Ruprecht-Karls-Universität Heidelberg Medizinische Fakultät Mannheim
Dissertations-Kurzfassung

Advanced hierarchical learning approach for microRNA and target prediction

Autor: Alisha Parveen
Institut / Klinik: Zentrum für Medizinische Forschung
Doktorvater: Prof. Dr. N. Gretz

microRNA (miRNA) is a small non-coding molecule that regulates gene expression by base-pairing with their target gene that leads to cleavage or repression of the target genes. miRNA target prediction plays an important role in several human diseases and regulatory pathways. Hence, identification of gene targets is important for functional characterization of miRNAs and it also gives further insights into biological and clinical processes. Experimental validation of miRNA and target prediction is time-consuming and has several other limitations. From the past decades, several sophisticated target prediction tools have been developed. However, there is a huge need for the improvement of the existing methods of miRNA-target predictions since complementarity of functional miRNA-target interactions is usually small, frequently just involving the seed region only. This often leads to low accuracy and high false-positive interactions. To address these issues, an advanced deep learning-based approach has been developed named as advanced hierarchical deep-rooted learning (AHDR) approach. This approach is based on a deep feedforward network which is the foundation of several deep learning models.

During data mining, a balanced dataset was constructed for the training of the model and was verified from Tarbase. Whereas, the negative dataset was generated using random mutations of 95% across human genome-wide transcripts. Several parameters were applied to distinguish properly between positive and negative datasets. Additionally, 95 features were identified extensively that exhibit sufficient information on miRNA target interaction. In a next step, a two-level of feature selection algorithm known as Least absolute shrinkage and selection operator has been applied to extract putative features of interactions. This results in a reduction of overfitting problems, is an easier to interpret model prediction and enables algorithms to work faster. Simultaneously, several other machine learning approaches have been trained for evaluation of statistical analysis from the AHDR.

This approach showed an excellent performance over other existing miRNA target prediction methods. AHDR analysis revealed that this algorithm was one of the best prediction tools of validated miRNA target predictions. Furthermore, AHDR has showed the highest number of miRNA interactions on top cancer genes as compared to other existing prediction tools.

In the present work it could be shown that a deep learning-based framework for the prediction of miRNA target interactions is very suitable. By using this technology, it is possible to combine very heterogeneous data from different sources to gain new insights into the development and course of diseases.