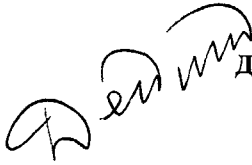


На правах рукописи



ДЕЙНЕКО Игорь Владимирович

**АНАЛИЗ РЕГУЛЯТОРНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
И ДИНАМИКИ МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКОЙ СИСТЕМЫ,
КОНТРОЛИРУЮЩЕЙ G1/S-ПЕРЕХОД
КЛЕТОЧНОГО ЦИКЛА ЭУКАРИОТ**

Генетика – 03.00.15

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Новосибирск, 2005

Работа выполнена в лаборатории молекулярно-генетических систем Института цитологии и генетики Сибирского отделения РАН, Новосибирск.

Научные руководители: доктор биологических наук, профессор
Вадим Александрович Ратнер
Институт цитологии и генетики СО РАН, Новосибирск

доктор биологических наук, профессор
Любовь Антоновна Васильева
Институт цитологии и генетики СО РАН, Новосибирск

Официальные оппоненты: доктор биологических наук
Леонид Владимирович Омелянчук
Институт цитологии и генетики СО РАН, Новосибирск

доктор технических наук
Ефим Яковлевич Фрисман
Институт комплексного анализа региональных
проблем ДВО РАН, Биробиджан

Ведущая организация: Государственный научный центр вирусологии и
биотехнологии «Вектор», Новосибирская обл.,
Кольцово

Защита диссертации состоится 26 мая 2005 года в 9 часов на
утреннем заседании Диссертационного совета Д-003.011.01 при Институте цитологии
и генетики СО РАН в конференц-зале по адресу: 630090, Новосибирск, пр. академика
Лаврентьева, 10. Факс: (3832) 33-12-78, e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке Института цитологии и генетики
Сибирского отделения РАН.

Автореферат разослан 15 апреля 2005 года.

Ученый секретарь
Диссертационного совета
доктор биологических наук



А.Д. Груздев

2005-4
45252

2050920

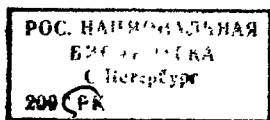
Актуальность проблемы. Последнее десятилетие характеризуется массовым секвенированием последовательностей ДНК, что дает широкие возможности для исследования функций, особенностей структуры и эволюции генетического материала. Эти задачи обуславливают широкое применение компьютерной техники, как для хранения получаемых данных, так и для их анализа, причем сложность методов и моделей увеличивается не только с ростом объемов знаний, но и с усовершенствованием современных компьютеров. Теоретические методы анализа генетических макромолекул (ДНК, РНК, белков) создали возможность предварительного выявления заведомо ложных вариантов опытов, формирования круга потенциальных мишеней и оптимального планирования экспериментов, что обеспечило широкую популярность этих методов.

Одной из важных задач исследования геномов является выявление регуляторных сигналов в ДНК, отвечающих за контроль транскрипции генов. Накопление новых экспериментальных данных о структурах регуляторных районов позволяет развивать новые методы поиска, основанные на композиционной и кластерной моделях регуляторных модулей. Однако, несмотря на имеющиеся подходы, эта задача остается до конца не решенной, в том числе из-за значительных ошибок недо- и перепредсказания.

Другой важной задачей является построение молекулярно-генетических регуляторных систем и изучение их динамических характеристик. Дополнение известных регуляторных систем потенциальными связями, выявленными на основе анализа регуляторных районов генов, и изучение их влияния на динамику системы может дать дополнительные доказательства наличия в реальной системе той или иной регуляторной связи. Такой всесторонний анализ генной регуляции, как на уровне транскрипции отдельных генов, так и на уровне регуляторных систем, представляется нам новым и актуальным направлением исследований.

Цели исследования. Целью данной диссертационной работы является комплексное исследование регуляторных последовательностей генов, кодирующих факторы группы AP-1, и генов машины клеточного цикла эукариот. Поскольку факторы транскрипции E2F и кодирующие их гены являются одними из ключевых звеньев в регуляции клеточного цикла, представляется необходимым, с одной стороны, исследовать регуляторные области ряда генов как возможных мишеней факторов группы E2F, с другой – влияние этих дополнительных генов и их продуктов на динамику поведения молекулярно-генетической системы, контролирующей прохождение G1-фазы и G1/S-перехода клеточного цикла эукариот.

Задачи исследования. Для достижения поставленной цели необходимо было решить следующие задачи:



- 1 Разработать новый метод выявления композиционных элементов (КЭ), основанный на комбинаторных матричных моделях известных композиционных элементов.
- 2 Разработать метод поиска кластеров регуляторных сайтов связывания транскрипционных факторов, основываясь на вероятностной модели распределения одиночных сайтов.
- 3 Исследовать регуляторные районы генов, кодирующих компоненты факторов группы AP-1, на наличие потенциальных одиночных сайтов связывания, композиционных элементов и кластеров сайтов связывания транскрипционных факторов.
- 4 Построить молекулярно-генетическую систему, контролирующую G1-фазу и G1/S-переход клеточного цикла эукариот, исследовать динамику поведения этой системы и влияние на нее предсказанных обратных связей к генам группы AP-1.

Научная новизна работы. Впервые в данной работе комплексно исследована регуляция транскрипции генов, участвующих в контроле G1-фазы клеточного цикла. С одной стороны, был проведен анализ регуляторных районов генов, кодирующих факторы AP-1, с другой – их влияние на динамику функционирования регуляторной системы управления клеточным циклом эукариот. В задаче по изучению регуляции генов эукариот оригинальными методами поиска потенциальных композиционных элементов и кластеров сайтов было показано: а) статистически значимое превышение частоты композиционных элементов Sp1/E2F-1 в промоторах генов факторов AP-1 и генов машины клеточного цикла; б) наличие кластеров сайтов связывания факторов группы E2F в промоторных районах большинства генов факторов AP-1. При анализе динамики регуляторной системы, контролирующей G1-фазу клеточного цикла эукариот, показано, что предполагаемая регуляция генов группы AP-1 транскрипционными факторами группы E2F качественно влияет на динамические режимы функционирования системы. Так, при учёте предсказанной регуляторной связи, молекулярно-генетическая система переходит в режим соответствующий пролиферации клетки, при значительно меньшем необходимом времени внешнего митогенного воздействия, а при некоторых параметрах вообще не нуждается во внешнем стимулировании для такого перехода.

Практическая ценность. Разработан метод распознавания композиционных элементов, использующий матричные комбинаторные модели экспериментально установленных композиционных элементов. Для этого было создано 265 таких моделей. Данный метод отличается не только лучшими характеристиками распознавания. В сравнении с существующими методами, он охватывает намного большее число различных типов КЭ, среди которых: AP1/Ets, AP1/NfκB, AP1/Sp1, AP1/Smad, CEBP/Ets, CEBP/NFκB, CREB/GATA, ER/Sp1, Ebox/Ets, Ets/AML, Ets/NFκB, Ets/SRF,

IRF/NFκB, Myb/AML, NFκB/HMG1Y, Sp1/NFκB, а также, подробно рассмотренные в данной работе, КЭ NFAT/AP-1 и Sp1/E2F-1. Метод распознавания реализован в виде пакета программ MatrixCatch, доступный по сети Internet (<http://compel.bionet.nsc.ru/cgi-bin/MatrixCatch/MatrixCatch.pl>). Разработан метод выявления кластеров регуляторных сайтов связывания транскрипционных факторов с помощью оценки вероятностных характеристик их совместного расположения на нуклеотидной последовательности. Отличительной особенностью предложенного подхода является широкий спектр рассматриваемых типов сайтов связывания, при этом, учитывается возможная корреляция расположения между отдельными сайтами, как одинаковых, так и различных типов. Показано, что расположение в геноме кластеров сайтов, найденных этим методом, коррелирует с позицией стартов транскрипции известных генов.

Апробация работы. Результаты данной работы были представлены на Второй международной конференции «Bioinformatics of Genome Regulation and Structure» Новосибирск, 2000 г., Германской конференции по биоинформатике «German Conference on Bioinformatics», Берлин, 2000 г., Международной конференции «Intelligent Systems for Molecular Biology» Эдмонтон, Канада, 2002 г. и Московской международной конференции «Moscow conference on computational molecular biology», Москва, 2003.

Публикации по теме работы. Всего по теме диссертации опубликовано одиннадцать научных работ, из них четыре – в рецензируемой литературе.

Структура работы. Диссертационная работа состоит из введения, обзора литературы (первая глава), двух глав, содержащих основные результаты, заключения, выводов и списка литературы. Работа изложена на 150 страницах, содержит 37 рисунков, 15 таблиц и 31 формулу, имеющих сквозную нумерацию.

Материалы и методы. Выявление одиночных сайтов связывания регуляторных факторов производилось с помощью разработанной нами программы FASTMATCH (<http://compel.bionet.nsc.ru/cgi-bin/MoMatch/fm.pl>) и весовых матриц, взятых из базы данных TRANSFAC (Wingender *et al.*, 2001). При обучении моделей композиционных элементов была использована база данных COMPEL (Kel-Margoulis *et al.*, 2000), при этом нуклеотидные последовательности композиционных элементов были получены из базы данных EMBL (Kulikova *et al.*, 2004). Для анализа промоторных районов генов машины клеточного цикла и генов группы AP-1 на содержание потенциальных композиционных элементов, была также использована разработанная ранее программа Catch и MatrixCatch (Kel-Margoulis *et al.*, 2002, <http://compel.bionet.nsc.ru/FunSite/CompelPatternSearch.html>, <http://compel.bionet.nsc.ru/cgi-bin/MatrixCatch/MatrixCatch.pl>). Поиск кластеров сайтов связывания транскрипционных факторов производился раз-

работанным нами методом (Deineko, Kel, 2002). Численный расчет динамики регуляторной системы управления клеточным циклом эукариот производился с помощью системы Mathcad 7.

Глава 1. Обзор литературы

В первой главе содержится обзор литературы по особенностям организации регуляторных районов генов эукариот. Описаны современные методы распознавания одиночных сайтов связывания транскрипционных факторов, КЭ и кластеров сайтов, а также методы, используемые для изучения динамики регуляторных систем. Рассмотрены математические модели клеточного цикла эукариот.

Глава 2. Поиск и анализ регуляторных районов

Распознавание композиционных элементов

Несмотря на значительное развитие экспериментальных методов выявления регуляторных единиц, данных о КЭ накоплено значительно меньше, по сравнению с данными об одиночных сайтах связывания транскрипционных факторов, и поэтому представления о вариабельности позиций и структуре КЭ являются весьма отрывочными. В данной работе, исходя из функционального строения КЭ, было предложено привнести дополнительную информацию о вариабельности сайтов, составляющих КЭ, на основе аппроксимации по наборам известных одиночных сайтов, связывающих те же транскрипционные факторы (по которым строятся весовые матрицы).

В нашем методе каждый экспериментально установленный композиционный элемент будет определять в своей комбинаторной модели только ее структуру, то есть тип связывающихся факторов и относительное расположение пары сайтов. Оба сайта связывания будут определяться с помощью весовых матриц для соответствующих транскрипционных факторов.

Таким образом, наша комбинаторная модель композиционного элемента состоит из:

- весовой матрицы, порога и ориентации для первого сайта связывания;
- расстояния между матрицами;
- весовой матрицы, порога и ориентации для второго сайта связывания.

Для обучения данной модели на распознавание определенного КЭ были использованы базы данных COMPEL, TRANSFAC и EMBL. Каждый композиционный элемент в базе данных COMPEL описывается своей нуклеотидной последовательностью, расположением относительно старта транскрипции и позициями обоих сайтов связывания транскрипционных факторов. В базе данных TRANSFAC нас будут интересовать весовые матрицы для определения потенциальных сайтов связывания транскрипционных факторов.

В целом, алгоритм построения модели конкретного КЭ следующий:

- Для фактора, связывающегося с первым сайтом, определяем набор весовых матриц;
- Используя нуклеотидную последовательность КЭ, выбираем ту весовую матрицу из набора, которая имеет наибольший вес на данной последовательности, запоминаем ее положение, ориентацию и значение веса;
- То же для второго сайта, составляющего КЭ;
- Вычисляем расстояние между матрицами.

Всего, описанным выше способом, было построено 265 моделей КЭ, охватывающих более 130 различных типов КЭ, наиболее представленные среди которых: AP1/Ets, AP1/NfκB, AP1/Sp1, AP1/Smad, CEBP/Ets, CEBP/NFκB, CREB/GATA, ER/Sp1, Ebox/Ets, Ets/AML, Ets/NFκB, Ets/SRF, IRF/NFκB, Myb/AML, NFκB/HMGIIY, Sp1/NFκB и, подробно рассмотренные в данной работе, NFAT/AP-1 и Sp1/E2F-1. На рисунке 1 показан один из 15 КЭ NFAT/AP-1 и построенная по нему модель. Для поиска потенциальных КЭ был разработан пакет программ MatrixCatch, доступный по сети Internet (<http://compel.bionet.nsc.ru/cgi-bin/MatrixCatch/MatrixCatch.pl>).

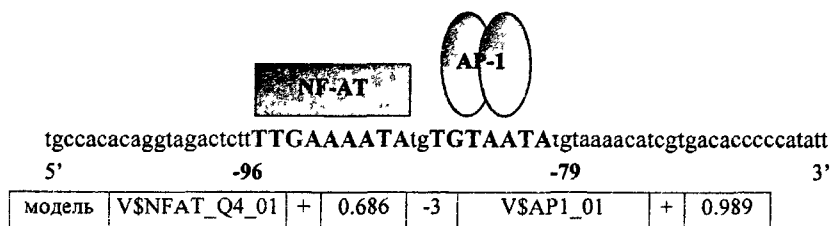


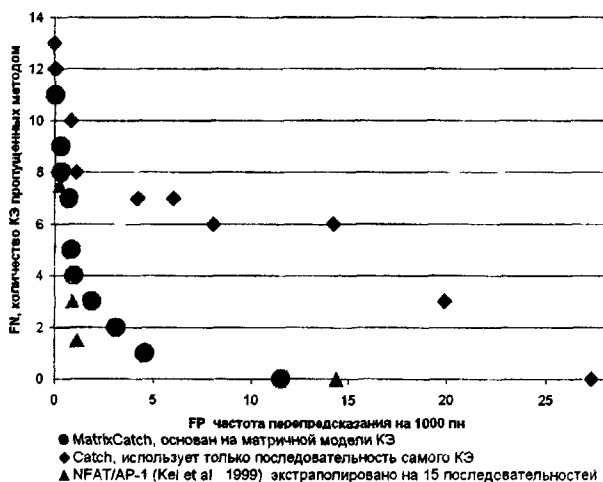
Рис. 1. Пример композиционного элемента NFAT/AP-1 в промоторном районе гена интерлейкина-2 мыши и построенная по нему модель.

Очевидно, что каждая построенная модель КЭ будет распознавать и сам КЭ, по которому она строилась. Для расширения круга распознаваемых потенциальных КЭ задаются параметры расслабления, такие как понижение порогов для обеих матриц и величина варьирования промежутка между сайтами. При этом естественно ожидать, что будут распознаваться последовательности, не являющиеся функциональными КЭ, то есть будет возникать ошибка перепредсказания.

Для оценки уровня ошибок при различных наборах параметров расслабления мы выбрали наиболее широко изученную группу композиционных элементов NFAT/AP-1. В базе данных COMPEL имеется 15 экспериментально выявленных КЭ этого типа. В качестве контрольных выборов мы взяли выборку вторых экзонов из генома человека Q-, а выборку

Q+ составили из последовательностей самих композиционных элементов из базы данных COMPEL. Подсчет ошибок первого и второго рода проводился с помощью метода «Jack knife». Таким образом, мы можем предполагать, что все найденные предложенным методом КЭ в Q- являются нефункциональными и составляют ошибку перепредсказания. При этом количество перепредсказанных КЭ нормировано на длину последовательности выборки вторых экзонов в тпн.

Из рисунка 2 видно, что предложенный метод, основанный на матричном представлении КЭ, значительно превосходит по точности распознавания раннее предложенный нами метод Catch (Kel-Margoulis *et al.*, 2002). Несмотря на то, что в области высокоспецифичного поиска (при малых параметрах расслабления для MatrixCatch и жестких ограничениях на несовпадения для Catch) методы показывают сопоставимую точность, в области низкоспецифичного поиска (высокая чувствительность метода), начиная с ошибки недопредсказания около 50 %, частота обнаруженных ложных сайтов отличается на порядок и более. Некоторые характерные значения ошибок и их соотношения приведены в таблице под рисунком 2.



Ошибка недопредсказания, FN	75 %	60 %	50 %	40 %	20 %	0 %
MatrixCatch	0,0408	0,3025	0,310	0,787	1,880	11,53
Catch	0,0657	0,3204	1,068	8,036	19,89	27,31
Отношение	1,607	1,059	3,437	10,201	10,577	2,367

Рис. 2. Зависимость ошибок распознавания КЭ NFAT/AP-1 программой MatrixCatch в сравнении с ранними результатами: разработанной нами программой Catch (Kel-Margoulis *et al.*, 2002) и методом, предложенным в (Kel *et al.*, 1999).

Дополнительно мы провели сравнение представленного метода с опубликованным ранее методом по определению потенциальных КЭ NFAT/AP-1 (Kel *et al.*, 1999). Точность распознавания показана на рисунке 2 (треугольники), при этом, значения FN и FP увеличены в 15/11 раза из-за различного количества последовательностей, участвующих в обучении методов и распознавании последовательностей. В целом, по графику видно, что предложенный нами метод и подход, рассмотренный в работе (Kel *et al.*, 1999), имеют сопоставимую точность. Однако, при низкоспецифичном поиске (Fp = 1 и 0), предложенный нами метод показывает несколько меньшую ошибку перепредсказания. Это можно объяснить тем, что наш метод, аналогично методу реализаций, использует для распознавания набор моделей (в данном примере 15), тогда как метод, предложенный в (Kel *et al.*, 1999), использует одно универсальное решающее правило. Стоит также отметить, что в нашей выборке имеется КЭ C00354 (AGGAAActctAACTACA), который значительно отличается от остальных. Так, например, «ближайшие» по последовательности к этому КЭ являются КЭ C00161 и C00149, которые отличаются по 4 позициям (по 2 для каждого сайта связывания).

Анализ промоторов генов клеточного цикла и генов факторов AP-1

Для проведения данного анализа в качестве теста мы составили выборку промоторов генов клеточного цикла (43 последовательности) и генов факторов AP-1 (30, взяты районы выше старта транскрипции длиной 1 тпн и 5 тпн). В качестве контрольных выборок мы взяли промоторы генов, характерно экспрессирующихся в Т-клетках (26), в нервных клетках (44) и промоторы мускульно-специфичных генов (24). Также составили выборку промоторов всех аннотированных генов человека (12931). Для полноты исследования мы также сравним частоты отдельных потенциальных сайтов и КЭ со среднегеномным уровнем (используя последовательность хромосомы 21 человека) и с уровнем на случайной последовательности (табл.).

Из рисунка 3 видно, что промоторы генов клеточного цикла обогащены потенциальными сайтами связывания E2F и Sp1. Так, например, частота встречаемости сайтов E2F в промоторах генов клеточного цикла в 14 раз превышает частоту их встречаемости в промоторах генов, специфично экспрессирующихся в Т-клетках.

Относительно всех промоторов генов человека, частота сайтов E2F в промоторах генов клеточного цикла превосходит почти в 5 раз. Это хорошо согласуется с ранее полученными данными о частотах этих сайтов (Kel *et al.*, 2001; Bina *et al.*, 2004).

Таблица. Модели, использованные для поиска композиционных элементов Sp1/E2F-1 с учетом параметров расслабления, дающих ошибку перепредсказания 1 КЭ на 10 тпн.

	Модель 1	Модель 2
Матрица	V\$SP1 Q6	V\$SP1_01
Вес	0.879	0.818
Ориентация	3'-5'	5'-3'
Минимальное расстояние	1пн	32пн
Максимальное расстояние	6пн	45пн
Матрица	V\$E2F1 Q4	V\$E2F_03
Вес	0.901	0.915
Ориентация	5'-3'	3'-5'
КЭ	CE00132	CE00226

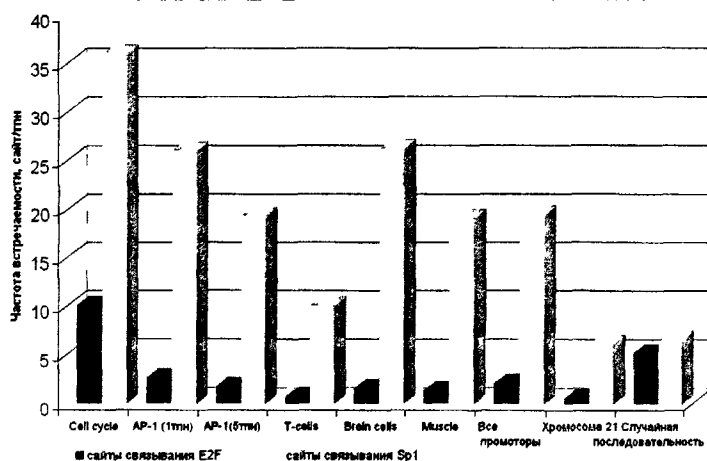


Рис. 3. Распределение частоты встречаемости одиночных сайтов связывания факторов E2F и Sp1 в различных последовательностях.

Распределение частот одиночных сайтов связывания в промоторах генов AP-1, хотя и превышает частоты на некоторых выборках (например, Т-клеток – в 3,8 раза для E2F и в 2,5 раза для Sp1), в целом, слабо отличается от большинства других выборок, в том числе и от выборки всех промоторов генов человека. Более того, частота этих сайтов сильно зависит от длины рассматриваемой промоторной области. Так, более протяженные регуляторные области генов AP-1 (5тпн) содержат заметно меньше сайтов E2F: 2,7 против 1,7 сайтов/тпн.

Частота КЭ Sp1/E2F-1 в промоторах генов клеточного цикла значительно превосходит частоты КЭ в других промоторах (рис. 4). При этом КЭ специфичны к этим промоторам именно как единая регуляторная единица, поскольку, как показывают расчеты, частота пар сайтов E2F и Sp1, ко-локализация которых обусловлена чисто случайными причинами, более чем в четыре раза меньше частоты найденных нами КЭ Sp1/E2F-1.

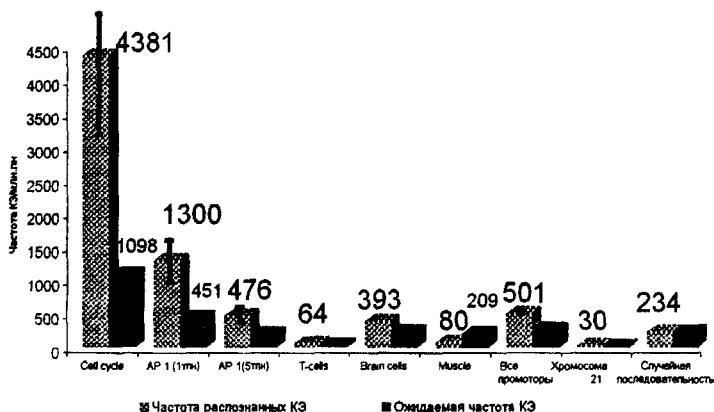


Рис. 4. Распределение частоты встречаемости КЭ Sp1/E2F-1 и частоты пар сайтов Sp1 и E2F-1, ожидаемые по случайным причинам.

Промоторы генов группы AP-1, также насыщены КЭ Sp1/E2F-1. При этом, несмотря на почти одинаковый уровень присутствия отдельных сайтов E2F и Sp1, промоторные районы этой группы генов значительно отличаются от других групп по содержанию композиционных элементов Sp1/E2F-1: в 20 раз превосходят промоторы генов экспрессирующихся в Т-клетках, в 15 раз промоторы мускульно-специфичных генов, более чем в три раза промоторы генов нервных клеток. Данные промоторные районы содержат в 2,5 раза больше потенциальных КЭ Sp1/E2F-1, чем, в среднем, по всем промоторам генов человека. Очевидно, что в промоторах генов, кодирующих факторы AP-1, одиночные сайты связывания факторов E2F и Sp1 располагаются неслучайным образом: они образуют пары, структурно схожие с экспериментально установленными КЭ типа Sp1/E2F-1. В случае независимого расположения рассматриваемых сайтов на последовательностях, ожидаемая частота этих «случайных» пар была бы почти в 3 раза меньше фактически наблюдаемой частоты (1300 против 451). При этом отличие распределения КЭ в выборке AP-1 (1тпн) от других выборок статистически значимо ($p < 0,01$).

Таким образом, можно предположить, что гены клеточного цикла в наибольшей степени подвержены регуляции композиционным элементом Sp1/E2F-1 (в соответствии с наибольшей частотой его встречаемости). Гены факторов AP-1, хотя и в меньшей степени, также предрасположены к подобной регуляции, тогда как мускульно-специфичные гены, наоборот, невосприимчивы к регуляции этой комбинацией факторов. Так как данный КЭ служит связующим звеном в управлении регуляцией прохождения стадий клеточного цикла эукариот и его активность максимальна в середине и конце G1-фазы (Lin *et al.*, 1996), то можно предположить, что гены факторов группы AP-1 регулируются факторами Sp1 и E2F и принимают участие в регуляции клеточного цикла.

Изучение промоторных районов генов *c-fos*

Для подробного анализа промоторных районов (-450...+50 пн), относительно содержания, расположения и консервативности КЭ Sp1/E2F-1, мы выбрали ген *c-fos* четырех организмов человека (*Homo sapiens*), мыши (*Mus musculus*), крысы (*Rattus norvegicus*) и хомячка (*Mesocricetus auratus*). Используя модели из таблицы, мы выявили потенциальные КЭ, которые нанесены на выравнивание этих промоторов (рис. 5).

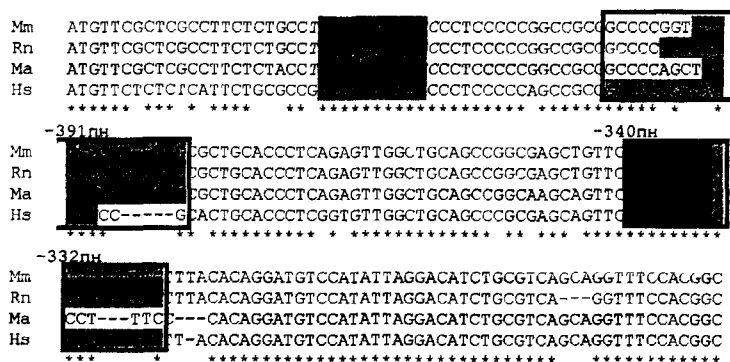


Рис. 5. Выравнивание четырех промоторов гена *c-fos* человека (*Hs*), мыши (*Mm*), крысы (*Rn*) и хомячка (*Ma*). Темным квадратом отмечены сайты связывания фактора E2F, серым SP1. Сайты связывания, образующие КЭ, заключены в рамку.

На этом рисунке видно, что все промоторы, кроме одного, которые образованы сайтом E2F и двумя альтернативными сайтами связывания фактора Sp1, имеют в своем составе два консервативных композиционных элемента в районе -340 пн и -390 пн выше старта транскрипции (обведены рамкой). Однако у хомячка формируется только один композицион-

ный элемент – за счет более далекого сайта Sp1. В этом районе произошли две замены С на Т и одна Т на С (ТСССТСССТС на ТСССТТТССС), что привело к понижению веса матрицы с 0,87 до 0,56 и возможной утрате сродства фактора к данной последовательности. Мы предполагаем, что именно за счет удаленного сайта достигается кооперативное действие факторов E2F и Sp1, тем самым компенсируется утраченная функция нижележащего сайта и сохраняется правильная транскрипционная регуляция гена *c-fos* у *Mesocricetus auratus*.

Распознавание кластеров сайтов связывания транскрипционных факторов в промоторах генов, кодирующих факторы группы AP-1

Исходя из того, что в промоторах генов машины клеточного цикла сайты связывания E2F с большой частотой образуют группы из двух, трех и более сайтов (Brooks *et al.*, 1996; Sears *et al.*, 1997), в этом разделе нашей задачей является проверка как близлежащих, так и удаленных от стартов транскрипции районов генов, входящих в группу AP-1, на наличие кластеров («плотных» групп) сайтов связывания E2F. В нашем подходе, аналогично (Wagner, 1999), в качестве меры значимости кластера была использована вероятность, вычисляемая из предположений о Пуассоновском законе распределения сайтов на последовательности, по формулам:

$$P(n) = \sum Q(k) = \sum q(k_1) \cdot q(k_2) \cdot \dots \cdot q(k_m) \quad \text{и} \quad q(k_i) = e^{-L\lambda_i} \frac{(L\lambda_i)^{k_i}}{k_i!},$$

где λ_i есть среднее количество сайтов связывания типа i на единицу длины, n – общее количество сайтов в кластере, k_i количество сайтов типа i , L – протяженность кластера, сумма производится по всевозможным комбинациям $(k_1, \dots, k_m)$ в сумме больших n . При этом в нашем подходе учитывается возможная корреляция между разными сайтами, проверяемая на протяженных случайных последовательностях.

Нами было показано, что расположение кластеров сайтов, найденных этим методом, коррелирует с позицией стартов транскрипции известных генов на 21 хромосоме человека (рис. 6); проанализированы районы у стартов транскрипции известных генов различной протяженности: -10 ппн, -2 ппн, -1 ппн, -300 ппн, +2 ппн, выборка вторых экзонов).

В результате проведенного анализа на наличие кластеров регуляторных сайтов (протяженность кластера варьировалась от 20 ппн до 500 ппн), нами было установлено, что большинство генов, кодирующих факторы группы AP-1 (гены семейств *jun*, *fos* и *maf*, *atf*, *Nfe2*, *nrfl*), содержат в своих регуляторных районах как удаленные, так и близкие к старту транскрипции кластеры сайтов связывания факторов E2F. Статистическая значимость

каждого из найденных кластеров оценивается величиной вероятности 10^{-6} . Нами было показано, что промоторы генов факторов Jun и Fos обогащены кластерами сайтов в сравнении с другими генами AP-1 (около двух третей всех выявленных кластеров приходится на промоторы гены этих семейств (8 из 26)).

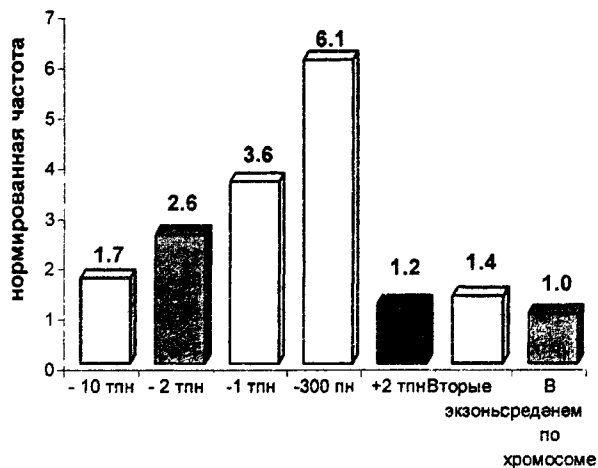


Рис. 6. Частотное распределение найденных кластеров на хромосоме 21 человека.

Таким образом, показанное нами наличие потенциальных композиционных элементов и кластеров регуляторных сайтов в промоторных районах генов, кодирующих факторы группы AP-1, может служить подтверждением выдвинутой нами ранее гипотезы о регуляторном воздействии факторов группы E2F на транскрипцию генов группы AP-1.

Построение и моделирование динамики МГСУ G1-фазой и G1/S- переходом клеточного цикла эукариот

Для выяснения возможного влияния предсказанной регуляторной связи на регуляцию клеточного цикла эукариот, нами была построена модель молекулярно-генетической системы управления (МГСУ) G1-фазой и G1/S-переходом клеточного цикла эукариот (рис. 7).

Анализ динамических свойств модели выявил два качественно различных устойчивых режима функционирования системы в зависимости от продолжительности действия митогенной стимуляции t_m (рис. 8). Если время воздействия $t_m = 60$ больше некоторого критического порога $t_{крит} = 55$, то наблюдается резкое увеличение концентраций основных

компонент: E2F и CysE/Cdc2. При этом дальнейший рост не зависит от наличия этого воздействия, что хорошо согласуется с данными, полученными с помощью микрочиповых технологий в экспериментах с фибробластами человека (Iyer *et al.*, 1999). Напротив, при меньшем времени внешнего стимулирования, компоненты имеют малые концентрации, которые стабилизируются на константном значении. Это состояние чувствительно к внешнему воздействию и может быть переведено в режим 1. Данное состояние системы можно интерпретировать как остановку деления и выход клетки в фазу G0 (рис. 86).

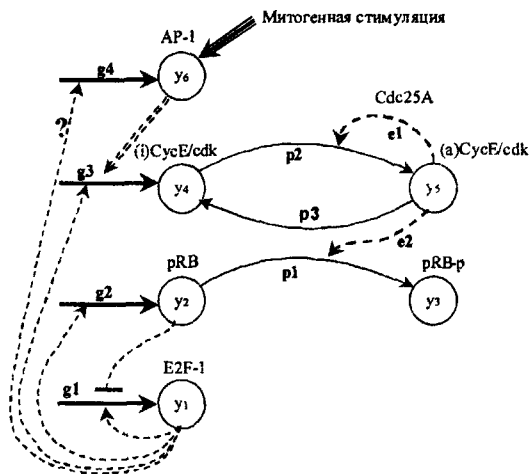


Рис. 7. Модель МГСУ G1/S-переходом клеточного цикла эукариот. Предсказанная теоретически активация генов семейства AP-1 фактором E2F показана стрелкой со знаком вопроса.

Интересным свойством модели является наличие точки бифуркации, при прохождении которой возможны два режима поведения. Мы изучили зависимость поведения системы от силы (F_m) и продолжительности (t_m) функции, моделирующей внешнее воздействие. Было выяснено, что модель переходит в один из режимов в зависимости от суммы накопленного сигнала. То есть, при более сильном воздействии ($F_m = 0,1$) необходимое время воздействия $t_{крит}$ уменьшается с 55 до 20 временных единиц, а при более слабом воздействии увеличивается. Однако временное положение самой точки бифуркации остается неизменным и равно приблизительно 105 единицам. Таким образом, рассматриваемая модель показывает наличие контрольной точки, в которой, в зависимости от накопленного внешнего воздействия, определяется дальнейшая динамика модели.

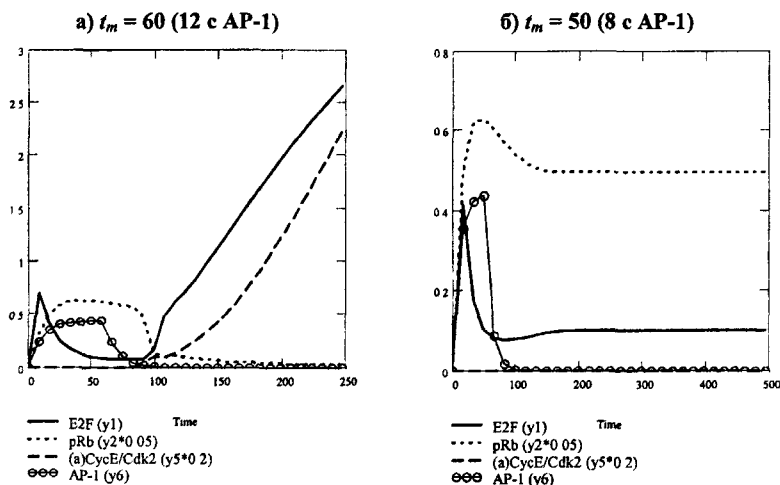


Рис. 8. Два режима поведения модели. Режим, соответствующий входу в S-фазу (а) и выходу клетки в фазу покоя G0 (б). В скобках показана продолжительность митогенной стимуляции при учетывании дополнительной регуляторной связи к генам факторов AP-1.

Изучение влияния предсказанной регуляторной связи на динамические свойства модели показало, что для перехода в режим дальнейшей пролиферации достаточно значительно меньшего времени воздействия внешних сигналов (рис. 8, $t_m = 12$ и 8), а пороговое значение времени воздействия $t_{крит}$ сократилось до ~ 10 единиц времени. Более того, в новой модели происходит постоянный рост концентрации AP-1 даже после прекращения действия внешних сигналов. Этот факт может коренным образом повлиять на следующую итерацию клеточного цикла. Поскольку транскрипция AP-1 не только зависит от внешних сигналов, но и положительно регулируется факторами E2F, то клетка может вступить в быстрое неконтролируемое деление. Можно предположить, что активация генов семейства AP-1 факторами E2F используется при необходимости быстрой клеточной пролиферации (раннее эмбриональное развитие, тканевая регенерация и т.п.), а также в случаях опухолей. Мы полагаем, что в норме этот механизм находится под контролем других регуляторов клеточного цикла, таких как белок p53, который, как известно, стоит на страже нормального прохождения клеточного цикла.

До настоящего времени регуляторная взаимосвязь между факторами E2F и всей группой генов AP-1 высших эукариот экспериментально не установлена. Активирование факторов AP-1 приводит к пролиферации клеток (Sylvester *et al.*, 1998; Iyer *et al.*, 1999), а блокирование этих факторов в активно пролиферирующих раковых клетках останавливает клеточный цикл

(Liu *et al.*, 2004). Можно предположить, что в последнем случае, в отсутствии внешнего сигнала, имеет место активация генов AP-1 факторами E2F, рассмотренная в модели с дополнительной регуляторной связью.

Рассматривая данную работу в рамках концепции молекулярно-генетических систем, можно заключить, что проведенный анализ регуляторных районов генов является топологическим анализом МГС эукариотического организма, на основании которого, мы можем предполагать дополнительные регуляторные связи между компонентами этой системы. Рассмотренная подсистема всего организма – МГС управления G1-фазой клеточного цикла, и проведенный анализ ее поведения есть динамический анализ МГС, на основании которого мы выяснили влияние дополнительных регуляторных связей на динамику поведения системы. В современной литературе данное направление исследований получило название системной биологии.

Выводы

1. Разработан метод распознавания композиционных элементов, использующий матричные комбинаторные модели экспериментально выявленных композиционных элементов. Создано 265 таких комбинаторных моделей, охватывающих более 130 различных типов КЭ. Показана более высокая точность распознавания КЭ в сравнении с ранее предложенными методами. Метод распознавания реализован в виде пакета программ MatrixCatch, который свободно доступен по сети Internet (<http://compel.bionet.nsc.ru/cgi-bin/MatrixCatch/MatrixCatch.pl>)
2. Проведен сравнительный анализ частот одиночных сайтов связывания транскрипционных факторов E2F и Sp1 и частот КЭ E2F/Sp1 в 6 выборках промоторов, включающих промоторы генов клеточного цикла, и генов, кодирующих факторы группы AP-1. Впервые показано статистически значимое превышение частоты композиционных элементов E2F/Sp1 в промоторах генов группы AP-1.
3. Проанализированы промоторы генов *c-fos* четырех организмов: человека, мыши, крысы и хомячка. Показано наличие потенциального композиционного элемента E2F/Sp1, консервативного среди всех рассматриваемых последовательностей.
4. Разработан пакет программ для выявления кластеров регуляторных сайтов. В качестве кластеров рассматривались группы сайтов связывания различных транскрипционных факторов, расположенных на близком расстоянии друг к другу. Построены статистические оценки достоверности кластеров сайтов. Показано наличие в большинстве (~90 %) промоторов генов факторов AP-1 статистически значимых (10^{-6}) кластеров сайтов связывания факторов E2F, установлено их преимущественное расположение в областях близких к стартам транскрипции.

5. Предложена модель молекулярно-генетической системы управления фазой G1 и G1/S-переходом клеточного цикла эукариот. Проанализирована динамика поведения этой системы и выявлены два устойчивых состояния, соответствующие дальнейшей пролиферации и выходу клетки из клеточного цикла, при этом установлено наличие в системе контрольной точки. Показана параметрическая устойчивость системы.
6. Показано качественное влияние на динамические режимы поведения МГС дополнительной регуляторной связи от факторов группы E2F к генам, кодирующим факторы AP-1, которое было предсказано на основе анализа частот одиночных сайтов связывания, композиционных элементов и кластеров.

Основные результаты диссертации опубликованы в работах

- Kel-Margoulis O.V., Kel A.E., Reuter I., **Deineko I.V.**, Wingender E. TRANSCompel: a database on composite regulatory elements in eukaryotic genes // *Nucleic Acids Res.* 2002. V. 30, N 1. P. 332–334.
- Дейнеко И.В.**, Кель А.Э., Кель-Маргулис О.В., Вингендер Э., Ратнер В.А. Моделирование динамики генных сетей, регулирующих клеточный цикл в клетках млекопитающих // *Генетика*. 2003. Т. 39, № 9. С. 1285–1292.
- Deineko I.V.**, Kel-Margoulis O.V., Ratner V.A., Kel A.E. Modeling of cell cycle gene regulatory network. A role of positive feedback loop implying potential E2F target sites in the regulatory regions of AP-1 // *Proc. of the second Intern. Conf. on Bioinformatics of genome regulation and structure*. Novosibirsk: IC&G SB RAS, 2000. V. 1. P. 226–229.
- Kel-Margoulis O.V., Romaschenko A.G., **Deineko I.V.**, Kolchanov N.A., Wingender E., Kel A.E. Database on composite regulatory elements in eukaryotic genes (compel) // *Proc. of the second Intern. Conf. on bioinformatics of genome regulation and structure*. Novosibirsk: IC&G SB RAS, 2000. V. 1. P. 45–48.
- Kel A., **Deineko I.**, Kel-Margoulis O., Wingender E., Ratner V. Modeling of gene regulatory network of cell cycle control. Role of E2F feedback loops. GCB 2000: Proceedings of the German Conference on Bioinformatics. E. Bornberg-Bauer, Berlin: Logos-Verl., 2000. P. 107–114.
- Kel-Margoulis O.V., **Deineko I.V.**, Reuter I., Wingender E., Kel A.E. TRANSCompel – a professional database on composite regulatory elements in eukaryotic genes // *Proc. of the German Conf. on Bioinformatics (GCB 2001)*.
- Deineko I.**, Kel A. Genome-wide search for composite clusters of transcription factor binding sites // *ISMB 2002. Abstract/ Edmonton, Canada 2002*. P. 124.
- Deineko I.**, Kel A. Probabilistic approach for revealing composite clusters of transcription factor binding sites in genomic scale. MCCMB 2003: *Proc. of the Intern. Moscow Conf. on computational molecular biology*. Moscow, Russia, 2003. P. 54–55.
- Deineko I.V.**, Kel-Margoulis O.V., Kel A.E. Mathematical model of the mitogen-dependent G1/S transition in mammalian cell cycle // *Proc. of the fifth Intern. Conf. on Systems Biology (ICSB 2004)*. Heidelberg, Germany, 2004. P. 385.
- Kolpakov F., **Deineko I.**, Kel A. Cyclonet a database on cell cycle regulation // *Proc. of the fifth Intern. Conf. on Systems Biology (ICSB 2004)*. Heidelberg, Germany, 2004. P. 385.
- Swat M., Kel A., Kel-Margoulis O., **Deineko I.**, Herzog H. Modeling the influence of feedback loops on the G1/S transition // *Proc. of the European Conf. on Computational Biology*, 2002 (ECCB 2002).

Подписано к печати 01 04 2005

Формат бумаги 60 x 90 Печ л 1 Уч изд л 0,7

Тираж 100 экз. Заказ 50

Ротапринт Института цитологии и генетики СО РАН
630090, Новосибирск, пр ак Лаврентьева, 10

РНБ Русский фонд

2005-4

45252

22 АПР 2005

2486