

# РАСПРЕДЕЛЕНИЕ ОДНОНУКЛЕОТИДНЫХ ПОЛИМОРФИЗМОВ В ГЕНОМЕ ЧЕЛОВЕКА<sup>1</sup>

Е.В.Игнатьева

*Учреждение Российской академии наук Институт Цитологии и Генетики Сибирского*

*Отделения РАН*

eignat@bionet.nsc.ru

В.Г.Левицкий

levitsky@bionet.nsc.ru

Н.С.Юдин

yudin@bionet.nsc.ru

## Аннотация

Известно, что частота возникновения мутаций вдоль молекулы ДНК существенно варьирует. Однако закономерности распределения однонуклеотидных полиморфизмов (ОНП) в масштабе генома до сих пор исследованы не достаточно. В настоящей работе на основе данных проекта «1000 Genomes» исследовано распределение ОНП в различных участках генома человека, а также контекстные характеристики ДНК в окрестностях ОНП, расположенных в соседних позициях. Выявлено, что ОНП, локализованы в геноме человека неравномерно. Около 1.3% от общего количества ОНП находится в соседних позициях («парные ОНП», расстояние между которыми составляет 0 нуклеотидов), что в 3 раза превышает количество, которое можно было ожидать по случайным причинам. Плотность ОНП зависит от их локализации в гене. Низкая плотность ОНП наблюдается в окрестностях старта транскрипции гена, на 5'- и 3'- границах интронов, что хорошо согласуется с важным функциональным значением этих районов. Во всех исследованных районах (экзоны, интроны, 5'-UTRs, 3'-UTRs) участки ДНК, включающих «парные» ОНП, имеют повышенное содержание CG динуклеотидов в центральных позициях, а их 5' и 3'-фланги (3-5 п.н.) являются АТ –богатыми. Данные о распределении ОНП в геноме очень важны для экспериментальных исследований, поскольку при совместной встречаемости двух и более ОНП в пределах анализируемого участка ДНК чувствительность и точность многих экспериментальных методик идентификации полиморфизмов значительно снижается.

---

<sup>1</sup> Работа выполнена при частичной поддержке Минобрнауки (госконтракт с ФАО №П721), СО РАН (проект №119), президиума РАН (программы №№ Б.25, Б.27, А.П.6)

## Введение

На современном этапе выполнения проекта «1000 Genomes», посвященном исследованию генетического разнообразия в популяциях человека, накоплены данные о 15 миллионах ОНП, 55% из которых идентифицировано впервые [1]. Известно, что частота возникновения мутаций вдоль молекулы ДНК существенно варьирует. Участки ДНК, в которых мутации наблюдаются с повышенной частотой, называются «горячими точками». Обнаружено, что возникновение горячих точек обусловлено определенными структурными характеристиками районов ДНК, на которых они расположены. В ряде работ [2] представлен анализ контекстных характеристики участков ДНК, соответствующих горячим точкам и выявлены, так называемые «мутабельные мотивы». Однако, закономерности распределения ОНП в масштабе генома до сих пор исследованы недостаточно. В частности, данные проекта «1000 Genomes» дают возможность ответить на вопрос, оказывают ли рекомбинационные события какой-либо локальный мутагенный эффект. Было сделано вывод о том, что хотя рекомбинации могут влиять на судьбу новых мутаций, (например, посредством геной конверсии), нет свидетельств тому, что они влияют на скорость возникновения мутаций *de novo*. Выявлено, что в геноме человека плотность ОНП в окрестностях (2-3 нуклеотида) границ моонуклеотидных повторов в два-четыре раза выше, чем в окружающих районах [3]. Данные о расстояниях между ОНП очень важны для экспериментальных исследований. Совместная встречаемость двух и более ОНП в пределах анализируемого участка ДНК значительно снижает чувствительность и точность многих экспериментальных методик идентификации полиморфизмов.

Целью работы являлось исследование расстояний между ОНП в контексте полного генома человека, особенностей распределения ОНП в различных участках генома (экзоны, интроны, промоторные районы и т.п.), а также контекстных характеристик ДНК в окрестностях ОНП, выявленных в соседних позициях («парных ОНП»).

Методы Анализ проводили на основе данных проекта «1000 Genomes», которые были экстрагированы из базы dbSNP (версия 131). Случайное распределение оценивали на основе равномерного распределения наблюдаемого количества ОНП по доступным позициям геномной ДНК.

## Результаты

Выявлено, что ОНП, идентифицированные и охарактеризованные в ходе выполнения проекта “1000 Genomes”, локализованы в геноме человека не равномерно. Исходя из того, что общая длина известной геномной последовательности ДНК человека составляет 2897310462 п.н. и в ней выявлено всего 11532812 ОНП, можно ожидать, что среднее

расстояние между ОНП будет составлять 251 нуклеотид. Вместе с тем, оказалось, что в 69 % случаев ОНП разделены расстоянием меньшим, чем 250 нуклеотидов.

Была определено количество ОНП (суммарно, по всем хромосомам), находящихся на различном расстоянии друг от друга (Рис. 1). Оказалось, что ситуации, когда ОНП находятся на близких расстояниях друг от друга (1-5 нуклеотидов) встречаются гораздо чаще, чем случаи, когда ОНП удалены друг от друга на большие расстояния. Около 1.3% от общего количества ОНП находится в соседних позициях (то есть расстояние составляет 0 нуклеотид), что в 3 раза превышает количество, которое можно было бы ожидать по случайным причинам на основе равномерного распределения. Количество случаев, когда ОНП находятся в соседних позициях («парные» ОНП) в 1.5 раз больше, чем когда ОНП разделены 1 нуклеотидом.

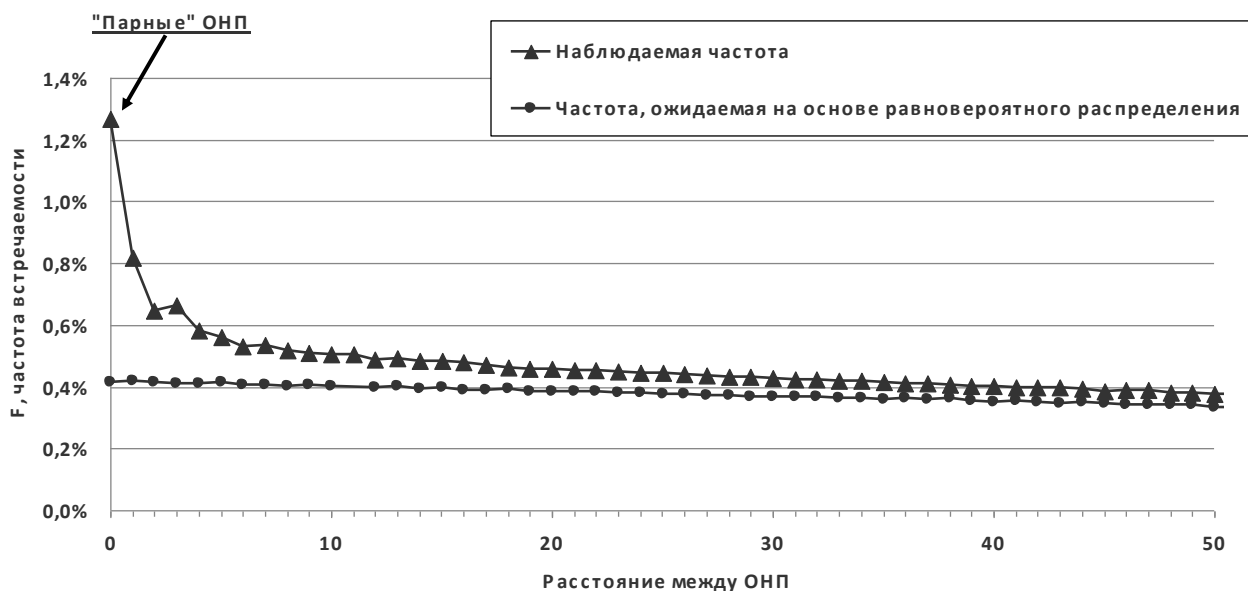


Рисунок 1. Расстояние между ОНП в геноме человека (ось OX) и частота встречаемости соответствующих случаев в геноме F (ось OY), где F = суммарное количество ОНП в геноме, находящихся на заданном расстоянии друг от друга / общее количество ОНП в геноме. «Наблюдаемая частота» - данные, полученные на основании dbSNP (версия 131); «Частота, ожидаемая на основе равномерного распределения» - данные, полученные на основании анализа теоретически предсказанного распределения ОНП.

Плотность ОНП зависит от локализации в гене. Пониженная плотность ОНП наблюдается в 5'-фланкирующих районах генов (-100 / -1 относительно начала транскрипта) (Рис. 2), а также на 5'- и 3'-границах интронов (Рис. 3). Эти наблюдения хорошо согласуются с тем, что эти районы взаимодействуют с белковыми комплексами, обеспечивающими инициацию транскрипции гена и сплайсинг пре-мРНК.

## Доля транскриптов

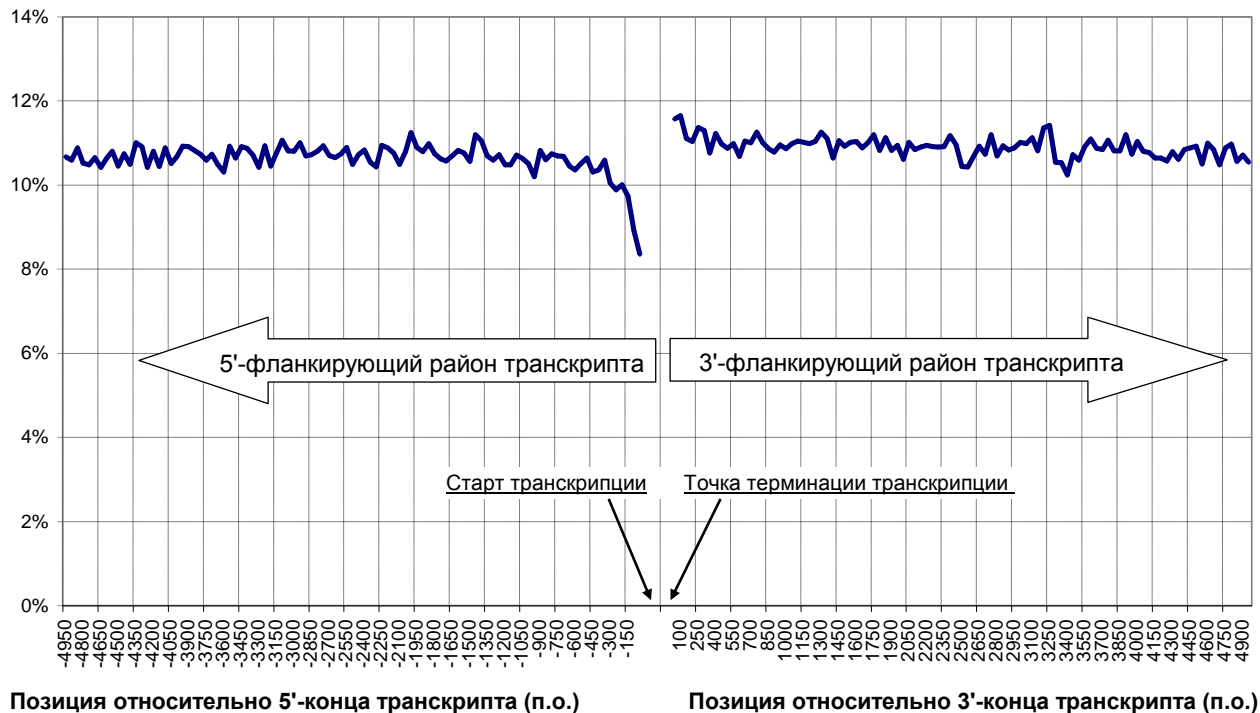


Рисунок 2. Плотность ОНП в окрестностях 5' и 3' границ транскриптов. По оси ОХ – координаты районов относительно 5' и 3' границ транскриптов (соответственно), по оси ОУ – доля транскриптов (в процентах), у которых обнаружены ОНП в районах, указанных по оси ОХ.

Выявлено, что встречаемость ОНП в экзонах имеет периодичность, кратную трем (на рисунках не представлено), что отражает тот факт, что в соответствии с триплетным кодом многие нуклеотидные замены в третьих позициях кодонов являются синонимичными, то есть не приводят к замене кодируемой аминокислоты.

Поскольку было обнаружено, что ОНП в соседних позициях встречаются неслучайно часто, нами были исследованы контекстные характеристики участков ДНК, включающих «парные» ОНП. Была проведена оценка моно- и динуклеотидного состава соответствующих участков ДНК. Обнаружено, что во всех исследованных районах (экзоны, интроны, 5'UTRs, 3'UTRs), центральные позиции содержат повышенную частоту CG динуклеотидов, в то время как 5' и 3'-фланкирующие участки являются AT-богатыми.

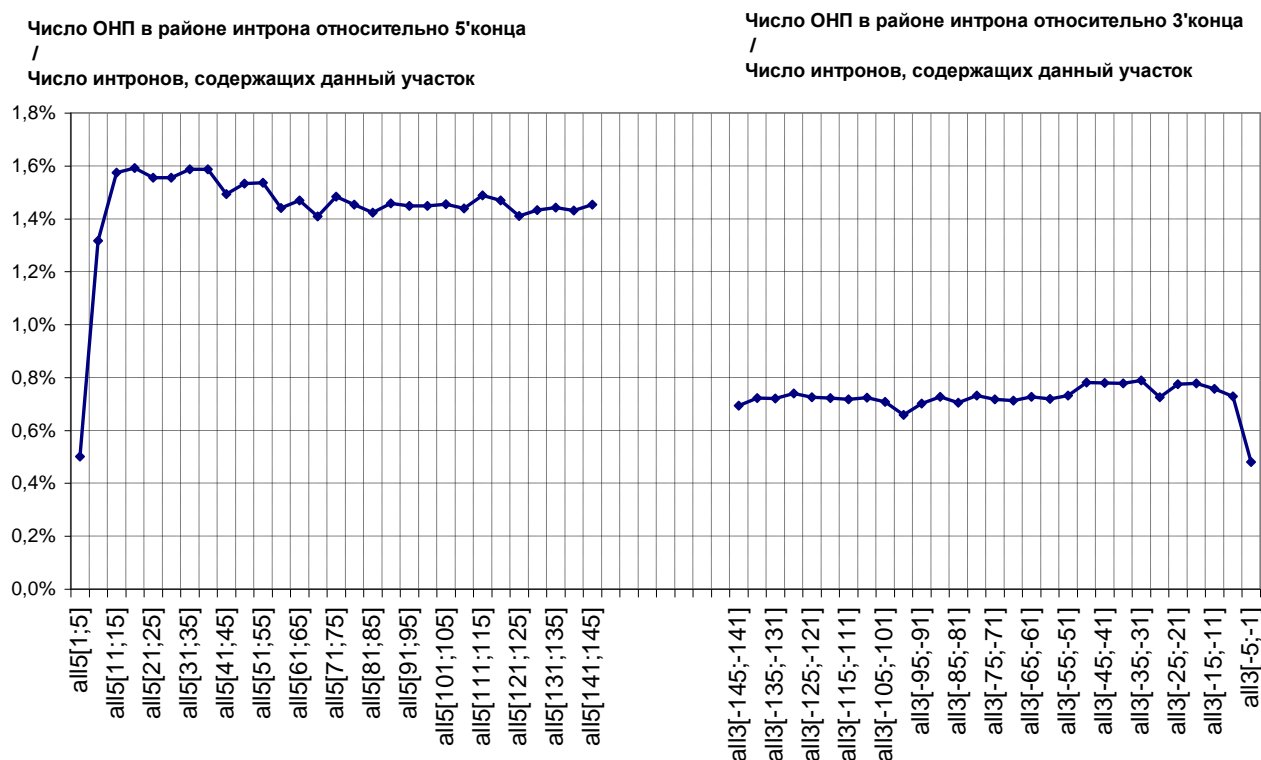


Рисунок 3. Плотность ОНП в окрестностях 5'- и 3'- границ интронов. По оси ОХ – координаты участков ДНК относительно 5'- и 3'- границ интронов (соответственно), по оси ОУ плотность ОНП в указанных по оси ОХ участках, рассчитанная как число ОНП в рассматриваемом районе / количество интронов, в которых было сделано наблюдение.

### Обсуждение

Анализ данных пилотной стадии проекта «1000 Genomes» показал, что распределение ОНП в геноме человека отличается от ожидаемого на основе равновероятного распределения, что хорошо согласуется с концепцией существования, так называемых, «горячих точек мутаций», обуславливающих предрасположенность определенных участков ДНК к мутациям. Наличие «горячих точек мутаций» принято объяснять несколькими причинами: 1) специфичностью взаимодействия мутагена с определенным участком ДНК; 2) специфичностью взаимодействия ферментов репарации с ДНК; 3) структурными либо функциональными особенностями кодируемого фрагмента белка либо РНК [2]. Например, выявлено, что плотность ОНП в промоторных районах с высоким CpG составом имеет периодичность, совпадающую с длиной участка ДНК, входящего в состав позиционированной нуклеосомы (146 п.о.) [4]. Этот факт хорошо соответствует тому наблюдению, что нуклеосомная укладка существенным образом влияет на доступность молекулы ДНК для ферментов репарации [5].

Полученные данные по частотам встречаемости и плотности ОНП в геноме человека

необходимо учитывать при оценке достоверности результатов экспериментов по генотипированию ОНП. Так, для методики, генотипирования на основе полиморфизма длин рестрикционных фрагментов длина анализируемого фрагмента ДНК составляет 4-8 п.н., а для широко используемой методики HRMA (high-resolution melting curve analysis ) размер фрагмента достигает 150-200 нуклеотидов [6]. Нами обнаружено, что в половине случаев расстояние между ОНП составляет менее 250 п.о. Полученный результат указывает на высокую вероятность существования дополнительных полиморфных позиций во фрагментах ДНК размером 150-250 п.о., что в методике HRMA соответствует размерам отжигаемых фрагментов. Возможность совместной встречаемости ОНП на близком расстоянии друг от друга должна быть учтена при разработке новых, более специфичных протоколов исследования. Например, если кривая плавления, определяемая в методике HRMA, содержит более одного домена плавления, должна быть дана рекомендация к ресеквенированию данного образца ДНК.

#### ЛИТЕРАТУРА

- [1]. 1000 Genomes Project Consortium et al. A map of human genome variation from population-scale sequencing *Nature*. 2010. V. 467, P. 1061-1073.
- [2]. Rogozin I.B. et al. Computational analysis of mutation spectra *Brief. Bioinform.* 2003. V. 4, P. 210-227.
- [3]. Siddle K.J. et al. Bases adjacent to mononucleotide repeats show an increased single nucleotide polymorphism frequency in the human genome *Bioinformatics*. 2011. V. 27, P. 895-898
- [4]. Higasa K., Hayash K. Periodicity of SNP distribution around transcription start sites *BMC Genomics*. 2006. V.7, P.66
- [5] Hinz J.M., Rodriguez Y., Smerdon M.J. Rotational dynamics of DNA on the nucleosome surface markedly impact accessibility to a DNA repair enzyme *Proc. Natl. Acad. Sci. USA*. 2010. V. 107, № 10. P. 4646–4651.
- [6] Vossen R.H. et al. High-resolution melting analysis (HRMA): more than just sequence variant screening *Hum. Mutat.* 2009. V.30, P. 860-866.
- [7] Sambrook J., Fritsch E.F., Maniatis T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY, 1989. Vol. 1.