Imputing chromatin state from GRO-seq data

Ulugbek Baymuradov - Elena Tuzhilina - Rafael Rafailov - Quinton Wessells

Motivation

- Chromatin landscapes provide critical insight into the transcriptional regulation of the genome
- Current methods involve multiple high-throughput genomic assays
  - ChIP-seq for each histone modification, DNase-seq etc.
- Profiling chromatin landscape from a single, cost-effective assay would be extremely valuable

Goal: Impute histone modification ChIP-seq and DNase seq profiles from GRO-seq data

Dataset

Cell types: GM12878 and K562

Data: GRO-seq, H3K4me3, H3K27ac, H3K27me3, DNase-seq

Binning: 10bp, 50bp, 100bp

Labels: generated using bigWig fold change in peak regions

Model

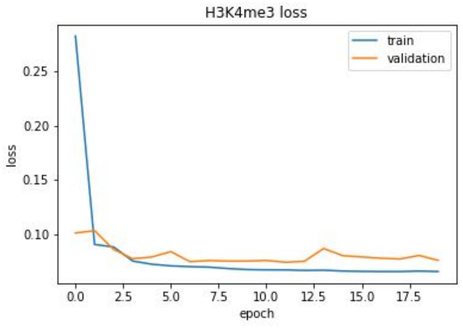
Convolutional neural net with dilated convolutions

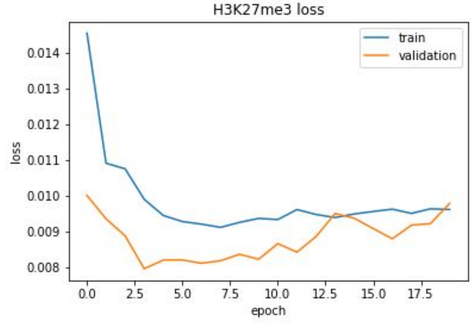
To manage the size of the receptive field we explored:

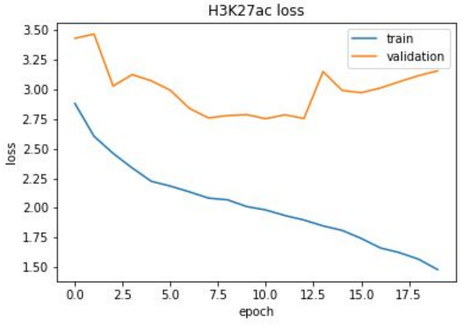
1. Using filters of different lengths
2. Deploying various levels of convolutional dilation
3. Adding additional convolutional layers
4. Adding skip connections to deeper networks
5. Used varied loss functions

Trained on chr1 through chr5; Validated on chr14

Results

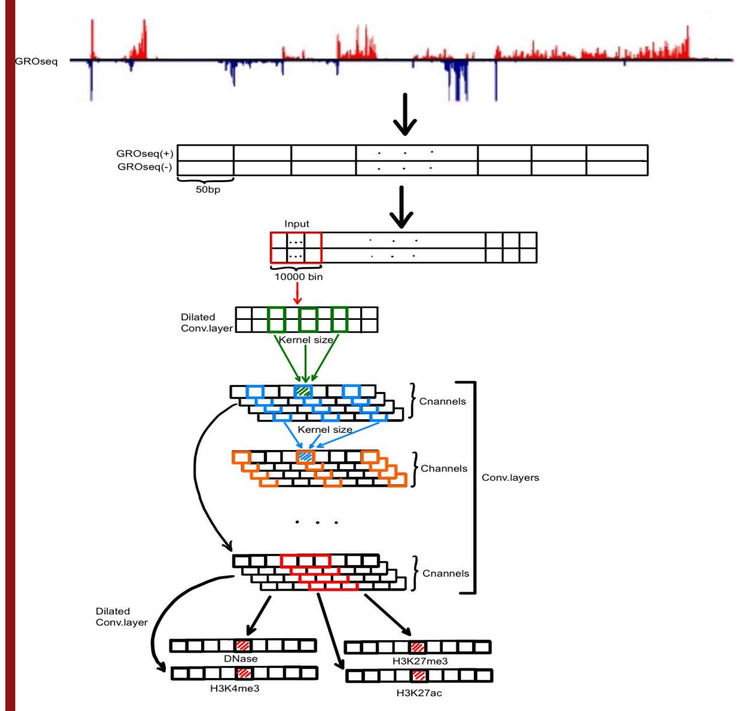






Possible interpretation: Our model fits the training data to some degree, but unfortunately does not learn transferable features between chromosomes.

Deep Learning Architectures



Discussion & Future Work

Either the hypothesis on imputing chromatin landscape from GRO-seq data is incorrect or our current models and/or data processing didn't capture the necessary information

Future work:

- Extend training with additional chromosomes
- Alternative data processing - include sequence data
- Different architectures and hyperparameters

References

[1] Kelley DR, Snoek J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res. 26(7), 990–999 (2016).

[2] Kelley DR, Reshef Y, Bileschi M, Belanger D, McLean CY, Snoek J. Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res. 2018 Mar.

[3] Danko, C.G. et al. (2015) Identification of active transcriptional regulatory elements from GRO-seq data. Nat. Methods 12, 433–438

[4] Danko, C.G. et al. (2018) Identification of regulatory elements from nascent transcription using dREG. BioRxiv.