# Поиск закономерностей в геномах

## Проблема

**Геном человека** — совокупность наследственного материала, заключённого в клетке человека. Кроме очевидной фундаментальной значимости, определение структуры человеческих генов является важным шагом для разработки новых медикаментов и развития других аспектов здравоохранения.

Математические методы применимы для выявления регуляторных и кодирующих участков в структуре геномов. Этот анализ основан на сопоставлении различных символьных последовательностей исследуемых геномов с паттернами ДНК, функции которых уже известны.

## Задача

С формальной точки зрения геном можно считать словом (или зацикленным словом) в алфавите мощности 4. Интересным является вопрос плотности упаковки генов.

## Особенности задачи

Содержательно ген – это последовательность, начинающаяся с так называемого начального кода и заканчивающаяся стоп-кодом. Если говорить о зацикленной структуре, одна буква исходного алфавита может находиться в шести генах одновременно.

Как только стало понятным, что ручной поиск примера генома либо не приведет к результату, либо окажется слишком долгим, было принято решение о написании оптимального алгоритма для поисков геномов с нужным свойством на языке Python.

## Что было сделано?

- Выявлены особенные комбинаторные свойства и критерии поиска генома
- Написана программа поиска на Python
- **Получен кратчайший геном, содержащий букву, одновременно входящую в 6 не пустых генов**

## Итог

**В работе показано, что ситуация реализуема, то есть возможна максимально плотная упаковка генов.**

**В дальнейшем планируется вычислить плотность упаковки на некоторых известных геномах, а также учесть тонкие ограничения на длины генов.**