Deep learning on single-cell ATAC-seq data to decipher enhancer logic

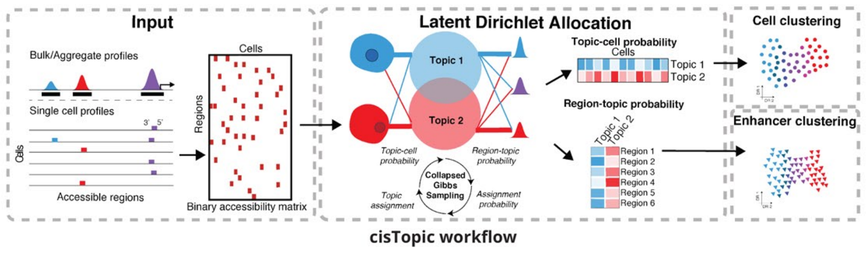
Ibrahim Ihsan Taskiran[1,2,], Liesbeth Minnoye[1,2], Zeynep Kalender Atak[1,2], Gert Hulselmans[1,2], Valerie Christiaens[1,2], and Stein Aerts[1,2]*

1. VIB Center for Brain & Disease Research, Leuven, Belgium
2. KU Leuven, Department of Human Genetics, Leuven, Belgium
* ibrahimihsan.taskiran@kuleuven.vib.be

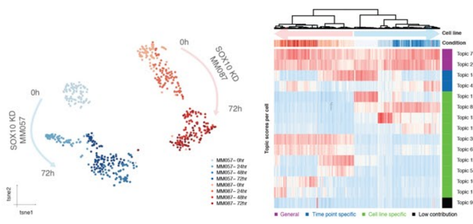
Deep Learning together with cisTopic on single-cell and bulk ATAC-seq on melanoma lines reveals melanoma enhancer logic

Single-cell ATAC-seq provides new opportunities to study gene regulation in heterogeneous cell populations such as complex tissues or dynamic processes. We recently developed a topic modeling approach to analyze bulk and single-cell ATAC-seq data, called **cisTopic**.

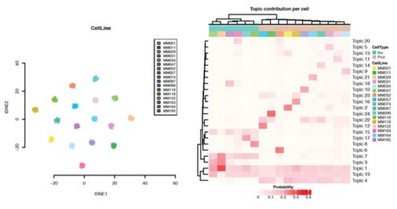
Here, we apply **deep learning** approaches to analyze sets of co-regulated enhancers found by topic modeling; using only the DNA sequence of the enhancers as input. We trained an Artificial Neural Network that is composed of Convolutional, Recurrent, and Dense layers.



cisTopic workflow



scATAC-seq on Melanoma Phenotype Switching



ATAC-seq on Personal Melanoma Genomes

Melanocyte-like cell state
SOX10 high
MITF high

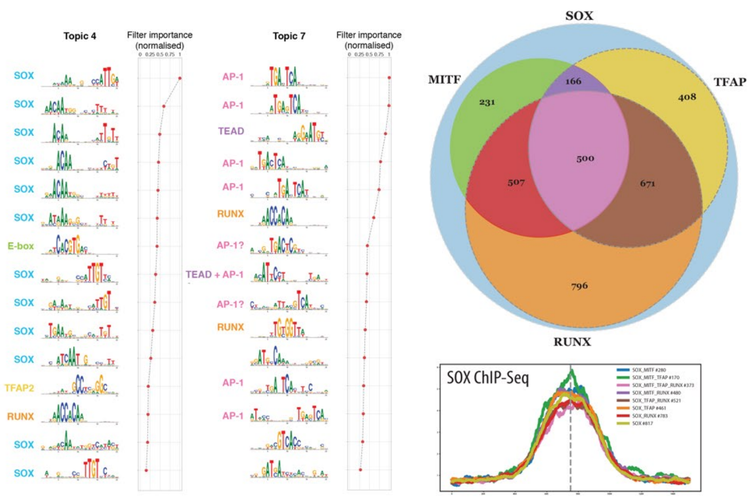
Mesencymal-like cell state
AP1 high
TEAD high

Melanoma is a type of skin cancer and malignant melanoma development is characterized by a bistable switch between different states called "**phenotype switching**". Melanocyte-like state is regulated by **MITF** and **SOX10**, while Mesenchymal-like state is regulated by **AP-1** and **TEAD** as the major transcription factors.

In this project, we study the enhancer logic of this state transition by using both bulk and single-cell ATAC-seq.

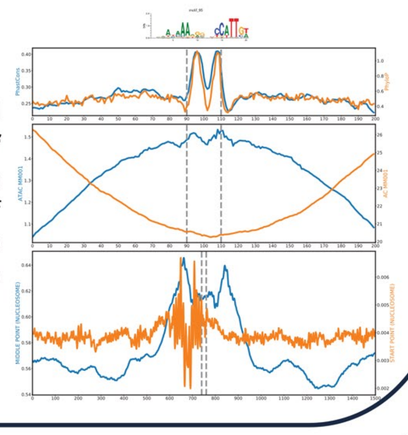
First layer of the model identifies Core Regulatory Complex

By applying this approach to a cohort of melanoma patient samples, we show that **key melanoma transcription factors** can be identified from the convolutional filters.



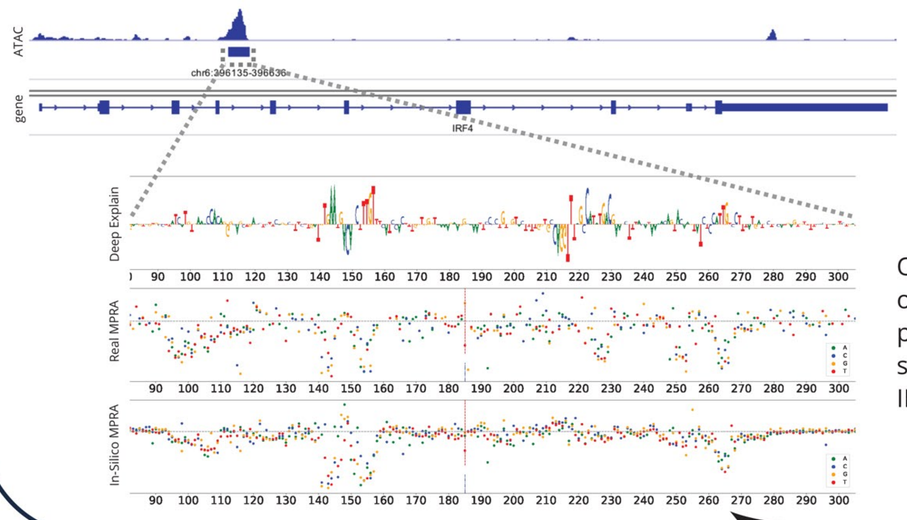
Combination of motifs composes **Core Regulatory Complex**

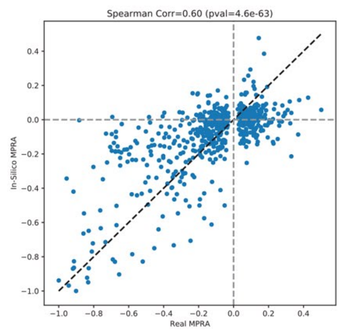
Next, we use the trained model to analyze the **motif architecture** in melanoma enhancers, such as motif combinations, binding site locations, and relationship to nucleosome preferences.



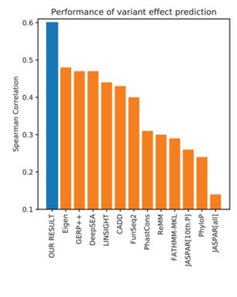
The code of disease related enhancers is revealed by the model and validated by MPRAs

To validate our models, we tested synthetic cell state specific enhancers using publicly available* and in-house **massively parallel enhancer reporter assays**.



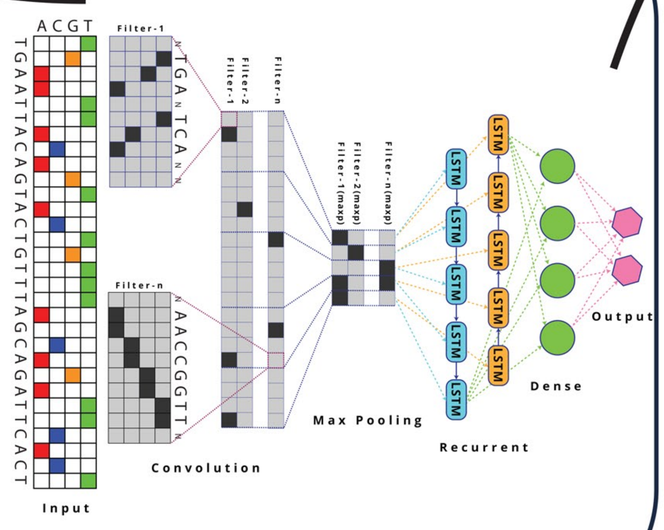


Our method **outperformed** other methods in terms of predicting the effect of a single nucleotide change on IRF4 enhancer



*Kircher et al., 2018

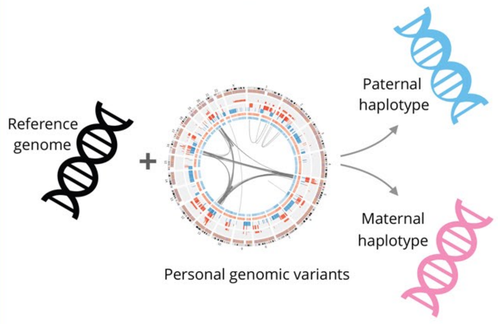
The model uses DNA sequence of the enhancers as input



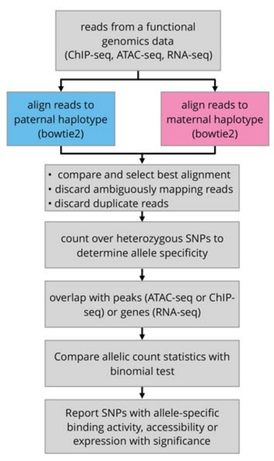
*Quang and Xie, 2016

Variants identified from Personalized Cancer Genomes are exploited by the model

We furthermore exploit network explaining methods* to predict the impact of **personal variants**, using patient matched whole genome-sequencing.



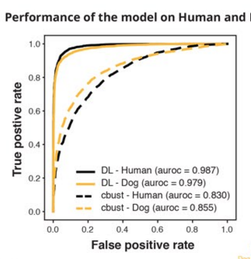
Personalized reference genomes are constructed using crossstich for each of the melanoma lines.

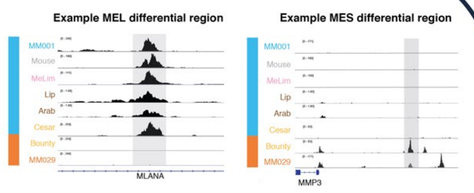


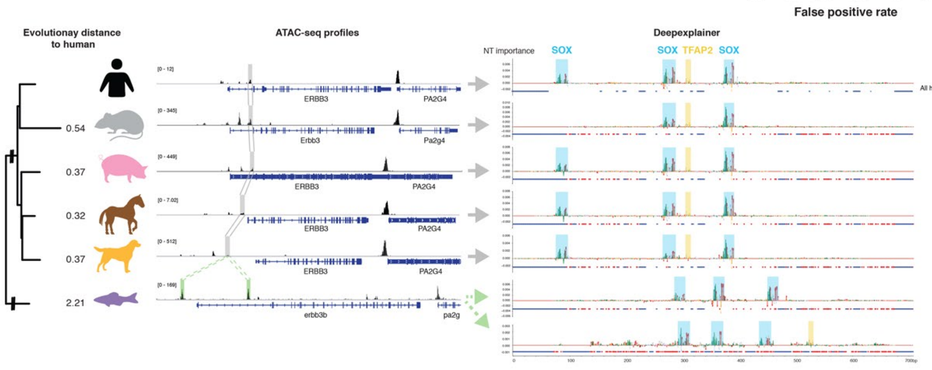
Pipeline to find impact of SNPs

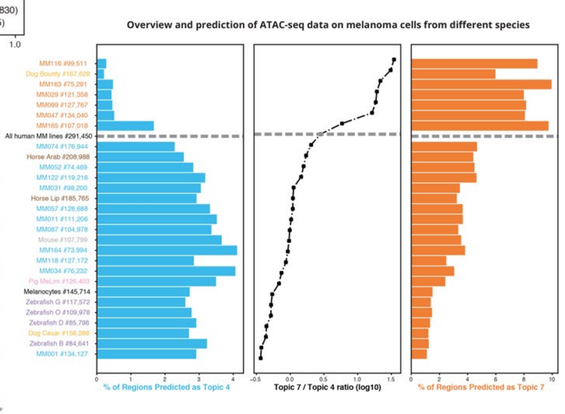
Cross-species analysis validates the model and uncovers conserved enhancer code

We use a cohort of in-house generated ATAC-seq on melanoma cells from different species including **human**, **mouse**, **pig**, **horse**, **dog**, and **zebrafish**.
Network explaining method applied on our model identifies conserved core elements of the same enhancer in different species.

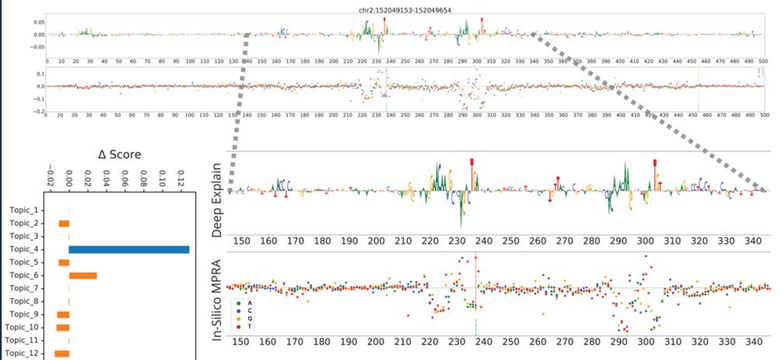








We thank Ghanem-Elias Ghanem for human, Karras Panagiotis and Chris Marine for mouse, Catherine André and Aline Primot for dog, Monika Seltenhammer and Leif Andersson for horse, Giorgia Egidy Maskos for pig, and Maurizio Fazio for zebrafish melanoma cell lines



Personal melanoma genomes reveal the existence of differential accesibility for the same enhancer. Mutations that cause accesibility gain/loss are identified from personal cancer genomes and it is validated by the model trained on personel melanoma ATAC profiles.

*Shrikumar, Greenside, and Kundaje, 2017

VIB-KU LEUVEN CENTER FOR BRAIN & DISEASE RESEARCH
KU LEUVEN
FWO Research Foundation Flanders Opening new horizons
erc European Research Council
SCIENCE MEETS LIFE