

ВВЕДЕНИЕ

Биоинформатика образована на стыке биологических направлений и информатики, как реакция на всё увеличивающийся объем данных, требующих сложных, быстрых и качественных алгоритмов для обработки и анализа. Разумеется, биоинформатикой занимаются как, непосредственно, информатики, которые обладают навыками программирования и могут реализовывать алгоритмы самостоятельно, так и биологи, которые отлично могут интерпретировать результаты работы алгоритмов и сами данные, но не имеют соответствующей подготовки для использования этих методов. Исследователям для более продуктивной работы нужны удобные и интуитивно понятные инструменты для анализа данных, которые бы хорошо покрывали все их потребности в реализованных методах и алгоритмах. На данный момент, таких инструментов достаточно мало, а в тех, что есть, неполноценный функционал. Таким образом, целью данной работы является разработка инструмента для полноценного анализа биологических данных.

ГЛАВА 1. ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ

В данной обзорной главе будет представлена предметная область биоинформатики, введено понятие экспрессии генов, рассмотрены существующие решения для анализа экспрессии генов, а также перечислен ряд технологий, которые используются или могут быть использованы для создания инструментов анализа экспрессии генов.

1.1. Биоинформатика

Биоинформатика — наука, объединяющая в себе методы прикладной математики, статистики, информатики для создания новых методов и алгоритмов для анализа разного рода биологических данных.

Биоинформатика занимается биохимией, биофизикой, экологией и многими другими областями биологии. Однако фокус в данной работе будет сосредоточен на конкретную задачу биоинформатики — анализ экспрессии генов.

1.1.1. Анализ экспрессии генов

Экспрессия генов — процесс преобразования наследственной информации от гена (в виде последовательности нуклеотидов ДНК) в функциональный продукт (РНК или белок).

Анализ экспрессии генов позволяет выяснить, как ведет себя каждый отдельный ген в разных условиях, тканях или организмах.

Экспрессия гена в образце характеризуется вещественным числом, которое также можно назвать некоторой мерой активности гена в данных условиях.

1.1.2. Используемые методы

Как было сказано ранее, биоинформатика использует в себе математику, информатику и статистику. Соответственно, задача анализа экспрессии генов сводится к исследованию путем статистических методов и алгоритмов числовой двумерной матрицы, где в виде вещественных чисел демонстрируется активность каждого гена в каждом образце. Пример такой матрицы можно увидеть в таблице 1.

Одним из первоочередных методов, применяемых для анализа экспрессии, является *визуальный анализ*. Числовая матрица представля-