The background features a collage of various computer mice and small mice. There are several computer mice in different colors (grey, red, black) and brands (Logitech, HP). Interspersed among them are several small, light-brown mice. The text is centered over this collage.

# Предсказание метаданных эксперимента по экспрессии генов

Команда #3

# Этапы

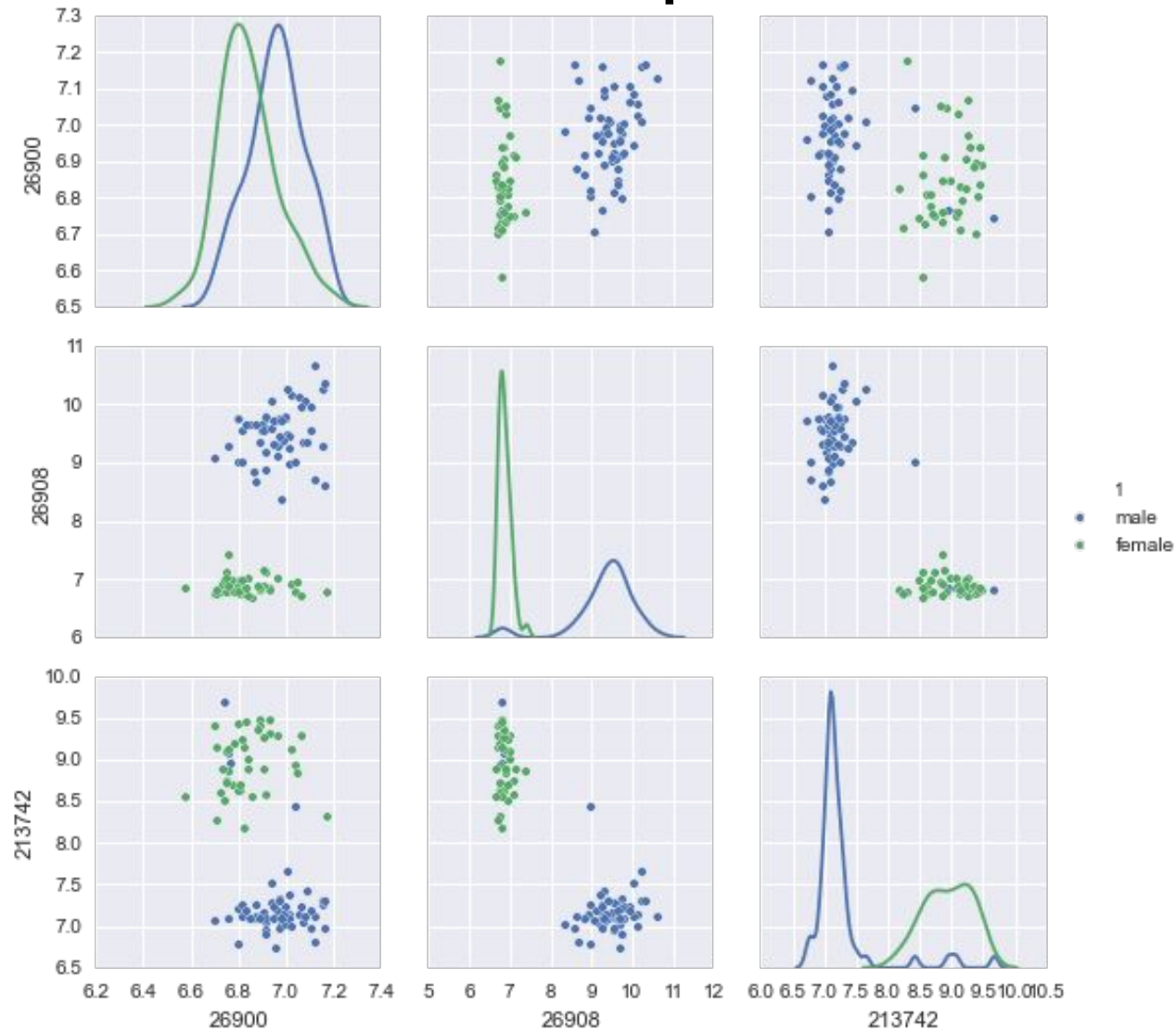


1. Нормализация входных данных
2. Разработка метода машинного обучения
3. Создание тренировочных данных для машинного обучения
4. Обучение машины на тренировочных данных
5. Предсказание аннотаций для тестовых данных

# Подготовка данных. Проблемы

1. Недостаточная выборка для кластеризации генов по платформам
2. Разная нормализация входных данных
3. Отсутствие некоторых генов в результатах микрочипирования образцов

# Определение пола мышей по данным экспрессии генов



Гены-детерминанты пола:

- *Xist* (213742) ♀
- *Ddx3y* (26900) ♂
- *Eif2s3y* (26908) ♂



# Определение типа ткани по экспрессии генов

Мышцы	Легкие	Мозг	Печень
<i>Actc1</i>	<i>Cavin2</i>	<i>Tubb3</i>	<i>Apoa4</i>
<i>Mybpc3</i>	<i>Sftpa1</i>	<i>Ywhah</i>	<i>C8b</i>
<i>Acta1</i>			
<i>Myod1</i>			



# Результаты. Определение пола

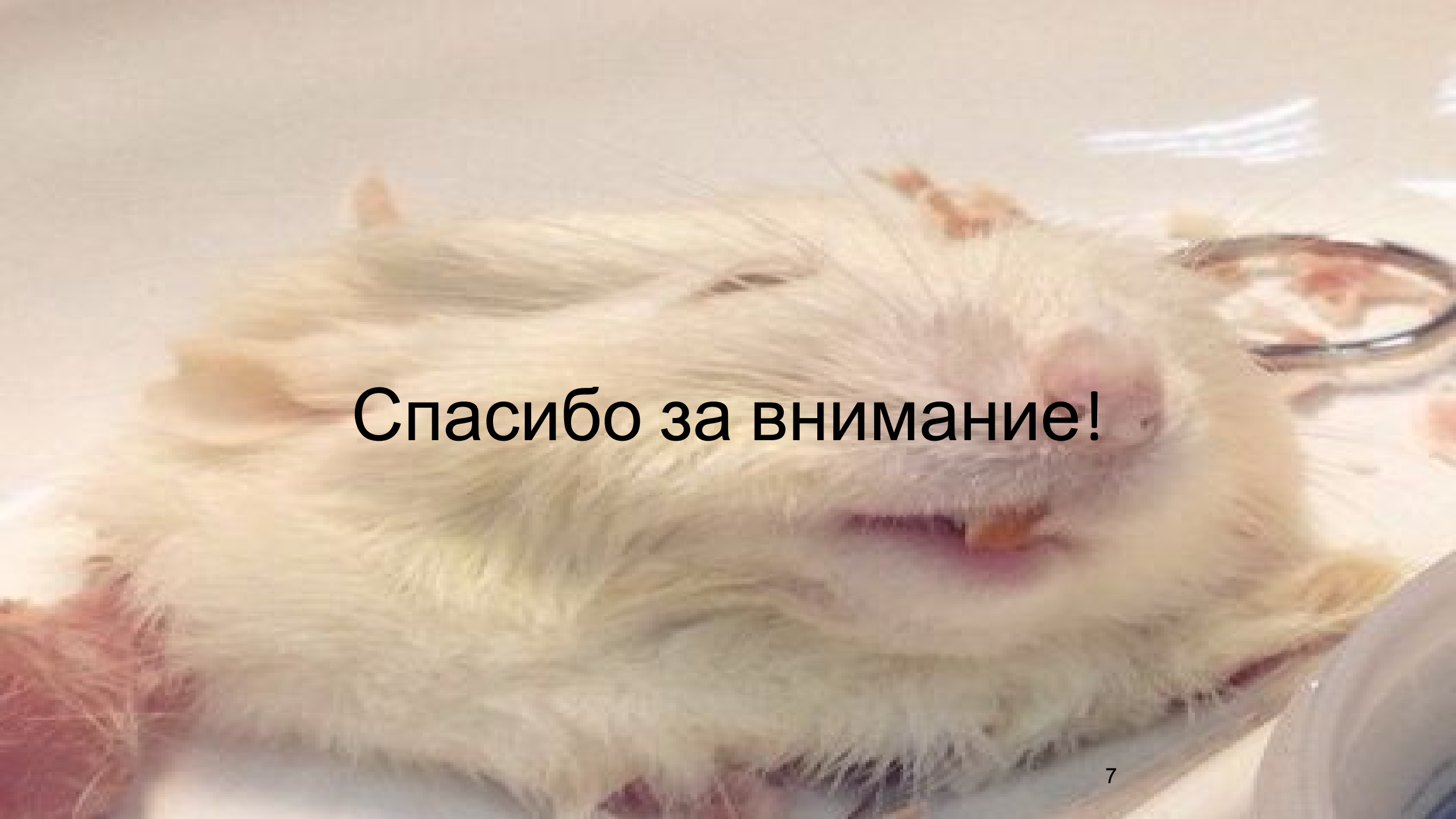
Метод: Random Forest

	dataset 1	dataset 2	dataset 3
training	93,2%	99,3%	93,2%
testing	78,1%	91,7%	65%

dataset 1 -- начальные данные для обучения, разделенные для обучения и тестирования (разные платформы, 98 образцов)

dataset 2 -- данные GSE31840 (одна платформа, 99 образцов)

dataset 3 -- начальные данные для обучения и тестирования (разные платформы, 97 и 98 образцов)



**Спасибо за внимание!**