

Deep learning to the rescue!

Дарья Балашова, Институт Биоинформатики
Научный руководитель: Олег Шпынов, JetBrains Research

План

Что такое ChIP-seq?

Ultra-Low-Input ChIP-Seq

Снижение уровня шума в ChIP-seq

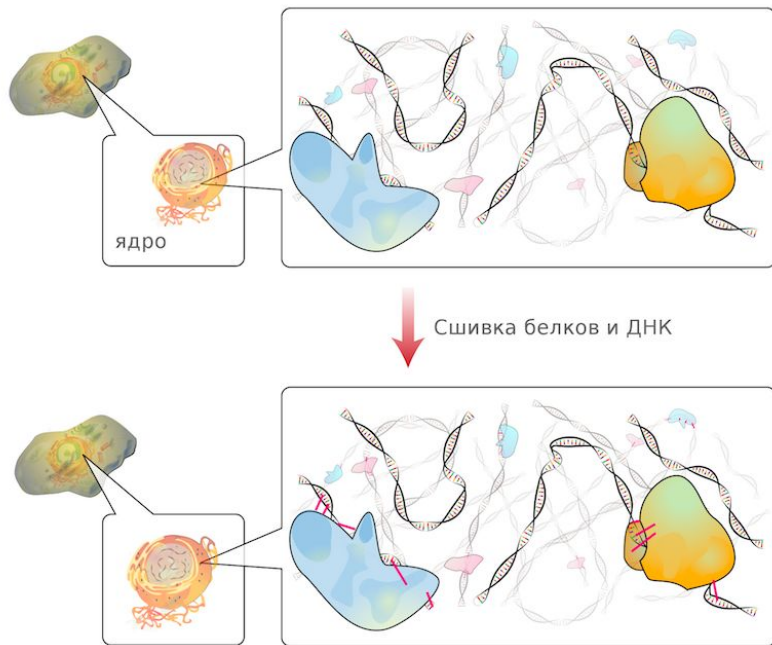
Оценка качества данных

Метрика качества данных

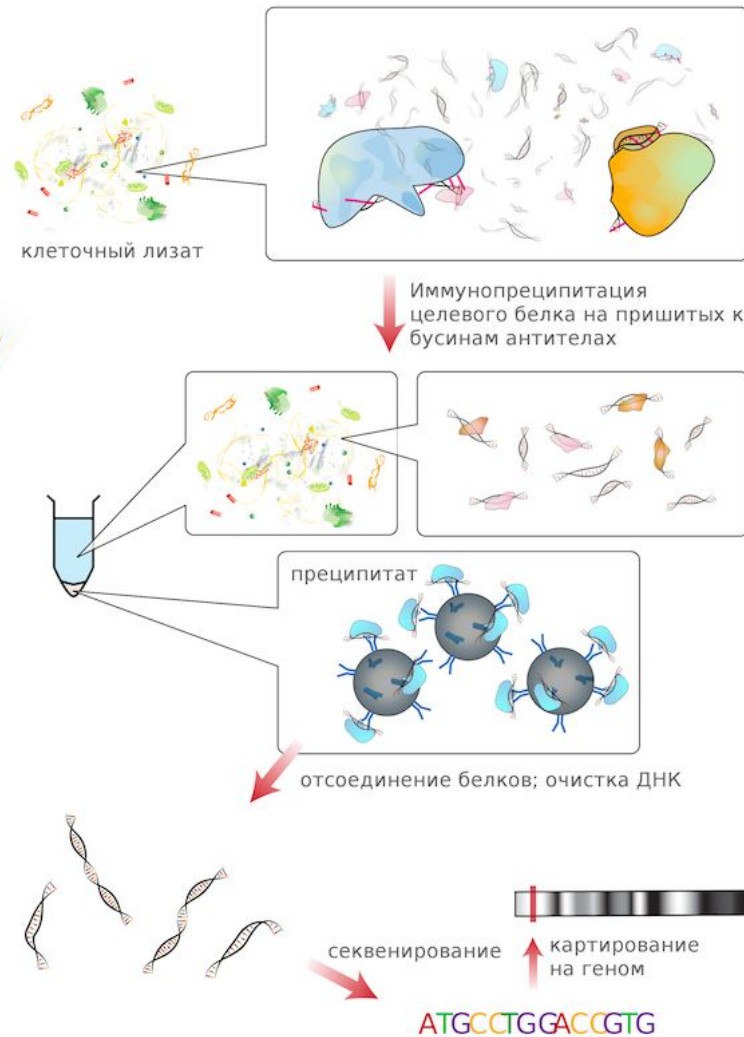
Нейронная сеть

Результаты

Что такое ChIP-seq?



Разрезание ДНК
ультразвуком



Ultra-Low-Input ChIP-Seq

Стандартный протокол **ChIP-seq** требует 1-5 миллионов клеток (или больше): серьезное ограничение для исследований (человеческие образцы, редкие популяции и т. д.)

Ultra-low-input ChIP-seq позволяет использовать до 100 000 клеток на образец. Недостатки:

- более высокий уровень шума
- меняющееся отношение сигнал/шум между образцами в течение одной подготовки
- сложности в последующей обработке

Снижение уровня шума в ChIP-seq

Denoising Genome-wide Histone ChIP-seq with Convolutional Neural Networks

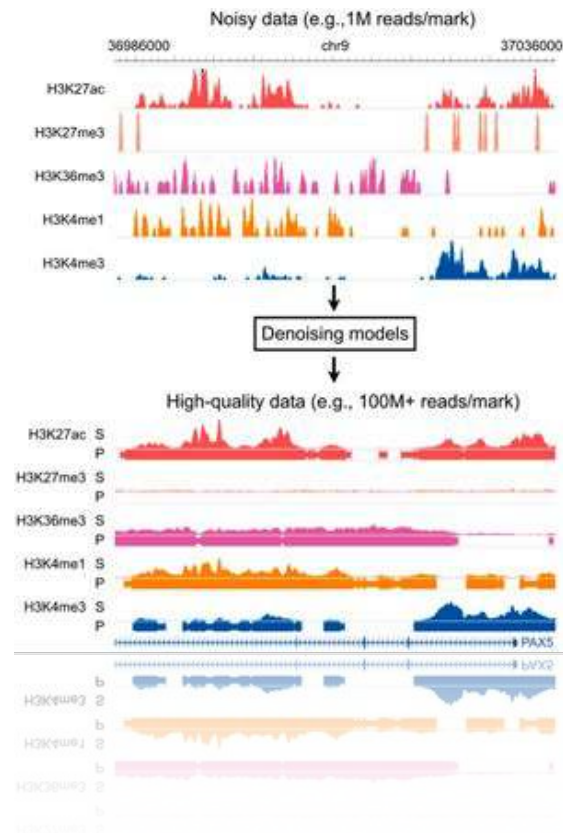
Pang Wei Koh^{1,2,†}, Emma Pierson^{1,†} and Anshul Kundaje^{1,2,*}

¹Department of Computer Science, Stanford University, Stanford, 94305, USA

²Department of Genetics, Stanford University, Stanford, 94305, USA.

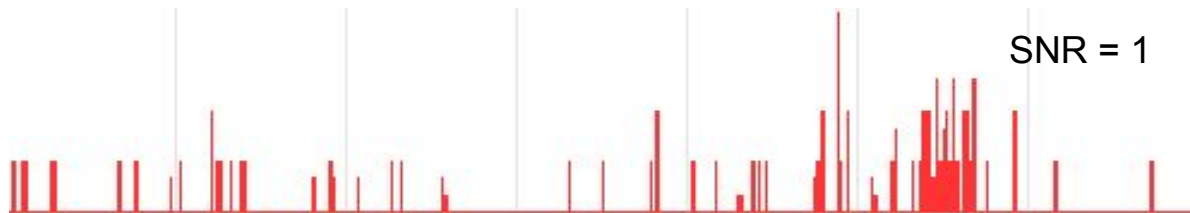
†These authors contributed equally to this work.

*To whom correspondence should be addressed.



Метрика качества данных (SNR)

SNR = квантиль 0.9 : квантиль 0.1

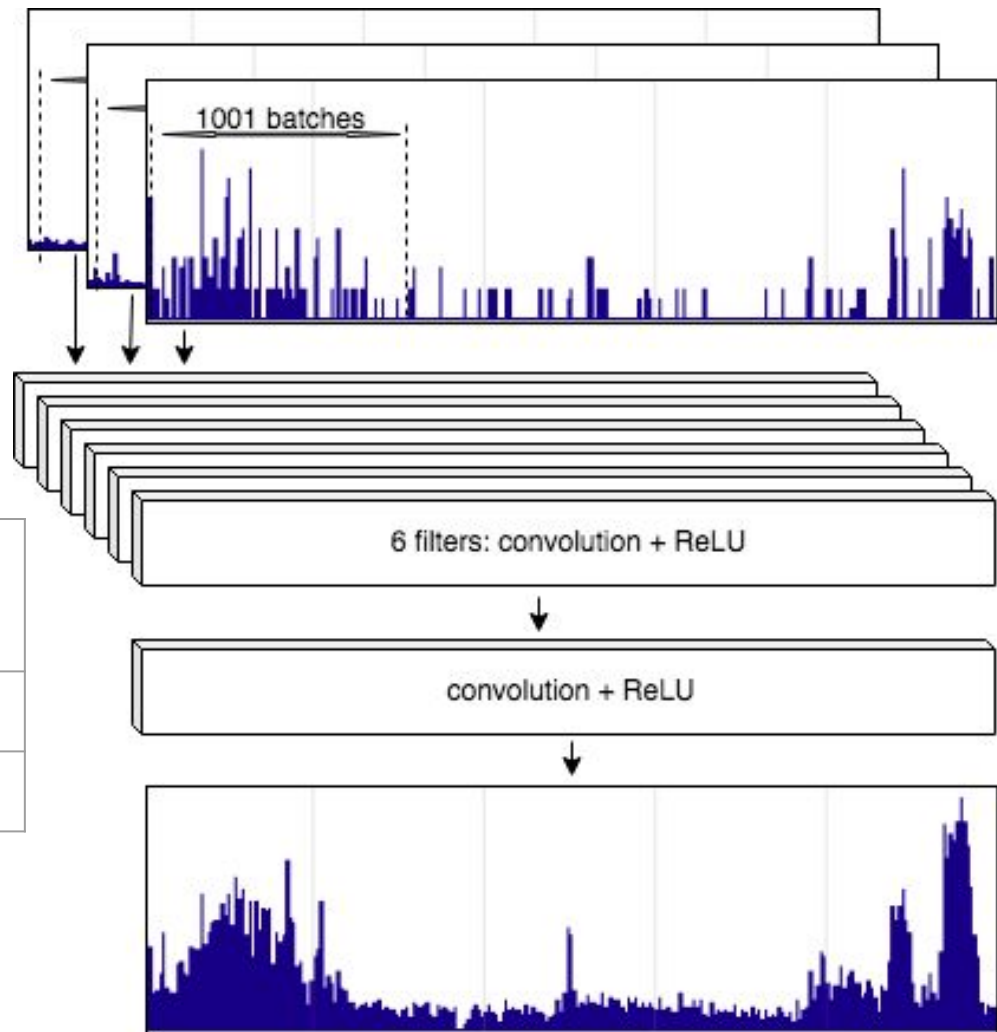


Нейронная сеть

<https://github.com/dashabalashova/DCNN>

Цель: улучшить качество данных
Ultra-low-input ChIP-seq

	низкое качество	результат применения CNN
H3K4me3	SNR = 3.0	4.19
H3K36me3	3.0	5.65



Результаты

- изучены протоколы ChIP-Seq и ULI ChIP-Seq
- рассмотрена оценка signal-to-noise ratio
- построена нейросеть, улучшено качество ULI ChIP-Seq на реальном эксперименте

Спасибо за внимание!