Abstract for the General Public

The main goal of the project is to provide a computational framework for the identification of effective and efficacious strategies for cellular reprogramming.

The quest of striving towards better health and long life has always accompanied the history of mankind. While the last few decades of biomedical research has led to the resolution of most of the 'simple' genetic disorders, complex diseases still pose a great challenge. This is largely because in a living cell genes and gene products, e.g., mRNA and proteins, function within a densely interconnected system of these molecular components, i.e., the gene regulatory network (GRN). A disease is seldom a disruption in a single gene; it arises as an aftermath of perturbations of this complex network.

Cellular reprogramming in general is a man-made process in which one mature, specialised cell type is changed into another. It has been drawing increasing research attention for its therapeutic potential for the most complex diseases by providing means of creating new cells to replace those whose death or damage causes disease symptoms, or by guiding cells from 'unhealthy' to the desired 'healthy' states. Classical reprogramming procedures are performed by modifying a specific combination of genes by turning each of them either ON (activating) or OFF (deactivating). However, in complex diseases it is more common that several subtle changes affect interactions between genes. In terms of modelling of GRNs, this means disruptions at the edge level, i.e., targeting selected interactions between genes. Compared to gene perturbations, edgetic perturbations provide much more fine-grained GRN manipulation capabilities. Hence, in order to derive effective GRN control strategies, interventions both on the gene and edge level should be considered.

Unfortunately, wet-lab experiments for finding proper combinations of reprogramming targets, i.e., genes or proteins, manipulation of which can trigger desired changes are very costly and time-consuming. These barriers can be lowered by developing computational approaches. EdgeCR will provide a computational framework for the identification of effective and efficacious strategies for cellular reprogramming by considering both genetic and 'edgetic' interventions. First, techniques to infer computational models of GRNs from experimental genomic data, specifically Boolean network (BN) models, will be developed. The new methods will provide means to incorporate characteristic structural and dynamical features of GRNs into the models. They will allow to construct more realistic and accurate BN models of GRNs.

Second, existing control algorithms and methods for mid-size BN models will be extended to include edgetic perturbations. This will fill the gap of currently missing computational toolkits for this type of control and will enable more comprehensive modelling and analysis of GRNs.

Next, EdgeCR will tackle the challenge that cellular systems are very complex and the number of reprogramming target combinations to be explored is immense, especially when both gene and edge perturbations are considered. Thus, for large-scale BN models of GRNs new control methods will be developed by exploiting artificial intelligence techniques, i.e., deep reinforcement learning (DRL), which has already shown its power on extremely complex decision problems such as winning the game of Go against a human top champion. The developed control methods will be combined into a framework, which will be implemented as a software tool.

Finally, the potential of the framework will be experimentally validated. To this aim, the *in vivo* application of cellular reprogramming for Parkinson's disease, which is the second most prevalent neurodegenerative disease, will be investigated. The disease is caused by the loss of dopaminergic neurons in the midbrain. Single-cell genomic data will be used to infer a BN model on human embryonic midbrain development and the framework will be applied to this model to identify the reprogramming targets and strategies for astrocytes to dopaminergic neurons conversion. The reprogramming candidates will be tested for efficiency and fidelity in the lab.

EdgeCR will advance the state-of-the-art in the field of cellular reprogramming of GRNs by providing scalable computational methods that facilitate the identification of reprogramming targets and control strategies. It will constitute a significant contribution by providing the more fine-grained edge-perturbation control dimension. Such computational techniques are still missing and their development is essential for the advancement of the understanding of gene regulation and its control. The new model inference pipeline combined with DRL-based gene- and edge-perturbation control methods will constitute an important contribution to the catalogue of computational frameworks and tools for effective and efficacious cellular reprogramming.