

Master thesis in Bioinformatics  
Professor Jan Komorowski  
Department of Cell and Molecular Biology  
Project proposal

**Identification of *de novo* Transcription Factor Binding Motifs Created by Cancer-related Mutations**

Cancer is a leading cause of morbidity and according to the World Health Organization (WHO), it accounts for nearly one in six deaths. Importantly, many cancers can be cured if detected early and treated effectively, preferably using a strategy of precision medicine directed against the cause of the disease.

Cancer is caused by genetic defects with most mutations acquired in differentiated cells and a small fraction of inherited variants. Currently the genome of many cancers is sequenced to detect mutations that may cause disease. The standard approach is to look for mutations in the coding regions that are relatively straightforward to interpret. The knowledge of gene regulatory regions has exploded in recent years making it possible to identify cancer mutations in them.

Transcription factors (TF) regulate gene activity by binding to transcription factor binding motifs (TFBMs) in gene regulatory regions and mutations there can change gene activity and thereby contribute to cancer. We have shown that many cancer mutations destroy TFBMs and prevent the TF from binding, which changes the activity of genes known to contribute to cancer and to new sets of genes.

In collaboration with Professor Claes Wadelius, we have developed a computational tool *funMotifs* that predicts and annotates functional TFBMs in the human genome. We applied it to mutation information from the international Pan-Cancer Analysis of Whole Genomes (PCAWG) project to predict the regulatory cancer mutations described above. However, many cancer mutations may also create new TFBMs. We would like to address this new challenge by expanding on existing methods with cutting-edge computational tools, such as mutation simulations, which would allow us to play nature and identify new potential mutation sites – called *de novo* mutations. Experts in computation, gene regulation and cancer genetics will support the proposed project with the aim of applying to a very large cohort of cancer. Thus, the results derived from this project may be useful in increasing the understanding of regulatory impact from cancer-related mutations.

Candidates are required to have strong programming skills in languages commonly used for data science, R, Python or Julia, as well as knowledge in machine learning algorithms. Experts in computation, molecular biology and machine learning will support and supervise the proposed project.

Contact:

Daniel Rivas: [daniel.rivas@icm.uu.se](mailto:daniel.rivas@icm.uu.se)

Jan Komorowski: [jan.komorowski@icm.uu.se](mailto:jan.komorowski@icm.uu.se)