

БОЛЬШИЕ ДАННЫЕ В ПОСТ-ГЕНОМНУЮ ЭРУ (26-813)

Аннотация

Под Большими данными понимаются генерация данных, разработка алгоритмов обработки данных, методов их хранения и доступа, собственно обработка данных.

Развитие технологий секвенирования порождает очень большие объемы данных – это многие тысячи терабайт информации. Эти не только новые геномные данные, но также целый ряд сопутствующих данных, которые отражают структуру и функционирование генома. Правильная обработка данных позволяет решать целый ряд фундаментальных и прикладных, в том числе медицинских, задач.

Несмотря на то, что в последнее время уже накоплено большое количество молекулярно-биологических данных, их генерация продолжается. Это связано, во-первых, с исследованием новых, ранее не исследованных, объектов и процессов, во-вторых, с разработкой новых методов исследования биологических объектов, и, наконец, с медицинскими исследованиями. Работа с большими данными требует разработки методов хранения, агрегации и доступа к этим данным. При этом встает несколько проблем. Во-первых, это просто гигантский объем данных, который требует развития инфраструктуры. Во-вторых, кроме хранения первичных данных необходимо обеспечить их пред-обработку. В-третьих, – однотипные данные порождаются в разных лабораториях по разным протоколам. Необходима агрегация данных, полученных в разных лабораториях. Наконец, необходимо разрабатывать пользовательские интерфейсы, которые позволили бы исследователям, далеким от ИТ, работать с ними.

Обработка Больших данных в биологии требует разработки новых методов анализа данных. Для анализа массовых данных разрабатываются новые методы статистического анализа данных, методы машинного обучения и методы машинного моделирования.

Кроме проблемы разработки новых методов анализа данных, весьма актуальной является задача анализа конкретных наборов данных для извлечения из них содержательных биологических результатов.

Наконец, машинное моделирование позволит проверять различные гипотезы, а также получать новые знания.

Согласно «Стратегии научно-технологического развития Российской Федерации», утвержденной Указом Президента РФ № 642 от 1 декабря 2016 года, проблема эффективной работы с Большими данными относится к категории «больших вызовов». В ходе решения данной проблемы необходимо создать систему эффективной обработки Больших данных с применением методов искусственного интеллекта, в том числе машинного обучения, для извлечения новых знаний,

ориентированных на переход к персонализированной медицине, высокотехнологичному здравоохранению и технологиям здоровьесбережения.

Прикладные работы, базирующиеся на результатах фундаментальных исследований в области Больших данных, не ограничиваются лишь возможностями открытия новых биомаркеров, диагностических средств, фармакологических мишеней, более безопасных и эффективных лекарств и персонализации терапии, они важны для выбора правильных направлений геномного редактирования для медицинских и сельскохозяйственных применений, а также для биобезопасности.

Рубрикатор

- 813.1.** Фундаментальные исследования, направленные на генерацию и анализ Больших геномных и пост-геномных данных.
- 813.2.** Фундаментальные исследования, направленные на разработку методов анализа Больших данных.
- 813.3.** Суперкомпьютерное моделирование структур и процессов в биологических системах.