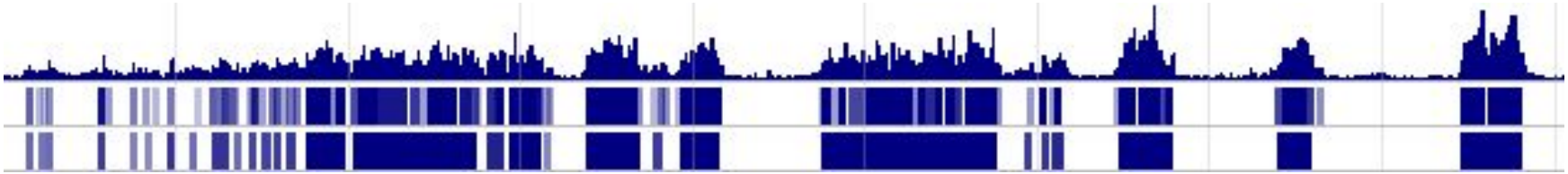
Noisy peak calling

Чаплыгина Дарья
Руководитель: Шпынов Олег
JetBrains Research

Peak calling

ChIP-seq -> peaks & noise



Impact of sequencing depth in ChIP-seq experiments

Impact of sequencing depth in ChIP-seq experiments

Youngsook L. Jung, Lovelace J. Luquette, Joshua W.K. Ho, Francesco Ferrari, Michael Tolstorukov, Aki Minoda, Robbyn Issner, Charles B. Epstein, Gary H. Karpen, Mitzi I. Kuroda, ... Show more

Nucleic Acids Research, Volume 42, Issue 9, 14 May 2014, Pages e74,
https://doi.org/10.1093/nar/gku178
Published: 05 March 2014 **Article history** ▾

- 20M прочтений -> алгоритмы работают по-разному
- 100M прочтений -> получаются похожие результаты
- зависимость и от биологических особенностей, и от качества данных

План

- Изучить протокол ChIP-Seq
- Разобраться со статьей
- Проанализировать влияние качества эксперимента на результаты работы алгоритмов MACS2, SICER, SPAN
- Обобщить на ATAC-Seq

@JetBrains. All rights reserved

План

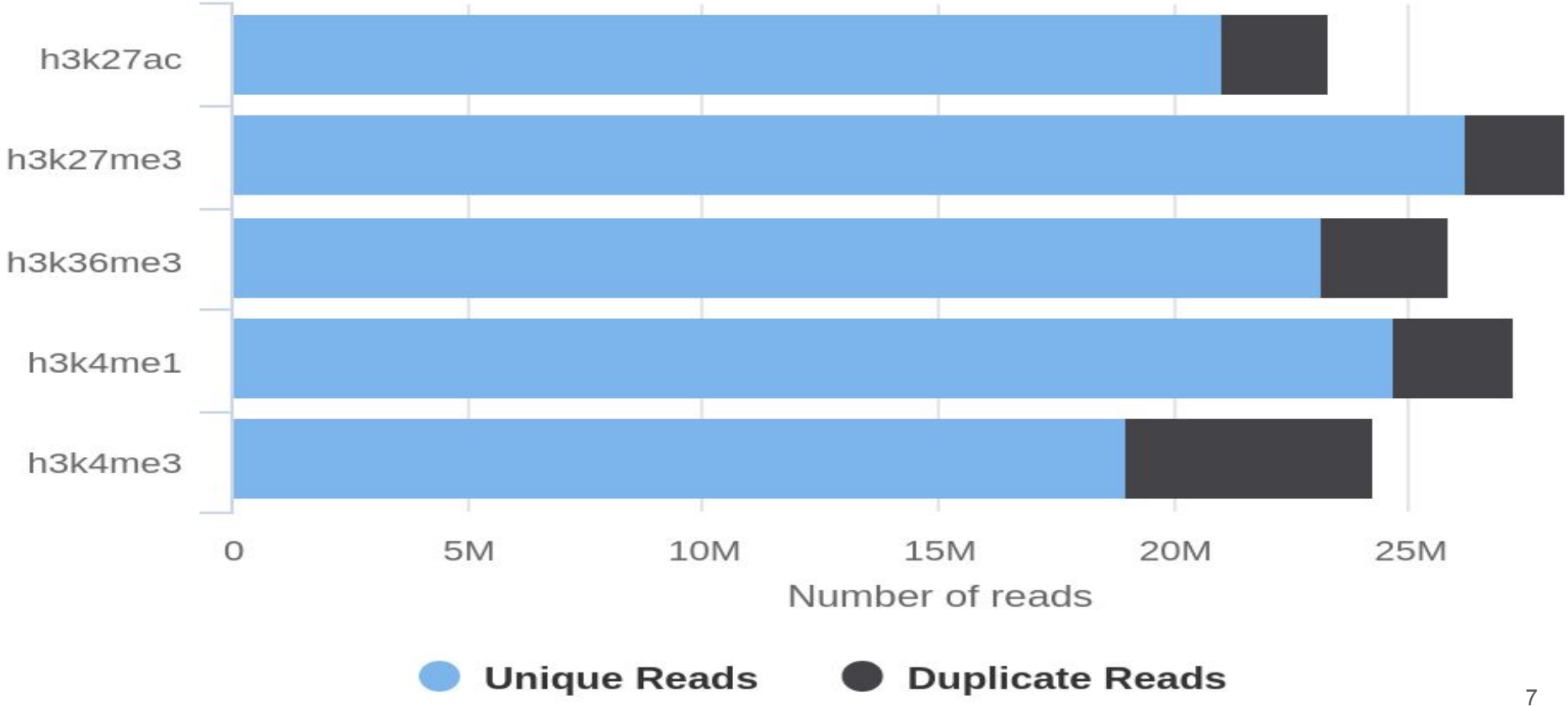
- Изучить протокол ChIP-Seq
- Разобраться со статьей
- Проанализировать влияние качества эксперимента на результаты работы алгоритмов MACS2, SICER, SPAN
- Обобщить на ATAC-Seq

@JetBrains. All rights reserved

Data for analysis

- **H3K4me1** (ENCODE accession number: ENCFF076WOE)
- **H3K4me3** (ENCODE accession number: ENCFF001FYS)
- **H3K27ac** (ENCODE accession number: ENCFF000CEN)
- **H3K27me3** (ENCODE accession number: ENCFF001FYR)
- **H3K36me3** (ENCODE accession number: ENCFF000CFB)

Sequence Counts (ChIP)



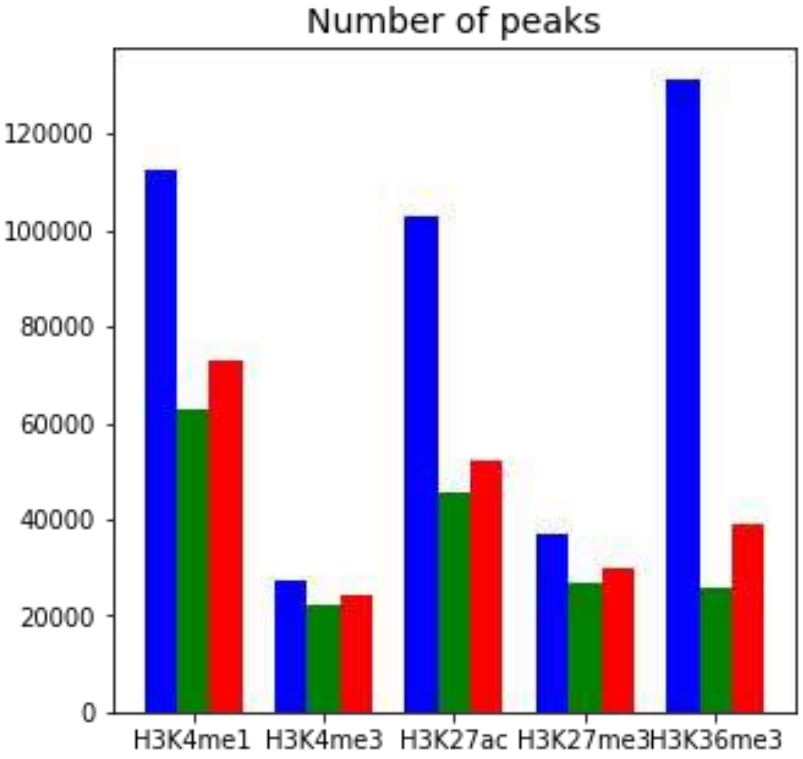
SAM / BAM

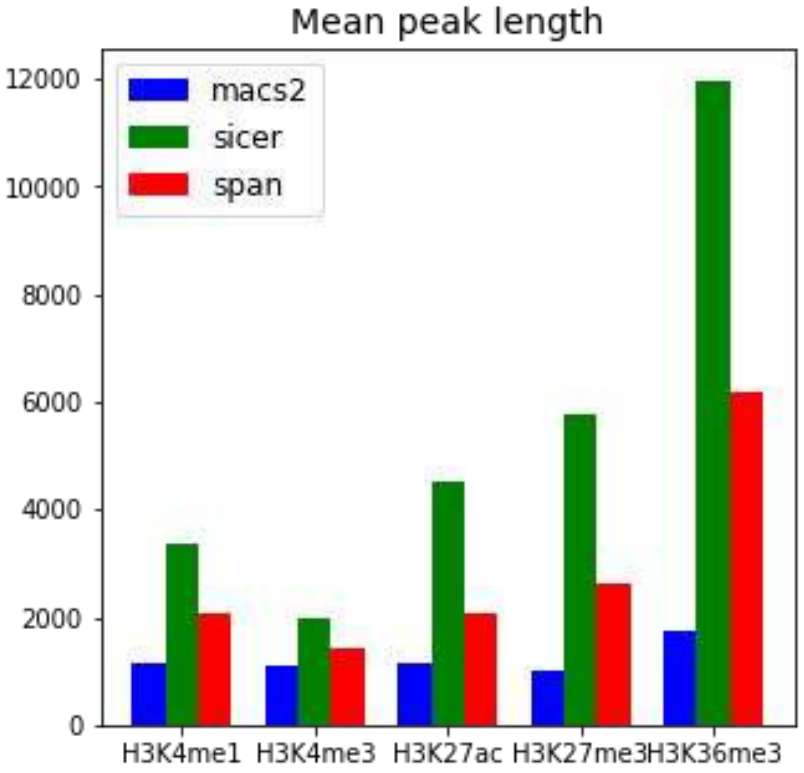
- выравнивание прочтений: **bowtie2**
- сортировка и индексирование файла: **samtools**

BW & BED

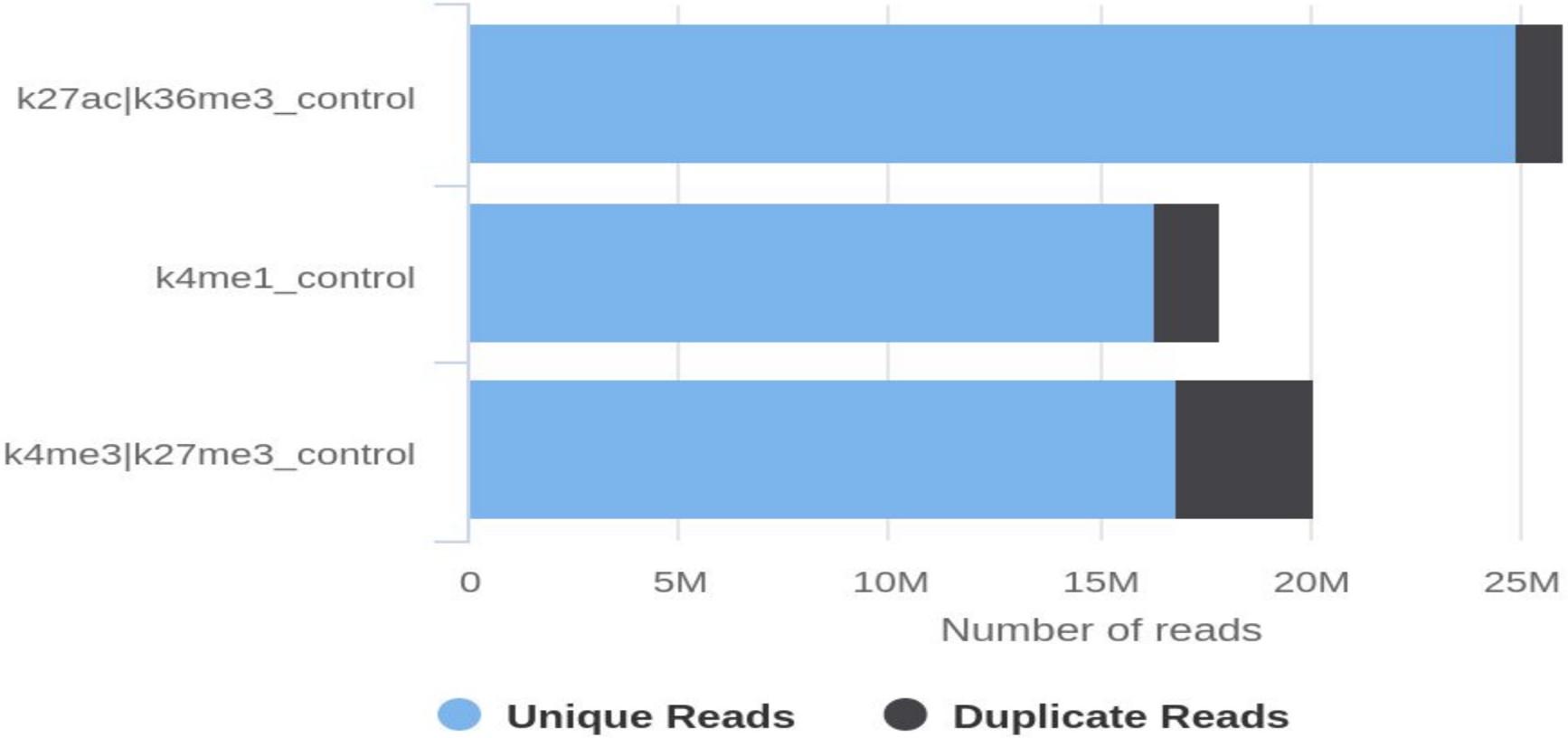
- Получение bigWig файла: **bamCoverage ; deeptools package**
- Peak calling: **macs2, span, sicer** (FDR: 1e-06)

Peaks statistics





Sequence Counts (control)



Искусственное зашумление данных

noize

chip-seq reads

70%

17.5M reads

Искусственное зашумление данных

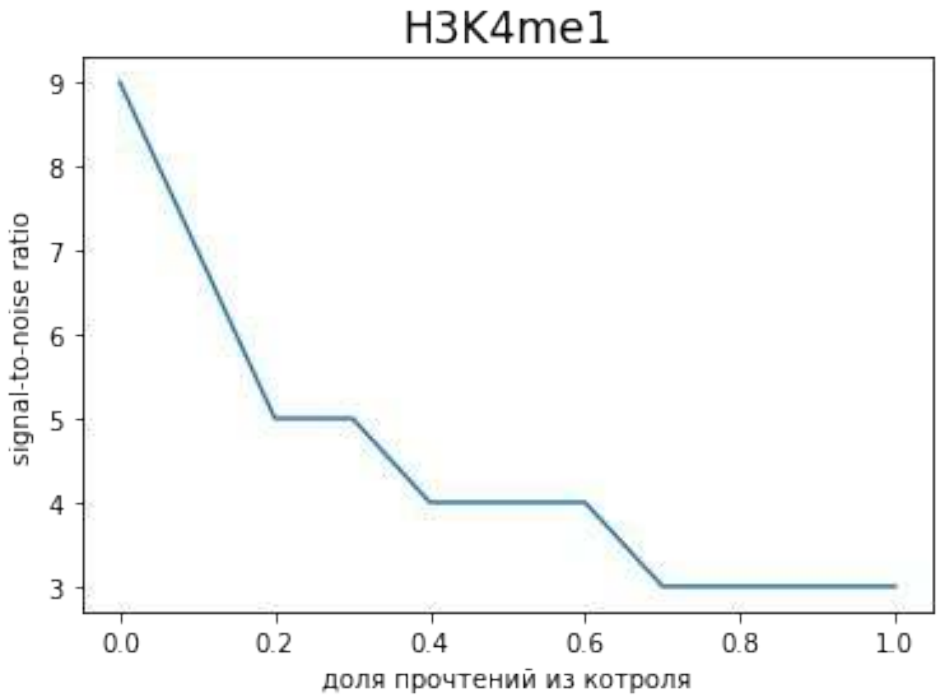
noize

chip-seq reads

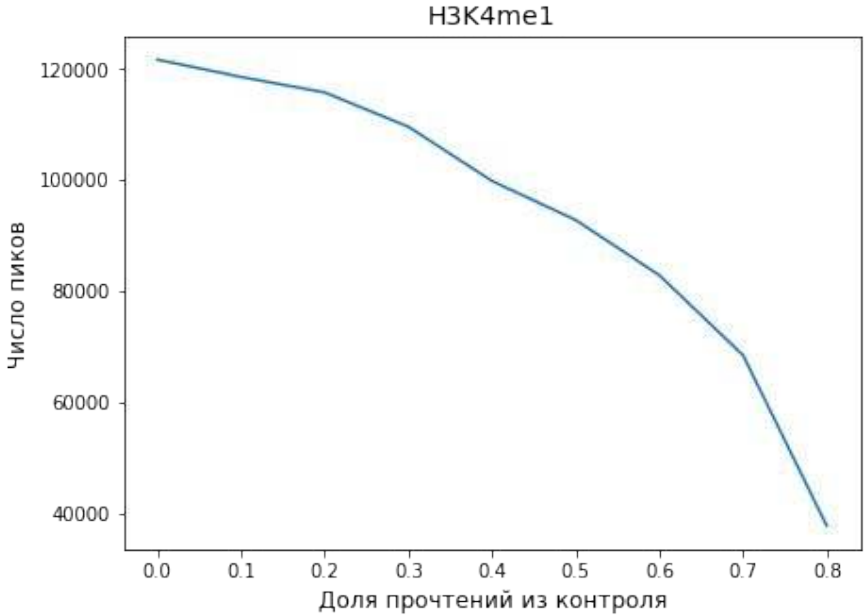
30%

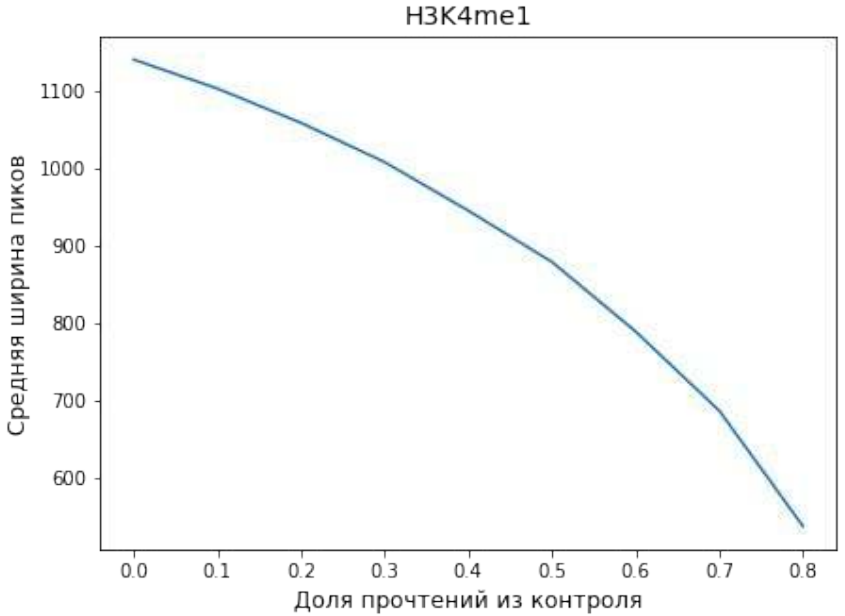
17.5M reads

SNR - отношение сигнал/шум



Изменение пиков в зависимости от пропорции шума





Спасибо за внимание!

Signal-to-noise ratio estimation (запасной слайд 1)

- сдвинуть прочтения на **fragment_size** / 2
- посчитать агрегированное покрытие в участках длины **bin_size**
- найти отношение **90** к **10** перцентили в распределении покрытия

Fragment sizes (запасной слайд 2)

	Estimated fragment size
H3K4me1	227
H3K4me3	189
H3K27ac	278
H3K27me3	222
H3K36me3	293