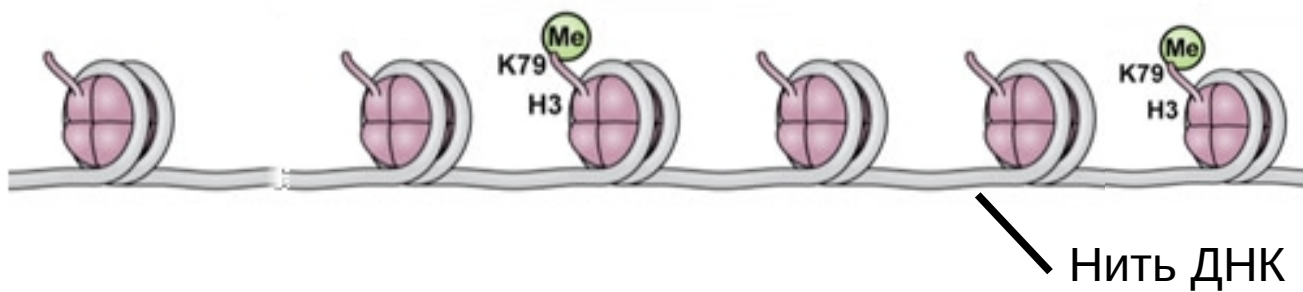


ИНСТИТУТ  
БИОИНФОРМАТИКИ

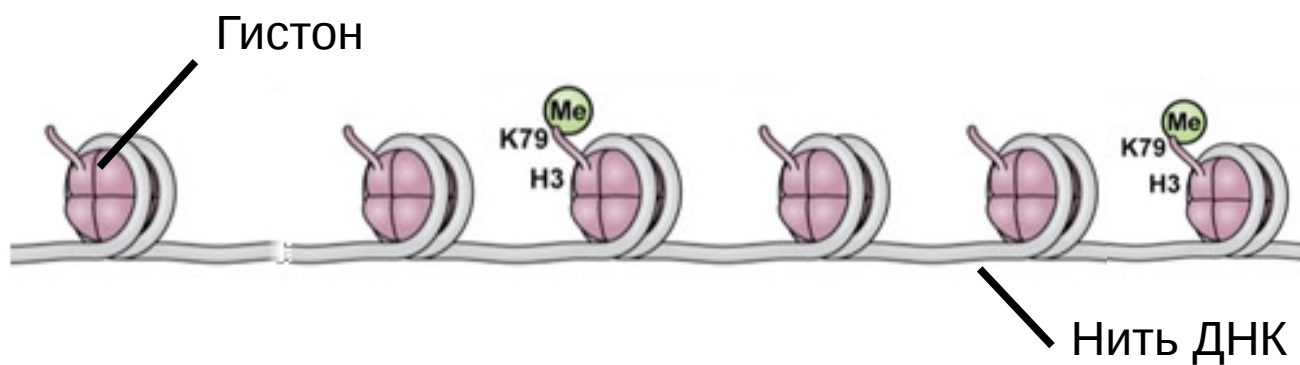
# Средство для сравнительного анализа результатов экспериментов ChIP-seq

Стажер: Евгений Бакин  
Руководитель: Олег Шпынов

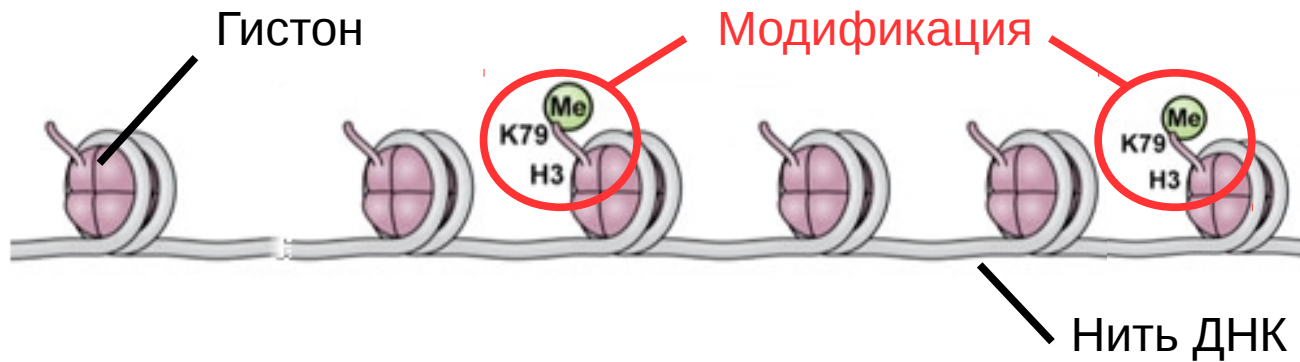
# Что такое ChIP-seq?



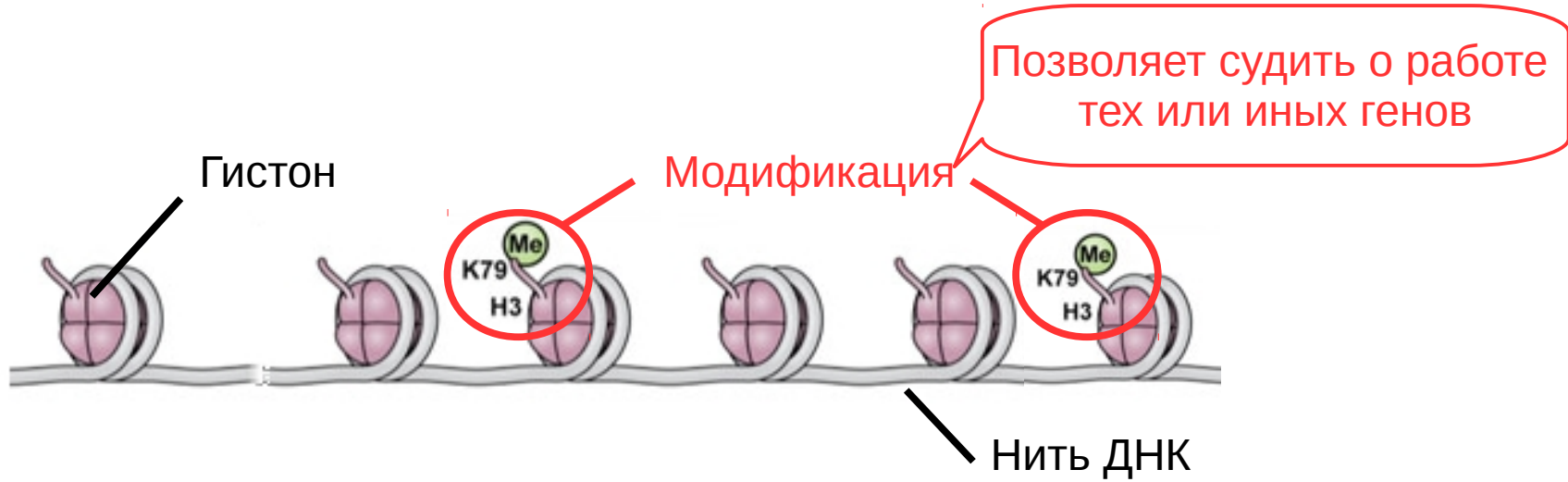
# Что такое ChIP-seq?



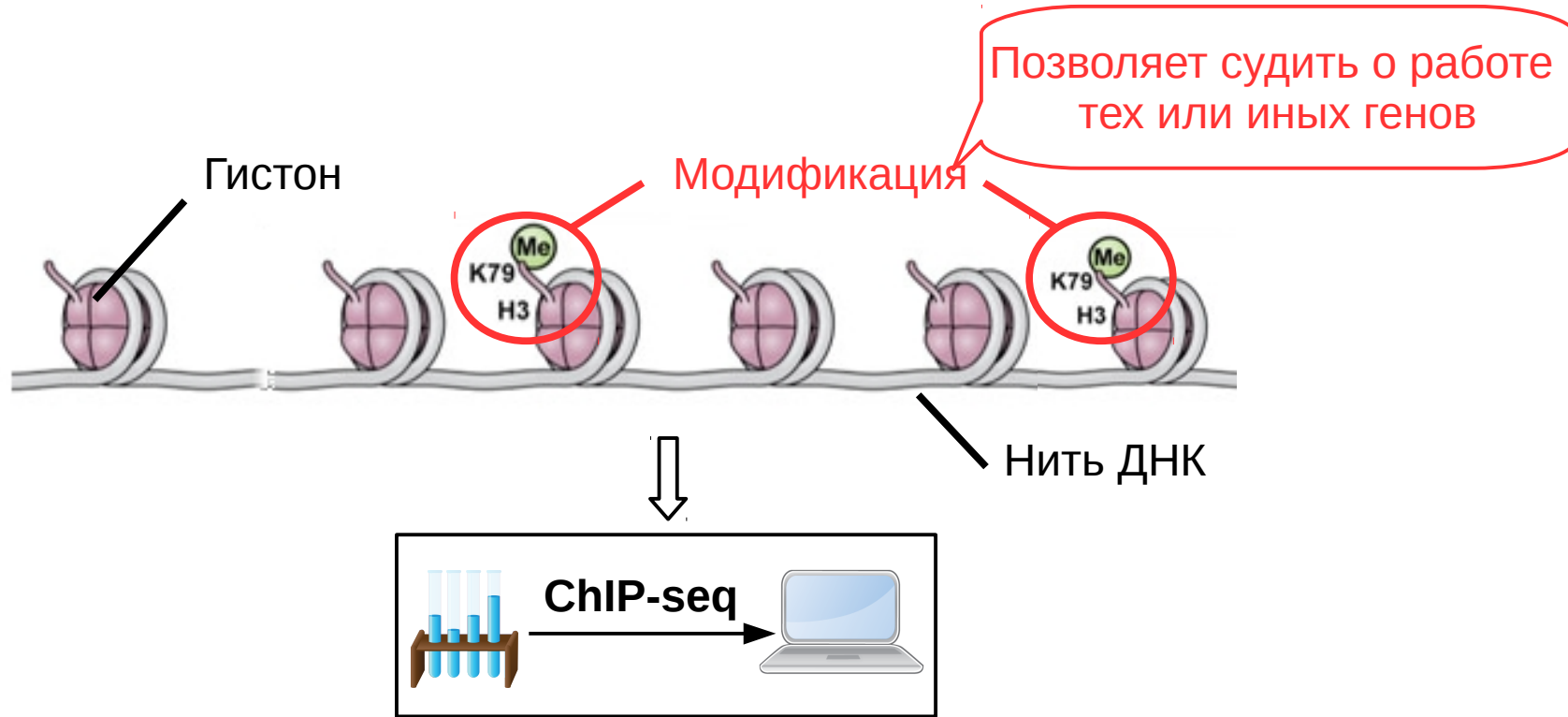
# Что такое ChIP-seq?



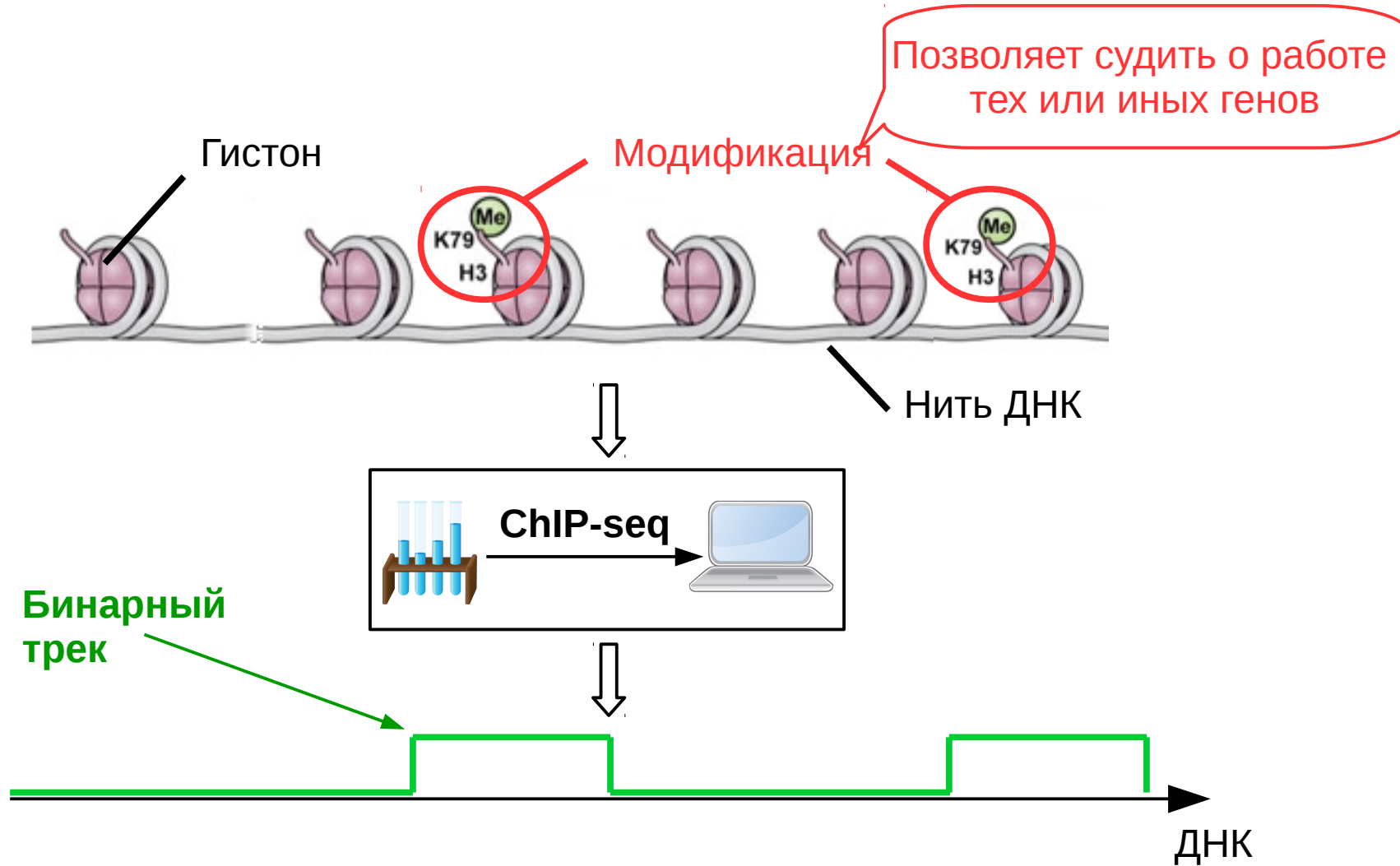
# Что такое ChIP-seq?



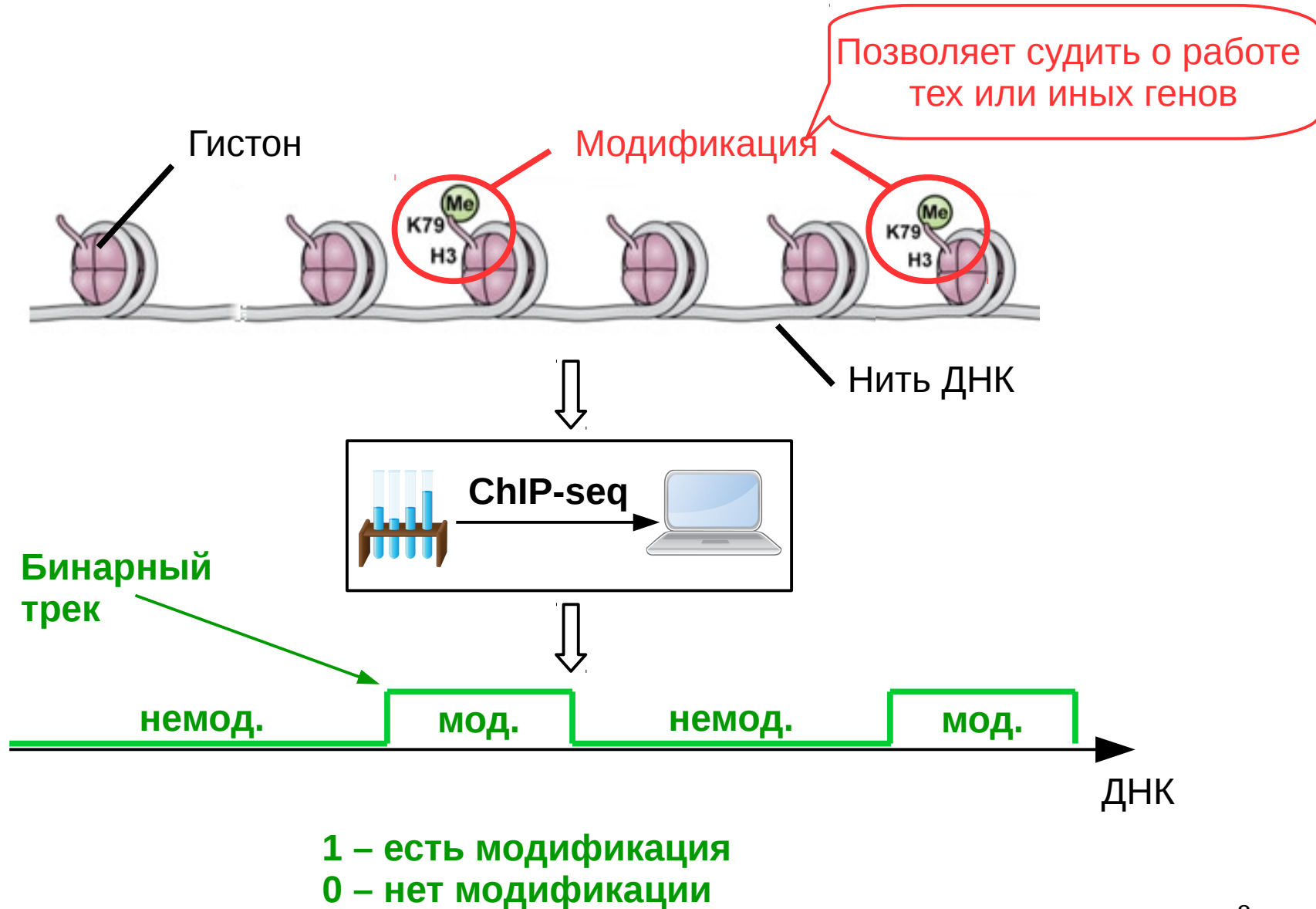
# Что такое ChIP-seq?



# Что такое ChIP-seq?



# Что такое ChIP-seq?





# Таким образом...

- В результате нового ChIP-seq эксперимента формируется **набор бинарных треков**, сохраненных в формате *.bed*

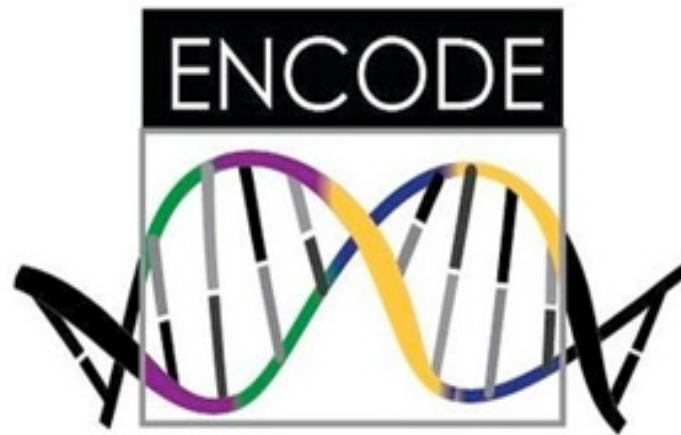
# Таким образом...

- В результате нового ChIP-seq эксперимента формируется **набор бинарных треков**, сохраненных в формате *.bed*
- На сколько эти данные соотносятся с **предыдущими подобными экспериментами?**

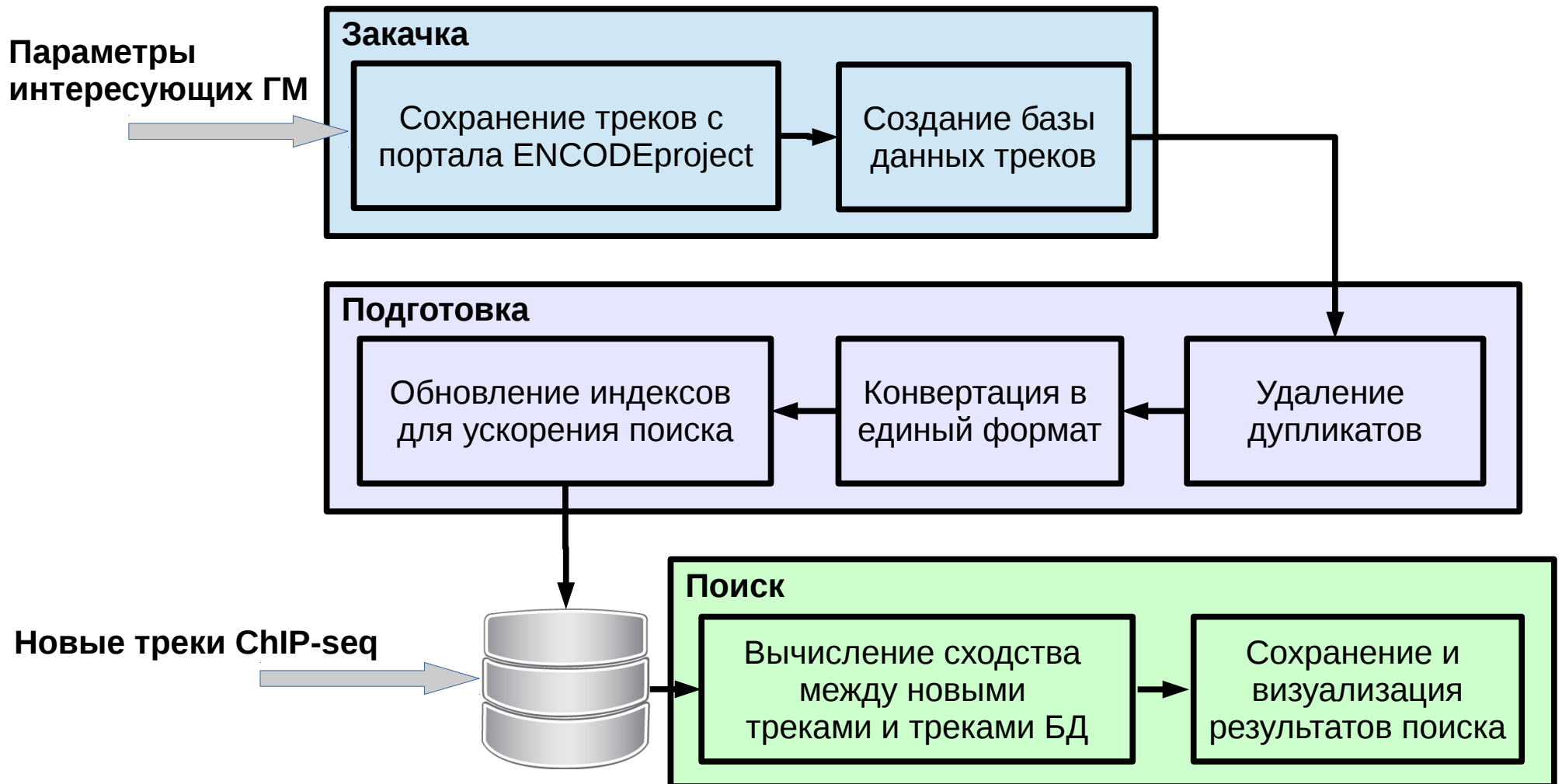
# Основной источник данных для сравнения: ENCODE-project

- ✓ ~ 5.5k треков ChIP-seq экспериментов
- ✓ Полноценные описания треков
- ✓ API для выгрузки данных

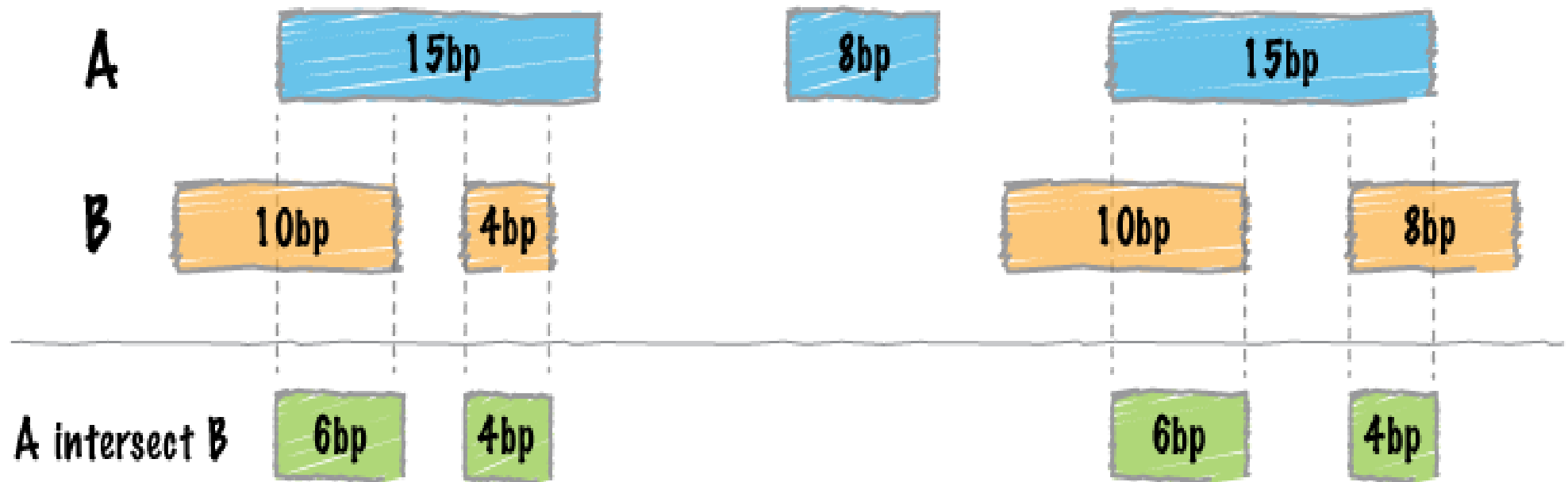
National Human Genome Research Institute



# ChIP-crawler: тул для сравнения НОВЫХ данных с данными ENCODE



# Основная мера сходства: индекс Жаккара

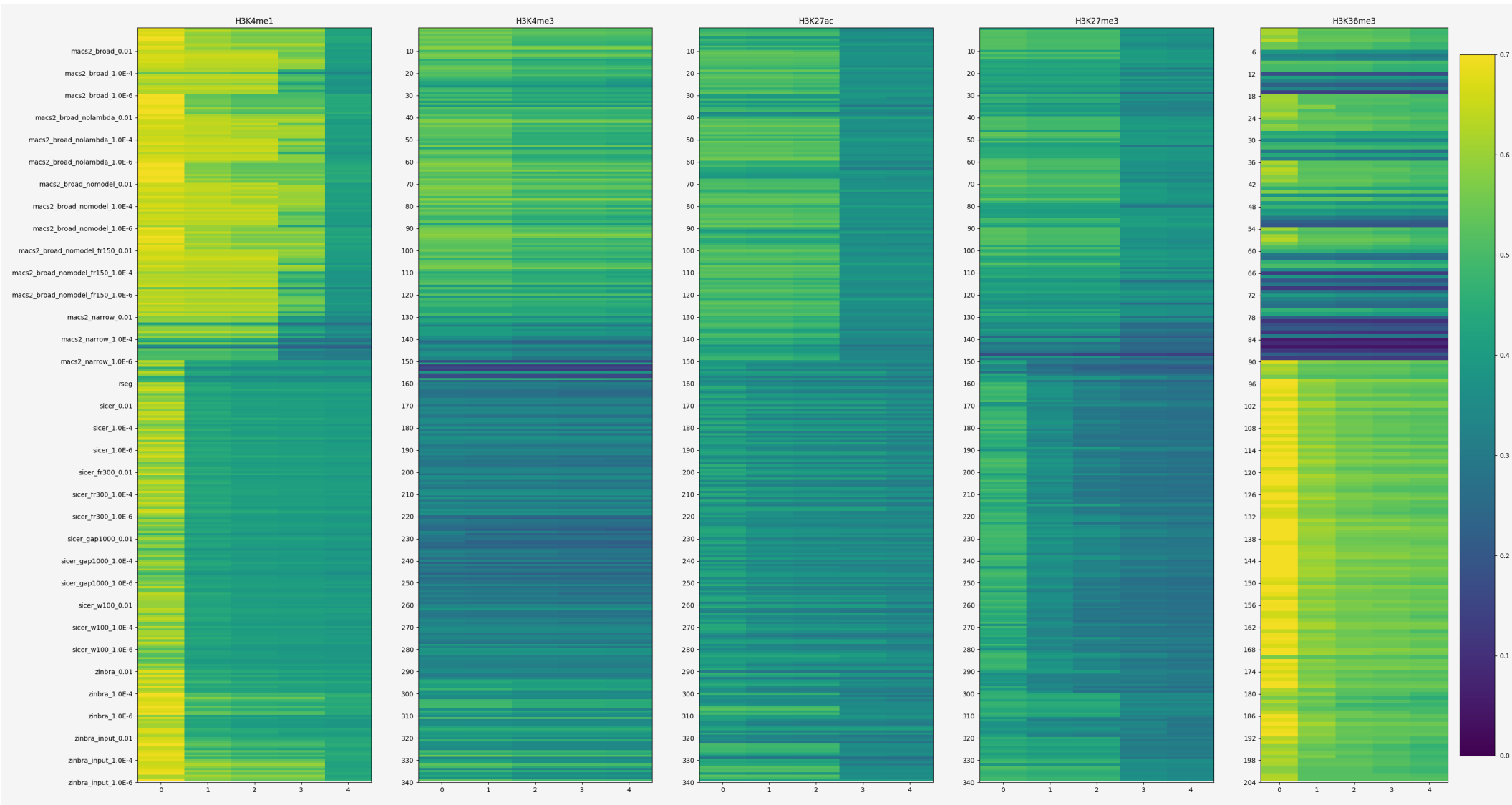


$$\text{Jaccard}(A,B) = \frac{\text{length(Intersection)}}{\text{length(Union)}} = \frac{6+4+6+4}{(15+8+15+10+4+10+8) - (6+4+6+4)} = \frac{20}{50} = 0.4$$

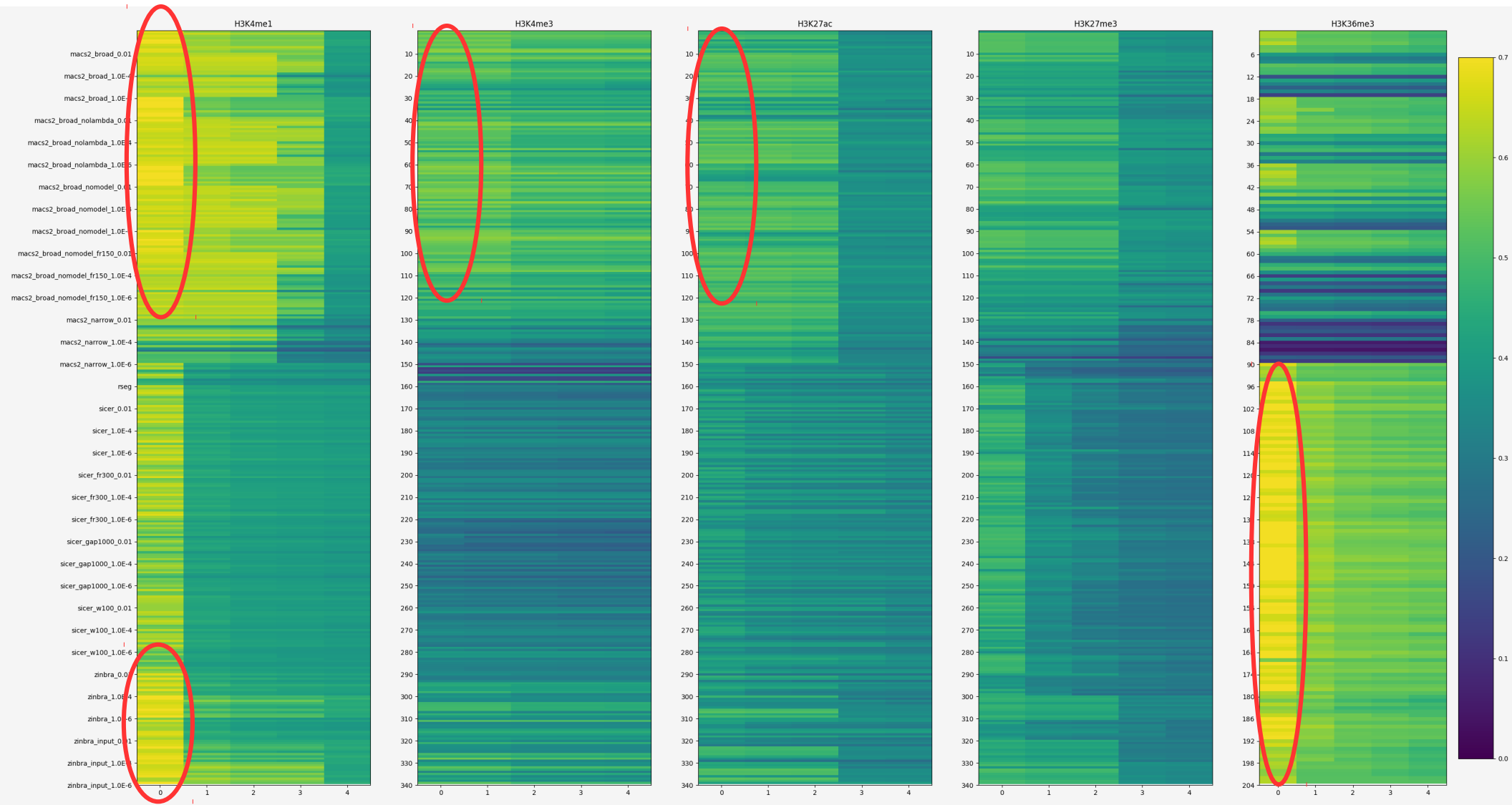
# Тестовый датасет (совместно с Wash-U)

- Исследовались эпигенетические маркеры старения в CD14 моноцитах
- Данные: 5 треков для молодых доноров и 5 треков для пожилых доноров
- Все 10 треков были обработаны 5-ю пикколерами с различными настройками (всего ~150 конфигураций)

# Результаты поиска (5 лучших совпадений)



# Результаты поиска (5 лучших совпадений)



**MACS broad, Zinbra**

**MACS broad**

**MACS broad**

**???**

**SICER, Zinbra**

**Best match: 0.73**

**0.58**

**0.56**

**0.55**

**0.75**



# Результаты поиска для одного трека

Category	Histone label	Closest accession	Description	Jaccard smilarity
Old	<b>H3K4me1</b>	<b>ENCFF658GEZ</b>	<b>H3K4me1</b> : CD14-positive monocyte male adult (21 year)	0.53
	H3K4me3	ENCFF280STO	H3K4me3: heart right ventricle male child (3 years)	0.35
	<b>H3K27ac</b>	<b>ENCFF658GEZ</b>	<b>H3K4me1</b> : CD14-positive monocyte male adult (21 year)	0.41
	H3K27me3	ENCFF592KAT	H3K27me3: CD14-positive monocyte male adult (21 year)	0.51
	H3K36me3	ENCFF060WSS	H3K36me3: CD14-positive monocyte male adult (21 year)	0.69
Young	<b>H3K4me1</b>	<b>ENCFF658GEZ</b>	<b>H3K4me1</b> : CD14-positive monocyte male adult (21 year)	0.58
	H3K4me3	ENCFF849MZZ	H3K4me3: iPS-15b female	0.34
	<b>H3K27ac</b>	<b>ENCFF658GEZ</b>	<b>H3K4me1</b> : CD14-positive monocyte male adult (21 year)	0.39
	H3K27me3	ENCFF592KAT	H3K27me3: CD14-positive monocyte male adult (21 year)	0.49
	<b>H3K36me3</b>	<b>ENCFF060WSS</b>	<b>H3K36me3: CD14-positive monocyte male adult (21 year)</b>	<b>0.73</b>



# Выводы

- Разработанный тул позволяет находить результаты, аналогичные полученным в предыдущих экспериментах
- Тул также может быть использован для сравнения различных колеров
- Необходимо добавить поддержку дифференциального ChIP-seq
- Представляет интерес создание открытого веб-сервиса на основе разработанного тула

**Спасибо за внимание!**