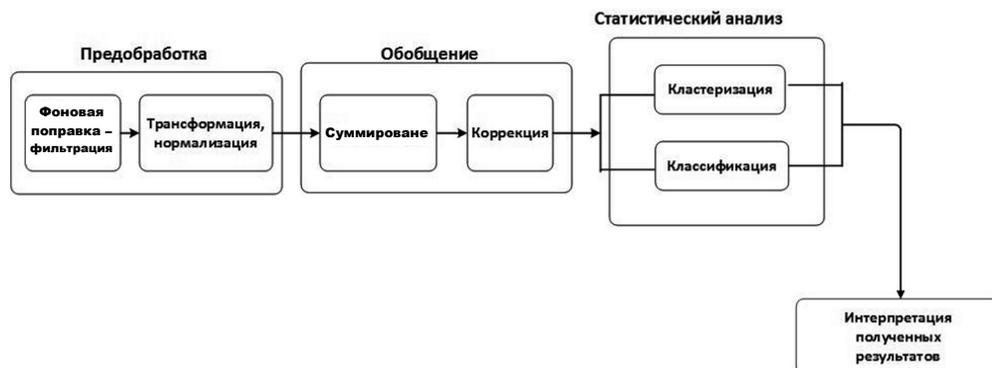


Рис. 5

рическим путем.  $S$  – абсолютное отклонение среднего, являющееся медианой следующей выборки:

$$\{|u_{1k} - M_k|, |u_{2k} - M_k|, \dots, |u_{ik} - M_k|\}.$$

Среднее взвешенное Тьюки вычисляется по формуле:

$$SB_k = \frac{\sum_{i=1}^n (w(\Delta u_{ik}) \cdot u_{ik})}{\sum_{i=1}^n w(\Delta u_{ik})}, \quad (4)$$

где  $w(\Delta u_{ik})$  – биквадратная функция:

$$w(\Delta u_{ik}) = \begin{cases} (1 - \Delta u_{ik}^2)^2, & \text{если } |\Delta u_{ik}| \leq 1, \\ 0, & \text{если } |\Delta u_{ik}| > 1. \end{cases} \quad (5)$$

Фоновая поправка для коррекции изображения вычисляется следующим образом:

$$\Delta I_{ik} = \begin{cases} MM_{ik}, & MM_{ik} \leq PM_{ik}, \\ \frac{PM_{ik}}{2^{SB_k}}, & MM_{ik} \geq PM_{ik}, SB_k > \tau_c, \\ \frac{PM_{ik}}{2^{\frac{\tau_c}{1 + (\tau_c - SB_k)/\tau_s}}}, & MM_{ik} \geq PM_{ik}, SB_k \leq \tau_c. \end{cases} \quad (6)$$

где  $\tau_c$  и  $\tau_s$  – константы различия и масштабирования, определяемые эмпирическим путем.

Итоговое значение интенсивностей соответствующей  $PM$ -пробы получают путем вычитания из исходной интенсивности элемента соответствующей ему фоновой поправки:

$$I_{ik} = I_{0ik} - \Delta I_{ik}.$$

Основной недостаток данного метода – высокий процент субъективизма при выборе эмпирических параметров и поправок, что повышает погрешность процесса предобработки исходного изображения микромассива ДНК.

В работе [41] представлен метод учета фоновой поправки, являющийся составляющей комплекса  $RMA$  (*Robust Multichip Average*) методов предобработки данных микромассивов ДНК. В данном методе используются только  $PM$ -пробы, что уменьшает затраты на подготовку микромассивов в отсутствие  $MM$ -проб. Анализируется каждый микромассив в отдельности. Интенсивность изображения микромассива ДНК представляется в виде суммы полез-

ного сигнала, распределенного по экспоненциальному закону, и шумовой составляющей, включающей в себя оптический шум, определяемый сканирующей системой, и биологический шум, определяемый биологическими процессами, протекающими в процессе создания микромассива:

$$I = S + Ns. \quad (7)$$

Предполагается, что шум распределен по нормальному закону. Очевидно, что задача заключается в извлечении из оригинального изображения  $I$  его полезной составляющей  $S$ .

Пусть  $\alpha$  – среднее значение экспоненциально распределенного сигнала;  $\mu$  и  $\sigma^2$  – математическое ожидание и дисперсия шумовой составляющей сигнала соответственно. Оценка полезного сигнала выполняется по следующей формуле:

$$S = a + b \frac{\varphi\left(\frac{a}{b}\right) - \varphi\left(\frac{I-a}{b}\right)}{\Phi\left(\frac{a}{b}\right) - \Phi\left(\frac{I-a}{b}\right) - 1}, \quad (8)$$

где  $a = I - \mu - \sigma^2\alpha$ ;  $b = \sigma$ ;  $\Phi$  и  $\varphi$  – функция распределения и плотность стандартного нормального распределения соответственно. Предложенный метод учета фоновой поправки также не оптимален по причине применения усредненных статистических характеристик к исходному изображению в целом. Кроме того, характер распределения интенсивностей на экспериментально полученном изображении может отличаться от характера распределения, предложенного в данном методе.

В работе [42] авторами представлена модель  $DFCM$  (*Distribution Free Convolution Model*) учета фоновой поправки, в которой интенсивность также представляется как сумма интенсивностей полезного сигнала и шумовой составляющей, но при этом не делается никаких предположений о характере распределения интенсивностей полезного сигнала и шумовой составляющей. Данный метод предполагает наличие  $PM$ - и  $MM$ -проб микромассива. Алгоритм работы данного метода состоит из следующих шагов:

Шаг 1. Выделяются наименьшие  $q_1$  процентов интенсивностей  $PM$ -проб. Значение  $q_1$  выбирается исходя из следующих соображений. Для повышения точности определения уровня экспрессии генов из микромассива желательнее удалить интенсивности, соответствующие не экспрессированным генам, которые можно считать одной из составляющих шумовой компоненты сигнала. В данном контексте параметр  $q_1$  рассматривается как количественная мера доли  $PM$ -проб, соответствующих не экспрессированным генам. Он эмпирическим путем подбирается так, чтобы доля  $MM$ -проб, интенсивности которых больше интенсивностей соответствующих им  $PM$ -проб, для наименьших  $q_1$  была приблизительно равна 50 процентам.

Шаг 2. Выбираются наименьшие  $q_2$  процента  $MM$ -проб, соответствующие  $PM$ -пробам, отобранным на шаге 1. Параметр  $q_2$  определяет вероятность того, что данная  $MM$ -проба вступает в неспецифическую гибридизацию, т.е. он – количественная мера фонового шума.

Шаг 3. Поиск распределения шума с использованием непараметрической оценки плотности распределения шума  $\hat{\mu}$ .

Шаг 4. Определяется стандартное отклонение шумовой компоненты  $\hat{\sigma}$ , как произведение стандартного отклонения шумовой составляющей, значение интенсивности которой меньше  $\hat{\mu}$ , на  $\sqrt{2}$ .

Шаг 5. Находится минимальное значение интенсивности  $PM$ -пробы ( $I_{\min}$ ).

Шаг 6. Определяется значение интенсивности  $k$ -й пробы  $i$ -го подмножества проб по формуле:

$$I_{ik} = \begin{cases} PM_{ik} - \hat{\mu}, & PM_{ik} \geq \hat{\mu} + 3\hat{\sigma}, \\ 1 + (PM_{ik} - I_{\min}) \left( \frac{2\hat{\sigma} - 1}{\hat{\mu} + 3\hat{\sigma} - I_{\min}} \right), & PM_{ik} < \hat{\mu} + 3\hat{\sigma}. \end{cases} \quad (9)$$

Данный метод, при условии выполнения гипотезы нормальности распределения, дает более высокую точность выделения полезного сигнала.

Каждый из перечисленных методов имеет свои достоинства и недостатки. Достоинство разностного метода  $MM$ - $PM$ -проб – невысокая

чувствительность к характеру распределения экспериментальных данных, а также возможность учесть эффект неспецифической гибридизации посредством введения  $MM$ -проб. Однако высокая степень субъективизма при введении поправок – его основной недостаток.  $RMA$ - и  $DFCM$ -методы достаточно хорошо работают при условии нормальности распределения данных, однако при отклонении характера распределения данных от нормального погрешность обработки данных возрастает.

### Трансформация, фильтрация и нормализация данных микромассива ДНК

Массивы данных, полученных посредством использования различных сканеров или каналов одного сканера, различаются диапазоном варьирования измеряемых величин. Для приведения их к диапазону, позволяющему провести сравнительный анализ, необходимо осуществить предварительную математическую обработку, т.е трансформировать данные так, чтобы они распределялись относительно выбранного центра в пределах одинакового диапазона. Вопросам трансформации данных посвящены работы [43, 44]. Трансформация данных состоит из процессов фильтрации, нормализации, центрирования и масштабирования. Классификация существующих методов фильтрации изображений представлена в работе [45]. Основу большинства используемых методов составляет Фурье-анализ, хорошо работающий для гладких дифференцируемых сигналов, но для сигналов с локальными особенностями его применение ограничено. В последнее время в различных областях обработки сигналов и изображений широкое применение нашел метод, основанный на вейвлетах [43–49]. Вейвлеты представляют собой локализованные в пространстве функции, способные отслеживать и необходимым образом обрабатывать локальные особенности сигналов и изображений. Эффективность использования вейвлетов для предварительной обработки микромассивов ДНК обусловлена характером распределения информационных данных на микромассиве. В процессе экспрессии гены распределяются на микромассиве так, что гены с разным уровнем экс-

прессии будут чередоваться с частотой, существенно более низкой частоты фоновой компоненты освещенности и частоты фоновых помех. Уровень экспрессии генов можно оценить по интенсивности освещенности в той или иной точке. В процессе вейвлет-декомпозиции из исходного сигнала выделяется его низкочастотная компонента, и при этом сохраняются локальные особенности этих данных. Основная идея вейвлет-декомпозиции сигнала заключается в следующем: исходное пространство интенсивности освещенностей, соответствующее различным уровням экспрессии генов, разлагается на систему подпространств так, что каждое последующее подпространство вложено в предыдущее:

$$I \subset I_1 \subset I_2 \subset I_3 \subset \dots \subset I_n. \quad (10)$$

Полученные подпространства при этом должны удовлетворять свойству:

$$\bigcap_{i=0}^n I_i = \{0\}. \quad (11)$$

Каждое из полученных подпространств разлагается, в свою очередь, с использованием вейвлет-функции и масштабирующей функции на пространства аппроксимирующих и детализирующих коэффициентов:

$$I_i = \{A_i, H_i, V_i, D_i\}, \quad (12)$$

где  $A_i$  – аппроксимирующие коэффициенты на  $i$ -м уровне вейвлет-декомпозиции сигнала,  $H_i, V_i, D_i$  – горизонтальные, вертикальные и диагональные детализирующие коэффициенты. Структурная схема вейвлет-декомпозиции сигнала представлена на рис. 6.

Аппроксимирующие коэффициенты несут информацию о низкочастотной составляющей сигнала, детализирующие коэффициенты содержат информацию о высокочастотной составляющей сигнала. В большинстве случаев полезный сигнал содержится в низкочастотной составляющей, а детализирующие коэффициенты несут информацию о локальных особенностях сигнала. Шумовая составляющая представляет собой высокочастотную компоненту, которая так же содержится в детализирующих

коэффициентах. Итак, можно сделать вывод, что обработав нужным образом детализирующие коэффициенты и восстановив сигнал из аппроксимирующих коэффициентов и обработанных детализирующих коэффициентов, получаем матрицу интенсивностей освещенностей при минимальном уровне шумовой компоненты, что повышает точность определения экспрессии генов.

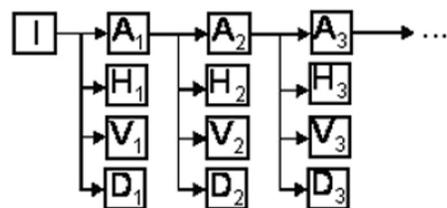


Рис. 6

Цель нормализации – приведение эмпирических данных микромассивов и их характеристик к одинаковому распределению, чаще всего к нормальному. В результате центрирования из массива данных удаляются данные, имеющие явно не свойственное данному распределению отклонения. После этого делается повторная операция масштабирования для выравнивания дисперсий данных различных микромассивов.

В общем случае все методы нормализации можно разделить на две подгруппы: методы, использующие опорное множество генов, и методы, использующие всю совокупность исследуемых данных. В первом случае методы нормализации подразделяются на линейные и нелинейные, во втором случае используют метод циклической локальной регрессии, метод контрастов и квантильную нормализацию [53, 54].

Алгоритм линейной нормализации применительно к анализу микромассивов ДНК предложен разработчиками *Affymetrix*. Он предполагает наличие следующих этапов:

- для каждого набора данных микромассива определяется усеченное среднее значение, при этом два процента минимальных и максимальных значений отбрасываются:

$$\bar{x}_k = \frac{\sum_{i=1}^{0,98N} x_{ik}}{0,98N}, \quad (13)$$

где  $k$  – количество массивов данных,  $N$  – количество данных в одном микромассиве;

- рассчитывается среднее значение данных в опорном микромассиве:

$$\bar{x}_{op} = \frac{\sum_{i=1}^N x_{iop}}{N}; \quad (14)$$

- нормализация данных во всех массивах осуществляется в соответствии с формулой:

$$x'_{ik} = \frac{\bar{x}_{op}}{\bar{x}_k} x_{ik}. \quad (15)$$

Недостатком метода линейной нормализации есть степень адекватности полученных данных, зависящих от выбора опорного микромассива. Кроме того, этот метод дает большую погрешность при наличии в данных нелинейных зависимостей, при этом применяется нелинейная нормализация [54]. В этом случае используют нелинейную функцию нормализации  $f(x)$ . В качестве функций нормализации могут использоваться сглаживающие сплайны, кусочно-линейная медиана, логистическая функция и т.д. Нормализованные значения в этом случае рассчитываются следующим образом:

$$x'_{ik} = f(x_{ik}). \quad (16)$$

Основной недостаток данного метода – это то, что его точность определяется выбором функции нормирования, которая в свою очередь определяется характером распределения исследуемых данных.

Основу метода циклической локальной регрессии составляет предположение, что между векторами  $M$  и  $A$  существует линейная регрессионная зависимость. Векторы  $M$  и  $A$  представляют собой совокупность следующих величин:

$$M_i = \log_2 \left( \frac{x_{ik}}{x_{in}} \right), \quad (17)$$

$$A_i = \log_2 (x_{ik} \cdot x_{in}), \quad (18)$$

где  $x_{ik}$  и  $x_{in}$  – значения интенсивности  $i$ -й пробы на  $k$ -м и  $n$ -м микромассивах соответственно. Нормирующая поправка вычисляется следующим образом:

$$\delta M_i = M_i - \hat{M}_i, \quad (19)$$

где  $\hat{M}_i$  – значение функции регрессии, соответствующее  $i$ -й пробе. Последний шаг – вычисление нормализованных значений интенсивностей проб:

$$x'_{ik} = 2^{\left( A_i + \frac{\delta M_i}{2} \right)}, \quad (20)$$

$$x'_{in} = 2^{\left( A_i - \frac{\delta M_i}{2} \right)}. \quad (21)$$

К недостаткам данного метода также можно отнести то, что при большом количестве микромассивов необходимо попарно обрабатывать всю совокупность экспериментальных данных, что увеличивает время проведения процесса нормализации. Кроме того, при большом разбросе данных использование линейной функции регрессии способствует искажению характера распределения исходных данных, что способствует повышению погрешности обработки результатов эксперимента.

Метод контрастов – в некотором роде альтернатива метода циклической локальной регрессии [55]. При использовании данного метода также строится регрессионная зависимость между векторами  $M$  и  $A$ , однако на первом этапе выбирается базовый микромассив и далее строятся регрессионные кривые для всех микромассивов по отношению к базовому. К преимуществам метода контрастов считается более высокая скорость обработки информации, однако ему присущи недостатки, свойственные методу циклической локальной регрессии.

Основная идея метода квантильной нормализации заключается в следующем [54]. Если предположить, что  $n$  проб имеют одинаковые распределения данных, то график квантилей в  $n$ -мерном пространстве будет представлять собой линию, лежащую вдоль диагонали, координаты

которой представлены вектором:  $\left( \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)$ .

Тогда для приведения  $n$  векторов к одному распределению, необходимо спроецировать точки  $n$ -мерного графика квантилей на данную диагональ. Пусть  $q_k = (q_{k1}, \dots, q_{kn})$  – вектор  $k$  квантилей для  $n$  микрочипов;

$d = \left( \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)$  –

вектор, задающий диагональ в  $n$ -мерном пространстве. Тогда проекции вектора квантилей на заданную диагональ рассчитываются следующим образом:

$$\text{proj}_d q_k = \left( \frac{1}{n} \sum_{j=1}^n q_{kj}, \dots, \frac{1}{n} \sum_{j=1}^n q_{kj} \right). \quad (22)$$

В работе [54] группой авторов проведен сравнительный анализ существующих методов нормализации на примере выборок данных, взятых с двух микромассивов ДНК. Результаты их исследований показали, что методы, не использующие эталон, дают лучшее качество нормализации данных. При попарной нормализации более эффективным оказался метод квантильной нормализации. Кроме того, он оказался самым быстрым из используемых методов.

#### **Критерии оценки качества предобработки данных**

В качестве *критериев* [56] для оценки степени вариации между массивами использовались: относительный логарифм экспрессии и нормализованная стандартизованная ошибка, входящие в качестве функций в пакет *affy-PLM*. Оценка качества РНК и процесса гибридизации выполняема с использованием следующих критериев, входящих в пакет *simpleaffy*: среднего значения фона; фактора нормализации; количества зондов на микромассиве, с которыми гибридизуется кРНК; соотношения интенсивностей сигналов, полученных вследствие взаимодействия 3'-участков ДНК глицеральдегидфосфатдегидрогеназы и  $\beta$ -актина, к интенсивности сигналов в зоне 5'-участков этих генов. Оценка исследуемых микромассивов с использованием перечисленных критериев показала, что в результате обработки все микромассивы сравнимы между собой и соответствуют общепринятым критериям качества. Однако следует отметить, что используемые в настоящее время критерии оценки качества микромассивов предполагают достаточно большие допуски на вариации признаков микромассива. При уменьшении допуска на вариации признаков для получения требуемой точности набор используемых критериев необходимо корректировать.

**Заключение.** Сегодня методы оценки уровня экспрессии генов получили значительное развитие. Методика подготовки образцов РНК производителями микромассивов и процесс создания на базе данной методики экспериментальной базы – достаточно высоко стандартизованы. В то же время компьютерные методы обработки экспериментальных данных дают большие возможности в выборе средств и методов обработки информации, от чего существенно зависит точность получения конечного результата. Производителями микромассивов созданы ряд пакетов, позволяющих осуществить предобработку экспериментальных данных с последующим проведением процессов классификации или кластеризации исследуемых объектов. Классификация используемого в настоящее время программного обеспечения представлена в работе [57]. Вместе с тем остается ряд нерешенных или частично решенных проблем, решение которых полностью позволит повысить точность прогноза при исследовании биологических объектов.

Существующие системы фильтрации полученных изображений микромассивов не обладают высокой эффективностью вследствие отсутствия эффективных алгоритмов удаления шумовой компоненты из высокочастотных данных сложной природы и отсутствия эффективной системы критериев, позволяющих оценить качество фильтрации микромассива.

Современные методы нормализации данных не позволяют получить распределения, способствующие адекватному разбиению исследуемых биологических объектов на классы или кластеры, что объясняется несовершенством существующих методов нормализации и алгоритмов проведения этого процесса.

Дальнейшее развитие методов предобработки данных микромассивов ДНК возможно путем усовершенствования алгоритмов обработки информации с использованием как существующих критериев оценки качества модифицированных данных, так и с учетом разработки новых критериев и их комбинаций, позволяющих выбрать оптимальный режим предобработки исследуемых данных.

1. Патрушев Л.И. Экспрессия генов. – М.: Наука, 2000. – 830 с.
2. Игнасимуту С. Основы биоинформатики. – М.: Ижевск, 2007. – 324 с.
3. Леск А. Введение в биоинформатику. – М.: Бином. Лаборатория знаний, 2009. – 326 с.
4. Baldi P., Brunak S. Bioinformatics: The machine learning approach. – Cambridge, Massachusetts, London, England, 2001. – 477 p.
5. Azuaje F., Dopazo J. Data analysis and visualization in genomics and proteomics // John Wiley and Sons, ltd. – 2005. – 269 p.
6. Ивахно С.С., Корнелюк О.И. Мікроареї: Огляд технологій та аналіз даних // Укр. біохім. журн. – 2004. – № 2(76). – С. 5–19.
7. Quantitative monitoring of gene expression patterns with a complementary DNA microarray / M. Schena, D. Shalon, R. Davis et al. // Science. – 1995. – **270**. – P. 467–470.
8. Blohm D.H., Guiseppi-Elie A. New developments in microarray technology // Curr. Opin. Biotechnol. – 2001. – **12**. – P. 41–47.
9. Lockhart D.J., Winzeler E.A. Genomics, gene expression and DNA arrays // Nature. – 2000. – **405**. – P. 827–836.
10. High-throughput variation detection and genotyping using microarrays / D. Cutler, M. Zwick, M. Carrasquillo et al. // Genome Res. – 2001. – **11**, N 11. – P. 1913–1925.
11. Kolchinsky A., Mirzabekov A. Analysis of SNPs and other genomic variations using gel-based chips // Hum. Mutat. – 2002. – **19**, N 4. – P. 343–360.
12. Scholtens D., A. van Heydebreck. Analysis of Differential Gene Expression Studies // Bioinformatics and Computational Biology Solutions Using R and Bioconductor Statistics for Biology and Health. – 2005. – P. 229–248.
13. Co-Regulation of Metabolic Genes Is Better Explained by Flux Coupling Than by Network Distance / R.A. Notabaart, Bas Teusink, R.J. Siezen et al. // Computational Biology. – 2008. – **4**. – P. 157–163.
14. Genome-wide association analysis identifies 20 loci that influence adult height / M.N. Weedon, H. Lango, C.M. Lindgren et al. // Nat. Genet. – 2008. – **40**. – P. 575–583.
15. DNA microarray-based detection and identification of fungal pathogens in clinical samples from neutropenic patients / B. Spiess, W. Seifarth, M. Hummel et al. // J. Clin. Microbiol. – 2007 Nov. – **45**, N 11. – P. 3743–3753.
16. Identification of human pathogens isolated from blood using microarray hybridization and signal pattern recognition / H. Wiesinger-Mayr, K. Vierlinger, R. Pichler et al. // BMC Microbiol. – 2007 Aug. – **14**, N 7. – P. 7–78.
17. Биомикрочипы и возможности их применения в дерматовенерологии / Н.В. Китаева, Н.В. Фриго, И.А. Волков и др. // Vestn. Dermatol Venerol. – 2009. – № 6. – С. 33–45.
18. Marton M.J. Drug target validation and identification of secondary drug target effects using DNA microarrays // Nature Med. – 1998. – **4**. – P. 1293–1301.
19. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high-density cDNA array / G. Pietu, O. Alibert, V. Guichard et al. // Genome Res. – 1996. – **6**. – P. 492–503.
20. Mount D.W. Bioinformatics: Sequence and Genome Analysis // Cold Spring Harbor Laboratory Press. – 2001. – Chapter 10. Genome Analysis. – P. 519–522.
21. Introduction to microarray data analysis / W. Dubitzky, M. Granzow, S. Downes et al. // Kluwer Acad. Publ. – 2003. – P. 1–46.
22. Sanger F., Coulson A. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase // J. Mol. Biol. – 1975. – **94**. – P. 444–448.
23. Sanger F., Niclein S., Coulson A. DNA sequencing with chain-terminating inhibitors // Proc. Natl. Acad. Sci. USA. – 1977. – **74**. – P. 5463–5467.
24. Maxam A., Gilbert W. A new method of sequencing DNA // Ibid. – P. 560–564.
25. Kleppe K. Studies on polynucleotides. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases // J. Mol. Biol. – 1971. – **56**. – P. 341–361.
26. Schena M., Davis R. Microarray biochip technology // Eaton Publishing. – 2000. – P. 1–18.
27. Microarray biochip technology / J. Worley, K. Bechtol, S. Penn et al. Ed. M. Schena // Ibid. – 2000. – P. 65–86.
28. Microarray biochip technology / L. Myles, Jr. Mace, J. Montagu et al. // Ibid. – 2000. – P. 39–64.
29. Handbook of Semiconductor Manufacturing Technology / Nishi Y. and Doering R. Eds. – New York: Marcel Dekker, 2000. – 1720 p.
30. Langdon W.B., Upton G.J., Harrison A.P. Probes containing runs of guanines provide insights into the biophysics and bioinformatics of Affymetrix Gene Chips // Briefings in bioinformatics. – 2009. – **10**, N 3. – P. 259–277.
31. Schermer M.J. DNA microarrays: a practical approach / Ed. M. Schena // Oxford University Press. – 1999. – P. 17–42.
32. Microarray biochip technology / T. Basarsky, D. Verdник, J. Zhai et al. // Eaton Publishing. – 2000. – P. 265–284.
33. Janesick J.R. Scientific Charge-Coupled Devices. Bellingham // WA: International Society for Optical Engineering. – 2001. – 899 p..
34. Baldi P., Hatfield G.W. DNA Microarrays and gene expression: From experiments to data analysis modeling // Cambridge University Press. – 2002. – P. 22–23.
35. Schena M. Microarray analysis. – New York: Wiley, 2002. – P. 240–251.

36. *Binnig G., Quate C.F.* Atomic Force Microscope // *Physical Review Letters*. – 1988. – N. 56(9). – P. 930–933.
37. *Bryant P.J., Miller R.G., Yang R.* Scanning tunneling and atomic force microscopy combined // *Applied Physics Letters*. – 1988. – **52**, Issue 26. – P. 2233–2235.
38. *Dip Pen Nanolithography / R. Pine, J. Zhu, F. Xu et al.* // *Science*. – 1999. – **283**. – P. 661–663.
39. *Data Preparation. Guide to Intelligent Data Analysis / M. Berthold, Ch. Borgelt, F. Hoppner et al.* – Springer-Verlag London Limited, 2010. – 394 p.
40. *A Novel Workflow for Microarray Data Analysis under Expression Level of Genes / Jianan Wu, Chunguang Zhou, Zhangxu Li et al.* // *J. of Information & Computational Sci.* – 2012. – **16**, N 9. – P. 4745–4754.
41. *Exploration, normalization, and summaries of high density oligonucleotide array probe level data / R. Irizarry, B. Hobbs, F. Collin et al.* // *Biostatistics*. – 2003. – **4**, N 2. – P. 249–264.
42. *Chen Z., McGee M., Liu Q.* Distribution-Free Convolution Model for background correction of oligonucleotide microarray data // *BMC genomics*. – 2009. – **10**. – P. 4–17.
43. *Data pre-processing issue in microarray analysis / N.A. Tinker, S.R. Lauriant, G. Butler et al.* // *Kluwer Acad. Publ.* – 2003. – P. 47–64.
44. *Causton H.C., Quackenbush J., Brazma A.* Microarray Gene Expression Data Analysis: A Beginner's Guide // *Wiley-Blackwell Science*. – 2003. – 184 p.
45. *From local kernel to nonlocal multiple-model image denoising / V. Katkovnik, A. Foi, K. Egiazarian et al.* // *Int J. Computer Vision*. – 2010. – **86**, N 8. – P. 1–32.
46. *Дьяконов В.П.* Вейвлеты. От теории к практике. – М.: СОЛОН-Р, 2002. – 448 с.
47. *Смоленцев Н.К.* Основы теории вейвлетов. Вейвлеты в *MATLAB*. – М.: ДМК, 2005. – 298 с.
48. *Добеши И.* Десять лекций по вейвлетам. – Ижевск.: НИЦ «Регулярная и хаотическая динамика», 2001. – 463 с.
49. *Уэлстид С.* Фракталы и вейвлеты для сжатия изображений в действии. – М.: Триумф, 2003. – 320 с.
50. *Поликар Р.* Введение в вейвлет-преобразование. – СПб.: АВТЭКС, 2001. – 260 с.
51. *Воробьев В.И., Грибунин В.Г.* Теория и практика вейвлет-преобразований. – СПб.: ВУС, 1999. – 210 с.
52. *Витязев В.В.* Вейвлет-анализ временных рядов: Учеб. пособие. – СПб.: Изд-во СПб ун-та, 2001. – 58 с.
53. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor / R. Gentleman, V. Carey, W. Huber et al.* – New York.: Springer-Verlag 2005. – 473 p.
54. *A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias / B. Bolstad, R. Irizarry, M. Astrand et al.* // *Bioinformatics*. – 2003. – **19**. – P. 185–193.
55. *Astrand M.* Contrast Normalization of Oligonucleotide Arrays // *J. of Computational Biology*. – 2003. – 10(1). – P. 95–102.
56. *Куклін А.В., Токовенко Б.Т., Оболенська М.Ю.* Критерії оцінки результатів мікромасив-експерименту з дослідження транскриптому гепатоцитів щура під впливом інтерферону альфа // *Biopolymers and Cell*. – 2013. – **29**, N 6. – P. 521–526.
57. *Leung Y.F., Lam D.S.C., Pang C.P.* Microarray Software review. – *Kluwer Acad. Publ.* – 2003. – P. 326–344.

Поступила 04.11.2014

Тел. для справок: 044 526-5589 (Херсон, Киев)

E-mail: [bsa63@mail.ru](mailto:bsa63@mail.ru), [pbidyuke@gmail.com](mailto:pbidyuke@gmail.com),

[kornelyuk@imbg.org.ua](mailto:kornelyuk@imbg.org.ua), [immun56@gmail.com](mailto:immun56@gmail.com)

© С.А. Бабичев, П.И. Бидюк, А.И. Корнелюк,

В.И. Литвиненко, 2015

## Внимание !

**Оформление подписки обязательно для желающих  
опубликовать статьи в нашем журнале.**

**В розничную продажу журнал не поступает.**

Подписной индекс 71008