

МАТЕМАТИЧЕСКАЯ ДОПУСТИМОСТЬ РАСШИРЕНИЯ МНОЖЕСТВА КОДОВ ГЕНОМА

Ю.А. Кузавко

ПАЭИ НАН Беларуси,
Московская 204, 224020 Брест, Беларусь
kuzavko@inbox.ru

В 2003 г. завершился грандиозный международный проект «Геном человека», длившийся более 10 лет и приведший к определению полной последовательности нуклеотидов ДНК человека, содержащей 3.2 млрд. пар нуклеотидов. Каждая из 20 аминокислот, из которых состоят все протеины (их около 100 тыс.), кодируются тройками нуклеотидов — кодонами (их 64). Математически кодировка означает способ расположить 64 кодона в 20 клетках. Универсальный код человека включает в себя три терминаторных кодона, при этом оставшиеся 61 кодируют 20 аминокислот неоднозначно, т. е. код является вырожденным. Важнейшим свойством кода является его универсальность: код одинаков для почти всех живых организмов. Однако уже сейчас обнаружены четырнадцать отклонений кода от стандартного K_0 . В частности, было обнаружено расширение числа кодируемых аминокислот у *E.coli*.

С точки зрения формальной математики показано, что кодирование ДНК и м-РНК и кодирование протеинов может происходить по разным сценариям, в т. ч. экспериментально не обнаруженным. Так код K_0 дополнен еще 18 возможными кодировками [1]. Здесь

аналогичные вычисления выполнены для рассмотрения всех обнаруженных 14 кодировок и обоснованы правила отбора сценариев кодирования, отличные от ранее предложенных [2].

Обсуждается кодирование в митохондриях клетки и ретровирусах, преимущественно онкологических. Записано уравнение преобразования генетической информации (у человека около 30 тыс. генов) для подсчета количества математически возможных кодов. Демонстрируется модель построения участков ДНК и соответствующих протеинов средствами молекулярного моделирования.

Отметим, что молекула ДНК состоит из двух полимерных цепочек и содержит всего 4 типа нуклеотидов: А (аденин), Т (тимин) и G (гуанин), С (цитозин), которые связаны между собой принципом комплиментарности. Если выстроить в одну линию все пары нуклеотидов ДНК, заключенные в клетке человека, то получится нить длиной 2 м, которая плотно упакована гистонами в хромосомах ядра. Делетация ДНК на отдельные органеллы с точностью до одного нуклеотида реализуется рестрикторами гель-электрофорезом. Несмотря на то, что фирма «Селера» под научным руководством Сэнгера использовала ряд новейших технологий, процесс распознавания генома человека сократился только до месяца. Нами предложено использование атомно- (магнитно-) силовой микроскопии для распознавания по зарядовой плотности нуклеотидов ДНК, которые по химическим формулам различаются слабо.

Вышеуказанная микроскопия позволяет распознавать один нуклеотид с низкой достоверностью $1-p$ ($p > 0.1$). Условие комплиментарности повышает достоверность распознавания до $1-p^2$. При получении n идентичных фрагментов ДНК достоверность составляет до $1-p^{2n}$. Определено число n необходимое для достижения необходимой ошибки (например, в один нуклеотид) при распознавании полной ДНК.

Литература

1. Кузавко Ю.А. О математической допустимости неединственности стандартного, нестандартных и искусственных кодов генома человека // Вестник БрГТУ. 2007. № 5(47). С. 75–77.
2. Козлов Н.Н., Базарев М.Г. Математический анализ одной биологической структуры // Препринт ИПМ им. М.В. Келдыша РАН. М., 2006. 28 с