

Моделирование генных сетей и клеточных ансамблей на платформе SETIES

Галимзянов А.В.*

Уфимский институт биологии УФИЦ РАН, Уфа, Россия

Аннотация. Генные сети, контролирующие экспрессию генов, играют первейшую роль в детерминации эпигенетических и биохимических свойств клеток. В то же время взаимосвязанность генных сетей с сетями других типов и уровней требует подходов, учитывающих эту комплексность. В статье данная проблема решается в аспекте вложенности генных сетей в клетки, которые сами являются узлами сети более высокого уровня, а именно клеточного ансамбля. Созданы информационные модели (объектная, реляционная и объектно-реляционная) управляющей генной сети над генетическими блоками, а также модели клеточного ансамбля как сети над клеточными управляющими генными сетями. Объединенная двухуровневая модель описывает генно-сетевые структуры в составе клеточных ансамблей с динамическими характеристиками основных молекулярно-генетических процессов. На основе моделей разработана платформа SETIES для исследования *in silico* молекулярно-генетических систем управления экспрессией генов про- и эукариот. Платформа включает в себя программные компоненты для дизайна и визуализации клеточных и межклеточных генных сетей, автоматического построения их математических моделей, моделирования сетевой динамики в бесклеточных системах, в клеточных линиях и клональных клеточных популяциях, а также в клеточных ансамблях с постоянной и динамической структурой. Счетные модули реализуют формализм обобщенных пороговых моделей, сочетающий в себе преимущества дискретных и непрерывных методов моделирования этих систем. Среди решаемых задач – построение диаграмм активности генов во времени, расчет кинетических кривых для концентраций мРНК и белков, выявление паттернов генной экспрессии как функциональных состояний генной сети, оценка параметрической устойчивости режимов, моделирование механизмов эпигенетического наследования, фенотипической гетерогенности и морфогенетического позиционирования. В целом, платформа реализует парадигму «сети сетей», служащую основой для перспективных исследований в области системной и синтетической биологии.

Ключевые слова: *сети сетей, системы управления, экспрессия генов, модели*

ВВЕДЕНИЕ

Сетевой принцип организации молекулярно-генетических систем фокусирует на себе передовые исследования в области системной биологии [1–3]. Важной особенностью биологических регуляторных сетей является их комплексность. Комплексность характеризуется сложными взаимодействиями между сетями и подсетями разных типов, слоев и уровней – взаимодействий как горизонтальных, так и вертикальных [4, 5]. Примером горизонтальных взаимодействий служат перекрывающиеся структуры, которые соединяются между собой связями через отдельные узлы или имеют в своем составе узлы, общие друг для друга [6, 7].

* galim@anrb.ru

Перекрывающиеся структуры могут создавать иерархии, основанные на связях между отдельными узлами двух и более смежных структур, внутри которых эти узлы связаны с ключевыми элементами или сами являются таковыми. Предельным случаем вертикальной связности выступают отношения восходящей / нисходящей вложенности. В этом случае структуры одного уровня оказываются сетью над элементами уровнем ниже и / или элементами, связанными сетью уровнем выше.

Так, полимерные молекулы ДНК, РНК и белков представимы как множество типовых узлов (нуклеотидов или аминокислотных остатков), соединенных химическими связями [8]. В свою очередь гены, РНК и белки образуют сети нескольких видов. Гены-регуляторы и гены-мишени с функциональными связями через кодируемые ими транскрипционные факторы формируют транскрипционные регуляторные сети [9]. Регуляторные взаимодействия на основе некодирующих РНК создают РНК-сети (RNA depended interactions, RDI-networks) [10, 11]. Взаимодействия белков при формировании белковых комплексов лежат в основе белок-белковых сетей (protein-protein interactions, PPI-networks) [12, 13]. Последовательности превращений метаболитов при участии ферментов организованы в метаболические сети, соединяющиеся с клеточными регуляторными процессами через глобальные и специализированные метаболиты-репортеры [14–16]. Совокупно межгенные, РНК-зависимые, белок-белковые и метаболические сети образуют клеточный интерактом. Ядром последнего является генная регуляторная сеть – группа совместно экспрессирующихся генов, связанная контурами прямых и обратных связей, как положительных, так и отрицательных [17]. Размер и сложность конкретной генной сети зависит от того, данные каких модальностей в нее включены. В одних случаях генная сеть тождественна транскрипционной регуляторной сети [18, 19], в других – объемлет длинный путь или сумму путей от физического или химического сигнала к гену, а от гена (генов) – к признаку [20–23].

В то время как клеточная генная сеть задает идентичность клетки в пространстве ее генома и регулома [24, 25], существуют ключевые гены (masters) базового контроля над состояниями самой сети [26–28]. В этой точке сетевая многослойность переходит на качественно иной уровень, а именно уровень клеточных ансамблей. Взаимодействия между клетками, содержащими однородные по своему составу генные сети, могут быть направлены как в сторону синхронизации их функциональных состояний (эпигенотипов), так и в сторону устойчивых различий, т.е. специализации. Хорошо известным механизмом синхронизации является чувство кворума у бактерий [29, 30]. К драйверам специализации можно отнести позиционирование клеток зародыша в системы морфогенетических координат [31, 32].

В рамках межклеточных взаимодействий на входы внутриклеточных генных сетей поступают регуляторные сигналы, которые формируются в контексте избирательной проницаемости клеточных мембран, активности в межклеточном пространстве внеклеточных ферментов, морфогенетических движений клеток, количества клеток и плотности клеточной популяции. Такие сигналы доминируют над внутрисетевыми взаимодействиями, управляя сетевой динамикой. Исходя из этого, можно говорить о клеточном ансамбле как сети над внутриклеточными генными сетями. Необходимость интеграции активности генов с активностью клеток в форме взаимодействия генных регуляторных сетей и клеточных регуляторных сетей обсуждается в работе [33].

Все перечисленные сетевые типы регуляторного аппарата клетки, взятые в том или ином сочетании, относятся к системам класса «сети сетей». Для изучения подобных систем актуальны комплексные многоуровневые информационные модели [34, 35], разработанные на их основе базы данных [36], инструменты визуализации [37–39], математическое обеспечение для идентификации и анализа сетевой динамики [40–44], а также средства для репрезентации и фондирования моделей [45]. Более того, конструируются модели с многослойной архитектурой, учитывающей иерархию всех

основных уровней биологической организации – генетического, метаболического, клеточного, популяционного и экологического [46].

В настоящей работе представлен единый формализм, интегрирующий микроуровень внутриклеточных генных сетей и макроуровень клеток как элементов клеточных ансамблей. Структура статьи такова. Сначала кратко излагаются используемые математические методы для анализа сетевой динамики [47–52]. Затем дается полное описание информационных моделей – объектной модели, модели «сущность – связь», положенной в основу реляционной модели, и специального текстового формата их репрезентации. Далее описывается разработанная на основе семейства математических и информационных моделей платформа SETIES для исследования природных и синтетических молекулярно-генетических систем. И, наконец, приводятся результаты расчетов модельных генных сетей в составе клеточных ансамблей трех типов.

МЕТОДЫ И АЛГОРИТМЫ

Обобщенные пороговые модели

Динамика генных сетей рассчитывается методом обобщенных пороговых моделей (ОПМ) [47–48]. В методе ОПМ генам сети сопоставляются формальные *генетические блоки*. Генетический блок – это ген, взятый в совокупности с механизмами транскрипции, процессинга, транспорта и депо конечного продукта. Входными переменными генетического блока являются концентрации регуляторных веществ (транскрипционных факторов) для соответствующего гена, а выходами – переменные, отражающие концентрации продуктов гена (мРНК и белка).

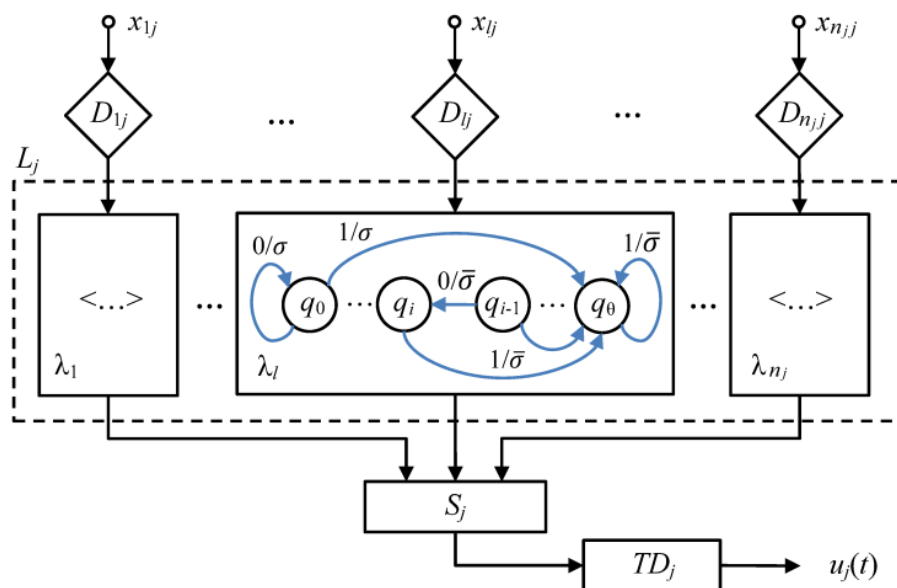


Рис. 1. Микроструктура управляющей подсистемы генетического блока. Дискриминатор D_{ij} перерабатывает непрерывные величины $x_{ij}(t)$ в дискретные бинарные величины $\varepsilon_{ij}(t)$ таким образом, что $\varepsilon_{ij}(t) = H[x_{ij}(t) - P_{kj}]$, где H – функция Хевисайда, $x_{ij}(t) = r_k(t)$, $r_k(t)$ – концентрация белкового продукта k -го блока, регулирующего j -й блок; P_{kj} – пороговое значение концентрации; λ_i – элементарный подавтомат автомата L_j , учитывающий образование комплекса «регуляторный белок – сайт ДНК» и длительность его существования (период рефрактерности θ_{ij}); S_j – комбинатор двоичных сигналов от элементарных подавтоматов; TD_j – элемент задержки; $u_j(t)$ – управляющая переменная. В диаграмме Мура для подавтомата: σ – собственное состояние гена (активен или неактивен); q_i – возраст комплекса регулятор-полимер

Информационную микроструктуру генетического блока G_j образуют дискриминаторы D_{ij} , элементарные подавтоматы λ_{ij} , комбинационная схема S_j и элемент временной задержки TD_j (рис. 1). В дискриминаторе D_{ij} величина x_{ij} ,

сопоставленная концентрации r_k белка блока-регулятора G_k , сравнивается с пороговым значением P_{kj} , при достижении которого происходит эффективное белок-ДНК взаимодействие. Элементарный подавтомат λ_{lj} описывает процесс связывания регуляторного белка с операторным участком. Комбинационная схема S_j , в зависимости от типа регуляторных отношений, вырабатывает интегральную реакцию при воздействии разных регуляторных молекул на функционирование данного генетического блока. Элемент TD_j учитывает временную задержку, необходимую для транскрипции соответствующей нуклеотидной последовательности и созревания молекулы мРНК. Дискретный логический элемент L_j , состоящий из элементарных подавтоматов λ_{lj} и комбинационной схемы S_j , в зависимости от концентраций регуляторных веществ формирует значения управляющей переменной $u_j = u_j(t)$ с учетом временной задержки.

Базовая модель управляющей подсистемы учитывает двух- и трехуровневую экспрессию генов. Многоуровневую экспрессию охватывает расширенная модель. Она описывает совокупность регуляторных сайтов гена как композиционный датчик, воспринимающий воздействия среды (регуляторных белков) (рис. 2). Последний предполагает разбиение обобщенной регуляторной зоны гена на локальные регуляторные фрагменты (регуляторные модули), приносящие свой вклад в совокупную интенсивность транскрипции гена. Вклады r -модулей в активацию транскрипции суммируются. При этом в случае достижимости своих порогов белками-репрессорами, специфичными ДНК-сайтам этих фрагментов, отключаются отдельные регуляторные модули, а не блокируется целиком инициация транскрипции гена.

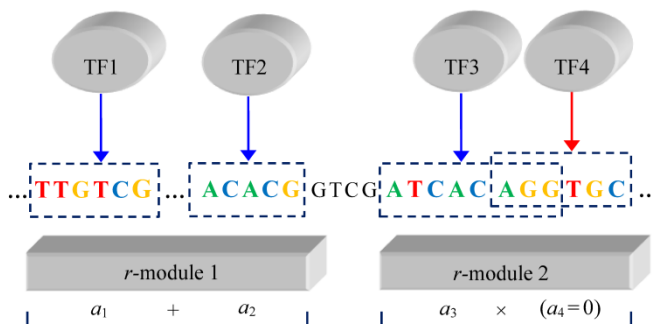


Рис. 2. Мультимодульность описания регуляторных районов

Регуляторный модуль можно формировать как в прямом соответствии подмножеству сайтов регуляторной последовательности гена, так и с учетом одной только логики регуляторного механизма, без явной привязки к районам ДНК. Таким образом, один фрагмент регуляторной последовательности гена может быть описан несколькими r -модулями, и наоборот, несколько таких фрагментов могут быть описаны одним r -модулем.

Управляющая переменная определяет режим функционирования исполняющего устройства, которое описывает механизмы синтеза мРНК и белка и задается системой кусочно-линейных дифференциальных уравнений.

В случае прокариотических систем уравнения динамики для концентраций мРНК и белков имеют следующий вид:

$$\dot{m}_j(t) = a_{1j}u_j(t) - b_{1j}m_j(t), \quad a_{1j} = \sum_{l=1}^{n_j} a_{1j}^l u_{lj}(t), \quad (1)$$

$$\dot{r}_j(t) = a_{2j}m_j(t) - b_{2j}r_j(t), \quad (2)$$

где $m_j(t)$ и $r_j(t)$ – концентрации мРНК и белка, соответственно, выраженные числом молекул на клетку; $u_{lj}(t)$ – компонента управляющего вектора; a_{1j} – интенсивность

транскрипции (сила промотора) (молек. · мин⁻¹); a_{1j}^l – вклад l -го (из n_j) регуляторного сайта в скорость транскрипции; a_{2j} – интенсивность трансляции (молек. мин⁻¹); b_{1j} и b_{2j} – скорости деградации, соответственно, мРНК и белка (мин⁻¹), вычисляемые как $\ln 2 / hl_{ij}$ ($i = 1$ или 2), где hl_{ij} – время полужизни вещества (мин).

В случае эукариотических систем уравнения динамики видоизменяются с учетом временной задержки на созревание мРНК и ее транспорт из ядра в цитоплазму.

В целом модель позволяет исследовать динамику генных сетей с учетом мультипорогового действия транскрипционных факторов и многоуровневой экспрессии генов.

Модель растущей клеточной популяции

Развитие клеточной популяции в экспоненциальной и стационарной фазах роста моделируется полным бинарным деревом клеточных делений [49]. В нем вершины – клетки, ребра – порождающие связи между материнской и двумя дочерними клетками. Для обхода дерева используется поиск в глубину, прямой (префиксный) его вариант. Для каждой клетки производится расчет индивидуальной ОП-модели с новыми начальными концентрациями мРНК и белков и особыми значениями отдельных параметров. Модель учитывает три различные функции роста клеток – экспоненциальную [50], линейную [51] и правило «noisy linear map» (NLM) [53], а также модуляцию скорости инициации транскрипции с учетом удвоения количества ДНК в клеточном цикле. Стохастическая компонента формируется за счет варьирования времени генерации клеток / времени удвоения популяции, начального и конечного объема клеток, скорости роста клеток, а также скоростей синтеза и деградации транскриптов и белков. Особый вклад вносит биномиальное распределение продуктов генов между дочерними клетками при делении материнской [49, 50].

В целом, подход позволяет отслеживать динамику генных сетей в последовательной ряду клеточных делений, выдавая на выходе профили экспрессии генов в клеточных линиях и процентное соотношение между функциональными режимами в клетках промежуточных и финального поколений.

Модель диффузии

Пусть дано клеточное множество C мощности I (количество клеток). На множестве C введена функция соседства Ψ , сопоставляющая каждому элементу $c_i \in C, i = \overline{1, I}$ множество C_i , элементы которого – суть соседние для c_i клетки ($C_i \subseteq C$). Под соседней клеткой понимается такая клетка, что между ней и элементом c_i существуют межклеточные взаимодействия (контакты). В частном случае функция Ψ может быть задана простым перечислением имен смежных клеток для каждой клетки из C . Контакты образованы каналами по веществам из S . Множество $S = (s_1, \dots, s_k, \dots, s_K)$ – перечень молекулярных компонент (РНК, белки, мультимерные комплексы), которые могут присутствовать в клетках, например, в результате синтеза или поступления извне.

Элемент $c_i \in C$ задается пятеркой:

$$\langle O_i(t), C_i(t), S_i, W_i, \overline{\xi}_i(t) \rangle \quad (3)$$

где $O_i(t) = (x_i(t), y_i(t), z_i(t))$ – координаты условного центра клетки в трехмерном пространстве; $C_i(t) = \{c_{i_1}, \dots, c_{i_j}, \dots, c_{i_{J_i}}\}$ – множество соседних клеток мощности J_i ; $S_i = \{s_{i_1}, \dots, s_{i_k}, \dots, s_{i_{K_i}}\}$ – множество внутриклеточных веществ из S (полагаем далее $S_i = S \forall i$); $W_i = (w_{kj}^i)_{K \times J_i}$ – матрица весовых коэффициентов, в которой элемент w_{kj}^i

означает силу взаимодействия (интенсивность межклеточного контакта) между клеткой c_i и ее j -м соседом по веществу s_k ; $\bar{\xi}_i(t) = (\xi_{i1}(t), \dots, \xi_{ik}(t), \dots, \xi_{iK}(t))$ – концентрации соответствующих молекулярных компонент в клетке в момент времени t .

Таким образом, клеточный ансамбль CE задается набором элементов вида (3). Совокупность множеств $C_i(t)$, а также матриц W_i определяют конфигурацию клеточного ансамбля. На рассматриваемом временном интервале будем считать, что: а) количество клеток остается одним и тем же; б) пространственные координаты всех клеток фиксированы (клетки не перемещаются); в) отношения соседства неизменны. Условия (а-в) означают, что конфигурация клеточного ансамбля постоянна.

Состояние $\Omega(t)$ системы CE в момент времени t описывается набором векторов $\bar{\xi}_i(t) (i = \overline{1, I})$, т.е. матрицей $\Xi(t) = (\xi_{ik}(t))_{I \times K}$, элемент $\xi_{ik}(t)$ которой означает концентрацию k -го вещества в i -й клетке в момент времени t . Матрица $\Xi(t)$ отражает паттерн распределения концентраций молекулярных компонент из множества S по клеткам из множества C в момент времени t . Начальные данные для системы CE задаются матрицей $\Xi(t_0)$. Таким образом, описание динамики системы CE заключается в получении последовательности матриц $\Xi(t_1), \Xi(t_2), \dots, \Xi(t_T)$ на интервале наблюдений $[t_1, t_T]$.

Принят ряд допущений относительно рассматриваемого процесса.

1. Будем считать, что молекулы вещества выходят из компартмента с большей концентрацией вещества (клетки-доноры, d -клетки) в компартмент с меньшей концентрацией (клетки-акцепторы, a -клетки) в сторону выравнивания концентраций в обоих компартментах. Статус клетки по веществу s_k : донор, акцептор.

2. Количество молекул вещества, переходящих из одной d -клетки в конкретную соседнюю a -клетку, зависит от значения весового коэффициента по этому веществу и от разности между уровнями концентраций вещества в этих двух клетках, и не может превышать величины $w_{ka}^d (\xi_{dk} - \xi_{ak})$, $w_{ka}^d \in [0, 0.5]$.

3. Перераспределение молекул вещества направлено в сторону выравнивания концентраций вещества между клеткой c_i и всеми соседними для нее a -клетками.

4. Количество молекул вещества, переходящих из одной d -клетки в конкретные соседние a -клетки, зависит от разности между уровнями концентраций вещества в них и весовых коэффициентов по этому веществу.

5. Перенос вещества из клетки-донора c_i в клетку-акцептор c_j возможен только в том случае, если разница между уровнем концентрации вещества в клетке c_i в предыдущий момент времени $(t - 1)$ и совокупным выходом молекул этого вещества в другие a -клетки в текущий момент (t) больше, чем концентрация вещества в c_j .

6. Количество вещества, убывающего из клетки c_i в момент времени $(t + 1)$, не может быть больше, чем количество вещества в ней на момент времени t .

Условие (1) непосредственно отражает направление процесса диффузии. В соответствии с ним, невозможно, в частности, попадание молекул вещества из a -клетки в d -клетку. Условием (2) вводятся локальные коэффициенты диффузии, смысл которых имеют весовые коэффициенты $w_{kj}^i [0, 0.5]$ (из условия 1). Условия (3–5) описывают особенности процесса перераспределения молекул вещества из d -клетки между соседними a -клетками с учетом контекста (уровни концентраций вещества в каждой из клеток относительно друг друга и d -клетки, значения весовых коэффициентов). Условие (6) устанавливает принцип сохранения количества вещества в системе CE на всем интервале наблюдений.

Далее будем рассматривать произвольную клетку c_i (назовем ее базовой), множество соседних для нее клеток C_i и вещество s_k .

Выделим из множества C_i для клетки c_i три непересекающихся подмножества:

$C_i^D(t, s_k) = \{c_{i_j} \mid \xi_{i,j,k}(t) > \xi_{ik}(t), j \in \{1, \dots, J_i\}\}$ – множество клеток-доноров;

$C_i^A(t, s_k) = \{c_{i_j} \mid \xi_{i,j,k}(t) < \xi_{ik}(t), j \in \{1, \dots, J_i\}\}$ – множество клеток-акцепторов;

$C_i^E(t, s_k) = \{c_{i_j} \mid \xi_{i,j,k}(t) = \xi_{ik}(t), j \in \{1, \dots, J_i\}\}$ – множество нейтральных клеток c_i .

Множества $C_i^D(t, s_k)$, $C_i^A(t, s_k)$, $C_i^E(t, s_k)$ имеют мощности J_i^D , J_i^A и J_i^E , соответственно, причем $J_i^D + J_i^A + J_i^E = J_i$.

Для удобства записи обозначим базовую клетку c_i – символом ζ^* , клетки из $C_i^A(t, s_k)$ – символами ζ_l ($l = \overline{1, L}$, где $L = J_i^A$), локальные концентрации вещества s_k в этих клетках в момент времени t – переменными ξ^* , $\xi_1, \dots, \xi_l, \dots, \xi_L$, соответственно.

Сформулируем задачу распределения молекул вещества между базовой и соседними клетками в терминах задачи линейного программирования (ЛП) со смешанными ограничениями [54]:

Максимизировать линейную форму $\Phi(\mathbf{x}) = \sum_{l=1}^L x_l$ на множестве векторов

$\mathbf{x} = (x_1, \dots, x_L)$, удовлетворяющих условиям:

$$1) x_l \geq 0 \quad (l = \overline{1, L});$$

$$2) x_l \leq w_l \left(\left(\xi^* - \sum_{j \neq l}^L x_j \right) - \xi_l \right) \quad (l = \overline{1, L}),$$

где x_l – выход вещества s_k из базовой клетки ζ^* в клетку ζ_l в момент времени t ; $w_l \in [0, 1/2]$ – весовой коэффициент (аналог коэффициента диффузии), означающий силу контактного взаимодействия между базовой клеткой и l -й клеткой-акцептором по веществу s_k в направлении от ζ^* к ζ_l .

Ограничения (1) и диапазон значений весовых коэффициентов w_l отражают направление процесса диффузии в сторону выравнивания уровня концентраций вещества в соседних компартментах. Согласно ограничениям (2), количество молекул вещества, переходящих из клетки-донора в конкретную соседнюю клетку-акцептор, зависит от значения весового коэффициента по этому веществу и от разности между уровнями концентраций вещества в двух клетках, и не может превышать величины $w_l (\xi^* - \xi_l)$. Кроме того, перенос вещества из ζ^* в ζ_l возможен только в том случае, если разница между уровнем его концентрации в базовой клетке и совокупным выходом молекул этого вещества в другие клетки-акцепторы ($C_i^A(t, s_k) \setminus \zeta_l$) на текущем такте t больше, чем $\xi_l(t)$.

Задача ЛП решается симплекс-методом [54] для каждой ячейки (клетки, компартмента) и ее окружения (совокупности соседних ячеек) по каждому генному продукту на очередном такте функционирования модели.

Принципиальная схема алгоритма

1. Каждая клетка c_i из C в ситуации $\Xi(t)$ последовательно выбирается в качестве базовой (ζ^*). Формируется соответствующее ей множество $C_i^D(t, s_k)$ клеток-доноров и множество $C_i^A(t, s_k)$ клеток-акцепторов.

2. Посредством решения задачи ЛП рассчитывается максимальный выход генного продукта из базовой клетки на t -м такте: $\phi_k^i = \max_{x_1, \dots, x_{J_i^A}} \Phi(\mathbf{x})$, а также порции этого продукта для соседних клеток-акцепторов – вектор $\mathbf{x}^*(t)$.

3. По окончании выполнения последовательности шагов (1–2) формируется набор чисел $\psi_k^i (i = \overline{1, I}) \mid \Psi_k^i = \sum_{j=1}^{J_i^D} \psi_k^{i,j}$, где $\psi_k^{i,j}$ – количество молекул вещества s_k , перешедших в клетку i из клетки $j \in C_i^D(t, s_k)$ на t -м такте. Таким образом, Ψ_k^i – общая прибыль вещества s_k в клетке i в результате его диффузии из всех соседних клеток-доноров на t -м такте.

4. Рассчитываются величины $\Delta_k^i(t) = \Psi_k^i(t) - \Phi_k^i(t)$ – изменение количества вещества s_k в клетке i в результате функционирования системы на t -м такте.

5. Проводится пересчет распределения молекул вещества между всеми элементами клеточного множества, то есть формируется метаболический профиль $\Xi(t+1)$:

$$\xi_{ik}(t+1) = \xi_{ik}(t) + \Delta_k^i(t) \quad (i = \overline{1, I}, k = \overline{1, K})$$

Критерий останова. Выполнение шагов (1–5) продолжается до тех пор, пока не будет выполнено условие: $\forall i, k (i = \overline{1, I}, k = \overline{1, K}) \Psi_k^i(t) = 0, \Phi_k^i(t) = 0$.

Интеграция алгоритма с ОПМ-формализмом позволяет моделировать динамику управляющих генных сетей с учетом межклеточных взаимодействий, основанных на явлении простой и облегченной диффузии.

МОДЕЛИ ДАННЫХ

Объектные модели

Разработаны две объектные модели генных сетей. Первая модель (O_{GN}) описывает произвольную генную сеть без привязки к математическому формализму, отражая только ее структуру и функциональные характеристики. К структурным характеристикам относится список генов, матрица связей между генами и тройки <связь, сайт, модуль>. К функциональным характеристикам относятся концентрации мРНК и белков, пороговые концентрации белков, периоды рефрактерности, временные задержки, длительности процессинга и транспорта мРНК, скорости деградации транскриптов и белков. Модель образована базовым классом (класс TGeneNet) и классами, описывающими сущности четырех типов: генетический блок (класс TGBlock), регуляторная связь (класс TRBond), регуляторный сайт (класс TRSite), регуляторный модуль (класс TRModule). Модель реализуется двунаправленным односвязным списком, что позволяет динамическим способом собирать ее структуру из произвольного количества элементов стандартных типов («Генетический блок», «Регуляторный район / операторный участок» и «Регуляторная связь»).

Вторая объектная модель (O_{GT}) базируется на математической модели молекулярно-генетических систем управления в методе обобщенных пороговых моделей. На основе описания элементов ОП-модели разработаны абстрактные классы T MGCS, TControlSystem, TControlledSystem, TTimeDelayElement и др. Класс TControlSystem реализует логику функционирования дискретных элементов математической модели, включая работу специальных конечных автоматов и вычисление значений логических функций (ДНФ). Класс TControlledSystem содержит функции, реализующие аналитические решения систем дифференциальных уравнений (1 и 2).

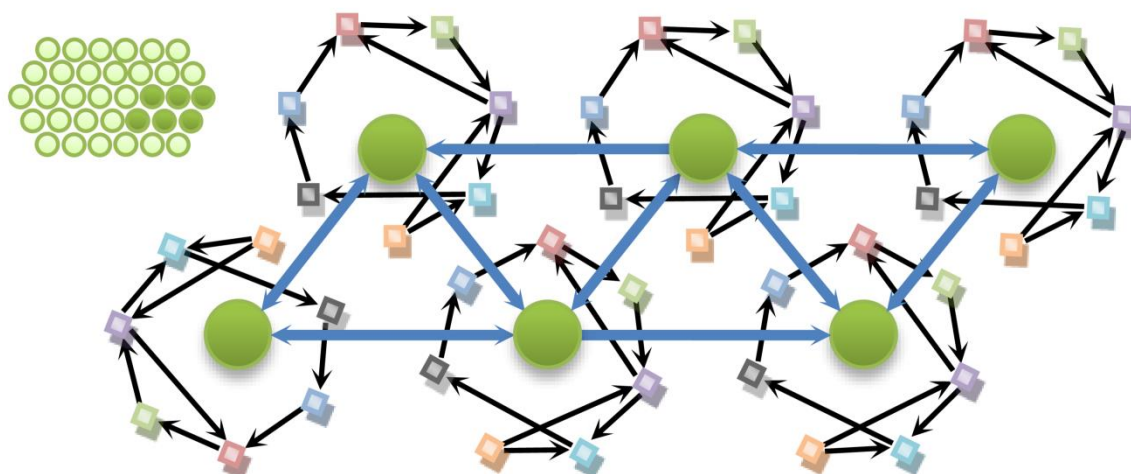


Рис. 3. Клеточный ансамбль как сеть над внутриклеточными генными сетями

Разработана также объектная модель интеграции двух уровней организации молекулярно-генетической информации – внутриклеточных генных сетей и клеточных ансамблей (*O_{CE}*) (рис. 3 и 4). Модель реализует концепцию «сети сетей», ядром которой является сеть над элементами, включающими в себя сетевые структуры различной природы. Сеть верхнего уровня (*C*) задается ориентированным мультиграфом, вершины которого соответствуют клеткам. Каждая вершина является интерактивным узлом и выступает как «белый ящик» со множествами входов и выходов; структуры «белого ящика» – генные и эпигенные сети (*G*). Совокупности связанных между собой вершин задают в пространстве клеток отношения соседства и описывают каналы межклеточных взаимодействий.

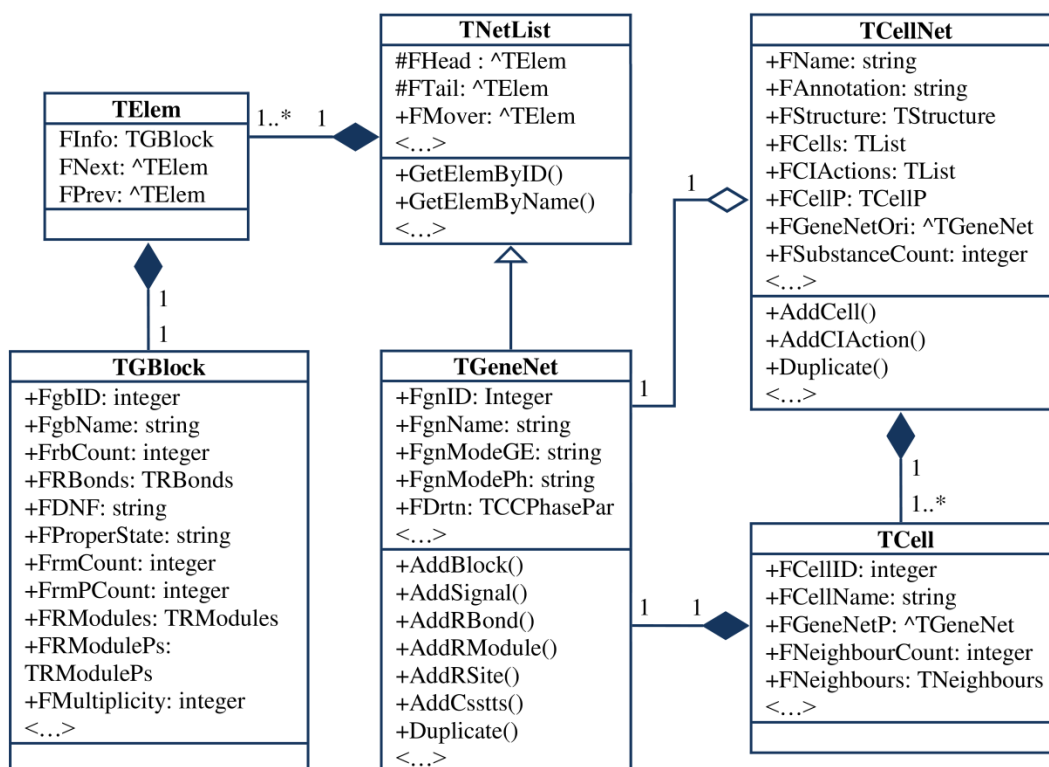


Рис. 4. Модель клеточного ансамбля: диаграмма классов верхнего уровня в нотации унифицированного языка моделирования UML [55]. Обозначения: линия с треугольником на конце – наследование, линия с не закрашенным или с закрашенным ромбом – агрегация или композиция, соответственно

Реляционная модель

Реляционные модели уровня генных сетей (I) и уровня клеточных ансамблей (II) базируются на модели «сущность – связь» [56] в нотации языка моделирования IDEF1X для разработки логических моделей данных [57]. Концептуальная схема модели геной сети в составе клеточного ансамбля представлена на рисунке 5.

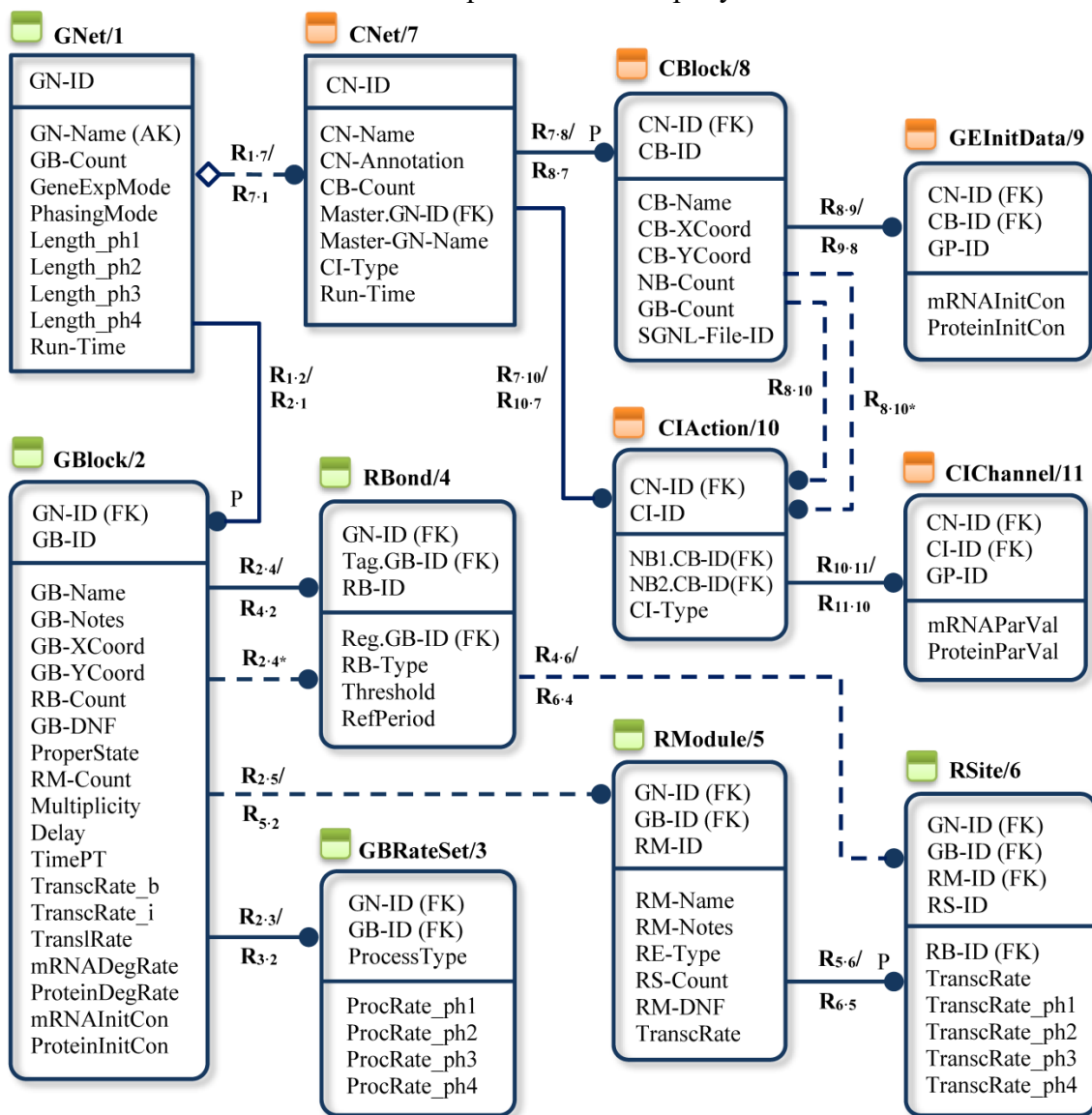


Рис. 5. Логическая модель данных в нотации IDEF1X. Обозначения: GN – геной сеть, GB – генетический блок, RB – регуляторная связь, RM – регуляторный модуль, RS – регуляторный сайт, RE – регуляторный элемент, Reg – регулятор, Tag – мишень, DNF – дизъюнктивная нормальная форма; CN – клеточная ансамбль, CB – клеточный блок, SGNL – формат машиночитаемого текста с описанием геной сети, NB – смежная клетка, CI – клеточное взаимодействие, GP – пара геной продуктов (мРНК и белок). Прямоугольники с острыми (закругленными) углами – независимые (зависимые) сущности, сплошная линия – идентифицирующая связь, пунктирная линия – неидентифицирующая связь, «•» – дочерняя сущность, FK – внешний ключ, AK – альтернативный ключ

В модель входят шесть сущностей первого уровня и пять сущностей второго. К первому уровню относятся сущности «Геной сеть», «Генетический блок», «Набор скоростных характеристик генетического блока», «Регуляторная связь», «Регуляторный модуль», «Регуляторный сайт». Ко второму уровню относятся сущности «Клеточная сеть», «Клеточный блок», «Начальные данные (профиль) геной экспрессии», «Межклеточное взаимодействие», «Канал межклеточного взаимодействия».

Практически все отношения между родительскими (P) и дочерними (D) сущностями можно лаконично охарактеризовать меткой «имеет» в PD-направлении и «принадлежит» – в DP-направлении. Определим детально каждое из них.

Уровень I

R_{1,2}: генная сеть содержит в себе генетический блок.

R_{2,1}: генетический блок входит в состав генной сети.

R_{2,3}: генетический блок характеризуется дополнительно набором скоростей синтеза и деградации мРНК и белка, специфичных к фазам клеточного цикла.

R_{3,2}: набор скоростей внутриблоковых процессов детализирует функциональные свойства генетического блока по фазам клеточного цикла.

R_{2,4}: генетический блок контролируется регуляторной связью. К блоку приписываются только те связи, в которых блок выступает в качестве мишени.

R_{4,2}: регуляторная связь управляет генетическим блоком.

R_{2,4*}: генетический блок участвует в регуляторной связи в качестве регулятора. Ключевой атрибут сущности «Генетический блок» мигрирует в сущность «Регуляторная связь» через два отношения, идентифицирующее и неидентифицирующее. В первом случае ему присваивается роль «мишень» (Tag), во втором – роль «регулятор» (Reg).

R_{2,5}: генетический блок включает в себя регуляторный модуль.

R_{5,2}: регуляторный модуль входит в генетический блок (в контролируемую его подсистему).

R_{5,6}: регуляторный модуль включает в себя регуляторный сайт.

R_{6,5}: регуляторный сайт входит в состав регуляторного модуля.

R_{4,6}: регуляторная связь может быть привязана к регуляторному сайту.

R_{6,4}: регуляторный сайт ассоциирован с регуляторной связью.

Уровень 2

R_{7,8}: клеточная сеть содержит в себе клеточный блок.

R_{8,7}: клеточный блок входит в состав клеточной сети.

R_{8,9}: клеточный блок характеризуется начальными данными генной экспрессии.

R_{9,8}: в клетке продукты гена присутствуют в различных концентрациях;

R_{7,10}: клеточная сеть включает в себя межклеточное взаимодействие (бинарное соединение двух смежных клеток);

R_{10,7}: межклеточное взаимодействие является частью клеточной сети;

R_{8,10}: клетка участвует в межклеточном взаимодействии как одна из двух смежных клеток (первый сосед).

R_{8,10*}: клетка участвует в межклеточном взаимодействии как одна из двух смежных клеток (второй сосед).

R_{10,11}: межклеточное взаимодействие опосредовано каналом переноса / трансформации молекулярных веществ. В текущей версии молекулярные вещества – продукты одинаковых генов (транскрипты и белки) двух разных клеток, содержащих типовые индивидуальные генные сети. Параметры канала – аналог коэффициентов диффузии по веществу.

R_{11,10}: канал специфицирует межклеточное взаимодействие по продуктам одинаковых генов двух связанных клеток (клеточных генных сетей).

Интеграция уровней I и II

R_{1,7}: генная сеть может являться сетью-шаблоном для всех клеток клеточной сети. Начальные данные для каждой внутриклеточной генной сети несет набор экземпляров сущности GEInitData.

R_{7.1}: клеточная сеть может иметь генную сеть в роли мастер-модели (шаблона) для индивидуальных генных сетей образующих ее клеток.

Альтернативой отношению R_{1.7}/R_{7.1} является интеграция, при которой сущность «Генная сеть» не участвует в сущности «Клеточная сеть». В том смысле, что у клеточной сети нет генно-сетевого шаблона, хранимого именно в таблицах. В этом варианте клетки могут содержать генные сети как с одинаковыми, так и с отличающимися структурно-функциональными характеристиками (с учетом мутагенеза / трансгенеза или общей вариативности генной экспрессии в однотипных клетках). Данные о генной сети каждой клетки хранятся в отдельном SGNL-файле. Идентификатор файла является неключевым атрибутом сущности «Клеточный блок». Экземпляры генных сетей в конкретных клетках аккумулируются в репозитории SGNL-моделей, в выделенном под модель клеточной сети разделе.

Преобразование концептуальной модели в реляционную выполнено по стандартной схеме, согласно которой сущности превращаются в таблицы, атрибуты становятся столбцами, а система ключей повторяет организацию первичных и внешних ключей на основе идентифицирующих атрибутов зависимых и независимых сущностей.

Язык спецификаций объектно-реляционных моделей генных сетей

Разработан язык спецификаций объектно-реляционных моделей генных сетей в составе клеточных ансамблей SETIES Standard Gene Network Language (SGNL), сочетающий в себе свойства командного языка и языка разметки.

SGNL – внутренний машиночитаемый формат представления данных по структурно-функциональной организации генных сетей. Назначение его – стандартизированное описание систем в целях их комплексной репрезентации, хранения, а также передачи данных между программными компонентами SETIES.

В рамках SGNL-формализма характеристики генной сети описываются набором командных строк (табл. 1).

Каждая командная строка имеет двойное назначение:

1) описывает поле одной из таблиц реляционной базы данных. По командным строкам одного SGNL-файла можно воссоздать в базе данных запись, соответствующую одной SGNL-модели.

2) содержит команду, при выполнении которой совершается одна элементарная операция по сборке объектной модели в среде программы (модуль «Конструктор информационных моделей генных сетей»).

Формат командной строки представлен четверкой вида:

«-<c1> <p1> <c2> <p2>»,

где *c1* – тип объекта; *p1* – значение первичного ключа объекта (в массиве данных набора моделей ГС); *c2* – параметр объекта (атрибут данных информационной модели ГС); *p2* – значение параметра объекта. Командная строка начинается с символа «-», элементы командной строки (поля) разделяются одиночным пробелом.

Собственно команду образуют два элемента командной строки:

<тип объекта> и <параметр объекта>.

Параметрами команды являются:

<значение первичного ключа объекта> и <значение параметра объекта>.

На уровне программной реализации командная строка предписывает интерпретатору получить в среде динамической сборки доступ к объекту типа *c1* с идентификатором *p1* и присвоить его параметру *c2* значение *p2*. При этом первое появление пары «-*c1 p1*» в тексте приводит к созданию соответствующего объекта.

Таблица 1. Список команд

| | Команда | Смысл параметров и переменных | Тип |
|---|---|---|----------------------|
| gn | -gn gnID name val | наименование генной сети | string |
| | -gn gnID gb_count val | количество g-блоков | integer |
| | -gn gnID gn_mode_ge val | двух- или многоуровневая экспрессия генов | (tl, ml) |
| | -gn gnID gn_mode_ph val | опция учета фаз клеточного цикла и митотических делений | (ph, un-ph) |
| | -gn gnID drtn_rt val | длительность наблюдений (run-time) | integer |
| | -gn gnID drtn_ph_g1 val | длительность G ₁ -фазы | integer |
| | -gn gnID drtn_ph_s val | длительность S-фазы | integer |
| | -gn gnID drtn_ph_g2 val | длительность G ₂ -фазы | integer |
| | -gn gnID drtn_ph_m val | длительность M-фазы | integer |
| gb | -gb gnID:gbID name val | наименование гена | string |
| | -gb gnID:gbID comm val | комментарий | string |
| | -gb gnID:gbID coord_x val | x-координата вершины граф-схемы | integer |
| | -gb gnID:gbID coord_y val | y-координата вершины граф-схемы | integer |
| | -gb gnID:gbID proper_state val | собственное состояние гена | (active, non-active) |
| | -gb gnID:gbID rb_count val | число регуляторных связей | integer |
| | -gb gnID:gbID rm_count val | число регуляторных модулей | integer |
| | -gb gnID:gbID dnf val | дизъюнктивная нормальная форма* | string |
| | -gb gnID:gbID multipl val | количество копий гена | integer |
| | -gb gnID:gbID delay val | задержка на синтез РНК | integer |
| | -gb gnID:gbID time_pt val | задержка на созревание и транспорт мРНК | integer |
| | -gb gnID:gbID transc_rate_b_rt val | скорость транскрипции базальная (run-time, un-ph) | real |
| | -gb gnID:gbID transc_rate_i_rt val | скорость транскрипции наведенная (run-time, un-ph) | real |
| | -gb gnID:gbID transl_rate_rt val | скорость трансляции (un-phased) | real |
| | -gb gnID:gbID rna_deg_rate_rt val | скорость деградации мРНК | real |
| | -gb gnID:gbID prt_deg_rate_rt val | скорость деградации белка | real |
| -gb gnID:gbID rna_init_concent val | начальная концентрация мРНК | real | |
| -gb gnID:gbID prt_init_concent val | начальная концентрация белка | real | |
| rb | -rb gnID:gbID:rbID gb_ctrl_id val | ID гена-регулятора | integer |
| | -rb gnID:gbID:rbID rb_type val | тип регуляторной связи | (pos, neg) |
| | -rb gnID:gbID:rbID threshold val | пороговое значение концентрации | real |
| | -rb gnID:gbID:rbID ref_period val | период рефрактерности | integer |
| rm | -rm gnID:gbID:rmID name val | наименование модуля | string |
| | -rm gnID:gbID:rmID comm val | комментарий | string |
| | -rm gnID:gbID:rmID rs_count val | число регуляторных сайтов | integer |
| | -rm gnID:gbID:rmID dnf val | дизъюнктивная нормальная форма | string |
| rs | -rs gnID:gbID:rmID:rsID rb_id val | ID r-связи в списке воздействий на g-блок | integer |
| | -rs gnID:gbID:rmID:rsID transc_rate_i_rt val | удельная единичная сила сайта | real |

*Пример записи дизъюнктивной нормальной формы в строковом формате: $X_1x_2 + X_3$, где знак сложения обозначает логическую операцию дизъюнкции, X_1x_2 – элементарная конъюнкция переменной X_1 и переменной X_2 , взятой с отрицанием (строчный символ обозначает операцию отрицания)

Типы объектов:

- gn – генная сеть (ГС),
- gb – генетический блок (g-блок),
- rb – регуляторная связь (r-связь),
- rm – регуляторный модуль (r-модуль),
- rs – регуляторный сайт (r-сайт).

Первичные ключи объектов состоят из идентификаторов объектов.

Используются пять типов идентификаторов (все – целые числа):

gnID – идентификатор ГС,

gbID – идентификатор *g*-блока в ГС,

rbID – идентификатор *r*-связи в *g*-блоке,

rmID – идентификатор *r*-модуля в *g*-блоке,

rsID – идентификатор *r*-сайта в *r*-модуле.

Число компонент в первичном ключе для ГС – один, для *g*-блока – два, для *r*-связи – три, для *r*-модуля – три, для *r*-сайта – четыре. Идентификаторы в составном первичном ключе разделяются символом «:».

К параметрам объектов относятся различные структурные, функциональные, количественные, семантические и визуальные характеристики компонент генных сетей (всего свыше 50 параметров).

На рисунке 6 приведен пример SGNL-описания одного генетического блока.

| <u>Common description</u> | <u>Regulator modules</u> | <u>Controlled subsystem</u> |
|-----------------------------|-------------------------------------|------------------------------|
| -gb 1:3 name g3 | -rm 1:3:1 name r-module1 | -gb 1:3 delay 1.0 |
| -gb 1:3 comm none | -rm 1:3:1 rs_count 1 | -gb 1:3 time_pt 0 |
| -gb 1:3 coord_x 1617.0 | -rm 1:3:1 dnf X1 | -gb 1:3 transc_rate_b_rt 0.5 |
| -gb 1:3 coord_y 2725.0 | -rm 1:3:2 name r-module2 | -gb 1:3 transc_rate_i_rt 2.0 |
| <u>Control subsystem</u> | -rm 1:3:2 rs_count 1 | -gb 1:3 transl_rate_rt 3.0 |
| -gb 1:3 proper_state active | -rm 1:3:2 dnf X1 | -gb 1:3 rna_deg_rate 0.34657 |
| -gb 1:3 rb_count 2 | <u>Regulator sites</u> | -gb 1:3 prt_deg_rate 0.13863 |
| -gb 1:3 rm_count 2 | <u>R-module ID: 1</u> | -gb 1:3 rna_init_con 0.0 |
| -gb 1:3 dnf X1X2 | -rs 1:3:1:1 rb_id 1 | -gb 1:3 prt_init_con 0.0 |
| -gb 1:3 multipl 1 | -rs 1:3:1:1 transc_rate_i_rt 1.0 | |
| <u>Regulatory bonds</u> | -rs 1:3:1:1 transc_rate_i_ph_g1 0.5 | |
| -rb 1:3:1 gb_ctrl_id 1 | -rs 1:3:1:1 transc_rate_i_ph_s 0.5 | |
| -rb 1:3:1 rb_type pos | -rs 1:3:1:1 transc_rate_i_ph_g2 0.5 | |
| -rb 1:3:1 threshold 10.0 | -rs 1:3:1:1 transc_rate_i_ph_m 0.5 | |
| -rb 1:3:1 ref_period 1.0 | <u>R-module ID: 2</u> | |
| -rb 1:3:2 gb_ctrl_id 2 | -rs 1:3:2:1 rb_id 2 | |
| -rb 1:3:2 rb_type pos | -rs 1:3:2:1 transc_rate_i_rt 1.0 | |
| -rb 1:3:2 threshold 20.0 | <...> | |
| -rb 1:3:2 ref_period 1.0 | | |

Рис. 6. Моногенный фрагмент SGNL-описания генной сети

Взаимосвязь моделей

Объектная модель O_{GN} , реляционная модель R_{GN} и объектно-реляционная модель OR_{GN} генной сети гомогенны по своей структуре.

Совокупность подмножеств связанных записей реляционных таблиц базы данных описывает один конкретный экземпляр коллекции – информационную R -модель управляющей генной сети. Для целостной ее репрезентации используется SGNL-формат.

В текстовом машиночитаемом SGNL-файле (модель OR_{GN}) характеристики генной сети кодируются набором командных строк. Каждая командная строка соответствует единственной строке (записи) одной из таблиц реляционной базы данных. Аналогом «виртуальной» объектно-реляционной базы моделей генных сетей в составе клеточных ансамблей является репозиторий машиночитаемых файлов в SGNL-формате. По командным строкам SGNL-файла посредством специального загрузчика и автоматического конструктора можно создать в базе данных запись, соответствующую одной объектной модели.

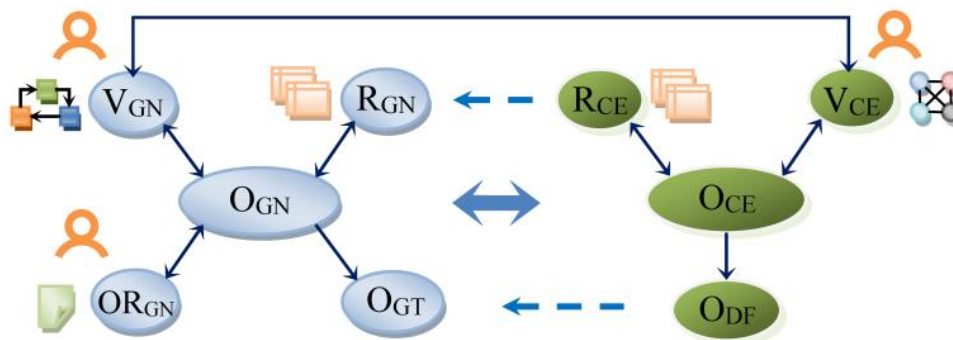


Рис. 7. Схема порождения экземпляров классов-моделей сетевых структур. Обозначения: O, R, OR, V – объектная, реляционная, объектно-реляционная (SGNL-формат), визуальная модели, соответственно; DF – диффузное поле

Для прямого и обратного преобразования объектных моделей генных сетей и клеточных ансамблей в реляционную модель используется ORM-посредник. Стратегия отображения – схемо-зависимая, с использованием паттернов отображения для простых атрибутов, ассоциаций между классами и атрибутов агрегатных типов (Attribute-Column, OneClass-OneTable, ForeignKeyAssociation, ClassAggregate и др.). Агрегатные атрибуты: клеточный ансамбль, клетка, клеточная генная сеть, генетический блок, ген копияности N (кассетный автомат), регуляторный модуль, регуляторный сайт, регуляторная связь.

Визуальная V-модель – граф-схема сети с интерактивным наполнением. V-модель и SGNL-модель доступны пользователю напрямую. Модели одного типа могут порождать модели другого типа. Это происходит благодаря компонентам-посредникам. O-модели скрыты от пользователя и создаются по командам визуального конструирования (через V-модель), по записям базы данных (по R-модели) или по строкам-инструкциям текстового файла (SGNL-модель). Информационные объектные модели порождают математические объектные модели. Для генных сетей это обобщенные пороговые модели O_{GT}, а для клеточных ансамблей – клеточные автоматы O_{DF} со специальным алгоритмом расчета диффузии. Взаимосвязанность моделей различных типов представлена на рисунке 7. Автоматизация переходов между моделями пяти типов обеспечивает конвейерный принцип обработки биоинформационных данных в программном комплексе SETIES.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Программный комплекс создан на языках программирования Pascal и Java в интегрированных средах разработки Borland Pascal и Borland JBuilder; для создания баз данных использовалась СУБД Paradox 7.x.

Комплекс состоит из 12 основных модулей (рис. 8). Вместе они образуют несколько крупных блоков работы с данными. Коровый блок сформирован компонентами GN-Core и CE-Core, содержащими базовые классы-модели генных сетей и клеточных ансамблей, соответственно.

В блоке визуального конструирования сетей, образованном модулями GN-Builder и CE-Builder, создаются и редактируются конкретные объекты (экземпляры классов) с произвольными структурными и функциональными характеристиками.

Блок репрезентации и хранения данных состоит из модуля реализации языка SGNL, репозитория SGNL-моделей генных сетей, а также баз данных Gene Networks и Cell Ensembles.

В счетный блок входят модуль реализации обобщенной пороговой модели генных сетей (MGCS/GTM), компонента для моделирования растущей клеточной популяции на

дереве клеточных делений (EpiGN / CP), а также модуль для расчета потоков вещества в доменах диффузного поля (DF-Modeler).

Через блок визуализации выходных данных осуществляется вывод результатов компьютерных экспериментов с компьютерными моделями, в частности кинетические кривые концентраций мРНК и белков, диаграммы активностей промоторов генов, профили параметрической устойчивости генных сетей при варьировании значений кинетических параметров, паттерны распределения молекулярных компонент генных сетей в пространстве, динамика наследования функциональных режимов в клеточных линиях и в ряду поколений клеточной популяции.

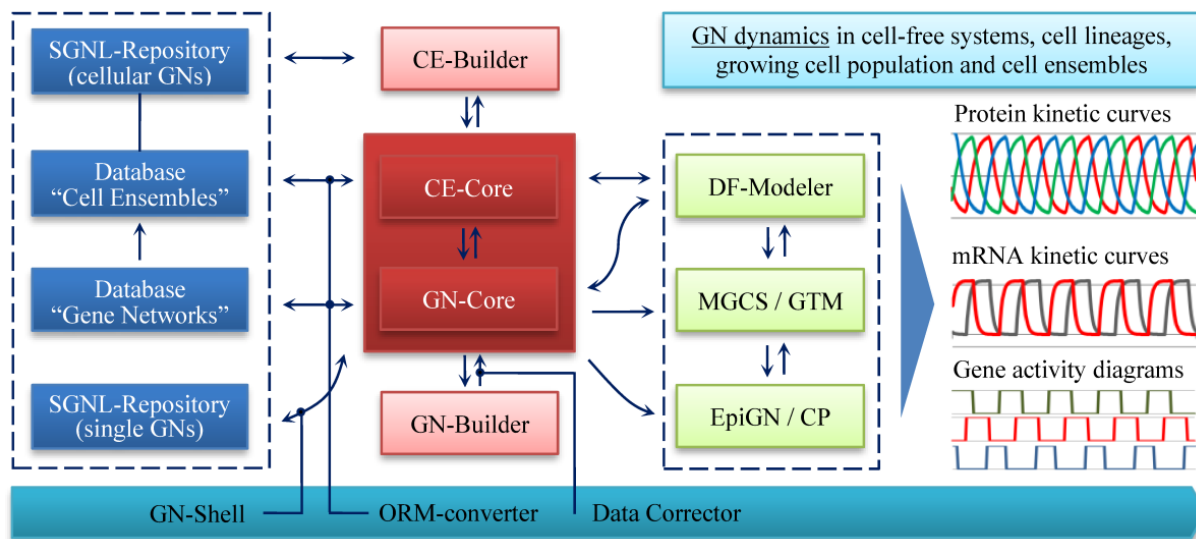


Рис. 8. Архитектура платформы SETIES

GN-Core

Содержит базовые классы объектно-ориентированной модели генной сети. Обеспечивает создание динамического экземпляра информационной модели генной сети, его модификацию, резервное копирование, а также тиражирование при построении дерева клеточных делений или генерации клеточных ансамблей. В качестве источника данных использует запросы пользователя через модуль GN-Builder, записи базы данных Gene Networks, репрезентации генных сетей в SGNL-формате.

CE-Core

Содержит базовые классы объектно-ориентированной модели клеточного ансамбля. Интегрирует данные по структурно-функциональной организации генных сетей и клеточных ансамблей. Обеспечивает генерацию *in silico* клеточных множеств произвольной конфигурации, при этом клетки содержат индивидуальные внутриклеточные генные сети (электронные модели) – типовые или производные от них (трансгенные или мутантные варианты).

GN-Builder

Предназначен для редактирования и просмотра пользователем данных по структурно-функциональной организации управляющих генных сетей про- и эукариот. Обеспечивает выполнение следующих функций: пошаговое построение и визуализацию граф-схем генных сетей с интерактивным наполнением; генерирование, посредством порождающих алгоритмов, типовых контуров (сетевых мотивов) транскрипционной регуляции экспрессии генов; модульную сборку сложных сетей из

других сетей; ввод и просмотр данных, описывающих регуляторные механизмы и кинетические характеристики индивидуальной экспрессии генов сети, на основе оконного интерфейса; автоматическое создание типовых отчетов в виде графических и текстовых файлов.

CE-Builder

Модуль предназначен для визуального конструирования информационных (объектных) моделей клеточных ансамблей. Представляет собой пользовательский интерфейс для построения и визуализации граф-схем межклеточных взаимодействий, определения значений атрибутов клеточных блоков и связей между ними, вместе предоставляющих возможность для формирования прототипов клеточных популяций.

База данных Gene Networks

База данных содержит описания схем управляющих генных сетей про- и эукариот с динамическими характеристиками основных молекулярно-генетических процессов. Доступ к базе осуществляется посредством конструктора информационных моделей генных сетей. Совместно с репозиторием машиночитаемых файлов обеспечивает статическое хранение объектов с возможностью автоматического воспроизведения и портретного моделирования в среде программного комплекса. В настоящий момент в базе представлены репрезентации теоретических и экспериментальных генетических конструкций с заданными управляемыми режимами функционирования.

База данных Cell Ensembles

База данных содержит описания структурных и функциональных характеристик клеточных ансамблей. Доступ к базе осуществляется посредством конструктора электронных моделей сетевых структур. Совместно с базой данных Gene Networks поддерживает начальный этап процесса конвейерного моделирования молекулярно-генетических систем в среде программного комплекса – первичную или воспроизводящую сборку объектов. В настоящий момент в базе представлены репрезентации гипотетических ансамблей клеток, содержащих генные сети с заданными управляемыми режимами функционирования.

Gene Networking Shell

Компонента предназначена для генерации в режиме пакетного интерфейса объектных информационных моделей клеточных управляющих генных сетей. Представляет собой командный интерпретатор, который построчно разбирает и исполняет текстовые файлы, содержащие объектно-реляционные описания генных сетей в SGNL-формате, сочетающего в себе свойства командного языка и языка разметки. Обеспечивает формирование машиночитаемых отчетов о структуре и динамических характеристиках реализованных *in silico* моделей генных сетей в целях их комплексной репрезентации, хранения, а также передачи данных между приложениями семейства SETIES и, потенциально, другими платформами в области системной биологии. Совместно с базами данных и визуальным конструктором моделей поддерживает процесс конвейерного моделирования молекулярно-генетических систем.

Объектно-реляционный посредник

Компонента предназначена для прямого и обратного преобразования объектных моделей генных сетей и клеточных ансамблей в реляционную модель.

MGCS / GTM

Программа предназначена для автоматического построения и анализа обобщенных пороговых моделей генных сетей с известной структурой. В моделях учитываются динамические характеристики основных молекулярно-генетических процессов, модульная организации регуляторных районов, многоуровневая экспрессия и мультикопийность генов. Среди решаемых задач – получение кинетических кривых для концентраций генных продуктов (РНК, белков), выявление паттернов активности генов во времени, оценка параметрической устойчивости, моделирование мутагенеза и трансгенеза, дискриминирование гипотез о возможных механизмах взаимодействия между генами-регуляторами и генами-мишенями.

DF-Modeler

Компонента предназначена для исследования генных сетей в составе клеточных ансамблей с диффузией. Клеточные ансамбли моделируются клеточными автоматами дискретно-непрерывного типа (как решеточными, так и нерешеточными).

EpiGN / CP

Компонента предназначена для изучения эпигенетических свойств генных сетей на дереве клеточных делений, моделирующем экспоненциальную и стационарную фазы роста клеточной популяции. Программа обеспечивает выполнение следующих функций:

- построение и обход дерева клеточных делений размерностью до 30 ярусов;
- наблюдение за динамикой генных и эпигенных сетей в клеточных линиях, с учетом фаз клеточного цикла, роста клеточного объема (линейный, экспоненциальный, правило NLM), биномиального распределения продуктов генов между дочерними клетками при делении материнской, с мультиплицированием сетей для клеток новой генерации;
- расчет профилей экспрессии генов в клетках промежуточных и финального поколений; выявление и запоминание режимов функционирования, в которые переходят экземпляры типовой сети в ряду клеточных генераций.

Общий сценарий работы с программой

Платформа реализует конвейерный сценарий компьютерного моделирования молекулярно-генетических систем (прил. 1).

Начальным этапом является привнесение в программную среду объектной информационной модели генной сети, не имеющей непосредственной привязки к структурам данных математического формализма. Это можно сделать четырьмя способами:

- 1) выбрать запись из локальной базы данных Gene Networks;
- 2) загрузить репрезентацию из репозитория SGNL-моделей;
- 3) создать экземпляр в конструкторе моделей GN-Builder;
- 4) собрать из модулей с помощью утилиты «Network Motifs» для генерации сетевых мотивов транскрипционной регуляции экспрессии генов.

Эти способы можно использовать вместе и в любой последовательности, наращивая модель из разных источников. Во всех случаях структура модельной генной сети будет представлена в виде граф-схемы с интерактивным наполнением. Созданный объект хранится в оперативной памяти компьютера и доступен пользователю через визуальные формы редактирования входных данных на протяжении всего сеанса работы с изучаемой системой.

Далее для решения задачи анализа динамики ГС по объектной модели ГС автоматически строится обобщенная пороговая модель ГС (сеть над генетическими блоками), и проводятся расчеты. После их завершения ОПМ-экземпляр уничтожается. Актуальную объектную модель можно сохранить в реляционной базе данных или в репозитории в виде машиночитаемого текстового SGNL-файла.

Для изучения генных сетей в составе клеточных ансамблей необходимо пройти вышеописанный первый уровень и перейти ко второму. Здесь, принимая объектную модель генной сети в качестве типовой модели, нужно задать структуру клеточного ансамбля. При создании новой клетки программа автоматически генерирует клеточную генную сеть как реплику с типовой, так что все клетки содержат индивидуальные экземпляры типовой объектной модели. Далее любую модельную клеточную сеть можно модифицировать, имитируя гетерогенность структурных или функциональных характеристик. При сохранении актуальной объектной модели клеточного ансамбля данные распределяются следующим образом. Конфигурация клеточной сети вносится в базу Cell Ensembles, типовая модельная генная сеть записывается в базу Gene Networks, а индивидуальные модельные генные сети клеток в виде отдельных SGNL-файлов передаются в подраздел репозитория, закрепленного за актуальной моделью клеточного ансамбля.

При моделировании клеточных линий или растущей клеточной популяции сначала также задается объектная модель генной сети. Этот экземпляр становится «родовым» для популяции. При делении материнской клетки каждая дочерняя клетка получает от нее функциональную наследственную информацию, в точности уникальные уровни концентраций генных продуктов (начальные данные). По конфигурации родовой генной сети в контексте этих данных строится и рассчитывается ОП-модель генной сети дочерней клетки. Далее уже ее выходы детерминируют входы для ОП-моделей генных сетей дочерних клеток, для которых она становится материнской. При этом новые входы могут учитывать стохастические изменения значений начального клеточного объема и кинетических параметров, а также индукцию эпимутационных доз транскрипционных факторов.

Таким образом, функциональные возможности SETIES позволяют оперировать данными по структурно-функциональной организации генных и клеточных регуляторных сетей и анализировать определяемую ими сетевую динамику в компьютерных экспериментах с большим количеством молекулярно-генетических опций.

ПРИМЕРЫ ПРИМЕНЕНИЯ

Ранее решенные задачи

С использованием SETIES ранее исследованы молекулярно-генетические системы управления морфогенезом цветка *Arabidopsis thaliana* и эмбриогенезом *Drosophila melanogaster*, для которых получены оценки параметрической устойчивости [58, 59]. В моделях синтетического генетического переключателя, функционирующего *in vivo* в клетках *Escherichia coli*, изучена динамика его стационарных и переходных режимов, обнаружено и описано особое метастабильное состояние, запускающее процесс самодифференцировки генетически идентичных клеток [48, 60]. В компьютерных экспериментах рассчитаны модели фазовых вариаций у бактерий на основе генных сетей нового класса – трингеров [50, 51]. Трингеры представляют собой кольцевые соединения бистабильных переключателей на основе генетического осциллятора на транскрипционных репрессорах. Показано, что в экспоненциальной фазе роста бактериальной популяции трингеры могут порождать фенотипическую гетерогенность колонии, связанную с альтернативной и обратимой экспрессией как генов трингеров, так и подчиненных им регулонов.

Генная сеть развития в составе клеточного ансамбля с дифференцировкой

Рассмотрим модельную 6-генную сеть в составе 100-клеточного ансамбля кольцевого типа (рис. 9,а). Генная сеть состоит из двух симметричных модулей, образованных генами G_1 – G_3 и G_4 – G_6 , кодирующими транскрипционные факторы друг для друга. Гены G_1 и G_2 являются авторегуляторными элементами и активируют ген G_3 при наличии пороговых доз белков обоих генов одновременно. В свою очередь ген G_3 подавляет активность генов G_4 и G_5 второго модуля. В этом модуле аналогично гены G_4 и G_5 активируют сами себя и общий для них целевой ген G_6 , также совместно. Ген G_6 подавляет экспрессию генов G_1 и G_2 первого модуля. Данная схема учитывает три распространенных структурных мотива генных сетей с функцией контроля развития: цепь положительной обратной связи, цепь взаимного подавления и цепь когерентной обратной связи [61].

Логика функционирования сети такова. В отсутствие внешних сигналов сеть находится в неактивном состоянии, когда оба гена-блокиратора выключены. Сеть имеет два нетривиальных стационарных состояния. В первом из них (ON) активны все три гена первого модуля и выключены гены второго, во втором (OFF), напротив, включены гены второго модуля и выключены гены первого. Таким образом, функционально сеть представляет собой бистабильный переключатель, у которого функции активации и блокирования активности альтернативных модулей разнесены по отдельным генам. При этом гены G_1 и G_2 (G_4 и G_5) можно считать стартерными, т.к. только при их активации сеть может выйти из неактивного состояния и эволюционировать к нетривиальным стационарным ON- или OFF-состояниям. Если неактивен хотя бы один стартерный ген каждого модуля, весь переключатель находится в неактивном состоянии. Для индукции переходного режима состояния системы нужны внешние сигналы на входы элементов G_1 / G_2 и G_4 / G_5 .

Клеточный ансамбль представляет собой замкнутую цепь попарно связанных клеток, содержащих одинаковые генные сети. Во всех клетках, кроме четырех полюсных, заданы нулевые начальные концентрации мРНК и белков всех генов. В полюсных клетках C_1 , C_{51} , C_{26} и C_{76} начальное содержание белков стартерных генов G_1 , G_2 , G_4 и G_5 , соответственно, равно 1000 молекул на клетку. Взаимодействие клеточных генных регуляторных сетей в составе клеточной сети строится на основе модели позиционной информации Вольперта [62–64].

Опишем динамику системы в заданных условиях (рис. 9,б и 9,в). Пороговая доза белка гена G_1 в полюсной клетке C_1 индуцирует экспрессию собственно гена G_1 . Синтезируемый в полюсной клетке белок диффундирует в смежные клетки и, достигнув пороговой концентрации, активирует в них новые копии G_1 , дающие ему прибавление. С каждым тактом (отрезок t_1 – t_3) область экспрессии / диффузии расширяется, при этом с противоположного полюса C_{51} движется поток диффундирующего белка гена G_2 . В верхней и нижней точке (полюсные клетки C_{26} и C_{27}) они сходятся (момент t_3), что вызывает экспрессию G_3 и, как следствие, диффузию белка этого гена влево и вправо. К этому времени потоки белков генов G_4 и G_5 с верхнего и нижнего полюсов сходятся в области левого и правого полюсов и запускают экспрессию G_6 . Кодируемый им белок блокирует G_1 , так что область экспрессии G_1 разрывается в области левого полюса (момент t_4), сужаясь в сторону полюсов верхнего и нижнего (но продолжает расширяться в направлении от двух последних). Аналогичным образом G_6 действует на G_2 в области правого полюса. Также симметрично гены G_1 , G_2 и G_3 работают по генам G_4 , G_5 и G_6 в областях верхнего и нижнего полюсов. Благодаря перекрестной репрессии стартерных генов генами-блокираторами, перераспределение белков во всех направлениях замедляется и далее прекращается, являя положение равновесия (момент t_5). Отметим особо, что в итоге области экспрессии генов обоих модулей, будучи альтернативными, совершили поворот на 90° градусов относительно сигналов, индуцировавших их формирование.

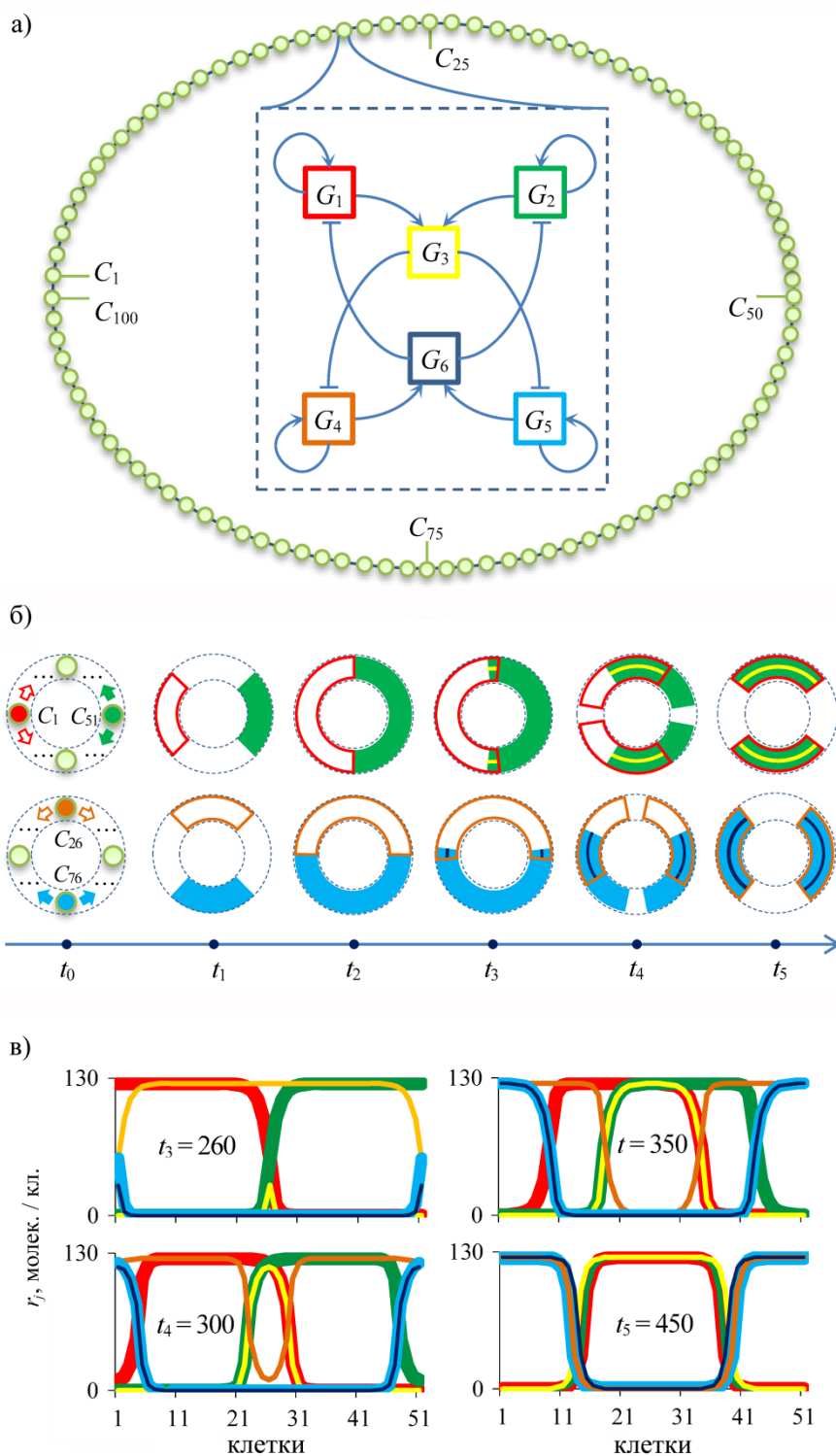


Рис. 9. Стоклеточный кольцевой ансамбль. **а)** Структура генной сети, вид клеточного ансамбля. Обозначения: C_i – клетки, G_j – гены (генетические блоки); связь со стрелкой (с чертой) – позитивная (негативная) регуляция. **б)** Динамика системы от начального состояния (t_0) до стационарного (t_5). Верхняя серия – паттерны экспрессии генов G_1-G_3 , нижняя – генов G_4-G_6 . Закрашенный кружок – клетка, индуцированная надпороговой концентрацией белка гена соответствующего цвета. **в)** Кинетические кривые для белков в клетках C_1-C_{51} (верхняя половина клеточного кольца). Значения параметров: $a_{1j} = 2 \text{ мин}^{-1}$, $a_{2j} = 3 \text{ мин}^{-1}$, $b_{1j} = 0.347 \text{ мин}^{-1}$, $b_{2j} = 0.13863 \text{ мин}^{-1}$, задержка = 1 мин, пороги для активаторов и репрессоров равны 20 и 30 молекул на клетку, соответственно; весовой коэффициент (аналог коэффициентов диффузии) = 0.1.

Генная сеть в составе клеточного ансамбля матричного типа

Рассмотрим систему из семи генов (рис. 10,а).

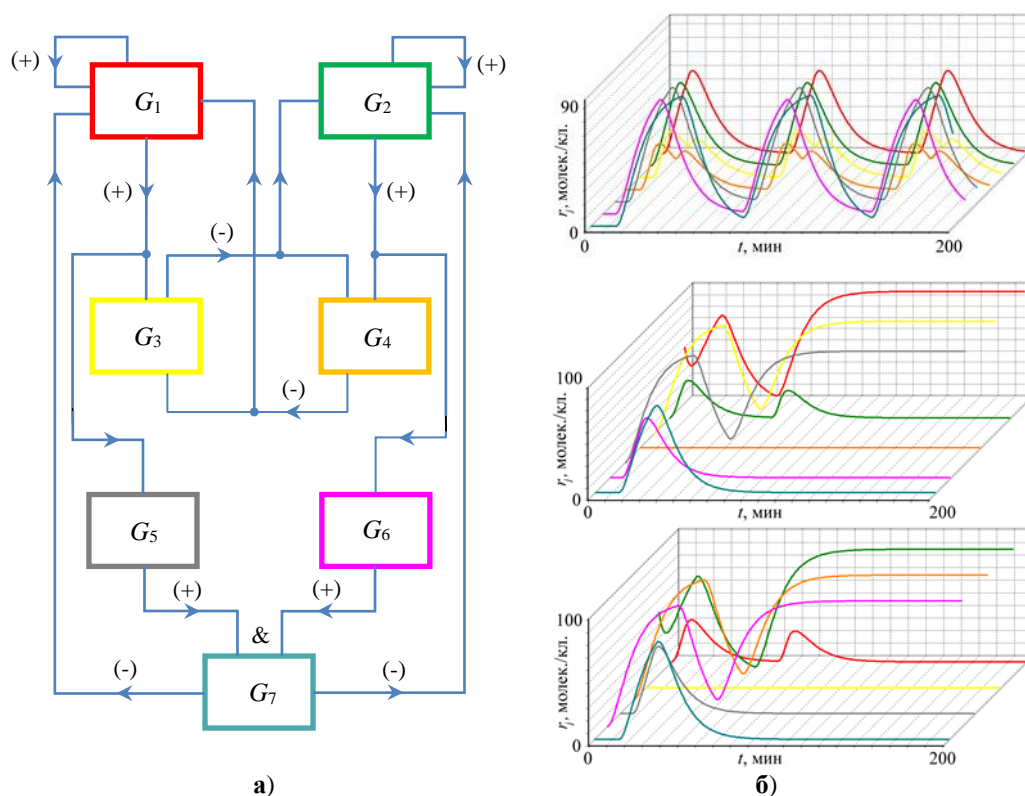


Рис. 10. Структура генной сети (а) и динамика в модели (б). Обозначения: G_j – гены (генетические блоки) ($j = 1..7$); связи между генами – транскрипционная активация (+) или репрессия (-). Верхний график – осцилляторный режим, два нижних – стационарные состояния

Данная генная сеть включает в себя следующие регуляторные контуры:

- 1) два положительных авторегуляторных элемента (гены G_1 и G_2);
- 2) одну циклическую дигенную систему с отрицательными обратными связями между генами G_3 и G_4 ;
- 3) кассету генов G_1 , G_3 и G_5 под общим для них положительным регулятором – продуктом гена G_1 и кассету генов G_2 , G_4 и G_6 также под общим положительным регулятором – продуктом гена G_2 ;
- 4) комбинаторный элемент G_7 , вырабатывающий интегральную реакцию на сигналы, поступающие от генов G_5 и G_6 ;
- 5) колебательный контур, образованный главными для него генами – G_1 , G_2 и G_7 (гены G_1 и G_2 опосредованно активируют ген G_7 , который подавляет их активность).

Система имеет один осцилляторный режим и два альтернативных стационарных состояния (рис. 10,б).

Изучим поведение генной сети в многоклеточной (многоядерной) системе с диффузией белков. Структура клеточного ансамбля представлена на рисунке 11.

Зададим три набора значений параметров, отличающихся активностью промоторов и временами полужизни РНК и белков отдельных генов. В первом наборе значения параметров у всех генетических блоков одинаковы (симметричный случай): $a_{1j} = 1$ молек. · мин⁻¹, $a_{2j} = 2$ молек. · мин⁻¹, $b_{1j} = 0.09$ мин⁻¹, $b_{2j} = 0.05$ мин⁻¹ (времена полужизни РНК и белков – 7.7 и 13.8 мин, соответственно), пороги $P_{ij} = 25$ молекул на клетку. Во втором наборе (табл. 2) времена полужизни РНК и белков генов G_1 , G_2 , G_3 и G_4 – 7.7 и 13.8 мин, РНК и белков генов G_5 и G_6 – 5.8 и 4.6 мин, РНК и белка гена G_7 – 3.8 и 3 мин, соответственно. Ключевое отличие третьего набора (табл. 3) от второго в

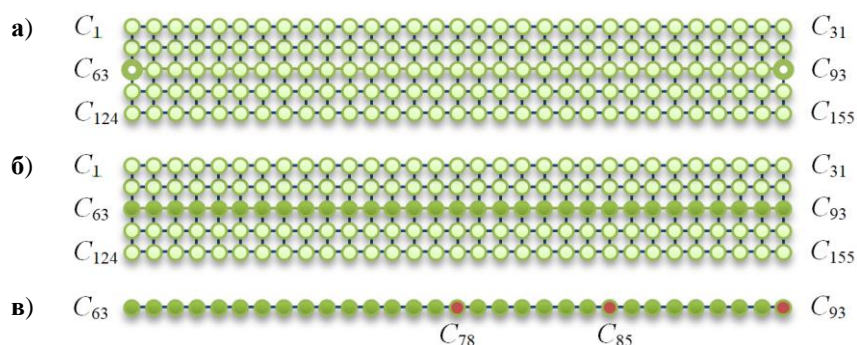


Рис. 11. Клеточный ансамбль: **а)** сеть и полусные клетки C_{63} и C_{93} ; **б)** сеть и контрольная линия C_{63} – C_{93} ; **в)** контрольная линия и клетки C_{78} , C_{85} , C_{93} для детализации динамики генной экспрессии

том, что в этом наборе активность промоторов генов G_1 и G_3 на 20 % выше, чем у генов G_2 и G_4 (асимметрия по элементам молекулярного триггера).

Таблица 2. Второй набор значений параметров генетических блоков

| | G_1 | G_2 | G_3 | G_4 | G_5 | G_6 | G_7 |
|----------|-------|-------|-------|-------|-------|-------|--------|
| a_{1j} | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| a_{2j} | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| b_{1j} | 0.09 | 0.09 | 0.09 | 0.09 | 0.12* | 0.12* | 0.18* |
| b_{2j} | 0.05 | 0.05 | 0.05 | 0.05 | 0.15* | 0.15* | 0.231* |

* Значения, отличающиеся от значений в первом наборе

Таблица 3. Третий набор значений параметров генетических блоков

| | G_1 | G_2 | G_3 | G_4 | G_5 | G_6 | G_7 |
|----------|--------|--------|--------|--------|-------|-------|--------|
| a_{1j} | 1,2** | 1 | 1,2** | 1 | 1 | 1 | 1 |
| a_{2j} | 2,5** | 2 | 2,4** | 2 | 2 | 2 | 2 |
| b_{1j} | 0.08** | 0.08** | 0.08** | 0.08** | 0.12* | 0.12* | 0.15** |
| b_{2j} | 0.05 | 0.05 | 0.05 | 0.05 | 0.15* | 0.15* | 0.231* |

** Значения, отличающиеся от значений во втором наборе

Начальные концентрации генных продуктов во всех клетках, за исключением C_{63} и C_{93} , – нулевые. В клетке C_{63} на старте присутствует 10 молекул мРНК гена G_1 , а в клетке C_{93} – 10 молекул мРНК гена G_2 , т.е. две эти клетки выбраны в качестве полусных. Коэффициент диффузии для всех белков $d_j = 0.1$.

Для визуализации результатов вычислений клетки выбираются как точно, так и вдоль линий, называемых *контрольными*. Выбранные клетки получают дополнительные номера согласно позиции на контрольной линии (рис. 11,б).

В компьютерных экспериментах выявлены три режима пространственно-временной динамики – автоколебательный в центральной области (рис. 12,а), стационарный с формированием трехслойной структуры (рис. 12,б) и реверсивный режим, в котором трехслойная структура эволюционирует в однослойную (рис. 13) (наборы значений параметров – первый, второй и третий, соответственно).

На рисунках 12 и 13 приведены профили пространственного распределения белков вдоль контрольной линии C_{63} – C_{93} (номера 1–31 для визуализации). На параллельных линиях C_1 – C_{31} , C_{32} – C_{62} , C_{94} – C_{124} , C_{125} – C_{155} профили схожи.

Опишем поведение системы и взаимосвязь трех выявленных вариантов пространственно-временной динамики.

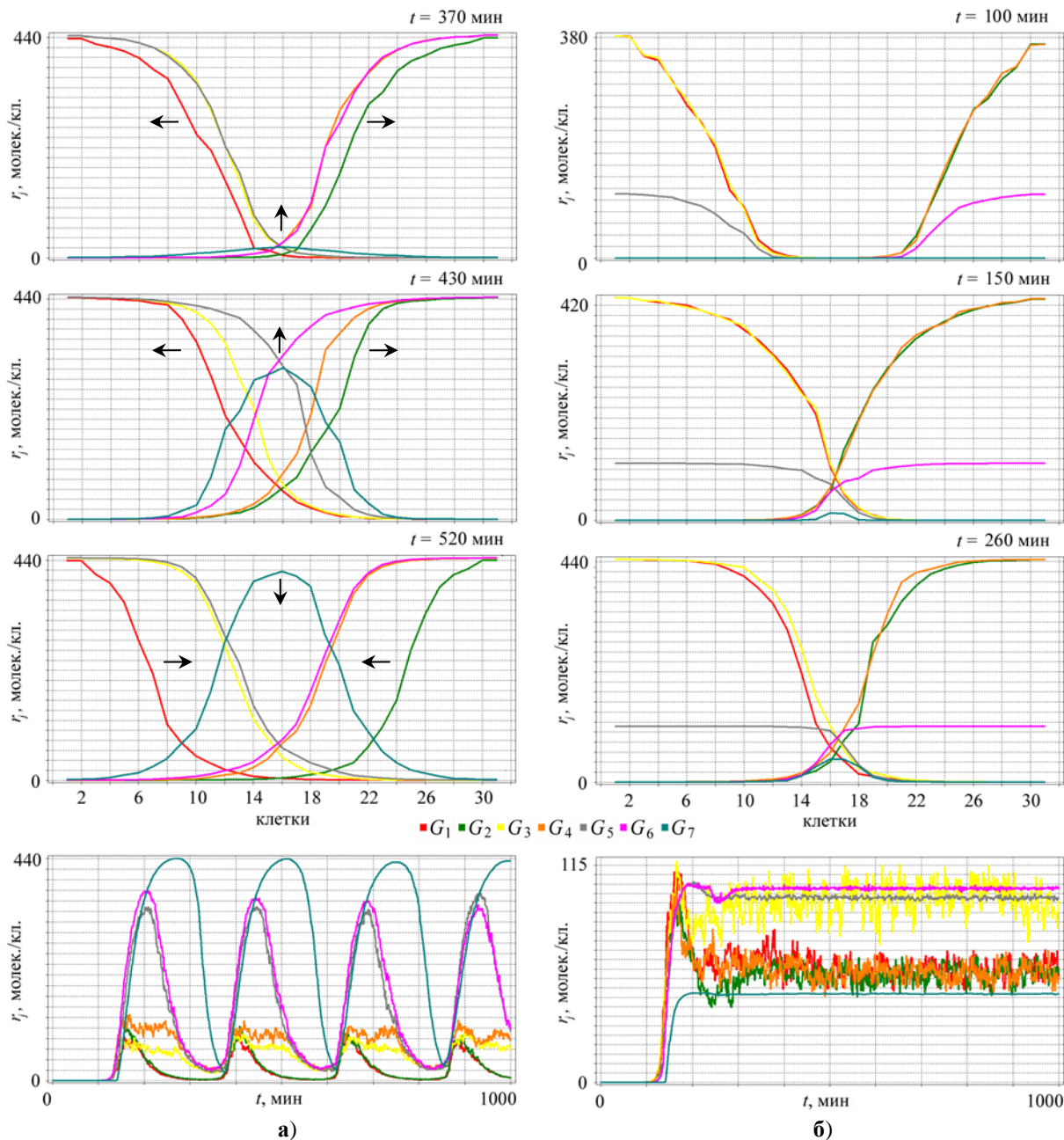


Рис. 12. Динамика в модели. **а)** Автоколебательный режим с центром в поперечной оси клеточного ансамбля. Сверху вниз: профили распределения белков вдоль контрольной линии на 370-й, 430-й и 520-й мин (стрелками показаны направления изменения значений концентраций); нижний график – кинетические кривые для белков в клетке № 16 (C_{78} на рисунке 11,в) в течение 1000 мин (период колебаний ~ 220 мин). **б)** Эволюция системы в стационарный режим. Сверху вниз: профили распределения белков вдоль контрольной линии на 100-й, 150-й и 260-й мин. Первый слой – область преимущественно высоких концентраций белков генов G_1 и G_2 , второй – белков генов G_2 и G_4 , третий – промежуточный слой, белок гена G_7 (появляется на 140-й мин, к 260-й мин устанавливается стационарное состояние)

На полюсах клеточного множества изначально присутствуют мРНК генов G_1 и G_2 в концентрациях, достаточных для «запуска» двух циклических моногенных систем – по одной с каждой стороны. Диффундируя в соседние клетки, белки авторегуляторных элементов активируют собственные гены. Вместе с ними происходит активация генов-мишеней, находящихся в одной кассете вместе с соответствующим геном-регулятором. В итоге области экспрессии-диффузии генов G_1 , G_3 , G_5 с одной стороны и генов G_2 , G_4 , G_6 – с другой расширяются навстречу друг другу.

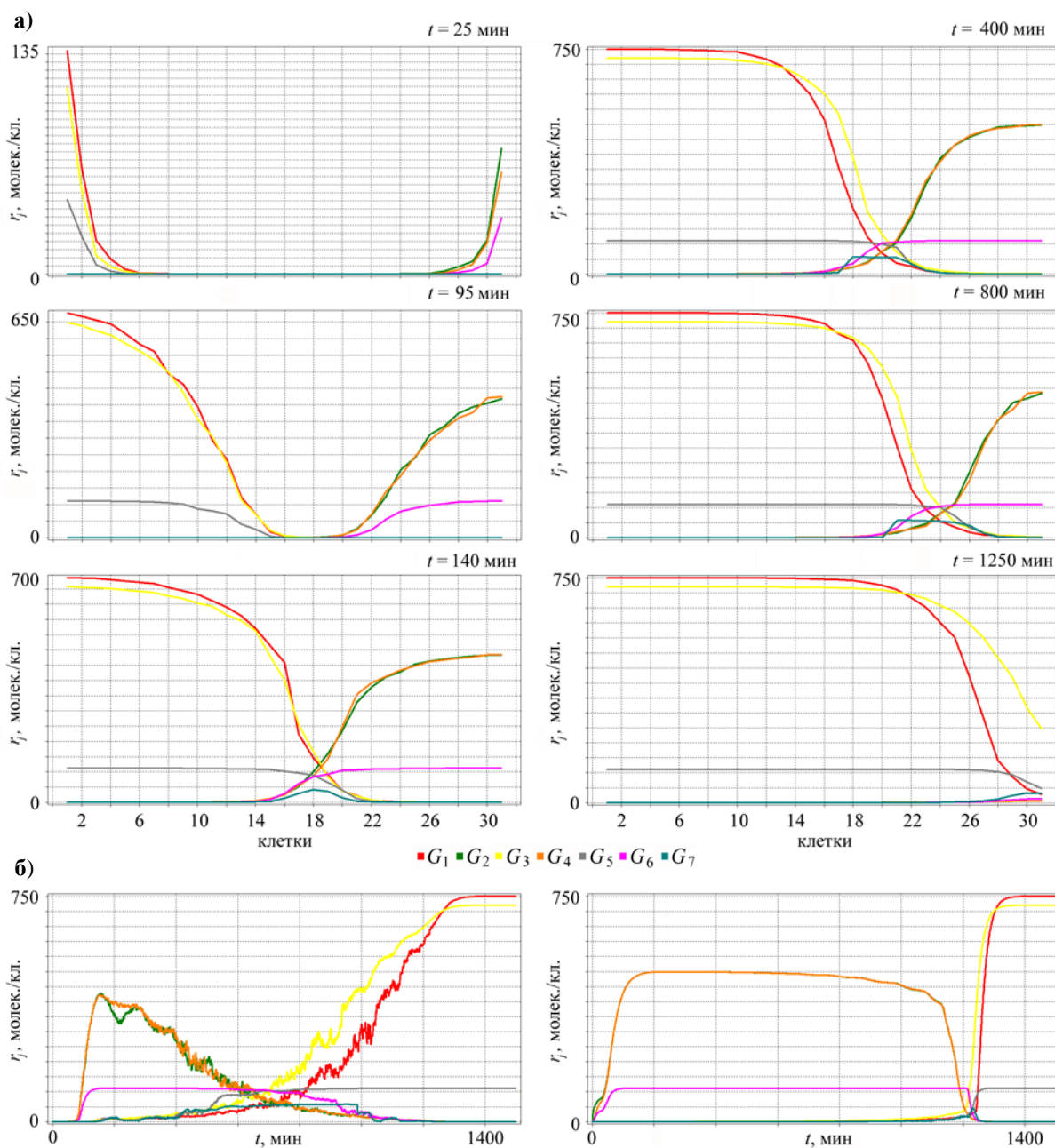


Рис. 13. Реверсивный режим в модели: **а)** пространственное распределение белков в ряду $t = 25, 95, 140, 400, 800$ и 1250 мин (к 1300-й мин остается только один слой – с белками генов G_1, G_3 и G_5); **б)** динамика генной сети в клетках №23 (слева) и №31 (справа) (соответственно, клетки C_{85} и C_{93} на рисунке 11,в)

В центральной части клеточного множества расширение такого рода прекращается, при этом области диффузии частично перекрыты, а области экспрессии не пересекаются, чему способствует перекрестная репрессия гена G_2 белком гена G_3 и гена G_1 – белком гена G_4 . Вместе с тем, белки генов G_5 и G_6 одновременно присутствуют здесь в концентрациях, достаточных для активации гена G_7 , так что с появлением этого белка формируется новый – промежуточный слой, в добавление к слоям 1 и 2.

В первых двух вариантах (рис. 12) промежуточный слой либо периодически сжимается-расширяется (первый набор значений параметров), либо фиксируется (второй набор). В третьем варианте (рис. 13) промежуточный слой, сохраняясь в размерах и сохраняя баланс в уровне концентраций белков, последовательно сдвигается слева направо, при этом на пути следования слой 1 прирастает за счет слоя 2. Механизм поглощения определяется превышением скоростей синтеза РНК и белка гена G_1 над

такowymi у гена G_2 , в результате чего при краткосрочном снятии репрессии генов G_1 и G_2 со стороны гена G_7 в режиме автоколебаний с малой амплитудой в промежуточном слое, продукты гена G_1 синтезируются быстрее и блокируют экспрессию гена G_2 .

Следует отметить, что любой из генов кассеты, продукты которых, по факту, вытесняются из финальной морфогенетической разметки, может быть также прямым или опосредованным активатором какого-либо дополнительного гена, область экспрессии которого установится на первом этапе (до возникновения промежуточного слоя) и сохранится в этом установленном виде и после исчезновения фактора активации. Кроме того, временное сдерживание экспрессии конкретного гена (здесь в составе более активной кассеты) может быть полезным для случая, когда оба гена должны экспрессироваться в заданной области, но одновременная инициация их экспрессии в одной области невозможна. Наконец, данный мобильный динамический паттерн может использоваться как способ морфогенетической разметки из глубины клеточного множества к его краям, наряду с известной у *D. melanogaster* разметкой от краев в глубину на ранней стадии эмбрионального развития [65].

В процессе моделирования были изучены и другие (близкие) варианты реализации регуляторной логики данной системы – 3- и 5-компонентные. При этом белковые продукты отдельных генов отличались функциональностью в плане количества одновременно регулируемых ими генов и типов регуляции. Все три варианта в многоядерной системе с диффузией регуляторных белков обладают схожей на качественном и количественном уровне динамикой.

Модель клеточного ансамбля с перемещениями клеток

Рассматривается абстрактный клеточный ансамбль, имеющий исходную структуру, представленную на рисунке 14. Это N-клеточная цепь, в которой соседние клетки попарно связаны одна с другой. Межклеточные взаимодействия базируются на диффузии белковых продуктов внутриклеточных генных сетей.



Рис. 14. Исходная структура клеточного ансамбля.

Каждая клетка ансамбля содержит типовую генную сеть, схема которой изображена на рисунке 15. Основным ее компонентом является циклическая дигенная система (гены G_3 и G_4) с отрицательными обратными связями, имеющая два альтернативных устойчивых стационарных состояния, т.е. являющаяся молекулярным триггером. Гены G_1 и G_2 кодируют белки-активаторы генов G_3 и G_4 , соответственно. Кроме того, последние два гена активируют сами себя. Пятый ген репрессируется любым из генов триггера, поэтому активен только в отсутствии пороговых доз обоих репрессоров.

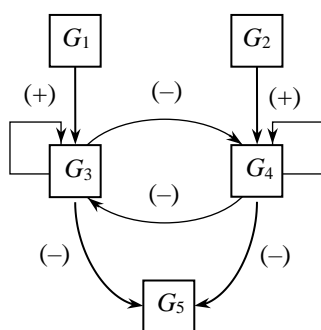


Рис. 15. Типовая схема индивидуальных генных сетей в клетках КА. В скобках при стрелках указан тип регуляторной связи – положительная (+) или отрицательная (-)

Таблица 4. Набор значений параметров генетических блоков

| | G_1 | G_2 | G_3 | G_4 | G_5 |
|----------|-------|-------|-------|-------|-------|
| a_{1j} | – | – | 1.5 | 1.5 | 2 |
| a_{2j} | – | – | 2 | 2 | 1.2 |
| b_{1j} | – | – | 0.05 | 0.05 | 0.05 |
| b_{2j} | 0.005 | 0.05 | 0.01 | 0.01 | 0.1 |

Примечание. Пороги: $P_{13} = P_{24} = 20$, $P_{33} = P_{44} = 50$, $P_{43} = P_{34} = 25$, $P_{35} = P_{45} = 40$

Пространственная динамика белковых продуктов клеточных генных сетей при неизменной конфигурации клеточного ансамбля приведена на рисунке 16. На противоположных КА-полюсах изначально присутствуют белки генов G_1 и G_2 («слева» и «справа», соответственно) в дозах, позволяющих запустить триггер в альтернативном состоянии на каждой условной стороне. Диффундируя в соседние клетки, белки генов G_1 и G_2 , равно как и белки генов триггера, расширяют (продвигают) область экспрессии подчиненных им генов.

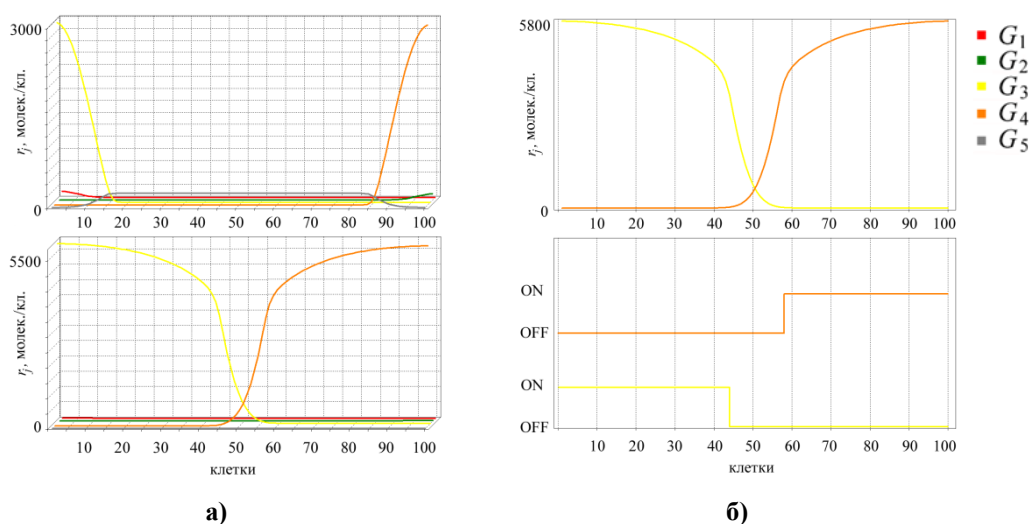


Рис. 16. Динамика в модели: **а)** распределение белков к 100-й и 450-й мин; **б)** сопоставление области экспрессии генов-антагонистов G_3 и G_4 и областей присутствия их белков. Значения параметров генной сети представлены в таблице 4; нач. концентрации белков генов G_1 и G_2 – 1000 молек./кл.; весовой коэффициент (аналог коэффициентов диффузии) = 0.1

Изменение структуры клеточного ансамбля происходит согласно паттерну, представленному на рисунке 17. Качественно его можно описать как сдвиг правой половины цепи вдоль левой с образованием двурядной структуры.

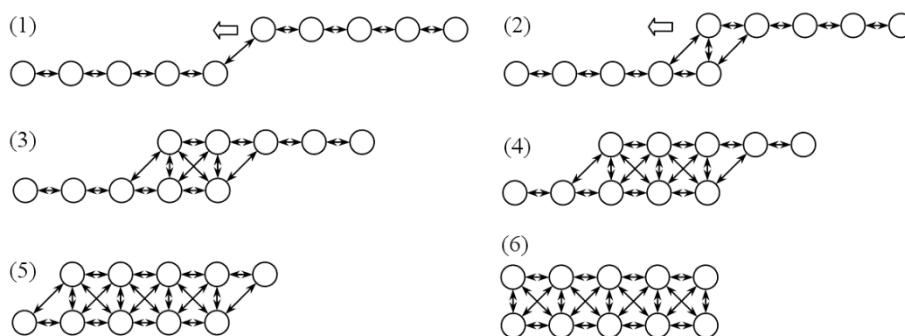


Рис. 17. Паттерн морфогенетического движения клеток

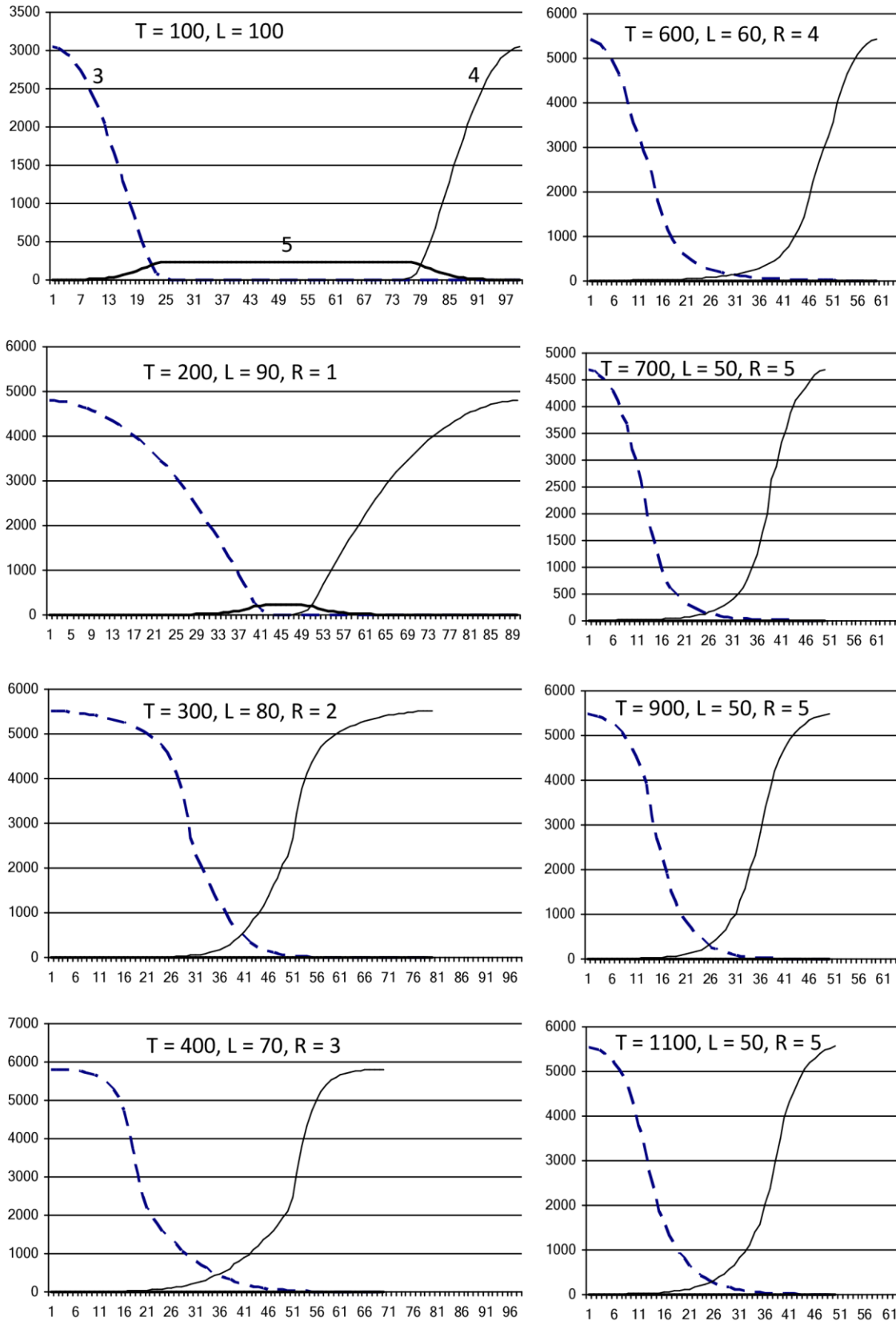


Рис. 18. Распределение белковых продуктов генов G_3 – G_5 в подцепях клеточного ансамбля в результате экспрессии генов, диффузии белков и перемещений клеток. Обозначения: T – время (мин), L – протяженность подцепи, R – раунд перемещения. По оси Ox – k -позиции клеток из КА в подцепи, образованной неподвижными клетками и клетками перемещающимися, но не совмещенными с первыми: при $k \leq \frac{1}{2}N$ в позиции k стоит клетка C_k ; при $k > \frac{1}{2}N$ (для случаев $\frac{1}{2}N < L \leq N$) в позиции k стоит клетка $C_{k+(N-L)}$. По оси Oy – концентрации белков (молек./кл.)

На заданном интервале наблюдений реконфигурация может производиться многократно с произвольным шагом. В качестве примера приведены расчеты для КА размером (N) и исходной протяженностью (L) в 100 клеток, функционирующего в течение 1100 мин при перемещениях клеток правой (верхней) части на 10 позиций к началу левой (нижней) в моменты времени $T = 101, 201, 301, 401, 601$ мин (раунды $R = 1, 3, 4, 5$, соответственно). Итоговая КА-структура аналогична варианту (6) на рисунке 17, при этом $L = 50$. Профили распределения белков в описанных условиях показаны на рисунке 18. Динамика индивидуальных генных сетей в клетках 25 и 26 (на границе двух областей с преобладанием альтернативных белков) представлена на рисунке 19.

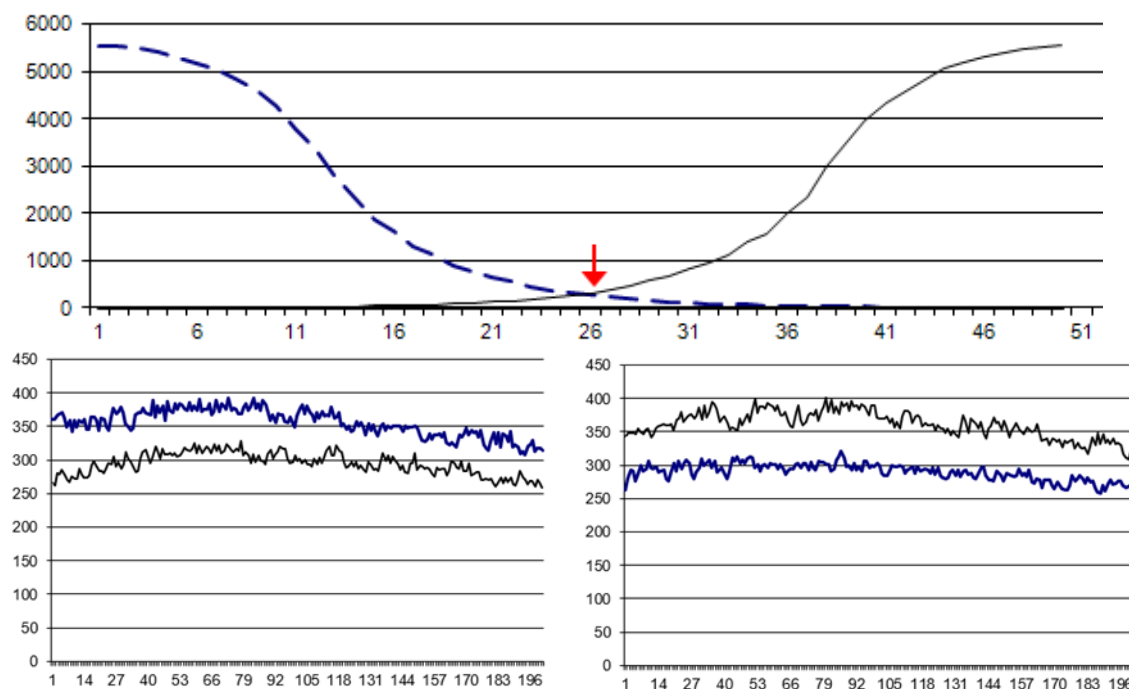


Рис. 19. Динамика генной сети в клетках на границе областей экспрессии генов. Верхний график – пространственное распределение белков генов G_3 и G_4 в клетках 1–50 в момент времени $T = 1100$ мин; нижние графики – кинетические кривые для белков (молек./кл.) в клетках 25 (слева) и 26 (справа) с 900-й (1-й) по 1100-ю (200-ю) мин.

Поведение системы, подобно системе из второго примера, характеризуется встречным расширением областей экспрессии генов G_3 и G_4 . При этом в силу отсутствия репрессивного воздействия на них белка гена-мишени G_5 эти области со временем стабильно фиксируются. Примененный двигательный паттерн может повлиять на этот баланс как количественно, так и качественно. В первом случае это может быть ускорение, а во втором – нивелирование. Ключевым фактором здесь выступает длина актуальной цепи, определяемая длиной начальной цепи и глубиной текущего сдвига. В модели обнаружено, что области экспрессии альтернативных генов и области преобладающего присутствия их белковых продуктов неравнозначны: первая меньше второй (рис. 16,б). Как итог, в центре формируется область с нулевой экспрессией генов-антагонистов (14 клеток внутри 100-клеточной цепи). Сдвиг подцепей сближает области экспрессии обоих генов и через шаг сужает каждую, продлевая во встречных направлениях области действия репрессоров. Критическое сближение может полностью обнулить экспрессию альтернативных генов во всем массиве клеток. После чего, с падением концентраций белков генов G_3 и G_4 ниже порогов, произойдет стабильное включение гена G_5 . Отметим, что в первом случае сдвиг ускоряет процесс, при котором ген G_5 выключается. Таким образом, движение типа «сдвиг» может приводить к активации одних генов и деактивации других.

ОБСУЖДЕНИЕ

Генные сети как группы взаимодействующих генов и клеточные ансамбли как системы взаимосвязанных клеток функционируют по собственным законам, влияя друг на друга. Современная вычислительная биология располагает целым арсеналом эффективных методов, моделей и программных средств, позволяющим интегрировать оба этих уровня организации живой материи.

Методы для анализа динамики генных сетей подразделяются на дискретные, непрерывные и смешанные. Первые базируются на булевой логике [17, 66], вторые – на дифференциальном исчислении [67], а третьи используют обе парадигмы [68, 69]. Существует большое количество биоинформационных проектов, в рамках которых развиваются как частные, так и общие модели генно-сетевой динамики [70]. Программный комплекс SETIES разработан с учетом достижений динамической теории молекулярно-генетических систем управления [48, 71], а его расчетный модуль реализует метод обобщенных пороговых моделей [47, 48], относящийся к смешанному типу. ОПМ-формализм учитывает дискретность транскрипционной регуляции генома, позволяя получать, наряду с диаграммами активности промоторов генов, кинетические кривые для генных продуктов и тем самым полноценно решать основную задачу анализа динамики. Подобное профилирование, равно как и возможность идентификации клеточных фенотипов через функциональные состояния генных сетей, естественным образом приводит к моделям многоклеточной генной экспрессии.

В исследованиях *in silico* клеточных ансамблей, будь то колонии микроорганизмов или ткани многоклеточных, применяют модели, которые в явном виде представляют отдельные клетки и взаимодействия между ними и основаны на агентах [72, 73]. Континуальные модели работают с плотностью популяции и концентрациями вещества и используют дифференциальные уравнения [74]. Модели гибридного типа сочетают в себе преимущества обоих подходов, избирательно масштабируя отдельные части популяции сообразно требуемой точности [72].

Дискретные модели клеточных ансамблей подразделяются на два типа – решеточные (lattice-based) и нерешеточные (lattice-free) [72, 73]. Решеточные модели ограничивают пространственное расположение клеток фиксированной сеткой – решеткой, узлы которой ассоциированы с клетками (один узел содержит одну клетку или много клеток или одна клетка занимает несколько узлов). К ним относятся клеточные автоматы (cell automata, CA) [75], решеточные газы (lattice-gas cell automata, LGCA) [76] и клеточная модель Поттса (cell Potts model, CPM) [77]. Решетки бывают регулярные (квадратная, сотовая) и нерегулярные (как производные от регулярных). В нерешеточных моделях положение клетки определяется либо ее центром (center-based model, CBM) [78], либо вершинами (vertex model, VM) [79].

В качестве примера платформ, развивающих такого рода модели, приведем следующие: CompuCell3D [80], Morpheus [81], Artistoo [82], Tissue simulation toolkit [83], Chaste [84], PhysiCell [85], Gell [86]. Области применения этих и аналогичных им программ: индивидуальное поведение клеток (рождение, рост, деление, гибель), коллективная миграция, клеточная адгезия, морфогенез и дифференцировка тканей, механотрансдукция, взаимодействие с патогенами.

Охарактеризуем методологию SETIES относительно перечисленных моделей. В методе ОПМ динамика внутриклеточных блоков описывается входными, внутренними и выходными переменными. Это концентрации регуляторных белков, управляющие переменные и концентрации синтезируемых в блоках генных продуктов, соответственно. Из них первые и третьи являются непрерывными величинами, а вторые – дискретными. Управляющие переменные формируются на выходе управляющей подсистемы, которая представляет собой конечный автомат с памятью (рис. 1). Это позволяет задавать правила межклеточных взаимодействий разной

сложности – можно использовать всю структуру и как простой клеточный автомат, и как более сложную агентную модель дискретно-непрерывного типа.

В SETIES узлы решетки содержат по одной клетке, они же реализуют систему морфогенетических координат. Представление клеточного ансамбля в виде граф-схемы позволяет модели быть опционально как решеточной, так и нерешеточной. Конфигурирование сети посредством добавления / удаления межклеточных связей (или изменением весовых коэффициентов при ребрах) обеспечивает для двумерного случая окрестности из четырех (окрестность Неймана), шести (соты) или восьми (окрестность Мура) ячеек. В целом возможны произвольные окрестности для разных клеток, что актуально для моделирования клеточных популяций с учетом их слоистости или внутренней очаговости. Перемещение клеток имитируется обновлением списков смежности вершин графа.

Отметим, что в работе [87] также используется сетевое представление многоклеточной системы. Авторы вводят для каждой клетки матрицу внутриклеточных межгенных связей, для каждой пары смежных клеток – матрицу весовых коэффициентов, характеризующих влияние генов одной клетки на гены другой, а для всех клеток – матрицу межклеточных взаимодействий. Далее по этим матрицам работает нейронная сеть Хопфилда, посредством которой для начальных транскрипционных профилей связанных между собой генных сетей индивидуальных клеток прогнозируется репертуар устойчивых тканевых паттернов.

Из нерешеточных алгоритмов модель SETIES в части представления клеток близка к моделям, основанным на центре. Координаты клетки на граф-схеме задаются именно как центр окружности. При необходимости им можно придать физический смысл и использовать в методах, расширяющих текущий функционал программы.

ЗАКЛЮЧЕНИЕ

Изучение генных регуляторных сетей методами вычислительной биологии предполагает описание их системных свойств как в абстрагированном виде, так и в контексте связности с другими сетевыми структурами. Вариантом такой комплексности можно считать «агрегацию по значению» копии генной сети в несущую ее клетку и представление группы взаимосвязанных клеток как регуляторной сети над генными регуляторными сетями. Данную концепцию в полной мере реализует семейство представленных в работе информационных моделей и разработанная на их основе платформа SETIES. Реализуя также формализм обобщенных пороговых моделей, платформа поддерживает решение основных задач портретного моделирования молекулярно-генетических систем. Среди них конструирование и визуализация схем произвольной конфигурации, анализ сетевой динамики в клетках, в клеточных линиях и в клеточных ансамблях, получение диаграмм активности промоторов и кинетических кривых для мРНК и белков, тестирование генных сетей на эпигенетические свойства, моделирование эпимутаций, изучение динамической памяти клеточной популяции. В целом, развитие платформы осуществляется на основе теории генных и эпигенных сетей и сопутствующей ей динамической теории молекулярно-генетических систем управления [71], вместе позволяющие решать задачи анализа и синтеза природных, искусственных и гипотетических генных сетей, управляющих как метаболическими, так и онтогенетическими процессами в живых системах.

Работа выполнена в рамках государственного задания Минобрнауки России № 075-03-2021-607 по теме № 122031100163-4.

СПИСОК ЛИТЕРАТУРЫ

1. Pržulj N., Malod-Dognin N. Simplicity within biological complexity. *Bioinform. Adv.* 2025. V. 5. No. 1. Article No. vbae164. doi: [10.1093/bioadv/vbae164](https://doi.org/10.1093/bioadv/vbae164)
2. Panditrao G., Bhowmick R., Meena C., Sarkar R.R. Emerging landscape of molecular interaction networks: Opportunities, challenges and prospects. *J. Biosci.* 2022. V. 47. No. 2. Article No. 24. doi: [10.1007/s12038-022-00253-y](https://doi.org/10.1007/s12038-022-00253-y)
3. Gysi D.M., Nowick K. Construction, comparison and evolution of networks in life sciences and other disciplines. *J. R. Soc. Interface.* 2020. V. 17. No. 166. Article No. 20190610. doi: [10.1098/rsif.2019.0610](https://doi.org/10.1098/rsif.2019.0610)
4. Kivelä M., Arenas A., Barthelemy M., Gleeson J.P., Moreno Y., Porter M.A. Multilayer networks. *J. Complex. Netw.* 2014. V. 2. No. 3. P. 203–71. doi: [10.1093/comnet/cnu016](https://doi.org/10.1093/comnet/cnu016)
5. Nicosia V., Bianconi G., Latora V., Barthelemy M. Growing multiplex networks. *Phys. Rev. Lett.* 2013. V. 111. No. 5. Article No. 058701. doi: [10.1103/PhysRevLett.111.058701](https://doi.org/10.1103/PhysRevLett.111.058701)
6. Wang W., Meng J., Li H., Fan J. Non-negative matrix factorization for overlapping community detection in directed weighted networks with sparse constraints. *Chaos.* 2023. V. 33. No. 5. Article No. 053111. doi: [10.1063/5.0152280](https://doi.org/10.1063/5.0152280)
7. Palla G., Derényi I., Farkas I., Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature.* 2005. V. 435. No. 7043. P. 814–818. doi: [10.1038/nature03607](https://doi.org/10.1038/nature03607)
8. Yan W., Hu G. Structural biology meets biomolecular networks: The Post-AlphaFold Era. *Current Bioinformatics.* 2022. V. 17. No. 6. doi: [10.2174/1574893617666220211115211](https://doi.org/10.2174/1574893617666220211115211)
9. Zheng H., Fu X., Shao J., Tang Y., Yu M., Li L., Huang L., Tang K. Transcriptional regulatory network of high-value active ingredients in medicinal plants. *Trends in plant science.* 2023. V. 28. No. 4. P. 429–446. doi: [10.1016/j.tplants.2022.12.007](https://doi.org/10.1016/j.tplants.2022.12.007)
10. Fradera-Sola A., Nischwitz E., Bayer M.E., Luck K., Butter F. RNA-dependent interactome allows network-based assignment of RNA-binding protein function. *Nucleic Acids Res.* 2023. V. 51. No. 10. P. 5162–5176. doi: [10.1093/nar/gkad245](https://doi.org/10.1093/nar/gkad245)
11. Papenfort K., Melamed S. Small RNAs, large networks: posttranscriptional regulons in gram-negative bacteria. *Annu. Rev. Microbiol.* 2023. V. 77. P. 23–43. doi: [10.1146/annurev-micro-041320-025836](https://doi.org/10.1146/annurev-micro-041320-025836)
12. Zhang H., Feng Z., Wu C. Refining protein interaction network for identifying essential proteins. *Current Bioinformatics.* 2023. V. 18. No. 3. P. 255–265. doi: [10.2174/1574893618666230217140446](https://doi.org/10.2174/1574893618666230217140446)
13. Meng X., Li W., Xiang J., Bedru H.D. Temporal-spatial analysis of the essentiality of hub proteins in protein-protein interaction networks. *IEEE Trans. Netw. Sci. Eng.* 2022. V. 9. No. 5. P. 3504–3514. doi: [10.1109/TNSE.2022.3185717](https://doi.org/10.1109/TNSE.2022.3185717)
14. Banos D.T., Trébulle P., Elati M. Integrating transcriptional activity in genome-scale models of metabolism. *BMC Syst. Biol.* 2017. V. 11. S. 7. Article No. 134. doi: [10.1186/s12918-017-0507-0](https://doi.org/10.1186/s12918-017-0507-0)
15. Soh K.C., Miskovic L., Hatzimanikatis V. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS Yeast Res.* 2012. V. 12. P. 129–143. doi: [10.1111/j.1567-1364.2011.00771.x](https://doi.org/10.1111/j.1567-1364.2011.00771.x)
16. Grüning N.M., Lehrach H., Ralser M. Regulatory crosstalk of the metabolic network. *Trends Biochem. Sci.* 2010. V. 35. No. 4. P. 220–227. doi: [10.1016/j.tibs.2009.12.001](https://doi.org/10.1016/j.tibs.2009.12.001)
17. Kauffman S.A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 1969. V. 22. No. 3. P. 437–467. doi: [10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0)

18. Zhou Z., Liu Y., Feng Y., Klepin S., Tsimring L.S., Pillus L., Hasty J., Hao N. Engineering longevity-design of a synthetic gene oscillator to slow cell aging. *Science*. 2023. V. 380. No. 6643. P. 376–381. doi: [10.1126/science.add7631](https://doi.org/10.1126/science.add7631)
19. Brophy J.A.N., Magallon K.J., Duan L., Zhong V. Synthetic genetic circuits as a means of reprogramming plant roots. *Science*. 2022. V. 377. No. 6607. P. 747–751. doi: [10.1126/science.abo4326](https://doi.org/10.1126/science.abo4326)
20. Hummel N.F.C., Zhou A., Li B., Markel K., Ornelas I.J., Shih P.M. The trans-regulatory landscape of gene networks in plants. *Cell Syst*. 2023. V. 14. No. 6. Article No. 501–511.e4. doi: [10.1016/j.cels.2023.05.002](https://doi.org/10.1016/j.cels.2023.05.002)
21. Akberdin I.R., Vertyshev A.Yu., Pintus S.S., Popov D.V., Kolpakov F.A. A mathematical model linking Ca²⁺-dependent signaling pathway and gene expression regulation in human skeletal muscle. *Mathematical biology and bioinformatics*. 2020. V. 15. No. 1. P. 20–39. doi: [10.17537/2020.15.20](https://doi.org/10.17537/2020.15.20)
22. Costanzo M., Kuzmin E., van Leeuwen J., Mair B., Moffat J., Boone C., Andrews B. Global genetic networks and the genotype-to-phenotype relationship. *Cell*. 2019. V. 177. No. 1. P. 85–100. doi: [10.1016/j.cell.2019.01.033](https://doi.org/10.1016/j.cell.2019.01.033)
23. Martinez-Pastor M., Tonner P.D., Darnell C.L., Schmid A.K. Transcriptional regulation in Archaea: from individual genes to global regulatory networks. *Annu. Rev. Genet.* 2017. V. 51. P. 143–170. doi: [10.1146/annurev-genet-120116-023413](https://doi.org/10.1146/annurev-genet-120116-023413)
24. Borriello E., Walker S.I., Laubichler M.D. Cell phenotypes as macrostates of the GRN dynamics. *J. Exp. Zool. B Mol. Dev. Evol.* 2020. V. 334. No. 4. P. 213–224. doi: [10.1002/jez.b.22938](https://doi.org/10.1002/jez.b.22938)
25. Roy S., Kundu T.K. Gene regulatory networks and epigenetic modifications in cell differentiation. *IUBMB Life*. 2014. V. 66. No. 2. P. 100–109. doi: [10.1002/iub.1249](https://doi.org/10.1002/iub.1249)
26. Nazarieh M., Wiese A., Will T., Hamed M., Helms V. Identification of key player genes in gene regulatory networks. *BMC Syst. Biol.* 2016. V. 10. No. 1. Article No. 88. doi: [10.1186/s12918-016-0329-5](https://doi.org/10.1186/s12918-016-0329-5)
27. Bakhteh S., Ghaffari-Hadigheh A., Chaparzadeh N. Identification of minimum set of master regulatory genes in gene regulatory networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2020. V. 17. No. 3. P. 999–1009. doi: [10.1109/TCBB.2018.2875692](https://doi.org/10.1109/TCBB.2018.2875692)
28. Liu Y.Y., Slotine J.J., and Barabási A.L. Controllability of complex networks. *Nature*. 2011. V. 473. No. 7346. P. 167–173. doi: [10.1038/nature10011](https://doi.org/10.1038/nature10011)
29. Moreno-Gómez S., Hochberg M.E., van Doorn G.S. Quorum sensing as a mechanism to harness the wisdom of the crowds. *Nat. Commun.* 2023. V. 14. No. 1. Article No. 3415. doi: [10.1038/s41467-023-37950-7](https://doi.org/10.1038/s41467-023-37950-7)
30. Efremenko E., Senko O., Stepanov N., Aslanli A., Maslova O., Lyagin I. Quorum sensing as a trigger that improves characteristics of microbial biocatalysts. *Microorganisms*. 2023. V. 11. No. 6. Article No. 1395. doi: [10.3390/microorganisms11061395](https://doi.org/10.3390/microorganisms11061395)
31. Espina J.A., Cordeiro M.H., Barriga E.H. Tissue interplay during morphogenesis. *Semin. Cell Dev. Biol.* 2023. V. 147. P. 12–23. doi: [10.1016/j.semcdb.2023.03.010](https://doi.org/10.1016/j.semcdb.2023.03.010)
32. Stapornwongkul K.S., Vincent J.P. Generation of extracellular morphogen gradients: the case for diffusion. *Nat. Rev. Genet.* 2021. V. 22. No. 6. P. 393–411. doi: [10.1038/s41576-021-00342-y](https://doi.org/10.1038/s41576-021-00342-y)
33. Gorfinkiel N., Martinez-Arias A. The cell in the age of the genomic revolution: Cell Regulatory Networks. *Cells Dev.* 2021. V. 168. Article No. 203720. doi: [10.1016/j.cdev.2021.203720](https://doi.org/10.1016/j.cdev.2021.203720)
34. Paton V., Türei D., Ivanova O., Müller-Dott S., Rodriguez-Mier P., Venafra V., Perfetto L., Garrido-Rodriguez M., Saez-Rodriguez J. NetworkCommons: bridging data, knowledge, and methods to build and evaluate context-specific biological

- networks. *Bioinformatics*. 2025. V. 41. No. 2. Article No. btaf048. doi: [10.1093/bioinformatics/btaf048](https://doi.org/10.1093/bioinformatics/btaf048)
35. Kolpakov F., Akberdin I., Kiselev I., Kolmykov S., Kondrakhin Y., Kulyashov M., Kutumova E., Pintus S., Ryabova A., Sharipov R. et al. BioUML-towards a universal research platform. *Nucleic acids res.* 2022. V. 50. No. W1. P. W124–W131. doi: [10.1093/nar/gkac286](https://doi.org/10.1093/nar/gkac286)
 36. Kolmykov S., Yevshin I., Kulyashov M., Sharipov R., Kondrakhin Y., Makeev V.J., Kulakovskiy I.V., Kel A., Kolpakov F. GTRD: an integrated view of transcription regulation. *Nucleic acids res.* 2021. V. 49. No. D1. P. D104–D111. doi: [10.1093/nar/gkaa105](https://doi.org/10.1093/nar/gkaa105)
 37. Le Novère N., Hucka M., Mi H., Moodie S., Schreiber F., Sorokin A., Demir E., Wegner K., Aladjem M.I., Wimalaratne S.M. et al. The systems biology graphical notation. *Nat. biotechnol.* 2009. V. 27. No. 8. P. 735–741. doi: [10.1038/nbt.1558](https://doi.org/10.1038/nbt.1558)
 38. Bergmann F.T., Czauderna T., Dogrusoz U., Rougny A., Dräger A., Touré V., Mazein A., Blinov M.L., Luna A. Systems biology graphical notation markup language (SBGNML) version 0.3. *J. Integr. Bioinform.* 2020. V. 17. No. 2–3. Article No. 20200016. doi: [10.1515/jib-2020-0016](https://doi.org/10.1515/jib-2020-0016)
 39. Rougny A., Balaur I., Luna A., Mazein A. StonPy: a tool to parse and query collections of SBGN maps in a graph database. *Bioinformatics*. 2023. V. 39. No. 3. Article No. btad100. doi: [10.1093/bioinformatics/btad100](https://doi.org/10.1093/bioinformatics/btad100)
 40. Noor M.S., Ferdous S., Salehi R., Gates H., Dey S., Raghunath V.S., Zargar M.R., Chowdhury R. Next-generation metabolic models informed by biomolecular simulations. *Curr. Opin. Biotechnol.* 2025. V. 92. Article No. 103259. doi: [10.1016/j.copbio.2025.103259](https://doi.org/10.1016/j.copbio.2025.103259)
 41. Cao Y., Pi W., Lin C.Y., Munzner U., Ohtomo M., Akutsu T. Common attractors in multiple boolean networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2023. V. 20. No. 5. P. 2862–2873. doi: [10.1109/TCBB.2023.3268795](https://doi.org/10.1109/TCBB.2023.3268795)
 42. Ren Y., Sarkar A., Veltri P., Ay A., Dobra A., Kahveci T. Pattern discovery in multilayer networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2022. V. 19. No. 2. P. 741–752. doi: [10.1109/TCBB.2021.3105001](https://doi.org/10.1109/TCBB.2021.3105001)
 43. Lakhova T.N., Kazantsev F.V., Lashin S.A., Matushkin Y.G. The finding and researching algorithm for potentially oscillating enzymatic systems. *Vavilov Journal of genetics and breeding*. 2021. V. 25. No. 3. P. 318–330. doi: [10.18699/VJ21.035](https://doi.org/10.18699/VJ21.035)
 44. Tyson J.J., Laomettachit T., Kraikivski P. Modeling the dynamic behavior of biochemical regulatory networks. *J. Theor. Biol.* 2019. V. 462. P. 514–527. doi: [10.1016/j.jtbi.2018.11.034](https://doi.org/10.1016/j.jtbi.2018.11.034)
 45. Keating S.M., Waltemath D., König M., Zhang F., Dräger A., Chaouiya C., Bergmann F.T., Finney A., Gillespie C.S., Helikar T. et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.* 2020. V. 16. No. 8. Article No. e9110. doi: [10.15252/msb.20199110](https://doi.org/10.15252/msb.20199110)
 46. Lashin S.A., Klimentenko A.I., Mustafin Z.S., Kolchanov N.A., Matushkin Yu.G. HEC 2.0: improved simulation of the evolution of prokaryotic communities. *Mathematical biology and bioinformatics*. 2014. V. 9. No. 2. P. 585–596. doi: [10.17537/2014.9.585](https://doi.org/10.17537/2014.9.585)
 47. Tchuraev R.N. A new method for the analysis of the dynamics of the molecular genetic control systems. I. Description of the method of generalized threshold models. *J. Theor. Biol.* 1991. V. 151. No. 1. P. 71–87. doi: [10.1016/s0022-5193\(05\)80144-4](https://doi.org/10.1016/s0022-5193(05)80144-4)
 48. Tchuraev R.N., Galimzyanov A.V. Gene and epigene networks: two levels in organizing the hereditary system. *J. Theor. Biol.* 2009. V. 259. No. 4. P. 659–669. doi: [10.1016/j.jtbi.2009.03.034](https://doi.org/10.1016/j.jtbi.2009.03.034)

49. Galimzyanov A.V. «GREENCE» technology for *in silico* analysis of gene network dynamics in individual cells of a clonal population. *Biophysics*. 2006. V. 51. S. 1. P. S66–S69. doi: [10.1134/S0006350906070141](https://doi.org/10.1134/S0006350906070141)
50. Galimzyanov A.V., Tchuraev R.N. Dynamic mechanism of phase variation in bacteria based on multistable gene regulatory networks. *J. Theor. Biol.* 2022. V. 549. Article No. 111212. doi: [10.1016/j.jtbi.2022.111212](https://doi.org/10.1016/j.jtbi.2022.111212)
51. Galimzyanov A.V. *In silico* model of epigene network capable of generating and maintaining phenotypical heterogeneity of a bacterial population. *Biomcs*. 2022. V. 14. No. 3. P. 278–284. doi: [10.31301/2221-6197.bmcs.2022-25](https://doi.org/10.31301/2221-6197.bmcs.2022-25)
52. Galimzyanov A.V. On one algorithm for modeling passive transport in cell sets of arbitrary configuration. *Proc. 4th BGRS*. Novosibirsk, Russia, 2004. V. 2. P. 38–41. URL: <https://conf.icgbio.ru/bgrs2004/en/> (дата обращения: 04.04.2025).
53. Tanouchi Y., Pai A., Park H., Huang H., Stamatov R., Buchler N.E., You L. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature*. 2015. V. 523. P. 357–360. doi: [10.1038/nature14562](https://doi.org/10.1038/nature14562)
54. Dantzig G.B. *Linear Programming and Extensions*. Princeton, NJ, USA: Princeton University Press, 1963.
55. *OMG Unified Modeling Language (OMG UML)*. Object Management Group, version 2.5, 2015. URL: <http://www.omg.org/spec/UML/2.5/> (дата обращения: 04.04.2025).
56. Chen P.P. The entity-relationship model – toward a unified view of data. *ACM Trans. Database Syst.* 1976. V. 1. No. 1. P. 9–36. doi: [10.1145/320434.320440](https://doi.org/10.1145/320434.320440)
57. Bruce T.A. *Designing Quality Databases with IDEFIX Information Models*. New York, NY, USA: Dorset House Publishing, 1992.
58. Tchuraev R.N., Galimzyanov A.V. On parametric stability of gene networks controlling ontogenetic processes. *Mol. Biol.* 2003. V. 37. No. 1. P. 81–88. doi: [10.1023/A:1022388930948](https://doi.org/10.1023/A:1022388930948)
59. Galimzyanov A.V., Tchuraev R.N. An investigation of the structural stability of *Drosophila* control gene subnetwork in computer experiments. In: *Bioinformatics of Genome Regulation and Structure*. Ed. Kolchanov N., Hofstaedt R. Boston, MA, USA: Springer, 2004. P. 301–309. doi: [10.1007/978-1-4419-7152-4_32](https://doi.org/10.1007/978-1-4419-7152-4_32)
60. Tchuraev R.N., Stupak I.V., Stupak E.E., Galimzyanov A.V. A new epigene property: metastable epigenotypes. *Dokl. Biol. Sci.* 2006. V. 406. No. P. 97–99. doi: [10.1134/s0012496606010285](https://doi.org/10.1134/s0012496606010285)
61. Peter I.S. The function of architecture and logic in developmental gene regulatory networks. *Curr. Top. Dev. Biol.* 2020. V. 139. P. 267–295. doi: [10.1016/bs.ctdb.2020.04.001](https://doi.org/10.1016/bs.ctdb.2020.04.001)
62. Wolpert L. Positional information and the spatial pattern of cell differentiation. *J. Theor. Biol.* 1969. V. 25. No. 1. P. 1–47. doi: [10.1016/s0022-5193\(69\)80016-0](https://doi.org/10.1016/s0022-5193(69)80016-0)
63. Vadde B.V.L., Roeder A.H.K. Can the French flag and reaction-diffusion models explain flower patterning? Celebrating the 50th anniversary of the French flag model. *J. Exp. Bot.* 2020. V. 71. No. 10. P. 2886–2897. doi: [10.1093/jxb/eraa065](https://doi.org/10.1093/jxb/eraa065)
64. Reinitz J., Vakulenko S., Sudakow I., Grigoriev D. Robust morphogenesis by chaotic dynamics. *Sci. Rep.* 2023. V. 13. No. 1. Article No. 7482. doi: [10.1038/s41598-023-34041-x](https://doi.org/10.1038/s41598-023-34041-x)
65. Jaeger J. The gap gene network. *Cell. Mol. Life Sci.* 2011 V. 68. P. 243–274. doi: [10.1007/s00018-010-0536-y](https://doi.org/10.1007/s00018-010-0536-y)
66. Cifuentes-Fontanals L., Tonello E., Siebert H. Node and edge control strategy identification via trap spaces in Boolean networks. *BMC Bioinformatics*. 2025. V. 24. Suppl. 1. Article No. 494. doi: [10.1186/s12859-025-06135-y](https://doi.org/10.1186/s12859-025-06135-y)
67. Swain M.T., Mande, J.J., Dubitzky W. Comparative study of three commonly used continuous deterministic methods for modeling gene regulation networks. *BMC Bioinformatics*. 2010. V. 11. Article No. 459. doi: [10.1186/1471-2105-11-459](https://doi.org/10.1186/1471-2105-11-459)

68. Mestl T., Plahte E., Omholt S.W. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.* 1995. V. 176. P. 291–300. doi: [10.1006/jtbi.1995.0199](https://doi.org/10.1006/jtbi.1995.0199)
69. Figueiredo D, Martins MA, Chaves M. Applying differential dynamic logic to reconfigurable biological networks. *Math. Biosci.* 2017. V. 291. P. 10–20. doi: [10.1016/j.mbs.2017.05.012](https://doi.org/10.1016/j.mbs.2017.05.012)
70. Le Novère N. Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 2015. V. 16. No. 3. P. 146–158. doi: [10.1038/nrg3885](https://doi.org/10.1038/nrg3885)
71. Galimzyanov A.V., Stupak E.E., Tchuraev R.N. Epigene networks: theory, models, and experiment. *Biol. Bull. Rev.* 2019. V. 9. No. 6. P. 484–490. doi: [10.1134/S2079086419060021](https://doi.org/10.1134/S2079086419060021)
72. Van Liedekerke P., Palm M.M., Jagiella N., Drasdo D. Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Comp. Part. Mech.* 2015. V. 2. P. 401–444. doi: [10.1007/S40571-015-0082-3](https://doi.org/10.1007/S40571-015-0082-3)
73. Metzcar J., Wang Y., Heiland R., Macklin P. A review of cell-based computational modeling in cancer biology. *JCO Clinical Cancer Informatics.* 2019. V. 3. P. 1–13. doi: [10.1200/CCI.18.00069](https://doi.org/10.1200/CCI.18.00069)
74. Kondo S. An updated kernel-based Turing model for studying the mechanisms of biological pattern formation. 2017. *J. Theor. Biol.* V. 414. P. 120–127. doi: [0.1016/j.jtbi.2016.11.003](https://doi.org/10.1016/j.jtbi.2016.11.003)
75. Deutsch A., Dormann S. Cell automaton modeling of biological pattern formation, characterization, examples, and analysis. New York: Springer, 2017.
76. Chopard B., Ouared R., Deutsch A., Hatzikirou H., Wolf-Gladrow D. Lattice-gas cell automaton models for biology: from fluids to cells. *Acta Biotheor.* 2010. V. 58. No. 4. P. 329–340. doi: [10.1007/s10441-010-9118-5](https://doi.org/10.1007/s10441-010-9118-5)
77. Nemati H., de Graaf J. The cell Potts model on disordered lattices. *Soft matter.* 2024 V. 20. No. 42. P. 8337–8352. doi: [10.1039/d4sm00445k](https://doi.org/10.1039/d4sm00445k)
78. Alert R., Trepast X. Physical models of collective cell migration. *Annu. Rev. Condens. Matter Phys.* 2020. V. 11. P. 77–101. doi: [10.1146/annurev-conmatphys-031218-013516](https://doi.org/10.1146/annurev-conmatphys-031218-013516)
79. Alt S., Ganguly P., Salbreux G. Vertex models: from cell mechanics to tissue morphogenesis. *Phil. Trans. R. Soc. B.* 2017. V. 372. Article No. 20150520. doi: [10.1098/rstb.2015.0520](https://doi.org/10.1098/rstb.2015.0520)
80. Swat M.H., Thomas G.L., Belmonte J.M., Shirinifard A., Hmeljak D., Glazier J.A. Multi-scale modeling of tissues using CompuCell3D. *Methods Cell Biol.* 2012. V. 110. P. 325–366. doi: [10.1016/B978-0-12-388403-9.00013-8](https://doi.org/10.1016/B978-0-12-388403-9.00013-8)
81. Starruß J., de Back W., Brusch L., Deutsch A. Morpheus: a user-friendly modeling environment for multiscale and multicell systems biology. *Bioinformatics.* 2014. V. 30. No. 9. P. 1331–1332. doi: [10.1093/bioinformatics/btt772](https://doi.org/10.1093/bioinformatics/btt772)
82. Wortel I.M., Textor J. Artistoo, a library to build, share, and explore simulations of cells and tissues in the web browser. *eLife.* 2021. V. 10. Article No. e61288. doi: [10.7554/eLife.61288](https://doi.org/10.7554/eLife.61288)
83. Daub J.T., Merks R.M.H. Cell-Based Computational Modeling of Vascular Morphogenesis Using Tissue Simulation Toolkit. In: Ribatti D. (eds) Vascular Morphogenesis. *Methods in Molecular Biology* V. 1214. New York: Humana Press, 2015. doi: [10.1007/978-1-4939-1462-3_6](https://doi.org/10.1007/978-1-4939-1462-3_6)
84. Mirams G.R., Arthurs C.J., Bernabeu M.O., Bordas R., Cooper J., Corrias A., Davit Y., Dunn S. J., Fletcher A. G., Harvey D. G. et al. Chaste: an open source C++ library for computational physiology and biology. *PLoS computational biology.* 2013. V. 9. No. 3. Article No. e1002970. doi: [10.1371/journal.pcbi.1002970](https://doi.org/10.1371/journal.pcbi.1002970)
85. Heiland R., Bergman D., Lyons B., Waldow G., Cass J., Lima da Rocha H., Ruscone M., Noël V., Macklin P. PhysiCell Studio: a graphical tool to make agent-based modeling more accessible. *GigaByte.* 2024. Article No. 2024:gigabyte128. doi: [10.46471/gigabyte.128](https://doi.org/10.46471/gigabyte.128)

86. Du J., Zhou Y., Jin L., Sheng K. Gell: A GPU-powered 3D hybrid simulator for large-scale multicell system. *PLoS One*. 2023. V. 18. No. 7. Article No. e0288721. doi: [10.1371/journal.pone.0288721](https://doi.org/10.1371/journal.pone.0288721)
87. Smart M., Zilman A. Emergent properties of collective gene expression patterns in multicell systems. *Cell Reports Physical Science*. 2023. V. 4. No. 101247. doi: [10.48550/arXiv.2110.07170](https://doi.org/10.48550/arXiv.2110.07170)

Рукопись поступила в редакцию 06.04.2025, переработанный вариант поступил 05.01.2026.
Дата опубликования 04.03.2026.

===== MATHEMATICAL MODELING =====

Modeling of Gene Networks and Cell Ensembles on the SETIES Platform

Galimzyanov A.V.

Ufa Institute of Biology of Ufa Federal Research Centre of RAS, Ufa, Russia

Abstract. Gene networks that control gene expression play a primary role in determining the epigenetic and biochemical properties of cells. At the same time, the interconnectedness of gene networks with networks of other types and levels requires approaches that consider this complexity. The article addresses this problem by examining the nesting of gene networks within cells, which themselves are nodes of a higher-level network, namely, a cell ensemble. Information models (object, relational, and object-relational) of the control gene network over genetic blocks are created, as well as a model of the cell ensemble as a network over cell control gene networks. The combined two-level model describes the gene network structures within cell ensembles with the dynamic characteristics of the main molecular genetic processes. Based on these models, the SETIES platform was developed for *in silico* studies of molecular genetic systems controlling gene expression in prokaryotes and eukaryotes. The platform includes software components for the design and visualization of cell and intercell gene networks, the automatic construction of their mathematical models, and the simulation of network dynamics in cell-free systems, cell lines, and clonal cell populations, as well as in cell ensembles with constant and dynamic structure. The computational modules implement the formalism of generalized threshold models, combining the advantages of discrete and continuous methods for modeling these systems. Among the tasks addressed are the construction of gene activity diagrams over time, the calculation of kinetic curves for mRNA and protein concentrations, the identification of gene expression patterns as functional states of the gene network, the assessment of parametric stability of regimes, and the simulation of mechanisms of epigenetic inheritance, phenotypic heterogeneity, and morphogenetic positioning. Overall, the platform implements the "network of networks" paradigm, which serves as the basis for advanced research in systems and synthetic biology.

Key words: *networks of networks, control systems, gene expression, models*