Can Neural Networks be used to Predict Cross-Species Annotations of Chromatin Regulation?

Maillard Noémien, Klopp Christophe, Donnadieu Cécile, Riquet Juliette, Fève Katia, Julie Demars, Mourad Raphaël

HAL Id: hal-04797361
https://hal.inrae.fr/hal-04797361v1

Submitted on 22 Nov 2024

Can Neural Networks be used to Predict Cross-Species Annotations of Chromatin Regulation ?

Maillard Noémien[1], Klopp Christophe[3], Donnadieu Cécile[4], Riquet Juliette[1], Fève Katia[1], Demars Julie[1], Mourad Raphaël[2]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet Tolosan, France

[2]MIAT, F-31326, Castanet Tolosan, France

[3]SIGENAE, Plate-forme Bio-informatique Genotoul, Mathématiques et Informatique Appliquées de Toulouse, INRAe, 31326 Castanet-Tolosan, France

[4]INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

Corresponding Author: noemien.maillard@inrae.fr

Keywords

Annotations, cross-species, deep learning

Abstract

A better knowledge of functional characterization of livestock species seems a lever to link genome to phenome. However, data describing gene regulation mechanisms and chromatin state in various experimental conditions are lacking compared to common animal models. To overcome this bottleneck, predictive biology seems a good alternative. Human and mouse are organisms phylogenetically close to pig, so we can assume that molecular mechanisms are similar. In addition, they offer much more data that is a condition to train powerful deep learning algorithms. Here we propose to evaluate the cross-species adaptability of 3 neural networks: Deepbind[1], DeepSEA[2] and Enformer[3] to predict cross-species annotations (transcription factor binding, chromatin opening, histone marks).

Here, predictions have been computed on reference genomes of organisms of interest and compared to observations. Firstly, human-trained neural network predictions have been performed on mouse reference genome to test a high variety of experiments (multiple transcription factors ChIP-seq, DNase, histone ChIP-seq). Secondly, we tested whether neural networks are also able to predict non-mammalian species annotations. For this, two species of agronomical interest have been used: the pig

(as mammalian reference) and the chicken. Finally, we evaluated the impact of genomic features (repeats, CpG islands, …) on the predictions.

To conclude, the 3 neural networks evaluated show a good capacity to predict annotations on other organism genomes but seem limited to mammals when they are trained with mammalian data. They also show some variability over considered genomic features so this should be taken into account for further analysis of predictions, like variant impact. Finally, we would like to compare the predictions between pig breeds to see how much genomic diversity changes predicted annotations.

References

1. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 2015;33:831-8.

2. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. Nat Methods 2015;12:931-4.

3. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods 2021;18:1196-203.