**Master thesis project**

Multi-tissue modeling of gene expression from DNA sequence by deep learning
Gagneur lab, Computational Biology, TU München

The lab of Julien Gagneur is looking for a student to perform his/her Master thesis. Our goal is an improved understanding of the genetic basis of gene regulation and its implication in diseases. To this end, we employ statistical modeling of 'omic data and work in close collaboration with experimentalists.
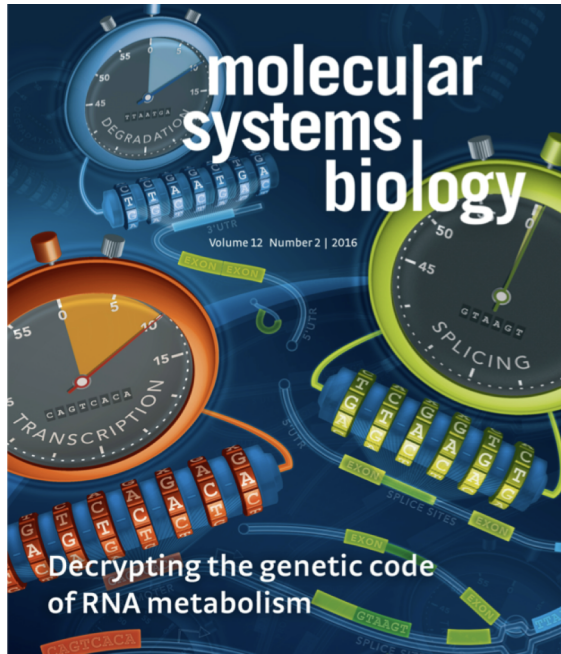
Gene expression is the process that dictates when, where, and how much cells read out genes and expresses their encoded proteins. We want to predict gene expression from genome sequence. The goal of this master thesis is to develop deep learning models [1,2] that models the regulatory code (Figure, ref. [3]) and its modulation across tissues. End-goal is the prediction of the effect of genetic mutations in any human tissue, to understand the mechanisms and the predisposition to genetic disorders and to develop prenatal diagnostics. The initial dataset will be 11,688 human RNA-seq samples and corresponding genotypes (https://gtexportal.org). You will extend and leverage Kipoi (https://kipoi.org), a collaborative initiative to define standards and to foster reuse of trained machine learning models in genomics that we co-develop [3].

The project is open to bioinformatics, statistics, physics, and computer science students. Fluency in English is essential. Our group is a lively, international and interdisciplinary group. We are located at the Informatics faculty of the Technical University of Munich. There is the perspective to join afterward as PhD student.

Apply with a brief motivation, CV, and transcript, to Julien Gagneur (gagneur@in.tum.de)
https://www.gagneurlab.in.tum.de

References

1. Avsec et al., Modeling positional effects of regulatory sequences with spline transformations increases prediction accuracy of deep neural networks, Bioinformatics, 2017
2. Cheng et al., Modular modeling improves the predictions of genetic variant effects on splicing, bioRxiv – winner model of the CAGI 2018 splicing challenge
3. Eser et al., Determinants of RNA metabolism in the *S. pombe* genome, Molecular Systems Biology, 2016
4. Avsec et al., Kipoi: accelerating the community exchange and reuse of predictive models for genomics, bioRxiv