Developing a Deep Learning Model to Predict Mutation Probabilities at Single Base Resolution

Nimród Grandpierre
Master's Thesis
June 15th, 2025

AARHUS
UNIVERSITY

ABSTRACT

Deep learning has undergone a resurgence over the past decade and a half, proving to be highly effective across a wide range of domains, including genomics. The mutational landscape - a topic within the field of genomics - has previously been investigated by applying both traditional statistical learning methods and neural networks. In this thesis, a new neural network-based approach is presented that aims to infer mutation outcome at single base resolution across the human genome. A key hypothesis in this domain that has gathered significant proof is that the most prevalent factor in a substitution being developed is the local context around the nucleotide site. This insight suggests that a *k*-mer representation of local nucleotide context captures the primary signal in mutation prediction. By utilizing both the local context and additional features that inform predictions about the characteristics of the broader region around the site, neural network models gain a comprehensive understanding of factors determining mutation outcome, leading to impressive predictive results.

CONTENTS