

# *Проекты в биоинформатике*

**Пржибельский Андрей**  
**[ap@bioinf.spbau.ru](mailto:ap@bioinf.spbau.ru)**

Лаборатория Алгоритмической Биотехнологии  
СПбГУ

# Что такое биоинформатика?

- Применение вычислительных методов для решения биологических задач
- Анализ больших объемов зашумленных данных
- Главный инструмент — компьютер

# Центральная догма молекулярной биологии

ДНК



Транскрипция

РНК



Трансляция

Белок



# “Чтение” молекул

ДНК



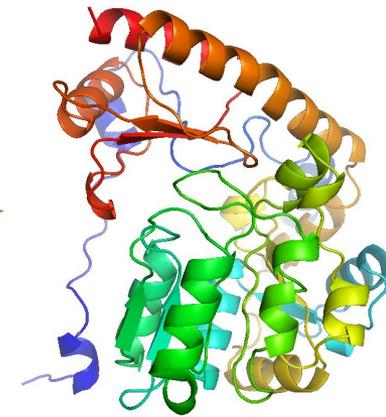
Транскрипция

РНК



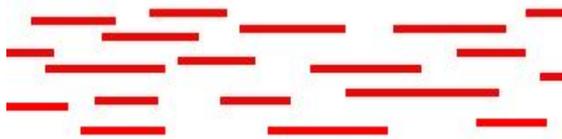
Трансляция

Белок

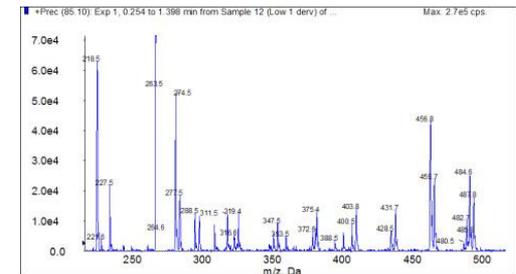
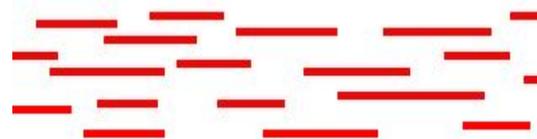


Масс-спектрометрия

Секвенирование



Секвенирование



# Секвенирование

...AACCCGTACGTTTTGCAAACGACCGT...

# Секвенирование

**GTACGTTTTGCA**

**GTTTTGCAAACG**

**CGTACGTTTTG**

**AACCCGTACGT**

**AACGACCG**

**...AACCCGTACGTTTTGCAAACGACCGT...**

# Секвенирование

GTA\_GTTTTGCA

GTTTTGCAAACG

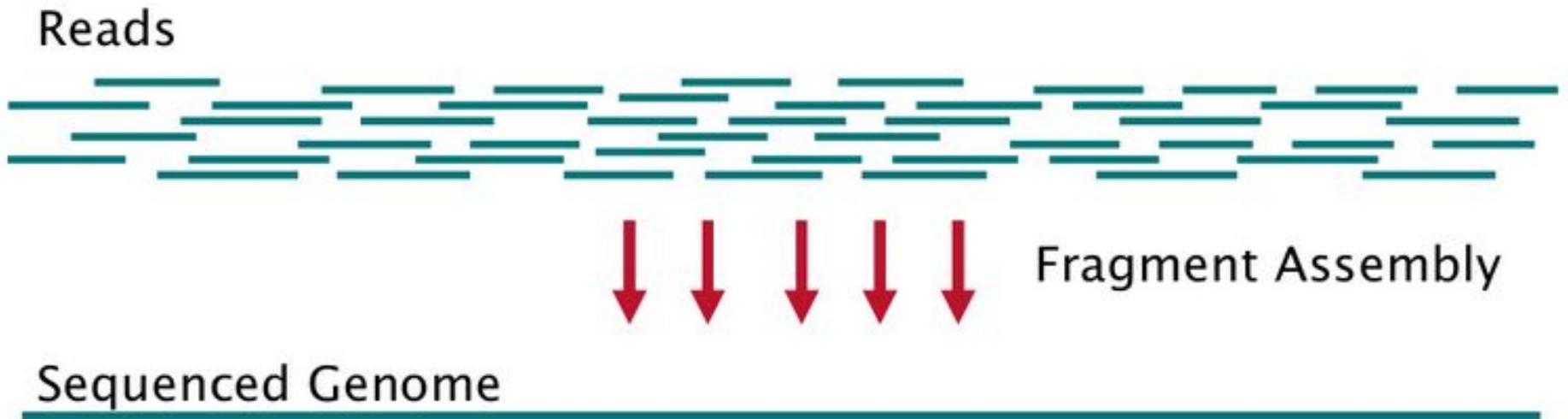
CGTACGTTTTC

AACCCGTTTCGT

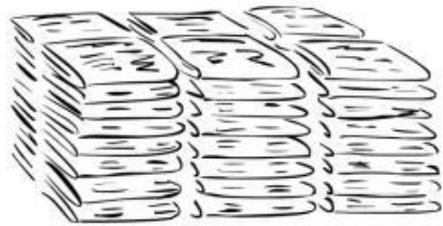
AACGACCG

...AACCCGTACGTTTTGCAAACGACCGT...

# Сборка генома



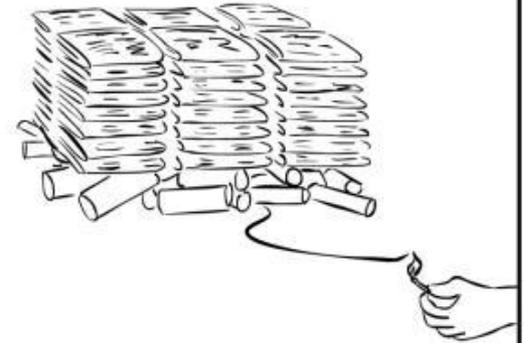
# Сборка генома



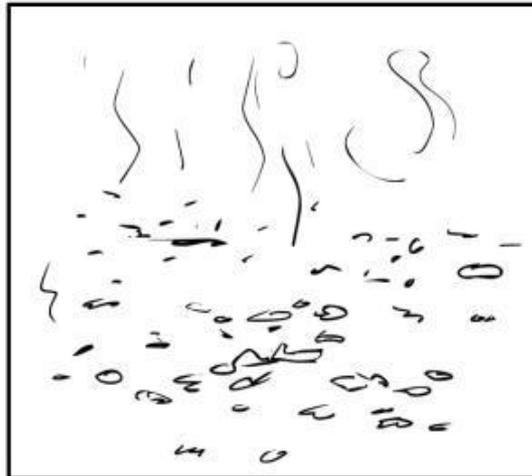
stack of NY Times, June 27, 2000



stack of NY Times, June 27, 2000  
on a pile of dynamite



this is just hypothetical



so, what did the June 27, 2000 NY  
Times say?

# Чем занимается лаборатория?

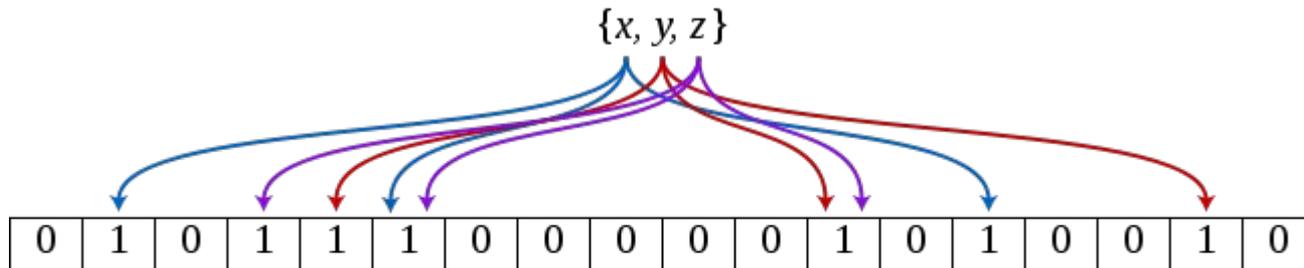
- Сборка по данным секвенирования
  - Геномика
  - Транскриптомика
- Протеомика
  - Восстановление белков по масс-спектрам
  - Поиск новых антибиотиков
- Иммуноинформатика

# Что нужно будет делать на проектах?

- Писать код (C++11, Python)
- Придумывать алгоритмы
- Разбираться в существующих алгоритмах и программах
- Разбираться в природе происхождения данных и понимать как их анализировать

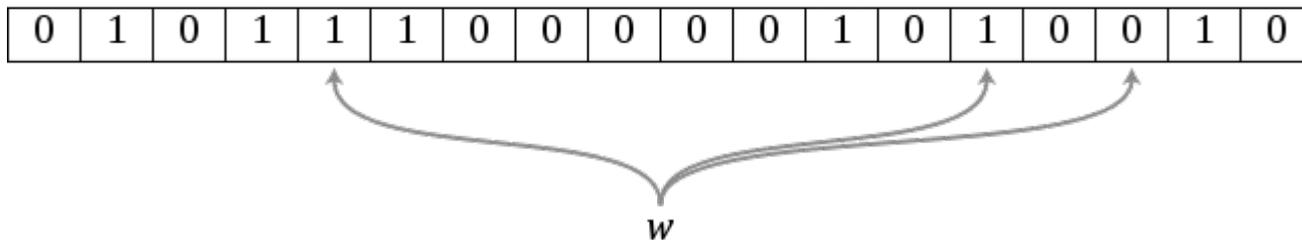
# Вloom-фильтр

- Битовый массив длины  $n$ ,  $k$  хеш-функций
- Добавление элемента  $x$ 
  - Установка битов  $\text{hash}_1(x), \dots, \text{hash}_k(x)$  в 1



# Вloom-фильтр

- Наличие элемента  $x$ 
  - Проверка битов  $\text{hash}_1(x), \dots, \text{hash}_k(x)$
  - Если хотя бы один 0, то элемента  $x$  нету
  - Если все 1, то может быть есть



# Список проектов

- Фильтр Блума (А. Коробейников, [anton@korobeynikov.info](mailto:anton@korobeynikov.info))
- Иммуноинформатика (А. Шлемов, [shlemovalex@gmail.com](mailto:shlemovalex@gmail.com))
  - Слияние парных ридов антител
  - Коррекция ошибок в ридах антител
- Построения графа связей геномных последовательностей по ридам РНК/ДНК (1-2 человека, Пржибельский Андрей, [ap@bioinf.spbau.ru](mailto:ap@bioinf.spbau.ru))
- Возможны дополнительные проекты по протеомике (К. Вяткина)