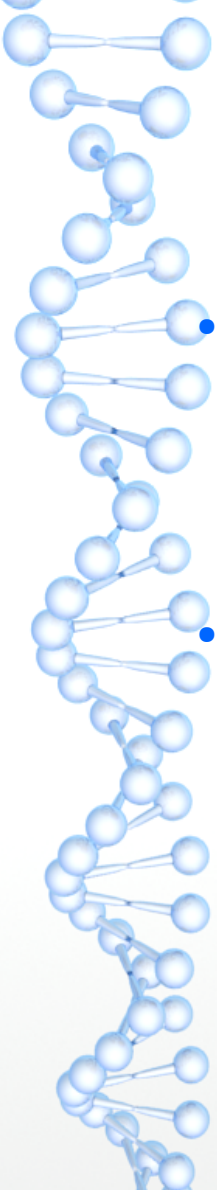


Предсказание метаданных эксперимента по экспрессии генов

Команда „Сказочники из Рэндом Форест“

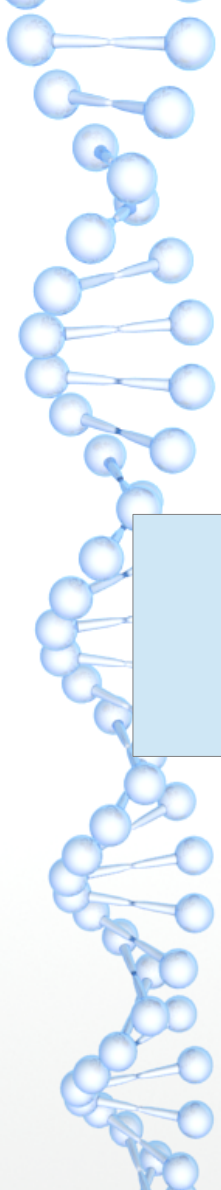


Цель:

- автоматическое предсказание некоторых типов аннотации по данным экспрессии генов (пол, тип клеток из которых взят материал и другие)

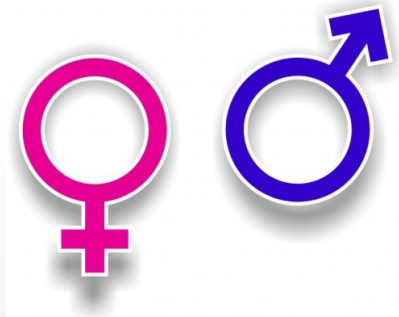
Задачи:

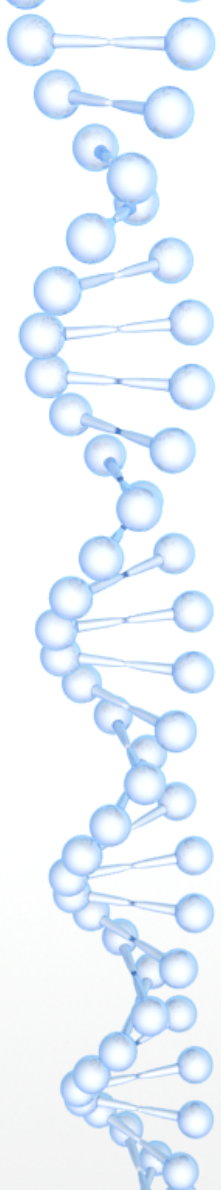
- сбор обучающего набора данных
- разработка метода машинного обучения для предсказания.



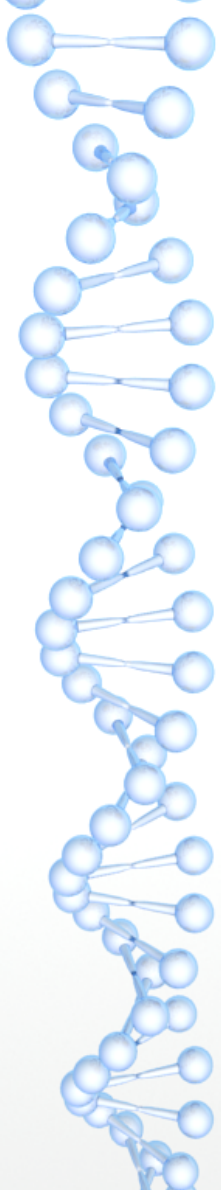
Ткань

Линия





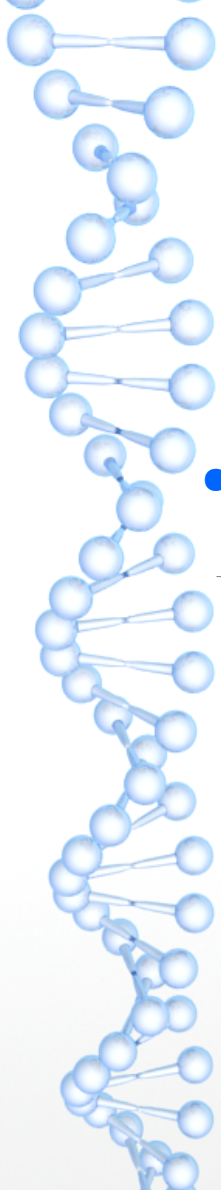
Анализ данных



- Фильтрация
 - 1 проба = 1 гену
- Логарифмирование
- Нормализация
 - Пакет программ Limma
 - (housekeeping genes, как множество генов, экспрессия которых не должна меняться в условиях эксперимента)



Процесс машинного обучения



- Supervised learning

- Classification



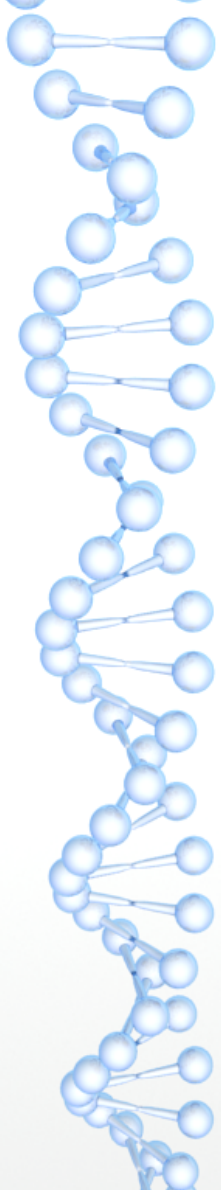
Gender

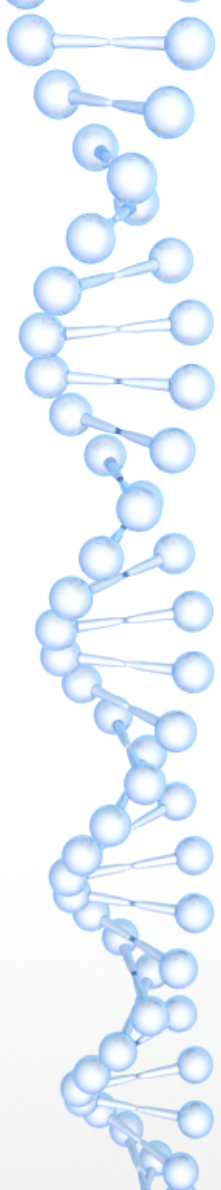
- Train dataset (70 s)
 - 5000 генов
 - 10 платформ
 - SVM 92,12%
 - Random forest 96,15%
 - Boosted logistic regression 96,15%
- Test dataset (98 s)
 - 3300 генов
 - 10 платформ
 - Random forest 62%



Tissue

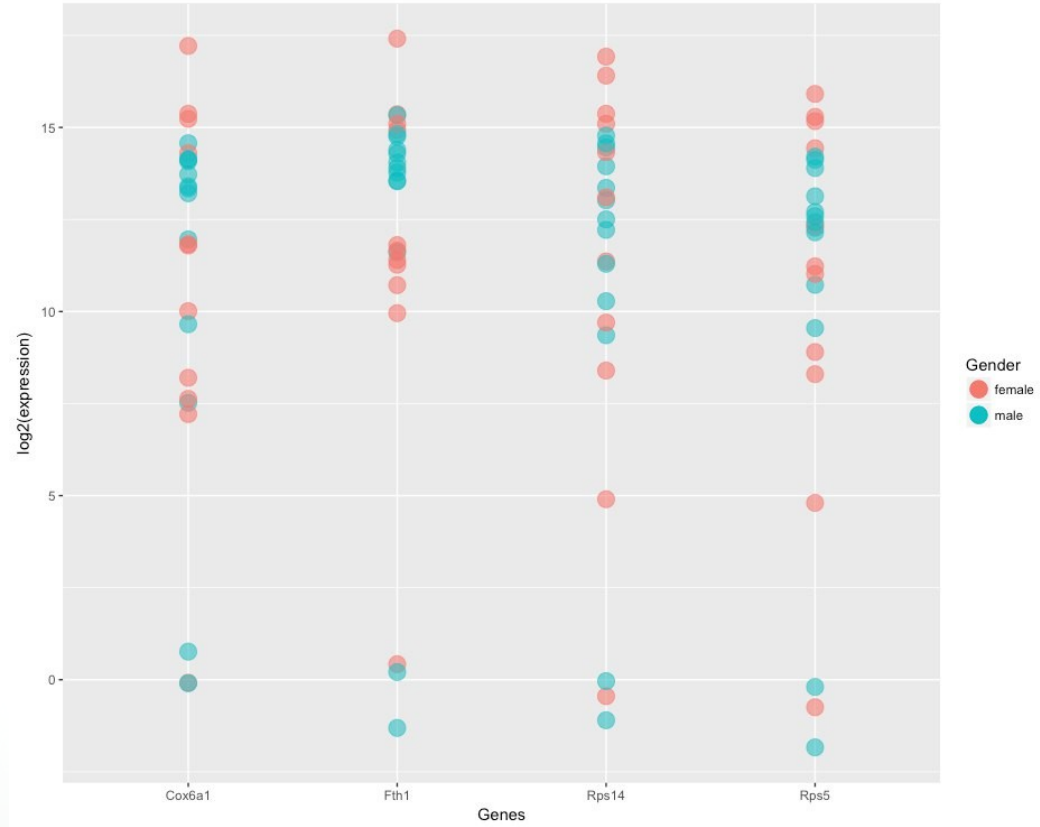
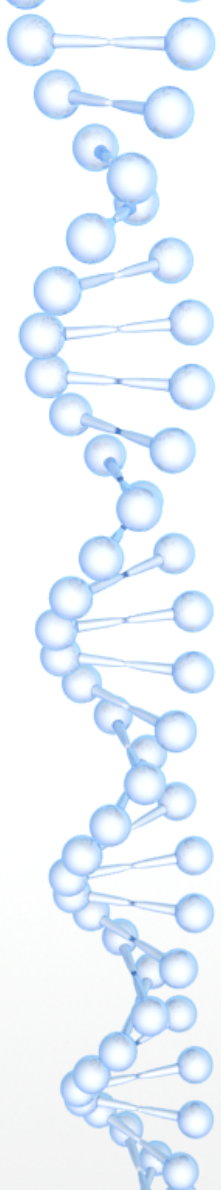
- Train dataset (34s)
 - 14000 генов
 - 10 платформ
 - SVM 98%
 - Random forest 90%
- Test dataset
 - 14000 генов
 - 10 платформ
 - Random forest 92%





Метод главных компонент

COX6a1
Fth1
Rps5
Rps14



СПАСИБО ЗА ВНИМАНИЕ

