

УДК: 575.852.112:575.852.113:577.214:577.217

АНАЛИЗ ВЫРОЖДЕННЫХ ОЛИГОНУКЛЕОТИДНЫХ МОТИВОВ В ПРОМОТОРАХ ГЕНОВ миРНК, ЭКСПРЕССИРУЮЩИХСЯ В РАЗЛИЧНЫХ ТКАНЯХ МЛЕКОПИТАЮЩИХ

О.В. Вишнеvский^{1,2}, К.В. Гунбин¹, А.В. Бочарников², Е.В. Березиков^{1,3}

¹Институт цитологии и генетики СО РАН, г. Новосибирск;

²кафедра информационной биологии факультета естественных наук Новосибирского государственного университета, г. Новосибирск;

³Hubrecht Institute, RNAAS, Utrecht, Netherlands; e-mail: oleg@bionet.nsc.ru)

Разработан метод выявления значимых олигонуклеотидных мотивов в регуляторных районах генов эукариот и проведен компьютерный анализ тканеспецифичных промоторов генов миРНК млекопитающих.

Ключевые слова: гены миРНК, промоторы, олигонуклеотидные мотивы.

МиРНК представляют собой короткие (~22 п.о.) последовательности РНК, способные связываться с 3'-нетранслируемыми районами (3'-НТР) мРНК генов-мишеней, приводя к репрессии трансляции и деградации мРНК. Около 30% генов человека являются мишенями для миРНК [1]. МиРНК играют важную роль в регуляции базовых биологических механизмов, управляя развитием организма и процессами дифференциации тканей. Нарушение экспрессии миРНК приводит к патологиям развития организма и тяжелым заболеваниям.

Большая часть миРНК транскрибируется полимеразой II в составе протяженной пре-мРНК, которая подобно другим транскриптам этой полимеразы кеппируется с 5'-конца и полиаденилируется с 3'-конца [2]. Была показана высокая ткане- и стадийспецифичность экспрессии миРНК с участием ряда транскрипционных факторов полимеразы II [3]. Выявлено присутствие мотивов канонического ТАТА-бокса в 5'-районе миРНК [4]. Масштабное применение высокопроизводительных технологий секвенирования и разработка новых компьютерных методов полногеномного анализа позволило существенно увеличить объем информации о миРНК. В то же время остаются до конца не ясными многие особенности организации регуляторных районов генов миРНК, обеспечивающие ткане- и стадийспецифичность их экспрессии. Задачей исследования было выявление функциональных сигналов в промоторах генов миРНК млекопитающих, экспрессирующихся в различных тканях.

Материалы и методы

Сформированы выборки регуляторных последовательностей генов миРНК 10 видов млекопитающих (*Homo sapiens*, *Pan troglodytes*, *Pongo pygmaeus*, *Gorilla gorilla*, *Macaca mulatta*, *Bos taurus*, *Equus ca-*

ballus, *Canis familiaris*, *Mus musculus*, *Rattus norvegicus*) в районе [-1000; +1] относительно старта транскрипции, экспрессирующихся в мозге, легких и выделительной системе. Для этого на первом этапе из базы данных Ensembl (<http://www.ensembl.org/>) была получена информация о локализации в геноме 1572 миРНК человека, предсказанных с использованием различных экспериментально-компьютерных подходов. Затем в 5'-районе ДНК на расстоянии не более 3000 нуклеотидов относительно начала миРНК осуществлен поиск наиболее близких стартов транскрипции, информация о которых извлекалась из базы данных FANTOM4 [5]. После этого из базы данных Ensembl Compara (<http://agd.vital-it.ch/info/software/compara/index.html>) с использованием Compara PerlAPI для 9 видов млекопитающих были получены гомологичные геномные участки длиной 1000 п.о.

В 40 построенных выборках последовательностей проводился поиск регуляторных сигналов, значимых для структурно-функциональной организации промоторов, с использованием разработанного нами метода выявления вырожденных олигонуклеотидных мотивов — коротких слов фиксированной длины, записанных в 15-буквенном IUPAC коде. Этот подход относится к так называемым методам без учителя, позволившим выявлять консервативные сигналы без предварительного обучения на экспериментально полученных и выравненных последовательностях. Для оценки сходства регуляторных районов генов на основе олигонуклеотидных мотивов, нами был разработан метод, основанный на сравнении представленности и характера распределения мотивов в рассматриваемых последовательностях. Методы реализованы в виде интернет-доступного пакета программ ARGON (<http://www.mgs.bionet.nsc.ru/argo/>) [6].

Результаты и обсуждение

Сравнение мотивов, обнаруженных с помощью системы ARGO, с известными сайтами связывания транскрипционных факторов (ССТФ) из баз данных TRANSFAC [7], TRRD [8] и MatBase (<http://www.genomatix.de/products/MatBase/>) показало, что большинство из них имеет сходство с ССТФ, участвующими в развитии организма и регуляции клеточного цикла. При этом в разных видах организмов и в различных тканях были обнаружены как мотивы, соответствующие повсеместно представленным базовым ССТФ, таким как ТАТА-бокс и ССААТ, так и мотивы, соответствующие ССТФ тканеспецифичных факторов, например MYT1 и NeuroD в выборке мозгспецифичных промоторов. Сравнительный анализ позволил выявить мотивы, соответствующие ССТФ, характерные только для выборки тканеспецифичных промоторов одного таксона, а также видоспецифичные мотивы. Так, ССТФ группы “Lim homeodomain factors” имеют достоверно более высокое количество сходных с ними мотивов из выборки промоторов выделительной системы гориллы, чем в соответствующих выборках всех других приматов. В то же время оказалось, что значительная часть полученных мотивов не может быть достоверно отнесена ни к одному из известных ССТФ. Видимо, такие мотивы могут соответствовать как сайтам связывания еще не исследованных видоспецифичных транскрипционных факторов, так и некоторым структурным особенностям промоторных районов эукариот, таким как поли-А/поли-Т тракты, которые способны приводить к локальному изгибу ДНК и повышению ее легкоплавкости.

Далее, мы провели оценку среднего олигонуклеотидного сходства промоторов человека с промоторами всех остальных видов с использованием разработанной нами системы ARGO. Для сравнения с помощью парного выравнивания для тех же выборок последовательностей была проведена оценка усредненного сходства. Затем оценки олигонуклеотидного сходства и усредненного сходства были проведены для выборок случайных последовательностей, сгенерированных с сохранением мононуклеотидного состава, характерного для реальных промоторов млекопитающих. После чего полученные ранее оценки нормировались. При этом в качестве верхней границы (1) рассматривалось соответствующее сходство промоторов человека с теми же промоторами человека, а нижней границей (0) явля-

лось сходство промоторов человека со случайными последовательностями. Анализ полученных таким образом нормированных оценок среднего олигонуклеотидного сходства (H_{Oli}) и усредненного сходства (H_{Align}) показал, что для всех видов и среди всех тканей наибольшими величинами и H_{Oli} и H_{Align} характеризуются промоторы мозга. Причем для приматов эти величины максимальны (0,77–0,92). Интересно отметить, что H_{Oli} в выборке промоторов собаки (0,77) значительно превышает значения H_{Oli} у других не-приматов (0,56–0,59) и достигает величины, характерной для макаки. Сравнение величин H_{Oli} и H_{Align} показывает, что для филогенетически близких к человеку видов (приматы) H_{Oli} и H_{Align} во всех тканях отличаются не более чем на 2%. Исключение составляет выборка промоторов выделительной системы гориллы, для которой при H_{Align} сходном с другими гоминидами (0,83) величина H_{Oli} была больше на 8% (0,91) и идентична олигонуклеотидному сходству промоторов мозга. Для промоторов эволюционно удаленных от человека видов млекопитающих (не приматы) характерны относительно низкие значения H_{Align} (0,09–0,27) при сохранении достаточно высоких H_{Oli} (0,29–0,77), при этом величина отклонения $D = H_{Oli} - H_{Align}$ варьировала от 20 до 45%. Можно предположить, что это объясняется сохранением в регуляторных районах тканеспецифичных генов эволюционно удаленных видов сходства только на уровне использования сходных регуляторных сигналов.

Заключение

С использованием разработанной нами компьютерной системы ARGO в тканеспецифичных промоторах генов мРНК млекопитающих выявлены и классифицированы потенциальные регуляторные сигналы, которые могут быть использованы для дальнейшего экспериментального анализа. Показано, что в тканеспецифичных промоторах генов эволюционно удаленных видов сходство на уровне регуляторных сигналов существенно выше средней гомологии последовательностей.

* * *

Работа поддержана Российским фондом фундаментальных исследований (грант № 09-04-01641-а); Интеграционными проектами СО РАН № 113, 119; Программами РАН № 22 (проект № 8) и 23 (проект № 29).

СПИСОК ЛИТЕРАТУРЫ

1. John B., Enright A.J., Aravin A., Tuschl T., Sander C., Marks D.S. Human microRNA targets // PLoS Biol. 2004. Vol. 2. N 11. P. 363.
2. Cai X., Hagedorn C.H., Cullen B.R. Human microRNAs are processed from capped, polyadenylated transcripts

that can also function as mRNAs // RNA. 2004. Vol. 10. N 12. P. 1957–1966.

3. Chen J.F., Mandel E.M., Thomson J.M., Wu Q., Callic T.E., Hammond S.M., Conlon F.L., Wang D.Z. The role of microRNA-1 and microRNA-133 in skeletal muscle pro-

liferation and differentiation // *Nature Genet.* 2006. Vol. 38. N 2. P. 228–233.

4. Houbaviy H.B., Dennis L., Jaenisch R., Sharp P.A. Characterization of a highly variable eutherian microRNA gene // *RNA.* 2005. Vol. 11. N 8. P. 1245–1257.

5. Kawaji H., Severin J., Lizio M., Waterhouse A., Kayama S., Irvine K.M., Hume D.A., Forrest A.R., Suzuki H., Carninci P., Hayashizaki Y., Daub C.O. The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation // *Genome Biol.* 2009. Vol. 10. N 4. P. 40.

6. Vishnevsky O.V., Kolchanov N.A. ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters // *Nucleic Acids Res.* 2005. Vol. 33. Web Server issue. P. 417–422.

7. Matys V., Kel-Margoulis O., Fricke E., Liebich I., Land S., Barre-Dirrie A., Reuter I., Chekmenov D., Krull M., Hornischer K., Voss N., Stegmaier P., Lewicki-Potapov B., Saxel H., Kel A., Wingender E. TRANSFAC(r) and its module TRANSCOMP(r): transcriptional gene regulation in eukaryotes // *Nucleic Acids Res.* 2006. Vol. 34. Database issue. P. 108–110.

8. Kolchanov N.A., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Stepanenko I.L., Merkulova T.I., Pozdnyakov M.A., Podkolodny N.L., Naumochkin A.N., Romashchenko A.G. Transcription Regulatory Regions Database (TRRD): its status in 2002 // *Nucleic Acids Res.* 2002. Vol. 30. N 1. P. 312–317.

Поступила в редакцию
19.04.10

ANALYSIS OF THE DEGENERATE MOTIVES IN PROMOTERS OF miRNA GENES EXPRESSED IN DIFFERENT TISSUES OF MAMMALIANS

O.V. Vishnevsky, K.V. Gunbin, A.V. Bocharnikov, E.V. Berezikov

The computer approach is developed to reveal the significant oligonucleotide motives in regulatory regions of eukaryotic genes. The regulatory signals, specific for promoter regions of miRNA genes, which are expressed in different tissues of mammals, are obtained and classified.

Key words: *miRNA genes, promoters, oligonucleotide motives.*

Сведения об авторах

Вишневский Олег Владимирович — канд. биол. наук, науч. сотр. Института цитологии и генетики СО РАН; НГУ, г. Новосибирск. Тел. (383)363-49-22; e-mail: oleg@bionet.nsc.ru

Гунбин Константин Владимирович — канд. биол. наук, науч. сотр. Института цитологии и генетики СО РАН. Тел. (383)363-49-23; e-mail: genkvg@bionet.nsc.ru

Бочарников Андрей Васильевич — студент Новосибирского государственного университета. E-mail: andrey.bocharnikov@gmail.com

Березиков Евгений Викторович — канд. биол. наук, вед. науч. сотр. Института цитологии и генетики СО РАН, Hubrecht Institute, RNAAS, Utrecht, Netherlands. E-mail: e.berezikov@niob.knaw.nl