Hi-C-LSTM: Learning representations of chromatin contacts using a recurrent neural network identifies genomic drivers of 3D genome conformation

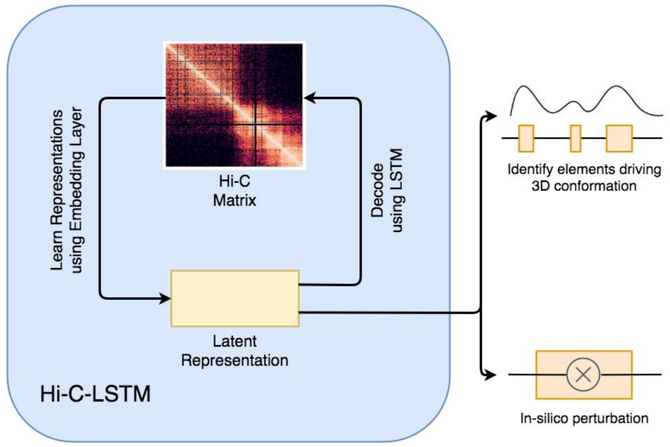
SFU SCHOOL OF COMPUTING SCIENCE

UBC Electrical and Computer Engineering

MRC London Institute of Medical Sciences

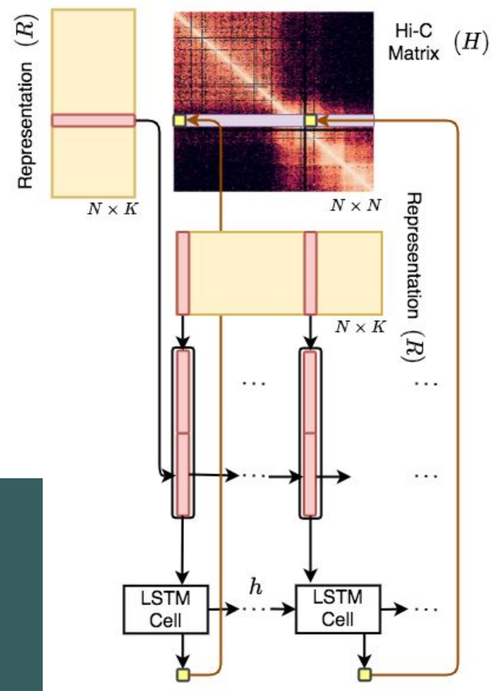
K. B. Dsouza, A. Maslova, E. Al-Jibury, M. Merkenschlager, V. K. Bhargava, M. W. Libbrecht

Need for Hi-C Representations



Goal: Understand of how the 1D genome influences 3D conformation

Hi-C-LSTM

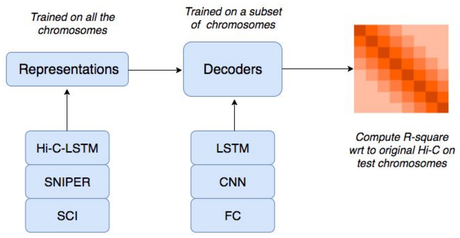


Salient Features of Hi-C-LSTM

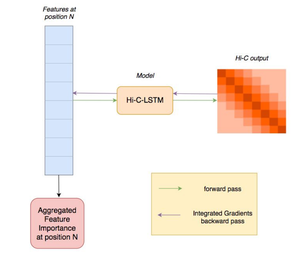
- Works on intrachromosomal contacts in contrast to most existing methods, which were designed for interchromosomal contacts
- Takes into account the sequential nature of the genome
- Representations have information needed to recreate the Hi-C matrix and recreation is more accurate using an LSTM as the decoder
- Captures a variety of genomic phenomena and distinguishes genomic regions that are known to cluster in 3D space
- Feature importance reveals association with transcription factors such as CTCF that are known to mediate chromatin conformation
- Enables us to perform in-silico knockout/insertion experiments as it behaves as a framework for contact-generation
- It can model only cis effects because trans-acting cellular machinery is captured within the Hi-C-LSTM decoder, which cannot be easily modified
- Recapitulates effects of known structural variants seen in certain developmental disorders

Evaluations

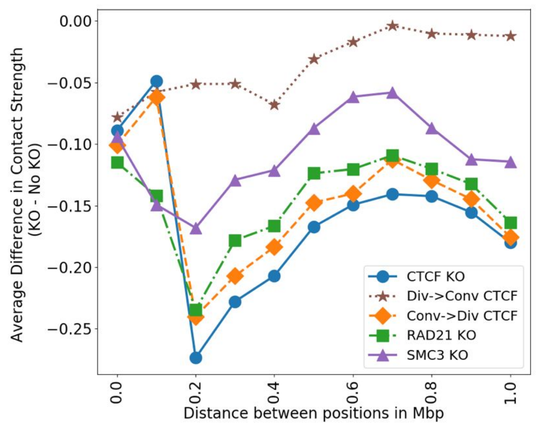
R-squared



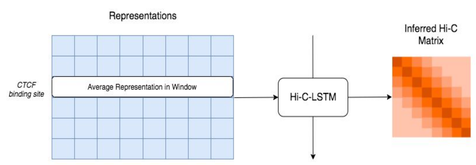
Feature Importance



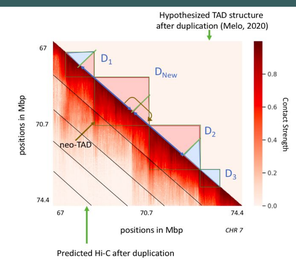
Hi-C-LSTM enables in-silico knockout experiments



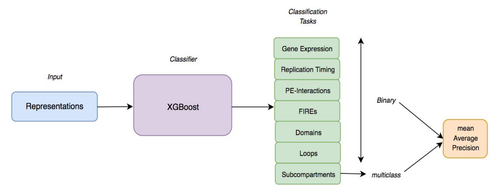
Knockout



Duplication

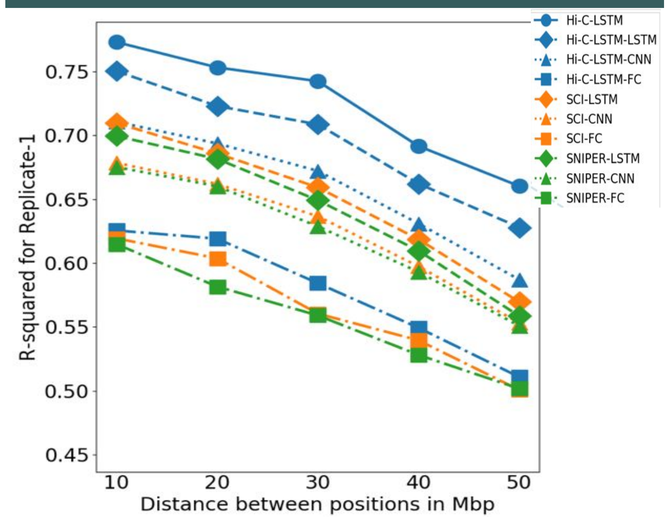


Classification

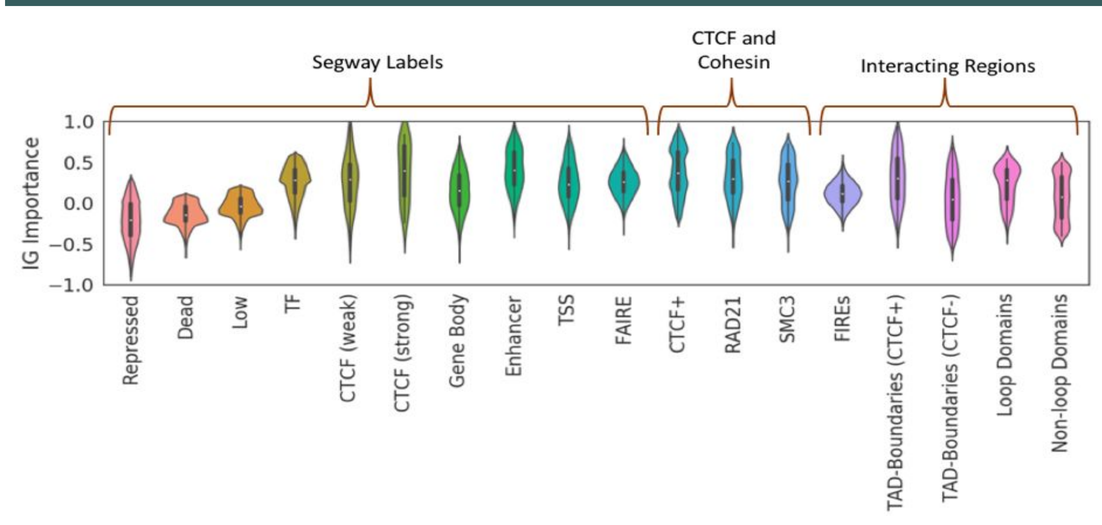


Results

Representations capture information needed to create the Hi-C-matrix



Feature Importance reveals association with genomic elements driving 3D conformation



Representations locate functional activity and regions that drive 3D conformation