Evaluation of Dissertation of Suren Davitavyan
CHARACTERIZATION OF MOLECULAR DIVERSITY OF BREAST CANCER AND GLIOMA WITH TRANSCRIPTOMIC, GENOMIC, AND EPIGENETIC DATA
seeking a Candidate of Biological Sciences (Ph.D.) degree

The study of cancer molecular diversity is a cornerstone of precision medicine. Advances in understanding cancer heterogeneity have significantly improved treatment efficacy, disease course prediction, and patient survival. High-throughput biomolecule measurement technologies and methods for -omic data analysis play an indispensable role in this progress. However, there is still ample room for new developments and improvements that will further advance this field.

The evolution of understanding cancer molecular diversity has evolved from stratification based on somatic genetic alterations (single nucleotide variations) to structural (copy number variations), epigenetic, transcriptomic, proteomic, and metabolomic features. The accumulation of large quantities of -omic data has made it evident that the future of understanding the specific molecular mechanisms driving cancer diversity and patient stratification lies in integrating multimodal -omics data and developing new bioinformatics and machine-learning approaches capable of analyzing such data.

This dissertation summarizes the candidate's achievements in characterizing the molecular diversity of cancers by developing and applying integrative multi-omic data analysis and knowledge transfer approaches. The candidate pursued several avenues to achieve this aim. First, he developed multi-SOM and supSOM machine learning-based techniques for multi-omic data integration, analysis, visualization, and projection. Both approaches are significant extensions of the self-organizing maps (SOM) algorithm, a popular method for analyzing big "biological" data. Self-organizing maps "stand out for their unprecedented capabilities in dimension reduction with minimal information loss, feature extraction, and visualization." The SOM portrayal method has been applied to solve a wide spectrum of biological questions, such as understanding the molecular basis of cancers, neurodegeneration, and aging, identifying disease molecular subtypes, performing patient stratification, and linking it to disease prognosis. However, the extensive usage of the SOM method pointed to its major drawbacks: the inability to integrate diverse -omic data and issues with extending new samples without retraining.

Suren Davitavyan successfully addressed these limitations by developing a multi-layer SOM that allows the combination of multiple -omic datasets in an elegant way to make it compatible with SOM training. The specific weightings balance the influence of each data type and offer an option to study the combined effects of genomic, transcriptomic, and epigenetic features for gene and sample clustering, as well as their interactions. Another significant advance is a supervised SOM (supSOM) approach for integrating new samples into existing SOM spaces without retraining the model, leveraging support vector machine regression. Combining these techniques offers new opportunities for deep analyses of interactions between different types of functional and regulatory features in the genome and making "context"-based comparisons, interpretations, and classifications.

Next, the candidate conducted a comprehensive study on the molecular diversity of breast cancer and low-grade glioma using multi-omic data (transcriptomic, genomic, and epigenetic). He addressed the complexity and heterogeneity of these cancers by applying the developed techniques. The candidate comprehensively characterized the molecular heterogeneity of PAM50 breast cancer subtypes, identified the core gene modules that are subtype-characteristic, and

thoroughly characterized their interactions and their association with clinical features of the disease. Notably, his work showed that the different features associated with cancer aggressiveness are perturbed in different -omic data layers (transcriptomic, genomic, and epigenetic). For example, the expression of estrogen receptor signaling distinguished between luminal A and other cancer subtypes, but the methylation of RNA splicing genes distinguishes between luminal and basal subtypes. The major takeaway of these results is that integrative analysis better describes breast cancer subtype diversity compared to single-omic analyses.

For low-grade glioma, integrating expression, methylation, and CNV data provided a more informative classification than genetic subtypes alone. The candidate demonstrated that methylation and CNV both affect expression independently, an insight hard to get in a single-omic study. Moreover, he showed that the contribution of CNV and methylation to gene expression is unequal in different genetic subtypes of low-grade glioma. Finally, the results clearly showed that integrating expression, methylation, and CNV data provided better resolution in classification than genetic alterations alone.

The dissertation presents the major findings of the work by Suren Davitavyan, which have been previously approved for publication in peer-reviewed journals by experts in the corresponding fields. The dissertation follows the classical thesis structure. The literature overview section provides a detailed overview of the molecular diversity of breast cancers and low-grade gliomas, discussing the significance of multi-omic data integration. It also covers machine learning techniques in biology, emphasizing the potential of SOM for clustering and visualizing high-dimensional data. The methods section describes the data sources, methods used to ensure consistency and comparability across samples, and theoretical and algorithmic foundations of multi-layer SOM and supSOM techniques. The results and discussion section demonstrates the performance of these techniques and their applications to breast cancer and glioma datasets. The inferences directly align with the aims and results of the research.

Overall, the body of work and the achievements in addressing the complexity and heterogeneity of cancers are impressive and present significant advancements in the field of cancer bioinformatics.

However, several areas could benefit from further scrutiny and improvement:

1. The thesis would benefit from a more thorough discussion of the computational resources required and analysis turnover time.
2. To increase the visibility and accessibility of the multi-layer SOM and supSOM techniques, it would be beneficial to deposit them on collaborative platforms such as GitHub.
3. Including more cancer datasets and other scenarios of cancer heterogeneity analysis would strengthen the thesis. In particular, it could be interesting to understand how these methods can be used for stratifying the tumor microenvironment and metastasis development.
4. The interpretation of the results could be more deeply explored. Specifically, the biological significance of the identified gene modules and their potential as therapeutic targets or biomarkers could be discussed in greater detail.
5. A more detailed analysis of the integration of different independent datasets and the potential impact of the batch effect on the analysis, especially in connection with the supSOM method, would be beneficial.
6. The dissertation would have also benefited from language editing by a native speaker, as there are some typographical and stylistic errors.

In summary, there is no doubt that the results of Suren Davitavyan's research are sound and novel and have contributed to further advancements in integrated multi-omic data analysis in general and the study of molecular diversity of cancers in particular. Therefore, I enthusiastically recommend awarding Suren Davitavyan the degree of Candidate of Biological Sciences (Ph.D.) in the field of 03.00.02, "Biophysics, Bioinformatics."

Dr. Artur Muradyan
Scientist, Deputy Head
Department of Tumor Cytogenetics | Human Genetics
Labor Berlin | Sylter Straße 2 | 13353 Berlin

Phone: +49 30 45 05 69 141
artur.muradyan@laborberlin.com

07.06.2024