

ФИЗИЧЕСКИЕ ХАРАКТЕРИСТИКИ РЕГУЛЯТОРНЫХ УЧАСТКОВ ГЕНОМА, ЭПИГЕНЕТИКА И КАНЦЕРОГЕНЕЗ**Нечипуренко Ю.Д.¹, Ильичева И.А.¹, Абдуллаев Э.Т.² Урошлев Л.А. ³, Гроховский С.Л.¹**¹ Институт молекулярной биологии им. В.А. Энгельгардта РАН
ул. Вавилова, 32, г. Москва, 119991, РФ; e-mail: nech99@mail.ru; imb_irina@rambler.ru² Институт молекулярной генетики им. Макса Планка
г. Берлин, 14195, Германия;³ Институт общей генетики им. Н.И. Вавилова РАН
ул. Губкина, 3, г. Москва, 119991, РФ

Поступила в редакцию: 07.08.2018.

Аннотация. Ранее мы показали, что при облучении ДНК ультразвуком происходят множественные разрывы её цепей, причем их интенсивность зависит как от типа динуклеотидов, между которыми происходит разрыв, так и от окружающих место разрыва последовательностей. На основании данных гелевого электрофореза нами были получены оценки средних относительных частот расщепления каждого из 16-ти динуклеотидов ДНК, а также 256 тетрануклеотидов. Оказалось, что относительные частоты расщепления комплементарных динуклеотидов различаются. Чаще других расщепляются межнуклеотидные связи, следующие за цитозином, причем наиболее часто в динуклеотиде CpG. Частота расщеплений убывает в ряду $d(\text{CpG}) > d(\text{CpA}) > d(\text{CpT}) \gg d(\text{CpC})$. Впоследствии мы показали, что эти соотношения очень близки соотношениям частот расщепления динуклеотидов геномной ДНК в процессе её фрагментирования при подготовке образцов к NGS-секвенированию. Мы связываем различие частот расщепления в разных позициях последовательности ДНК с различием интенсивности локальных конформационных движений, главным образом с особенностями псевдотворения фуранозного цикла. Мы пришли к выводу о существовании второго уровня кодирования информации молекулой ДНК – о существовании «конформационного кода». Анализ большого числа физических и структурных характеристик ДНК, усредненных по репрезентативным выборкам разных видов животных, растений и прокариот показал, что профили индексов, характеризующие расщепление ДНК ультразвуком имеют характерные свойства в области, которая соответствует у многоклеточных организмов последовательности между ТАТА боксом и началом сайта инициации транскрипции (TSS). Поскольку эти индексы несут информацию о конформационных движениях в отдельных нитях молекулы ДНК, можно утверждать, что интенсивность конформационных движений в комплементарных цепях на этих участках изменяется в противофазе вдоль оси двуспиральной молекулы. Наиболее выражены такие закономерности на профилях индексов расщепления в регуляторных областях ДНК из клеток млекопитающих. На интенсивность расщепления динуклеотида CpG влияет также метилирование цитозина по 5-му положению пиримидинового кольца. Мы оценили влияние метилирования цитозина на расщепление динуклеотидов в тканях человека, а также на расщепление динуклеотидов в CpG-островах в контрольных и раковых тканях (лимфоме и гепатокарциноме). Оказалось, что на участках скопления метилированных CpG-динуклеотидов геномная ДНК расщепляется значительно чаще, чем на аналогичных участках, но с низким уровнем метилирования. Более того, было установлено различие в расщеплении CpG-островов в раковых и контрольных образцах, и это также связано с различиями в уровне их метилирования. Поэтому появляется возможность выявления предракового состояния тканей на основании оценки частот расщепления CpG-островов различных клеточных линий по данным фрагментирования ДНК при NGS-секвенировании. Обнаруженные нами физико-химические свойства ДНК могут быть использованы как в теоретических исследованиях, так и при диагностике заболеваний.

Ключевые слова: локальная структура и конформация ДНК, специфичное к последовательности расщепление ДНК ультразвуком, промоторы РНК полимеразы II, метилирование оснований ДНК, эпигенетика, канцерогенез.

В 2006 году С.Л. Гроховский обнаружил, что в результате ультразвукового облучения молекул ДНК происходят множественные разрывы её цепей, причем их интенсивность зависит от типа динуклеотидов, между которыми происходит разрыв, а также от окружающих их последовательностей, от температуры, pH и ионной силы раствора, но не зависит от используемой частоты ультразвука [1]. Наиболее часто двойная спираль разрывается по динуклеотиду d(CpG). Было показано также, что фосфатная группа остается на 5'-конце образующихся фрагментов, а характер процесса имеет черты механохимической реакции.

Статистический анализ интенсивностей полос, соответствующих фрагментам ДНК, движущихся в полиакриламидном геле и образованных при разрывах в известных местах последовательностей, позволил нам получить относительные частоты разрывов центральных межнуклеотидных связей в каждом из 16 вариантов

динуклеотидов, и в 256 тетранулеотидах [2, 3]. Особенности методики их получения детально описаны также в работах [4-6], а построение модели механохимического разрыва цепей ДНК проведено в работе [7].

При современных методах секвенирования вся геномная ДНК разбивается на короткие куски (так называемые "риды"), которые все прочитываются, и потом эти риды, как пазлы, собираются в единый текст, используя перекрывающиеся концевые участки. В автоматических секвенаторах довольно большой процент ошибок считывания, и для надежного перекрывания концевых последовательностей при сборке этих "пазлов", требуется, чтобы суммарная длина прочитанных ридов в несколько десятков раз превосходила длину анализируемого генома (таким образом, чтобы вся последовательность «перекрывалась» ридами много раз). После секвенирования фрагментов ДНК производится компьютерная сборка всего генома с многократным наложением перекрывающихся участков. В основе такого подхода к секвенированию лежит предположение, что разрывы ДНК на фрагменты происходят случайно и не зависят от последовательности нуклеотидов. Однако при фрагментации ДНК геномов при помощи методов, основанных на разрыве ДНК под действием гидродинамических сил (ультразвук, небулизация и «Коварис»), обнаружено, что фрагменты из одних областей присутствуют в избытке, а из других - в недостатке. Это свидетельствует о неслучайном характере разрывов.

Мы провели анализ 5'-концевых участков фрагментов, выровненных на соответствующие референтные геномы, и обнаружили, что частоты разрывов всех вариантов ди- и тетрануклеотидов на концах фрагментов находятся в хорошем соответствии с данными ультразвукового расщепления рестриктных фрагментов ДНК [8].

Мы предположили, что относительные частоты ультразвукового расщепления можно использовать для физического картирования генома, то есть для определения участков последовательности, имеющих аномалии по ряду физических характеристик [9], которые связаны, как правило, с регуляторными участками генов. В работе [10] на основании анализа большого числа физических и структурных характеристик ДНК, и построения их профилей по усредненным репрезентативным выборкам разных видов животных, растений и одноклеточных грибов для промоторов РНК полимеразы II (Pol II), мы показали, что профили относительных частот ультразвукового расщепления ди- и тетрануклеотидов ДНК (далее - индексов расщепления) имеют характерные свойства на участках последовательностей, которые соответствуют у многоклеточных организмов области между ТАТА боксом и началом сайта инициации транскрипции (TSS). Индексы, характеризующие расщепление ДНК ультразвуком, несут информацию о свойствах отдельных нитей молекулы ДНК (в отличие от других индексов). Полученные данные показывают, что наряду с локальными характеристиками структуры двойной спирали ДНК, которые определяются взаимным расположением комплементарных пар оснований в динуклеотидах и влияют, в частности, на ширину бороздок, наша характеристика – индекс расщепления ДНК ультразвуком, связанная с динамикой конформационных движений фуранозных циклов в комплементарных цепях молекулы ДНК, помогает выявить динамические особенности регуляторных участков ДНК и обнаружить сходные свойства промоторов у разных видов. Действительно, эти индексы несут информацию о конформационных движениях в отдельных нитях молекулы ДНК, а полученные нами профили показывают, что интенсивность конформационных движений в комплементарных цепях на участках от ТАТА-бокса до начала транскрипции изменяются в противофазе вдоль оси двойной спирали молекулы ДНК. Наиболее выражены такие закономерности на профилях индексов расщепления в регуляторных областях ДНК из клеток млекопитающих [10]. Итак, относительные частоты ультразвукового расщепления могут быть использованы для картирования геномов. В будущем обнаруженные нами физико-химические свойства молекулы ДНК могут быть использованы в генетической инженерии для искусственной модуляции силы промоторов [11].

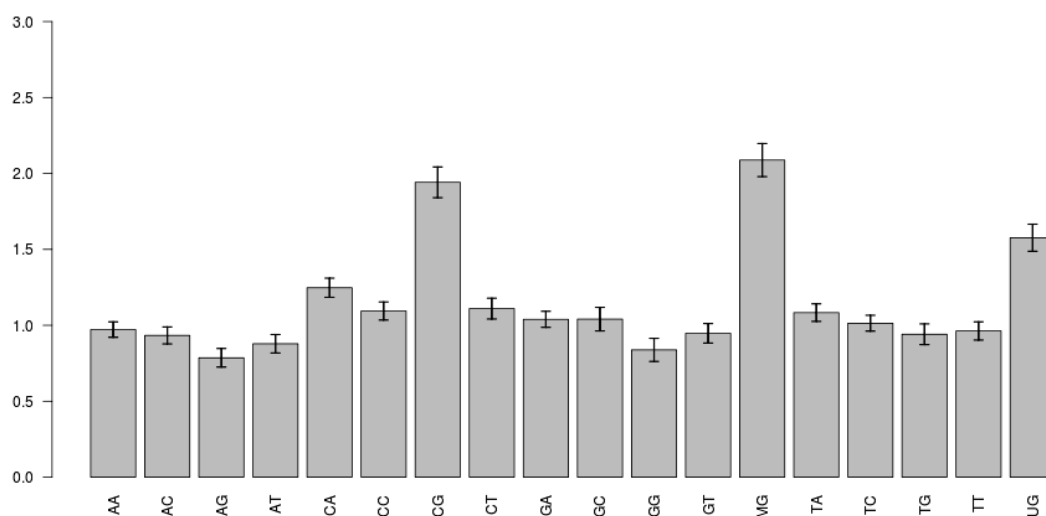


Рисунок 1. Относительные частоты расщепления динуклеотидов. Показаны данные для 18 вариантов динуклеотидов. CG – данные по расщеплению смеси метилированных и неметилированных динуклеотидов; UG и MG – данные для динуклеотидов, в первом положении которых стоит цитозин неметилированный и метилированный, соответственно

В ДНК человека значительный процент цитозинов метилирован, что определяет эпигенетические свойства ткани, в частности, паттерны экспрессии генов ткани, из которой взята ДНК для анализа. Однако ранее не проводилось исследования, насколько метилирование влияет на вероятность расщепления динуклеотида d(CpG). Мы привели такое исследование при помощи анализа ридов, статус метилирования которых известен - и показали, что метилирование приводит к значимому увеличению вероятности расщепления (рис. 1).

Таким образом, появляется возможность на основании ридов, биоинформатически определять статус последовательности ДНК и её эпигенетические свойства. Мы провели анализ расщепления CpG островов для ряда тканей, где имеются данные о расщеплении ДНК в случае раковой опухоли и в контроле - и обнаружили, что паттерны расщепления таких образцов значимо отличаются. Поэтому на основании анализа ридов можно определить не только общий статус метилирования ткани, но и находить гены, которые могут быть ответственны за злокачественное перерождение ткани.

Работа поддержана Программой фундаментальных исследований государственных академий наук на 2013-2020 годы (темы №01201363818 и № 01201363820) и также грантом Президиума РАН по Молекулярной и клеточной биологии.

Список литературы / References:

1. Гроховский С.Л. *Молекулярная биология*, 2006, т. 40, с. 317-325. [Grokhovsky S.L. *Molecular biology*, 2006, vol. 40, pp. 317-325. (In Russ.)]
2. Гроховский С.Л., Ильичева И.А., Нечипуренко Д.Ю., Панченко Л.А., Полозов Р.В., Нечипуренко Ю.Д. *Биофизика*, 2008, т. 53, с. 417-425. [Grokhovsky S.L. [et al.] *Biophysics*, 2008, vol. 53, pp. 417-425. (In Russ.)]
3. Grokhovsky S.L., Il'icheva I.A., Nechipurenko D.Yu., Golovkin M.V., Panchenko L.A., Polozov R.V., Nechipurenko Y.D. *Biophysical Journal*, 2011, vol. 100, no. 1, pp. 117-125.
4. Нечипуренко Ю.Д., Головкин М.В., Нечипуренко Д.Ю., Ильичева И.А., Панченко Л.А., Полозов Р.В., Гроховский С.Л. *Журнал структурной химии*, 2009, т. 50, с. 1040-1047. [Nechipurenko Yu.D. [et al.] *J. Strukt. Khimii*, 2009, vol. 50, pp. 1040-1047. (In Russ.)]
5. Grokhovsky S., Il'icheva I., Nechipurenko D., Golovkin M., Taranov G., Panchenko L., Polozov R., Nechipurenko Yu. in *Gel Electrophoresis - Principles and Basics*, Dr. Sameh Magdeldin (Ed.), ISBN: 978-953-51-0458-2, InTech. 2012.
6. Grokhovsky S.L., Il'icheva I.A., Nechipurenko D.Yu., Golovkin M.V., Panchenko L.A., Polozov R.V., Nechipurenko Yu.D. *Advances in Engineering Research*, 2014, vol. 8. pp. 213-236. ISSN: 2163-3932. ISBN: 978-1-63321-282-4.
7. Нечипуренко Д.Ю., Ильичева И.А., Ходыков М.В., Попцова М.С., Нечипуренко Ю.Д., Гроховский С.Л. *Биофизика*, 2014, т. 59, с. 1061-1070. [Nechipurenko D.Yu. [et al.] *Biophysics*, 2014, vol. 59, pp. 1061-1070. (In Russ.)]
8. Poptsova M.S., Il'icheva I.A., Nechipurenko D.Y., Panchenko L.A., Khodikov M.V., Oparina N.Y., Polozov R.V., Nechipurenko Y.D., Grokhovsky S.L. *Sci Rep.*, 2014, vol. 4, pp. 4532.
9. Нечипуренко Ю.Д., Нечипуренко Д.Ю., Ильичева И.А., Головкин М.В., Панченко Л.А., Полозов Р.В., Гроховский С.Л. *Компьютерные исследования и моделирование*, 2010, т. 2, с. 419-428. [Nechipurenko Yu.D. [et al.] *J. Computer Res. and Modelling*, 2010, vol. 2, pp. 419-428. (In Russ.)]
10. Il'icheva I.A., Khodikov M.V., Poptsova M.S., Nechipurenko D.Yu., Nechipurenko Yu.D., Grokhovsky S.L. *BMC genomics*, 2016, vol. 17, p. 973.
11. Нечипуренко Ю.Д., Ильичева И.А., Попцова М.С., Гроховский С.Л. Физико-химические свойства ДНК в регуляторных участках генома. *Актуальные вопросы биологической физики и химии*, 2017, т. 2, с. 339-341 [Nechipurenko Y.D., Il'icheva I.A., Poptsova M.S., Grokhovsky S.L. Physico-chemical properties of DNA in regulatory sites of the genome. *Russian Journal of Biological Physics and Chemistry*, 2017, vol. 2, pp. 339-341. (In Russ.)]

PHYSICO-CHEMICAL PROPERTIES OF DNA IN REGULATORY SITES OF GENOMES, EPIGENETICS AND CANCEROGENESIS

Nechipurenko Y.D.¹, Il'icheva I.A.¹, Abdullaev E.T.², Uroshlev L.A.³, Grokhovsky S.L.¹¹ Engelhardt Institute of Molecular Biology, Russian Academy of Sciences

119991 Moscow, Russia

² Max Planck Institute for Molecular Genetics

Berlin 14195, Germany

³ Vavilov Institute of General Genetics RAS

119991, Moscow, Russia

Abstract. Earlier we have shown that ultrasound irradiation causes multiple breakages in DNA and cleavage intensities in different positions of sugar-phosphate chains depend on dinucleotide types and the surrounding sequence. The relative intensities of cleavage of the central phosphodiester bond in 16 dinucleotides and 256 tetranucleotides were determined on the data of polyacrylamide gel electrophoresis. We observed a remarkable enhancement of the mean values of the relative intensities of cleavage (cleavage rates) in phosphodiester bonds following deoxycytidine, which diminished in the row of dinucleotides: $d(\text{CpG}) > d(\text{CpA}) > d(\text{CpT}) \gg d(\text{CpC})$. The cleavage rates for all pairs of complementary dinucleotides were significantly different from each other. Subsequently, we showed that the frequencies of dinucleotides cleavage of genomic DNA in the process of its fragmentation during NGS sequencing is very close to our cleavage rates, obtained earlier. These differences in cleavage intensity in different positions of DNA sequence may be caused by the differences of local conformational motions, mainly with singularities of pseudorotation of the furanose cycle. Thus, we came to the conclusion that there is a second level of information coding by a DNA molecule - the existence of a "conformational code". Analysis of a large number of physical and structural characteristics of DNA averaged over representative samples of different species of animals, plants and prokaryotes has shown that the index profiles characterizing DNA cleavage by ultrasound have characteristic properties in the areas that corresponds to region between the TATA box and the beginning of the transcription initiation site (TSS) in multicellular organisms. Since these indices carry information on conformational motions in individual strands of the DNA molecule, it can be asserted that the intensity of the conformational motions of the complementary chains in these regions varies in antiphase along the axis of the double-helix molecule. The most pronounced is this type of index profiles in the regulatory regions of DNA from mammalian cells. The intensity of cleavage of the CpG dinucleotide is also affected by the methylation of the cytosine at the 5-position of the pyrimidine ring. We evaluated the effect of methylation of cytosine on the cleavage of dinucleotides in human tissues, as well on the cleavage of CpG islands in control and cancer tissues (lymphoma and hepatocarcinoma). It turned out that on the sites of methylated CpG-dinucleotide clustering, genomic DNA is cleaved much more often than in similar sites, but with a low level of methylation. Moreover, a difference in the cleavage of CpG islands in cancer and control samples was established, and this is also due to differences in the level of their methylation. Therefore, it is possible to identify the precancerous state of tissues based on an estimate of the cleavage frequencies of CpG islands of different cell lines from DNA fragmentation data in NGS sequencing. The physical and chemical properties of DNA discovered by us can be used, both in theoretical studies and in the diagnostics of diseases.

Key words: Local DNA structure and conformation, sequence-specific ultrasonic cleavage, RNA polymerase II promoter sequences, methylation of heterocyclic basis in DNA, epigenetics, cancerogenesis.