10. Lyubetsky V.A., Gershgorin R.A., Seliverstov A.V., Gorbunov K.Yu. Algorithms for Reconstruction of Chromosomal Structures. BMC Bioinformatics, 2016, vol. 17, no. 40, pp. 1–23.

11. Lyubetsky V.A., Gershgorin R.A., Gorbunov K.Yu. Chromosome structures: reduction of certain problems with unequal gene content and gene paralogs to integer linear programming. BMC Bioinformatics, 2017, vol. 18, no. 537, pp. 1–18.

МОДЕЛЬ ГЕННОЙ РЕГУЛЯТОРНОЙ СЕТИ

Горбунов К.Ю.

Дискретная булевская *генетическая сеть* [1] – это конечное множество $g_1,g_2,\ldots,g_n$ из n генов, где каждый ген может находиться в одном из двух состояний (0 – пассивное и 1 – активное). Каждой упорядоченной паре генов (g,g') (возможно, $g=g'$) сопоставлено рациональное число – интенсивность воздействия гена g на ген g', когда g находится в активном состоянии. Оно может быть положительным (активация) или отрицательным (репрессия). Также, каждому гену сопоставлена некоторая интенсивность стационарного воздействия, независимая от состояний других генов, и неотрицательное рациональное число – порог активности. Функционирование сети происходит по шагам: в каждый следующий момент $t+1$ времени состояние $g(t+1)$ каждого гена g определяется суммой интенсивностей воздействий на него в момент t. Если эта сумма превышает порог его активности, ген g переходит в активное состояние, иначе – в пассивное (независимо от значения $g(t)$). *Стационарным состоянием* (или *аттрактором*) сети называется такой набор состояний всех генов, который не меняется со временем, т.е. для каждого гена g выполняется $g(t+1)=g(t)$. Последовательность состояний сети называется её *эволюционированием*. Первый повтор состояний в ней является либо стационарным состоянием (если повторилось предыдущее состояние), либо предельным циклом (если повторилось более раннее состояние). Задача поиска аттракторов (и предельных циклов заданной длины) в описанной сети является известной NP-трудной задачей, один из возможных подходов к ней описан в [1] и основан на использовании временной логики GRN. Мы предлагаем два других подхода: первый основан на сведении задачи к целочисленному линейному программированию (ЦЛП), второй – на обобщении задачи и сведении её к линейному или квадратичному программированию. Число переменных и ограничений в этих задачах линейно зависит от числа генов в сети.

ОПИСАНИЕ МОДЕЛИ. СВЕДЕНИЕ ЗАДАЧИ К ЦЛП

Каждому гену g_i сопоставим одноименную булеву переменную, выражающую его состояние. Пусть $g_1,g_2,\ldots,g_m$ – гены, активирующие ген g_i, а

числа $w_{1,i}, w_{2,i}, \ldots, w_{m,i}$ – коэффициенты, отражающие их интенсивности воздействия на g_i. Пусть также $g_{m+1}, g_{m+2}, \ldots, g_k$ – гены, репрессирующие ген g_i, соответственно, с интенсивностями $w_{m+1,i}, w_{m+2,i}, \ldots, w_{k,i}$ (по абсолютным значениям). Тогда суммарное воздействие $F(i)$ на ген g равно $b_i + \sum_{j=1}^{m} w_{j,i} g_j - \sum_{j=m+1}^{k} w_{j,i} g_j$, где b_i – интенсивность стационарного воздействия на ген g_i. Условия, определяющие стационарное состояние: $c_{1,i} g_i \geq F(i) - t_i$, $c_{2,i}(g_i - 1) \leq F(i) - t_i$, где t_i – порог активности гена g_i, $c_{1,i}$ – сумма всех положительных интенсивностей воздействий на g_i, $c_{2,i}$ – сумма модулей всех отрицательных интенсивностей воздействий на g_i и его порога активности. Чтобы отделить новые стационарные состояния от уже найденных, накладываем условия: для каждого найденного стационарного состояния $(g'_1, \ldots g'_m)$ выполняется $\sum_i \alpha(g_i) > 0$, где $\alpha(g_i) = g'_i$, если $g'_i = 0$ и $\alpha(g_i) = 1 - g'_i$ иначе.

Если ко всем переменным g_i добавить индекс $j = 1, \ldots m$, определяющий положение в предельном цикле, можно искать циклы заданной длины m, налагая условия $c_{1,i} g_{i,j+1} \geq F(i,j) - t_i$, $c_{2,i} g_{i,j+1} \leq F(i,j) - t_i$ для $j = 1, \ldots, m-1$ и $c_{1,i} g_{i,1} \geq F(i,m) - t_i$, $c_{2,i} g_{i,1} \leq F(i,m) - t_i$, а также аналогичное дополнительное условие на среднее число активных генов вдоль цикла. Чтобы отделить предельные циклы от найденных стационарных состояний, накладываем условия: для каждого стационарного состояния $(g_1, \ldots g_m)$ выполняется $\sum_i \alpha(g_{i1}) > 0$, где $\alpha(g_{i1}) = g_{i1}$, если $g_i = 0$ и $\alpha(g_{i1}) = 1 - g_{i1}$ иначе. Аналогично, отделяем найденные предельные циклы от новых.

Для решения задачи ЦЛП мы использовали известный пакет программ IBM CPLEX, который можно найти по следующей ссылке: https://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html.

Пример 1**.** Рассмотрим сеть из двух генов, изображённую на рис.1.



Рис. 1. Регуляторная сеть, содержащая два гена. Линией со стрелкой на конце обозначается активация, с палочкой на конце – репрессия

Каждый из генов активирует сам себя и репрессирует другой ген. Пусть интенсивности воздействия генов на себя равны 0.1, на другой ген – –0.3, интенсивности стационарного воздействия равны 0.7, пороги актив-

ности генов равны 0.6. Задача поиска стационарного состояния формулируется в виде четырех неравенств:

$$0.8g_1 \ge 0.7+0.1g_1-0.3g_2-0.6,\ 0.8g_2 \ge 0.7+0.1g_2-0.3g_1-0.6,$$
$$0.9(g_1-1) \le 0.7+0.1g_1-0.3g_2-0.6,\ 0.9(g_2-1) \le 0.7+0.1g_2-0.3g_1-0.6.$$

В этой сети найдено два стационарных состояния: (0,1) и (1,0). Если уменьшить интенсивности стационарного воздействия на гены, положив их равными 0.3, система примет вид

$$0.4g_1 \ge 0.3+0.1g_1-0.3g_2-0.6,\ 0.4g_2 \ge 0.3+0.1g_2-0.3g_1-0.6,$$
$$0.9(g_1-1) \le 0.3+0.1g_1-0.3g_2-0.6,\ 0.9(g_2-1) \le 0.3+0.1g_2-0.3g_1-0.6.$$

Теперь стационарное состояние лишь одно: (0,0).

Задача ЦЛП поиска предельного цикла длины 2 при первом варианте параметров содержит 8 основных неравенств:

$$0.8g_{12} \ge 0.7+0.1g_{11}-0.3g_{21}-0.6,\ 0.8g_{22} \ge 0.7+0.1g_{21}-0.3g_{11}-0.6,$$
$$0.9(g_{12}-1) \le 0.7+0.1g_{11}-0.3g_{21}-0.6,\ 0.9(g_{22}-1) \le 0.7+0.1g_{21}-0.3g_{11}-0.6;$$
$$0.8g_{11} \ge 0.7+0.1g_{12}-0.3g_{22}-0.6,\ 0.8g_{21} \ge 0.7+0.1g_{22}-0.3g_{12}-0.6,$$
$$0.9(g_{11}-1) \le 0.7+0.1g_{12}-0.3g_{22}-0.6,\ 0.9(g_{22}-1) \le 0.7+0.1g_{21}-0.3g_{11}-0.6.$$

Дополнительные ограничения $g_{11}+(1-g_{21})>0$ и $(1-g_{11})+g_{21}>0$ исключают из рассмотрения найденные стационарные состояния (0,1) и (1,0), и позволяют найти единственный предельный цикл: [(0,0),(1,1)]. При втором варианте параметров предельных циклов в сети нет.

Пример 2**.** Рассмотрим сеть из четырех генов, изображённую на рис.2.

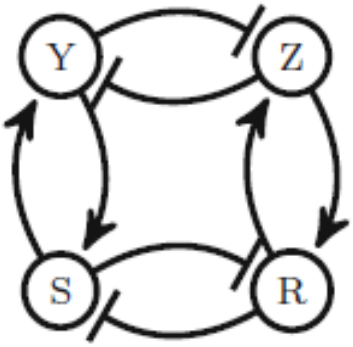


Рис. 2. Регуляторная сеть, содержащая четыре гена

Обозначим $Y=g_1$, $S=g_2$, $Z=g_3$, $R=g_4$. Пусть интенсивности активации генов равны 0.4, интенсивности репрессии – –0.6, интенсивности стационарного воздействия равны 0.4, пороги активности генов равны 0.3. Задача поиска стационарного состояния формулируется в виде восьми основных неравенств:

$$0.8g_1 \ge 0.4+0.4g_2-0.6g_3-0.3,\ 0.8g_2 \ge 0.4+0.4g_1-0.6g_4-0.3,$$
$$0.8g_3 \ge 0.4+0.4g_4-0.6g_1-0.3,\ 0.8g_4 \ge 0.4+0.4g_3-0.6g_2-0.3,$$
$$0.9(g_1-1) \le 0.4+0.4g_2-0.6g_3-0.3,\ 0.9(g_2-1) \le 0.4+0.4g_1-0.6g_4-0.3,$$

$$0.9(g_3-1)\le 0.4+0.4g_4-0.6g_1-0.3,\ 0.9(g_4-1)\le 0.4+0.4g_3-0.6g_2-0.3.$$

В этой сети найдено четыре стационарных состояния: (0,0,1,1), (1,1,0,0), (0,1,1,0), (1,0,0,1).

Задача ЦЛП поиска предельного цикла длины 2 содержит 16 основных неравенств:

$$0.8g_{12}\ge 0.4+0.4g_{21}-0.6g_{31}-0.3,\ 0.8g_{22}\ge 0.4+0.4g_{11}-0.6g_{41}-0.3,$$
$$0.8g_{32}\ge 0.4+0.4g_{41}-0.6g_{11}-0.3,\ 0.8g_{42}\ge 0.4+0.4g_{31}-0.6g_{21}-0.3,$$
$$0.9(g_{12}-1)\le 0.4+0.4g_{21}-0.6g_{31}-0.3,\ 0.9(g_{22}-1)\le 0.4+0.4g_{11}-0.6g_{41}-0.3,$$
$$0.9(g_{32}-1)\le 0.4+0.4g_{41}-0.6g_{11}-0.3,\ 0.9(g_{42}-1)\le 0.4+0.4g_{31}-0.6g_{21}-0.3;$$
$$0.8g_{11}\ge 0.4+0.4g_{22}-0.6g_{32}-0.3,\ 0.8g_{21}\ge 0.4+0.4g_{12}-0.6g_{42}-0.3,$$
$$0.8g_{31}\ge 0.4+0.4g_{42}-0.6g_{12}-0.3,\ 0.8g_{41}\ge 0.4+0.4g_{32}-0.6g_{22}-0.3,$$
$$0.9(g_{11}-1)\le 0.4+0.4g_{22}-0.6g_{32}-0.3,\ 0.9(g_{21}-1)\le 0.4+0.4g_{12}-0.6g_{42}-0.3,$$
$$0.9(g_{31}-1)\le 0.4+0.4g_{42}-0.6g_{12}-0.3,\ 0.9(g_{41}-1)\le 0.4+0.4g_{32}-0.6g_{22}-0.3$$

и четыре дополнительных:

$$g_{11}+g_{21}+(1-g_{31})+(1-g_{41})>0,\ (1-g_{11})+(1-g_{21})+g_{31}+g_{41}>0,$$
$$g_{11}+(1-g_{21})+(1-g_{31})+g_{41}>0,\ (1-g_{11})+g_{21}+g_{31}+(1-g_{41})>0.$$

Найден цикл [(0,0,0,0),(1,1,1,1)]. Для поиска других циклов накладываем два дополнительных ограничения: $g_{11}+g_{21}+g_{31}+g_{41}>0$ и $g_{12}+g_{22}+g_{32}+g_{42}>0$. Находим следующий цикл: [(1,1,1,0),(0,1,0,0)]. Продолжая таким же образом, находим ещё 4 цикла: [(0,1,0,1),(1,0,1,0)], [(0,1,1,1),(0,0,1,0)], [(1,1,0,1),(1,0,0,0)], [(1,0,1,1),(0,0,0,1)]. Циклов длины больше 2 в этой сети не найдено.

Пример 3. Рассмотрим сеть из пяти генов, изображённую на рис.3.

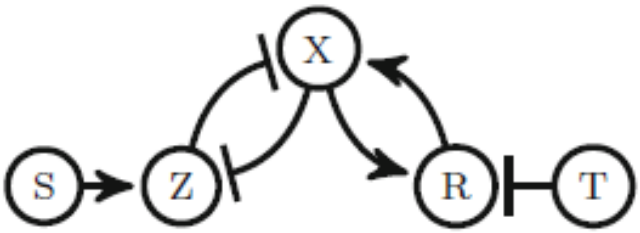


Рис. 3. Регуляторная сеть, содержащая пять генов

Обозначим $S=g_1$, $Z=g_2$, $X=g_3$, $R=g_4$, $T=g_5$. Пусть интенсивности активации генов равны 0.4, интенсивности репрессии – –0.6, интенсивности стационарного воздействия равны 0.4, пороги активности генов равны 0.3. Задача поиска стационарного состояния формулируется в виде десяти неравенств:

$$0.4g_1\ge 0.4-0.3,\ 0.8g_2\ge 0.4+0.4g_1-0.6g_3-0.3,$$
$$0.8g_3\ge 0.4+0.4g_4-0.6g_2-0.3,\ 0.8g_4\ge 0.4+0.4g_3-0.6g_5-0.3,$$
$$0.4g_5\ge 0.4-0.3,\ 0.3(g_1-1)\le 0.4-0.3,\ 0.9(g_2-1)\le 0.4+0.4g_1-0.6g_3-0.3,$$

$$0.9(g_3-1)\le 0.4+0.4g_4-0.6g_2-0.3,\ 0.9(g_4-1)\le 0.4+0.4g_3-0.6g_5-0.3,$$
$$0.3(g_5-1)\le 0.4-0.3.$$

В сети найдены два стационарных состояния: (1,0,1,0,1) и (1,1,0,0,1), а также один предельный цикл длины 2: [(1,0,0,0,1),(1,1,1,0,1)].

ОБОБЩЕНИЕ МОДЕЛИ. СВЕДЕНИЕ К ЛИНЕЙНОМУ И КВАДРАТИЧНОМУ ПРОГРАММИРОВАНИЮ

Не будем ограничивать степень экспрессии гена значениями 0 и 1, а разрешим ей принимать любые неотрицательные значения. Каждому гену g_i сопоставим одноимённую переменную, выражающую степень его экспрессии или уровень концентрации в клетке, что эквивалентно с точностью до пропорциональности. Пусть $g_1,g_2,\ldots,g_m$ – гены, активирующие ген g_i, а числа $c_{1,i},c_{2,i},\ldots,c_{m,i}$ – коэффициенты, отражающие их интенсивности (в частности, интенсивности ДНК-сайтов посадки регуляторных белков). Пусть также $g_{m+1},g_{m+2},\ldots,g_k$ – гены, репрессирующие ген g_i, соответственно, с интенсивностями $c_{m+1,i},c_{m+2,i},\ldots,c_{k,i}$. Суммарное воздействие $S(g_i)$ этих генов на ген g зададим выражением $\sum_{j=1}^{m} c_{j,i}g_j - \sum_{j=m+1}^{k} c_{j,i}g_j$. Пусть x_i – интенсивность стационарного воздействия на ген g_i, отражающая, в частности, степень интенсивности промотера; эту величину считаем неизвестной, но с известным ограничением снизу: $x_i \ge b_i \ge 0$ и сверху: $x_i \le d_i$. Тогда в идеальном случае в стационарном состоянии для всех i выполняются равенства $g_i = x_i + S(g_i)$. С учётом того, что в реальных ситуациях точные равенства могут не выполняться, возникает две задачи условной минимизации.

Задача 1. Естественно считать, что «в одну сторону» неравенства $g_i \le x_i + S(g_i)$ выполняются, а точные равенства не выполняются в силу некоторых конструктивных или иных факторов, ограничивающих уровень экспрессии генов. Отсюда возникает задача: при заданных коэффициентах c_i минимизировать значение $\sum_{i=1}^{n}(x_i + S(g_i) - g_i)$ по переменным x_i и g_i при условиях $x_i \ge b_i$, $x_i \le d_i$, $g_i \ge 0$ и $g_i \le x_i + S(g_i)$. Это задача линейного программирования.

Задача 2. Если допустить, что неравенства могут не выполняться в обе стороны (например, в силу неточного задания коэффициентов $c_{j,i}$, b_i, d_i), то возникает задача: при заданных коэффициентах $c_{j,i}$ минимизировать значение $\sum_{i=1}^{n}(x_i + S(g_i) - g_i)^2$ по переменным x_i и g_i при условиях $x_i \ge b_i$, $x_i \le d_i$, $g_i \ge 0$. Это задача квадратичного программирования.

Для решения задач 1 и 2 мы использовали известный пакет программ оптимизации MOSEK (www.mosek.com). Кроме поиска стационарных состояний, мы осуществляли также уточнение значений $c_{j,i}$, b_i и d_i, варьируя их и решая каждый раз две сформулированные задачи. При этом критериями служили значения минимумов минимизируемых функций и степень отличия решений двух задач друг от друга. Практика показала, что эти задачи имеют единственное биологически осмысленное решение лишь тогда, когда значения $c_{j,i}$, b_i, d_i известны достаточно точно, а отрезки $[b_i,d_i]$ малы. В противном случае нулевые значения функционалов в задачах 1 и 2 достигаются при многих наборах значений переменных, и из них не удаётся выбрать биологически разумного варианта. В этих случаях вместо задач 1 или 2 имеет смысл решать следующую задачу 3.

Задача 3. Применяется следующее правило выбора значений $c_{j,i}$: при каждом активаторе гена g коэффициент равен единице, деленной на количество активаторов (или положительных регуляций) гена g, а при каждом репрессоре гена g – соответственно, единице, деленной на количество репрессоров (или отрицательных регуляций) гена g. Кроме того на значение переменных g_i накладывается ограничение $g_i \le 1$. Далее, выберем приближенные значения переменных x_i (обозначим эти значения как x_i'), например, если гену g_i соответствует булевская функция (т.е. есть хоть один репрессор или активатор), то x_i' разумно положить ее значению при всех переменных, равных нулю; если же репрессоров и активаторов нет, то положим $x_i'=0.5$. Опишем задачу 3 квадратичного программирования. Переменные: g_i и x_i. Минимизируемый функционал – сумма квадратов отклонений x_i от x_i'. Линейные ограничения: $0 \le g_i \le 1$, $0 \le x_i \le 1$, и равенства $g_i = x_i + S(g_i)$, полученные из соответствующих неравенств задачи 1. Легко видеть, что задача 3 состоит в поиске проекции данной точки на пересечение 0-1-куба и (проходящего через ноль) линейного подпространства, а потому всегда имеет единственное решение.

Пример 1. Пусть генетическая сеть соответствует изображенной на рис. 1 в [2], то есть состоит из трёх генов g_1, g_2, g_3, которые в равновесном состоянии удовлетворяют формулам $g_1 = \neg g_2 \vee g_3$, $g_2 = g_1 \vee g_2 \vee \neg g_3$, $g_3 = g_1 \wedge \neg g_2$. Тогда для гена g_1 ген g_2 является репрессором, а ген g_3 – активатором, коэффициенты при них равны 1, $x_1'=1$. Для гена g_2 гены g_1 и g_2 являются активаторами, а ген g_3 – репрессором, коэффициенты при них равны, соответственно, 0.5, 0.5 и 1, $x_2'=1$. Для гена g_3 ген g_1 является активатором, а ген g_2 – репрессором, коэффициенты при них равны 1, $x_3'=0$. При таких значениях $c_{j,i}$ и x_i' решение задачи 3 дало следующий результат: значение целевой функции равно 0.35; $x_1=0.88$, $x_2=0.54$, $x_3=0.35$, $g_1=1$, $g_2=0.62$, $g_3=0.73$. Отметим, что с биологической точки зрения полученные значения g_1, g_2, g_3 более разумны, чем решение $g_1=0$, $g_2=1$, $g_3=0$ упомянутой булевской системы равенств.

Пример 2. Рассмотрим генетическую сеть в *Bacillus Subtilis*, изображенную на рис. 6 в [3]. В ней 85 вершин, включающих σ-факторы и транскрипционные факторы. Мы удалили из сети вершины, не связанные с основной частью сети, а также те вершины, в которых интенсивность транскрипции не влияет на интенсивности в других вершинах и однозначно по ним восстанавливается. После этого в сети осталось 31 вершина, соответствующая σ-факторам SigK, SigB, SigX, SigH, SigE, SigF, SigG, SigW и транскрипционным факторам CcpA, CodY, Med, DegU, YvrH, Mta, ComK, Spo0A, AbrB, YfiA, HutP, ExuR, PhoP, YrhI, SenS, ResD, CcpB, SpoVT, Hpr, LytR, FNR, SinR, ArfM. Значения переменных x_i' выбирались по следующему правилу: если вершина g не подвергается ни положительной, ни отрицательной регуляции, то в ней полагали g'=0.5, если только положительной, то g'=0, а если только отрицательной, то g'=1. Если же на g воздействуют, скажем, n положительных регуляторов и m отрицательных, то значение g' разбивало отрезок [0,1] в отношении m на n, например, если n=3, m=1, то g'=0.25. В сети имелась одна регуляция, знак которой не удалось установить авторам статьи [3]: это воздействие транскрипционного фактора ComK на сигма-фактор SigH. поэтому мы решали задачу 3 дважды: при гипотезе положительного воздействия и при гипотезе отрицательного. При гипотезе положительного воздействия значение минимизируемой функции составило 12.87, а значения интенсивностей транскрипций оказались следующими:
SigK=0.5, CcpA=0.5, CodY=0.5, SigB=0.62, Med=0.5, DegU=0.5, YvrH=0.5, Mta=0.5, ComK=0.5, AbrB=0.29, SigX=0.5, SigH=0.71, Spo0A=0.71, SigE=0.71, SigF=0.96, SpoVT=0.21, YfiA=0.37, HutP=0.46, ExuR=0.75, PhoP=0.28, SigG=0.79, YrhI=0.57, SigW=0.78, SenS=0.99, ResD=0.28, CcpB=0.79, Hpr=0.64, LytR=0.49, FNR=0.28, SinR=0.91, ArfM=0.39.

При гипотезе отрицательного воздействия значение минимизируемой функции составило 12.1, а значения интенсивностей транскрипций оказались следующими:
SigK=0.5, CcpA=0.5, CodY=0.5, SigB=0.65, Med=0.5, DegU=0.5, YvrH=0.5, Mta=0.5, ComK=0.41, AbrB=0.41, SigX=0.5, SigH=0.59, Spo0A=0.59, SigE=0.59, SigF=0.12, SpoVT=0.57, YfiA=0.25, HutP=0.58, ExuR=0.67, PhoP=0.24, SigG=0.43, YrhI=0.53, SigW=0.62, SenS=0.71, ResD=0.24, CcpB=0.43, Hpr=0.56, LytR=0.45, FNR=0.24, SinR=0.77, ArfM=0.34.

Как видно, при последней гипотезе интенсивности в среднем стали значительно меньше, в частности, максимальная интенсивность составила 0.77, тогда как при первой гипотезе три интенсивности превышали 0.9. В силу принципа максимальной экономии, гипотеза отрицательного воздействия транскрипционного фактора ComK на σ-фактор SigH выглядит более правдоподобной.

Пример 3. Рассмотрим генетическую регуляторную сеть (рис. 1 из [4]), содержащую гены транскрипционных факторов, участвующие в уста-

новлении двух режимов Th1 и Th2 иммунной защиты клеток. В ней 51 вершина. Значения переменных x_i' выбирались по тому же правилу, что и в примере 2. Значение минимизируемой функции составило 12.41, а значения интенсивностей транскрипций оказались следующими:
IL-13=0.35, IRF4=0.32, NFAT=0.43, MAF=0.39, GATA3=0.24, IL-5=0.35, TBET=0.64, IFN-γR=0.57, STAT1=0.54, IFN-γ=0.57, IL-7R=0.5, STAT4=0.45, IRAK=0.73, STAT6=0.39, IL-7=0.5, IL-18R=0.73, IL-18=0.5, IL-4R=0.27, IFN-αR1=0.5, IFN-α=0.5, IL-4=0.35, SOCS1=0.58, CTLA4=0.5, CD80=0.5, SHP1=0.5, LCK=0.5, CD45=0.5, CD4=0.5, ZAP70=0.24, TCR=0.5, CD3=0.5, SLP76=0.25, VAV1=0.5, ITK=0.41, CD28=0.5, PLCPG=0.41, ANTIGEN=0.5, TNFRSF4=0.5, TNFSF4=0.5, NFKB=0.26, IKBKB=0.5, PI3K=0.5, ICOS=0.5, AKT1=0.5, COT=0.5, NIK=0.5, CD86=0.5, JAK1=0.4, JAK3=0.27, IL-12R=0.5, IL-12=0.5.

Основной целью в [4] является опровержение распространенного мнения о том, что два режима Th1 и Th2 являются взаимоисключающими, т.е. когда активны гены, обеспечивающие режим Th1 (это гены IFN-γ, IFN-γR, SOCS1, STAT1, TBET), то неактивны гены, обеспечивающие режим Th2 (это гены GATA3, IL-13, IL-4, IL-4R, IL-5, IRF4, JAK1, JAK3, MAF, NFAT, STAT6) и наоборот. Для этого авторы рассмотрели упомянутую генетическую сеть в качестве булевской дискретной сети, где ген на каждом следующем шаге активен, если и только если на текущем шаге активен хоть один его активатор и не активно ни одного его репрессора. Был обнаружен цикл из трех состояний сети, которые в совокупности включали активное состояние как тех, так и других генов (кроме JAK1). Как видно из нашего результата, переход к действительнозначной сети позволяет сделать вывод о совместимости двух режимов на основании найденного стационарного состояния. В нем активны как гены из режима Th1, так и гены из режима Th2: средняя активность первых равна 0.58, а средняя активность вторых равна 0.34. Одно из возможных объяснений разницы активности состоит в том, что генов из Th1 меньше, поэтому для достижения требуемого эффекта они должны быть более активны.

Для подтверждения своего результата авторы статьи [4] рассмотрели ту же сеть с удаленным геном. Как они отмечают, если бы гены TBET и GATA3 были несовместимы, удаление гена TBET давало бы аттрактор с активным геном IL-4 и неактивным геном IFN-γ. Как их, так и наша модель не подтверждают гипотезу о несовместимости этих генов. Так, у нас после удаления гена TBET активность гена IFN-γ равна 0.53, а активность гена IL-4 равна 0.59.

В случаях, когда из биологических экспериментов известны дополнительные данные о концентрациях белков или мРНК, становится возможным решать задачу 1. При этом часто удается получить точные равенства, и задача сводится к решению системы уравнений.

Пример 4. Рассмотрим сеть из пяти генов, связанных с метаболизмом азота: *ntcA, glnA, glnN, gifA* и *gifB*, у которых концентрации соответствующих белков зависят от параметра a, равного относительной концентрации аммония в клетке. В [5] приведены некоторые данные об этой зависимости и описаны известные биологические механизмы рассматриваемой генетической сети. Напомним, что ген *ntcA* кодирует транскрипционный фактор, активирующий транскрипцию генов *ntcA, glnA, glnN* и репрессирующий транскрипцию *gifA* и *gifB*, а белки GifA (IF7) и GifB (IF17) инактивируют ферменты GlnA и GlnN. При этом любой из них инактивикует каждый из двух ферментов. Введем для рассматриваемых факторов переменные: x – относительная концентрация белка NtcA, y_1 – GlnA, y_2 – GlnN, z_1 – GifA (IF7) и z_2 – GifB (IF17). Фактор NtcA активирует сам себя, причем сила активации тем больше, чем меньше концентрация аммония. Считая, что концентрации принимают значения от 0 до 1, получаем уравнение $x=b_1+s_1(1-a)x$, где b_1 – базовый уровень концентрации NtcA, s_1 – интенсивность сайта посадки белка (поскольку NtcA всегда существует в форме димеров, т.е. сдвоенных молекул, под концентрацией NtcA везде понимается концентрация димеров NtcA). Обозначая b_0 – уровень концентрации NtcA при нулевой концентрации аммония, т.е. при $a=0$, имеем $b_0=b_1+s_1b_0$, и после замены b_1 на b_0 уравнение приобретает вид $x=b_0-s_1b_0+s_1(1-a)x$, откуда получаем

$$x=\frac{b_0(1-s_1)}{1-s_1(1-a)}.$$

Отсюда можно сделать первый вывод: при увеличении концентрации аммония в клетке концентрация транскрипционного фактора NtcA убывает.

В статье [6] приведены данные о регуляции NtcA у *Synechocystis* sp. PCC 6803 (здесь и далее до конца примера рассматриваем только этот вид, для других видов уравнения аналогичны, но с другими константами). Соответствующая мРНК транскрибируется с двух промоторов, при этом один из этих промоторов регулируется, а другой не регулируется. Частота транскрипции с регулируемого промотора возрастает в три раза при недостатке аммония и нитратов. Существование нерегулируемого промотора обеспечивает высокий базовый уровень концентрации белка NtcA, уменьшающийся лишь примерно на 10% при увеличении концентрации аммония. Таким образом, $\frac{b_0}{b_1}=\frac{1}{1-s_1}=\frac{10}{9}$, откуда $s_1=0.1$.

Ген *glnA* активируется белком NtcA на уровне транскрипции и репрессируется белками GifA и GifB на уровне белковых взаимодействий (каждая их молекула связывается с молекулой GlnA и ингибирует ферментативную активность). Следовательно, уравнение для GlnA имеет вид $y_1=b_2+s_2(1-a)x-k_1z_1-k_2z_2$, где b_2 – базовый уровень концентрации GlnA, s_2 – интенсивность посадки белка NtcA на сайт перед *glnA*, k_1 и k_2 – интенсив-

ности связывания фермента GlnA, соответственно, с факторами инактивации GifA и GifB. На уровне концентраций матричной РНК два последних слагаемых отсутствуют (поскольку GifA и GifB инактивируют только белки) и уравнение приобретает вид $y_1'=b_2'+s_2(1-a)x$, где штрихи указывают, что соответствующая величина относится к мРНК. В соответствии с данными из [5, рис. 6] b_2' примерно равно 1/3 (у *Synechocystis* sp. PCC 6803).

Если x – концентрация белка, а x' – концентрация соответствующих мРНК, то при отсутствии инактивирующих воздействий (учитываем только распад) имеем уравнение $\frac{dx}{dt}=c_1x'-c_2x$, решение которого, $x=\frac{c_1}{c_2}x'+const\cdot\exp(-c_2t)$, при больших t в пределе имеет вид $x=\frac{c_1}{c_2}x'$. Тогда производная величины b_2 равна $c_1b_2'-c_2b_2$, где c_1 и c_2 – положительные числа. Отсюда следует, что величины b_2 и b_2' пропорциональны друг другу в стационарном состоянии, но так как это относительные величины, то можно считать, что они равны, то есть $b_2=\frac{1}{3}$. По тем же данным, при некоторой ненулевой концентрации NtcA (обозначим ее c) и нулевой концентрации аммония концентрация мРНК *glnA* максимальна, т.е. равна 1. Подставляя эти значения в уравнение, получаем $1=\frac{1}{3}+cs_2$, откуда $cs_2=\frac{2}{3}$. Белки GifA и GifB гомологичны и оказывают близкое действие на GlnA и GlnN, поэтому будем считать, что их концентрации совпадают, т.е. $z_1=z_2$. Обозначим $z=z_1=z_2$, $k=k_1+k_2$. Из данных [5, рис. 2] следует, что при максимальной концентрации аммония концентрации белков GifA и GifB близки к максимальным, а концентрация GlnA примерно равна 0.25. Подставляя эти значения в уравнение для GlnA получим, что $\frac{1}{4}=\frac{1}{3}-k$, откуда $k=\frac{1}{12}$. Таким образом, итоговое уравнение для GlnA приобретает вид

$$y_1=\frac{1}{3}+\frac{2}{3}\frac{x}{c}(1-a)-\frac{1}{12}z\,.$$

В этом и в дальнейших уравнениях считаем, что если значение концентрации (в данном случае y_1) получается отрицательным, то его полагаем равным нулю.

При выводе уравнения для GlnN следует учесть результат работы [7], где показано, что у цианобактерии *Synechocystis* PCC 6803 при недостатке аммония транскрипционный фактор NtcA активирует транскрипцию гена *rpoD2-V* (называемого также *sll1689*), кодирующего одну из нескольких σ-субъединиц. В свою очередь, эта σ-субъединица играет существенную роль в транскрипции гена *glnN*. Таким образом, при рассмотрении *glnN* нужно одновременно учитывать и *rpoD2-V*. Транскрипция *rpoD2-V* почти полностью подавлена при достаточном количестве аммония, то есть отно-

сительная концентрация кодируемых этим геном σ-субъединиц равна 1–a. Транскрипция гена *glnN* также близка к нулю при избытке аммония, но значительно возрастает при его недостатке. При этом у мутанта, не имеющего σ-субъединицы RpoD2-V, транскрипция также возрастает, но лишь до уровня 80% от дикого типа. Очевидно, транскрипция *glnN* осуществляется двумя разными σ-субъединицами, вклады от которых составляют 80% (фон) и 20% (RpoD2-V).

Таким образом, транскрипция *glnN* регулируется двумя факторами NtcA и RpoD2-V, то есть соответствующее уравнение становится нелинейным. Для относительной концентрации мРНК *glnN* имеем уравнение $y_2'=(0.8+0.2(1-a))s_3x(1-a)=(1-0.2a)(1-a)s_3x$, где s_3 – интенсивность посадки белка NtcA на сайт перед геном *glnN*. Согласно данным из [5, рис. 6]) при концентрации NtcA, равной c, и нулевой концентрации аммония концентрация мРНК *glnN* равна 1, но в отличие от *glnA* базовый уровень экспресии *glnN* равен нулю. Подставляя эти значения в уравнение, получим, что $cs_3=1$.

Далее, относительная концентрация активного фермента (при учёте ингибиторов GifA и GifB, действия которых считаются равными) составит $y_2=(1-0.2a)(1-a)s_3x-(1/12)z$. Если допустить, что для GlnN величина k та же самая, что и для GlnA, то аналогично получим уравнение для GlnN:

$$y_2=(1-0.2a)\frac{x}{c}(1-a)-\frac{1}{12}z$$

(фактически коэффициент $1-0.2a$ перед первым слагаемым введён для учета вклада от гена *rpoD2-V*).

Благодарность. Работа поддержана грантом РФФИ № 18-29-13037.

Л И Т Е Р А Т У Р А

1. Giacobbe M., Guet C.C, Gupta A., Henzinger T.A., Paixao T., Petrov T. Model checking the evolution of gene regulatory networks // Acta Informatica. – 2017. – V.54. – P.765-787.

2. Melkman A.A., Tamura T., Akutsu T. Determining a singleton attractor of an AND/OR Boolean network in O(1.587n) time // Information Processing Letters. – 2010. – V.110. – P.565-569.

3. Fadda A., Fierro A.C., Lemmens K., Monsieurs P., Engelen K., Marchal K. Inferring the transcriptional network of Bacillus Subtilis // Molecular BioSystems. – 2009. – V.5(12). – P.1840-1852.

4. Pedicini M., Barrenäs F., Clancy T., Castiglione F., Hovig E., Kanduri K., Santoni D., Benson M. Combining network modeling and gene expression microarray analysis to explore the dynamics of Th1 and Th2 cell regulation // PloS Computational Biology. – 2010. – V.6(12). – e1001032.

5. Garcia-Domingues M., Reyes J.C., Florencio F.J. NtcA represses transcription of gifA and gifB, genes that encode inhibitors of glutamine synthetase

type I from Synechocystis sp. PCC 6803 // Molecular Microbiology. – 2000. – V.35(5). – P.1192-1201.

6. Alfonso M., Perewoska I., Kirilovsky D. Redox control of ntcA gene expression in Synechocystis sp. PCC 6803. Nitrogen availability and electron transport regulate the levels of the NtcA protein // Plant Physiol. – 2001. – V.125(2). – P.969-981.

7. Muro-pastor A.M., Herrero A., Flores E. Nitrogen-regulated group 2 sigma factor from Synechocystis sp. strain PCC 6803 involved in survival under nitrogen stress // Journal of Bacteriology. – 2001. – V.183(3). – P.1090-1095.

НАРУШЕНИЕ МИНЕРАЛЬНОГО БАЛАНСА ОРГАНИЗМА ЧЕЛОВЕКА ПРИ АЛОПЕЦИЯХ

Алтаева А.А., Рустамбекова С.А., Горшков В.В.

Дисбаланс макро- и микроэлементов в организме человека негативно сказывается как на здоровье в целом, так и на состоянии волос. Определение макро/микроэлементного состава в волосах является важным исследованием в диагностической тактике при алопециях, позволяя индивидуально подойти к лечению пациента и обоснованно выбрать средства фармакотерапии. При проведении аналогичного исследования у 45 пациентов, страдающих различными видами алопеций, и группы здоровых добровольцев выявлены тенденции и достоверные отклонения в макро/микроэлементном статусе организма. в исследовании участвовали 45 пациентов: 35 женщин и 10 мужчин в возрасте от 18 до 65 лет, без соматической патологии по результатам первичного приема. Пациенты были разделены на нозологические группы по 9 человек: 1-я группа — пациенты, страдающие хронической диффузной телогеновой алопецией (L65.00); 2-я группа — пациенты, страдающие диффузным телогеновым выпадением волос с наличием дополнительных признаков андрогенизации при трихоскопии (L65.00+L64.80); 3-я группа — пациенты с классическими проявлениями андрогенетической алопеции (L64.80); 4-я группа — пациенты, страдающие гнездной алопецией (L64.80); 5-я группа — группа контроля (лица, не предъявляющие жалобы на выпадение волос).

Ключевые слова: выпадение волос, хроническая диффузная телогеновая алопеция, андрогенетическая алопеция, гнездная алопеция, макро- и микроэлементы при алопециях.

ВВЕДЕНИЕ. Здоровье наших волос является зеркалом здоровья нашего организма в целом. Рост патологии в популяции отражает увеличение количества пациентов, обращающихся на прием к трихологам, косметологам и дерматовенерологам с проблемами выпадения волос, ухудше-