REVIEW OF THE DIPLOMA THESIS

Author: Mukhiddin Yusupov
Title: Utilization of methylation data in phenotype molecular models

Reviewer: Doc. Ing. Jiří Kléma, Ph.D.
              Department of Computer Science
              Czech Technical University in Prague
              e-mail: klema@fel.cvut.cz
              tel.: +420 224 357 608

High-throughput molecular screening techniques generate large amounts of omics (genomics, proteomics, etc.) data in parallel. These data can be used to draw attention to both simple and structured phenotype markers as well as for construction of phenotype predictive models. As the number of samples is typically small and the number of features is large, the above-mentioned analysis often suffers from multiple comparisons when performing statistical tests and overfitting when constructing predictive models. That is why it is important to employ the existing known interactions among the measured features and potentially automatically discover previously unknown interactions. This thesis aims at understanding and utilization of relationships between gene expression and DNA methylation data in the particular domain of myleodysplastic syndrome (MDS).

The author focuses on implications of known regulation of transcriptional activity by methylation at CpG sites. He first empirically studies the inverse relationship between gene expression and DNA methylation by means of correlation of methylation and expression profiles. Then, he proposes, implements and tests a new integrative method for MDS phenotype prediction. The method is based on signatures (extracted meta-features) that integrate interacting differentially expressed genes and differentially methylated sites. Mr. Yusupov shows that the new method in some tasks outperforms the benchmark that simply concatenates gene expression and methylation features. By these experiments, the main assignment has been successfully accomplished.

Mr. Yusupov worked on his thesis systematically for more than one year. He proved ability to understand the domain, analyze the data and finally formalize the predictive workflow. On the other hand, the work on the thesis has not been entirely smooth as it interacted with his other study requirements and deadlines. As a result, the thesis was finished in a big rush which considerably affected its quality both in terms of formal issues (readability, completeness of description, grammar) and the scope of verification of the proposed method (the experimental part is relatively brief). To summarize, the important requirements has been met but with certain reservation.

In his diploma thesis, Mukhiddin Yusupov proved the ability for inquisitive and independent work. I assess the diploma work by the grade

<div align="center">

C — good.

</div>

In Prague, May 26, 2015   Doc. Ing. Jiří Kléma, Ph.D., reviewer