



**MARMARA UNIVERSITY**  
**INSTITUTE FOR GRADUATE STUDIES**  
**IN PURE AND APPLIED SCIENCES**



**MULTI-OMICS DATA INTEGRATION IN  
THE PREDICTION OF POTENTIAL  
BIOMARKERS AND THERAPEUTICS IN  
HUMAN CANCERS**

---

**GİZEM GÜLFİDAN YILDIZ**

**Ph.D. THESIS**

Department of Bioengineering

**ADVISOR**

Prof. Kazım Yalçın ARGA

**CO- ADVISOR**

Assoc. Prof. Beste TURANLI

ISTANBUL, 2024



**MARMARA UNIVERSITY**  
**INSTITUTE FOR GRADUATE STUDIES**  
**IN PURE AND APPLIED SCIENCES**



**MULTI-OMICS DATA INTEGRATION IN  
THE PREDICTION OF POTENTIAL  
BIOMARKERS AND THERAPEUTICS IN  
HUMAN CANCERS**

---

**GİZEM GÜLFİDAN YILDIZ**  
724218001

**Ph.D. THESIS**

Department of Bioengineering

**ADVISOR**

Prof. Kazım Yalçın ARGA

**CO- ADVISOR**

Assoc. Prof. Beste TURANLI

ISTANBUL, 2024

## **ACKNOWLEDGEMENT**

Firstly, I would like to express my deepest gratitude to my thesis supervisor Prof. Dr. Kazım Yalçın ARGA, the best-loved teacher with lots of patience, understanding and sense of justice. Many thanks to him for giving me the opportunity to study in such a valuable field. I have always had the continual support and encouragement of him throughout my undergraduate and master studies. I am deeply grateful for the motivation and guidance he has provided.

Also, I would like to offer my special thanks to my dear co-advisor Assoc. Prof. Beste TURANLI for providing help whenever I need assistance, for sharing her experience and knowledge with me and for her contribution and comments on my thesis. I would like to count her as a friend of mine for the rest of my life.

Moreover, I thank Prof. Fatih BAYRAKLI and Assist. Prof. Pınar PİR for serving as committee members during my PhD, and other jury members who will take part in my defense Assist. Prof. Saliha Ece Acuner Zorluuysal, Assist. Prof. Elif KUBAT ÖKTEM, and Assist. Prof. Ceyhun BEREKETOĞLU.

In addition, I would like to thank The Scientific and Technological Research Council of Turkey (TUBITAK) who provided me financial supports through the grant BİDEB 2211-A National PhD Scholarship Program and the grant the Council of Higher Education (YÖK) 100/2000 Doctoral Fellowship Program.

My biggest thanks go to my beloved family Gönül GÜLFİDAN, Cengiz GÜLFİDAN, and Korcan GÜLFİDAN who have provided me all kind of moral and material support at any moment of my life. I could come to this day because of their endless opportunities and teachings. I am deeply grateful for everything they have done for me. I am so glad to have the family like that!

Lastly, I would like to thank the most precious person in my life, Barış YILDIZ, for always taking my side and listening to me patiently whenever I am nagging when I was in trouble and distress. My life would not have been this much good without him.

**İstanbul, 2024**

**Gizem GÜLFİDAN YILDIZ**

## TABLE OF CONTENTS

|                                                                                                                                                                              |      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| ACKNOWLEDGEMENT .....                                                                                                                                                        | i    |
| TABLE OF CONTENTS .....                                                                                                                                                      | ii   |
| ÖZET.....                                                                                                                                                                    | vi   |
| ABSTRACT .....                                                                                                                                                               | viii |
| CLAIM OF ORIGINALITY .....                                                                                                                                                   | x    |
| ABBREVIATIONS.....                                                                                                                                                           | vii  |
| LIST OF FIGURES.....                                                                                                                                                         | x    |
| LIST OF TABLES .....                                                                                                                                                         | xiii |
| 1 INTRODUCTION.....                                                                                                                                                          | 1    |
| 1.1. Omics Space .....                                                                                                                                                       | 4    |
| 1.1.1. Major databases storing omics data .....                                                                                                                              | 5    |
| 1.1.2. Comprehensive multi-omics data projects/initiatives .....                                                                                                             | 6    |
| 1.2. Multi-Omics Data Integration .....                                                                                                                                      | 7    |
| 1.2.1. Principle approaches.....                                                                                                                                             | 8    |
| 1.2.2. Featured method and bioinformatics tools for data integration.....                                                                                                    | 8    |
| 1.2.3. Featured web-based portals for data interpretation and visualization.....                                                                                             | 9    |
| 1.3. Applications of Multi-Omics Data Integration in Healthcare .....                                                                                                        | 10   |
| 1.3.1. Prediction of biomarkers for the diagnosis, monitoring, and treatment of diseases. 10                                                                                 |      |
| 1.3.2. Subtyping and classification of diseases based on multi-omics profiles.....                                                                                           | 11   |
| 1.3.3. Illumination of disease mechanisms .....                                                                                                                              | 11   |
| 1.3.4. Risk and clinical assessment predictions.....                                                                                                                         | 12   |
| 1.4. Challenges and Future Predictions in Multi-omics Data Integration .....                                                                                                 | 12   |
| 1.4. Current and Emerging Applications of AI in Genomic Medicine.....                                                                                                        | 19   |
| 1.5. Limitations and Grand Challenges.....                                                                                                                                   | 21   |
| 1.6. Aims and Scope .....                                                                                                                                                    | 24   |
| 1.7. How to Read This Thesis?.....                                                                                                                                           | 25   |
| 1.8. Background.....                                                                                                                                                         | 26   |
| 1.8.1. Integration of multi-omics data in prediction of potential biomarkers and<br>therapeutics in glioblastoma.....                                                        | 26   |
| 1.8.2. Differential protein interactome in esophageal squamous cell carcinoma offers<br>novel systems biomarker candidates with high diagnostic and prognostic performance . | 27   |

|                                                                                                                                                                             |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.8.3. Systems biomarkers for papillary thyroid cancer prognosis and treatment through multi-omics networks.....                                                            | 29 |
| 2. MATERIALS AND METHODS .....                                                                                                                                              | 32 |
| 2.1. Integration of multi-omics data in prediction of potential biomarkers and therapeutics in glioblastoma .....                                                           | 32 |
| 2.1.1. Data collection and preprocessing .....                                                                                                                              | 32 |
| 2.1.2. Reducing data dimension and clustering analysis.....                                                                                                                 | 33 |
| 2.1.3. Multi-omics integration of clusters .....                                                                                                                            | 34 |
| 2.1.4. Comparisons of glioblastoma subtypes .....                                                                                                                           | 34 |
| 2.1.5. Differential analysis based on mRNA, lncRNA and methylation .....                                                                                                    | 35 |
| 2.1.6. Gene set enrichment analysis of mRNA, lncRNA and methylation.....                                                                                                    | 36 |
| 2.2. Differential Protein Interactome in Esophageal Squamous Cell Carcinoma Offers Novel Systems Biomarker Candidates with High Diagnostic and Prognostic Performance ..... | 37 |
| 2.2.1. Gene expression data.....                                                                                                                                            | 37 |
| 2.2.2. PPIs data .....                                                                                                                                                      | 37 |
| 2.2.3. Identification of differential interactome .....                                                                                                                     | 37 |
| 2.2.4. Categorization of DIPs .....                                                                                                                                         | 38 |
| 2.2.5. Functional overrepresentation analysis.....                                                                                                                          | 38 |
| 2.2.6. Analysis of diagnostic performance .....                                                                                                                             | 39 |
| 2.2.7. Analysis of prognostic power .....                                                                                                                                   | 39 |
| 2.2.8. Evaluation of the secretion levels of candidate biomarkers in body fluids and tissues .....                                                                          | 40 |
| 2.2.9. Identification of candidate drugs through drug repositioning.....                                                                                                    | 40 |
| 2.2.10. Molecular docking analysis for repurposed drugs .....                                                                                                               | 40 |
| 2.2.11. Evaluating repurposed drugs in human ESCC cell lines.....                                                                                                           | 41 |
| 2.2.12. Cell lines and treatments .....                                                                                                                                     | 41 |
| 2.2.13. Cell proliferation (MTT Assay) .....                                                                                                                                | 41 |
| 2.3. Systems Biomarkers for Papillary Thyroid Cancer Prognosis and Treatment through Multi-Omics Networks .....                                                             | 42 |
| 2.3.1. Gene expression datasets of papillary thyroid carcinoma.....                                                                                                         | 42 |
| 2.3.2. Identification of differentially expressed genes.....                                                                                                                | 42 |
| 2.3.3. Identification of reporter receptors, transcription factors, and miRNAs.....                                                                                         | 43 |
| 2.3.4. Identification of reporter metabolites associated with thyroid cancer .....                                                                                          | 43 |
| 2.3.5. Gene set overrepresentation analyses .....                                                                                                                           | 43 |

|        |                                                                                                                                                                        |    |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.3.6. | Reconstruction of protein-protein interaction networks and topological analysis                                                                                        | 44 |
| 2.3.7. | Prognostic features of reporter molecules and hub proteins.....                                                                                                        | 44 |
| 2.3.8. | Determination of candidate drugs through drug repositioning.....                                                                                                       | 44 |
| 2.3.9. | Molecular docking analysis for repurposed drugs .....                                                                                                                  | 45 |
| 3.     | RESULTS AND DISCUSSION .....                                                                                                                                           | 46 |
| 3.1.   | Integration of multi-omics data in prediction of potential biomarkers and therapeutics in glioblastoma .....                                                           | 46 |
| 3.1.1. | Identification of GBM subtypes through clustering.....                                                                                                                 | 46 |
| 3.1.2. | Comparisons of identified GBM subtypes.....                                                                                                                            | 48 |
| 3.1.3. | Differential gene expression and methylation probes.....                                                                                                               | 53 |
| 3.1.4. | Gene Set Enrichment Analysis.....                                                                                                                                      | 55 |
| 3.1.5. | Discussion.....                                                                                                                                                        | 56 |
| 3.2.   | Differential Protein Interactome in Esophageal Squamous Cell Carcinoma Offers Novel Systems Biomarker Candidates with High Diagnostic and Prognostic Performance ..... | 57 |
| 3.2.1. | Differential interactome in ESCC.....                                                                                                                                  | 57 |
| 3.2.2. | Biological characterization of DIP-centered modules.....                                                                                                               | 61 |
| 3.2.3. | Diagnostic performance of DIP-centered modules.....                                                                                                                    | 64 |
| 3.2.4. | Prognostic performance of DIP-centered modules .....                                                                                                                   | 65 |
| 3.2.5. | Evaluation of biomarker potential of significant modules in body fluids and tissues                                                                                    | 67 |
| 3.2.6. | Drug repositioning considering prognostic modules as potential drug targets..                                                                                          | 70 |
| 3.2.7. | Molecular docking simulations of potential drug target interactions .....                                                                                              | 72 |
| 3.2.8. | Validation of growth inhibition by risperidone and clozapine in human ESCC cell lines                                                                                  | 73 |
| 3.2.9. | Discussion .....                                                                                                                                                       | 75 |
| 3.3.   | Systems Biomarkers for Papillary Thyroid Cancer Prognosis and Treatment through Multi-Omics Networks .....                                                             | 78 |
| 3.3.1. | The transcriptomic signatures of papillary thyroid cancer .....                                                                                                        | 78 |
| 3.3.2. | The proteomic signatures of papillary thyroid cancer .....                                                                                                             | 81 |
| 3.3.3. | The metabolic signatures of papillary thyroid cancer .....                                                                                                             | 83 |
| 3.3.4. | The reporter regulatory and signaling elements of papillary thyroid cancer .....                                                                                       | 84 |
| 3.3.5. | Identification of candidate repurposed therapeutic agents and molecular docking simulations.....                                                                       | 87 |

|                        |     |
|------------------------|-----|
| 3.3.6. Discussion..... | 88  |
| 4. CONCLUSION.....     | 95  |
| REFERENCES.....        | 97  |
| APPENDIX.....          | 112 |



## ÖZET

### **TEZ BAŞLIĞI: KANSERLERDE POTANSİYEL BİYOBELİRTEÇ VE TERAPÖTİKLERİN TAHMİNİNDE ÇOKLU OMİK VERİLERİN ENTEGRASYONU**

Geçtiğimiz yüzyılın en önemli hastalıklarından biri olan ve dünya çapında en yüksek ölüm oranlarından birine neden olan kanserden milyonlarca insan acı çekiyor veya ölüyor. Uluslararası Kanser Araştırma Ajansı'nın (IARC) yayınladığı Küresel Kanser İstatistikleri 2022'ye göre, küresel kanser yükünün 2022 yılına göre %77 artışla 2050 yılında 35 milyon vakaya ulaşması bekleniyor. Hastalık bu kadar ciddi bir tablo ortaya koysa da, moleküler mekanizması henüz anlaşılamamıştır ve bazı türlerin erken tanısında zorluklar yaşanmaktadır. Bu kanser türlerinden bazıları, bu çalışmada analiz edilen glioblastoma (GBM), papiller tiroid karsinomu (PTC) ve özofagus skuamöz hücreli karsinomdur (ESCC). Bu çalışmanın amacı, klinikteki soruları yanıtlamak için farklı omics düzeylerinden verileri sistem biyolojisi perspektifiyle birleştirerek GBM, PTC ve ESCC'nin tanı veya tedavisinde kullanılabilecek biyobelirteçleri, ilaç hedeflerini ve ilaç adaylarını belirlemektir. Burada, çoklu omik yöntemlerini, omik verilerini genom ölçeğinde biyolojik ağlarla entegre eden makine öğrenimi yaklaşımlarıyla zenginleştirerek çoklu omik verilerinin entegrasyonu için en uygun stratejiyi geliştirdik. Bu çalışma, TP53 ve NF1 genlerinin, GBM kümeleri arasında önemli ölçüde farklılaşan çoklu mutasyonlara sahip olduğunu gösterdi. GBM sınıflandırmasına göre küme 1'in pronöral alt tipe, küme 2'nin mezenkimal alt tipe ve küme 3'ün klasik alt tipe daha benzer olduğu bulundu. Mezenkimal alt tipe ait hastalar, pronöral alt tipe ait hastalara göre temazolomid tedavisine daha duyarlıydı. Ayrıca, PTC ile ilişkili hub proteinleri, reseptör proteinleri, TF'ler ve miRNA'lardan oluşan biyomolekülleri literatürde ilk kez belirledik. Bu sonuçlar sadece PTC'nin ilerlemesinin altında yatan mekanizmalara ışık tutmakla kalmıyor, aynı zamanda PTC'nin tedavisi için bazı potansiyel hedeflere de ışık tutuyor. Ek olarak, bir ilaç yeniden konumlandırma stratejisi uyguladık ve üç ESCC hücre hattında risperidon ve klozapinin büyümeyi inhibe edici etkilerini ilk kez doğruladık. Ayrıca, ESCC hastalarının tedavisinde kesin ve kişiselleştirilmiş tıp için etkili bir strateji geliştirmek amacıyla, yalnızca antipsikotiklerin değil, aynı zamanda çalışmamızda listeye alınan diğer ilaçların da yeni ilaç adayları olarak araştırılmasını öneriyoruz. Bu

alışmanın sonuçlarının muazzam nemine rağmen, bulguların deneysel ve klinik olarak doęrulanması iin daha fazla abaya ihtiya vardır. alışmanın sonuçları yukarıda adı geen kanserlerin altında yatan mekanizmaya ışık tutmakta ve literatürdeki bilgileri desteklemektedir. Ayrıca bu alışma, kanserin teşhis ve tedavisine yönelik yöntem, kit ve cihazların geliştirilmesi iin gerekli bilgileri sağlamaktadır.



## ABSTRACT

### **THESIS TITLE: MULTI-OMICS DATA INTEGRATION IN THE PREDICTION OF POTENTIAL BIOMARKERS AND THERAPEUTICS IN HUMAN CANCERS**

Millions of people suffer or die from cancer, one of the most important diseases of the last century, causing one of the highest mortality rates worldwide. According to the Global Cancer Statistics 2022 published by the International Agency for Research on Cancer (IARC), the global cancer burden is expected to reach 35 million cases in 2050, a 77% increase from 2022. Although the disease presents such a serious picture, its molecular mechanism is not yet understood, and several types have difficulties in early diagnosis. Some of these cancer types are glioblastoma (GBM), papillary thyroid carcinoma (PTC) and esophageal squamous cell carcinoma (ESCC), which were analyzed in this study. The aim of the study is to identify biomarkers, drug targets and drug candidates that can be used in the diagnosis or treatment of GBM, PTC and ESCC by combining data from different omics levels through a systems biology perspective to answer the questions in the clinic. Here, we developed an optimal strategy for the integration of multi-omics data by enriching multi-omics methods with machine learning approaches that integrate omics data with genome-scale biological networks. This study showed that the TP53 and NF1 genes have multiple mutations that are significantly differentiated among GBM clusters. Based on GBM classification, cluster 1 was found to be more similar to the proneural subtype, cluster 2 more similar to the mesenchymal subtype and cluster 3 more similar to the classical subtype. Patients belonging to the mesenchymal subtype were more sensitive to temazolomide treatment than patients belonging to the proneural subtype. Furthermore, we determined for the first-time reporter biomolecules consisting of hub proteins, receptor proteins, TFs and miRNAs associated with PTC. These results not only shed light on the mechanisms underlying the progression of PTC, but also on some potential targets for the treatment of PTC. In addition, we performed a drug repositioning strategy and validated for the first time the growth inhibitory effects of risperidone and clozapine in three ESCC cell lines. We also recommend investigating not only the antipsychotics, but also other drugs shortlisted in our study as new drug candidates to develop an effective strategy for precise and personalised medicine in the treatment of ESCC patients. Despite the enormous

significance of the results of this study, further efforts are needed to validate the findings experimentally and clinically. The results of the study shed light on the mechanism underlying the aforementioned cancers and support the knowledge in the literature. In addition, this study provides the necessary information for the development of methods, kits and devices for the diagnosis and treatment of cancer.



## **CLAIM OF ORIGINALITY**

### **THESIS TITLE: MULTI-OMICS DATA INTEGRATION IN THE PREDICTION OF POTENTIAL BIOMARKERS AND THERAPEUTICS IN HUMAN CANCERS**

An integrative approach using multi-omics data is a powerful strategy to decipher the mechanistic details of the information flow in a cell. Currently, there are a wide array of tools and methods available in the public domain to integrate multiomics data sets to derive meaningful insights. Nevertheless, there is a still need to have a uniform framework that can effectively process and analyze multiomics data in an end-to-end manner along with easy and user-friendly visualization and interpretation. The novel aspect of this study is to develop an optimal strategy for multi-omics data integration by enriching different methods with machine learning approaches that integrate sequencing data based on new-generation technologies with genome-scale biological networks.

The interaction list of genes, molecules and drugs that are important for the specified cancers (tumorigenesis, metastasis and treatment) will be brought to the literature with this study. Moreover, the proposed study offers direction in the development of new and accurate diagnostic methods and tools and has the potential to aid in the design of personalized medical approaches to complex human diseases in the future.

**İstanbul, 2024**

**Prof. Kazım Yalçın ARGA**

**Gizem GÜLFİDAN YILDIZ**

## **ABBREVIATIONS**

**AMI:** Adjusted Mutual Information

**CCA:** Canonical Correlation Analysis

**CGP:** Cancer Genome Project

**Cmap:** Connectivity Map

**CPI:** Cluster Prediction Index

**CPTAC:** The Clinical Proteomic Tumor Analysis Consortium

**CPTC:** Classic PTC

**CT:** Computed Tomography

**DEA:** Differentially expressed analysis

**DEGs:** Differentially Expressed Genes

**DElncRNAs:** differential expressed lncRNAs

**DEmRNAs:** differential expressed mRNAs

**DGB:** Drug Gene Budger

**DIPs:** Differentially Interacting Proteins

**DMPs:** differentially methylated CpG probes

**DOR:** Diagnostic Odds Ratio

**DPPIs:** Differential PPIs

**ECM:** Extracellular Matrix

**EMBL-EBI:** European Molecular Biology Laboratory, European Bioinformatics Institute

**EPTC:** Encapsulated PTC

**ESCC:** Esophageal Squamous Cell Carcinoma

**FDR:** False Discovery Rate

**FGA:** fraction genome altered

**FGG:** fraction genome gain

**FGL:** fraction genome loss

**FM:** Fowlkes-Mallows

**FVPTC:** Follicular Variant of PTC

**GBM:** Glioblastoma

**GDC:** Genomic Data Commons

**GENN:** Grammatical Evolution Neural Networks

**GEO:** Gene Expression Omnibus

**GO:** Gene Ontology

**GSEA:** Gene Set Enrichment Analysis

**HR:** Hazard Ratios

**HSP:** Heat Shock Protein

**HTRIDB:** Human Transcriptional Regulation Interactions Database

**IARC:** The International Agency for Research on Cancer

**INMF:** Nonnegative Matrix Factorization

**JI:** Jaccard Index

**JIVE:** Joint And Individual Variation Explained

**LIMMA:** Linear Models for Microarray Data

**LncRNAs:** Long Noncoding RNAs

**MDI:** Multiple Dataset Integration

**METABRIC:** The Molecular Taxonomy of Breast Cancer International Consortium

**MFA:** Multiple Factor Analysis

**MOFA:** Multi-Omics Factor Analysis

**MRI:** Magnetic Resonance Imaging

**MS:** Mass Spectrometry

**MSigDB:** The Molecular Signatures Database

**OS:** Overall survival

**PARADIGM:** The Pathway Recognition Algorithm Using Data Integration on Genomic Models

**PCA:** Principal Component Analyses

**PCR:** Polymerase Chain Reaction

**PDB:** Protein Data Bank

**PI:** Prognostic Index

**PID:** Pathway Interaction Database

**PPIs:** Protein-Protein Interactions

**PSDF:** Patient-Specific Data Fusion

**PTC:** Papillary Thyroid Carcinoma

**PTM:** Post-Translational Modification

**RI:** Rand Index

**RMA:** Robust Multiarray Average

**RPPAs:** Reverse Phase Protein Arrays

**SNF:** Similarity Network Fusion

**TCGA:** The Cancer Genome Atlas

**TCPA:** Cancer Proteome Atlas

**TFs:** Transcription Factors

**THCA:** Thyroid Carcinoma Project

**TiTv:** Transitions and Transversions

**TMB:** total mutation burden

**TMZ:** temozolomide

**TSPs:** Tumor Suppressor Proteins

**WGCNA:** Weighted Gene Co-Expression Network Analysis

## LIST OF FIGURES

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1. Technological dimensions in designing innovation systems that can harness the full potential of AI in genomic medicine and planetary health. AI, artificial intelligence.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 20 |
| Figure 1.2. Beyond the laboratory: societal, critical governance, and policy dimensions for AI-driven innovation ecosystems. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 22 |
| Figure 1.3. Sub-topics and publications from this thesis. Book chapter and review (blue), research articles (orange) are presented in the format of the thesis. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 25 |
| Figure 2.1. Graphical abstract indicating the multi-omics data integration method. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 33 |
| Figure 3.1. Identification of GBM subtypes (A) Identification of optimal cluster number by calculating CPI (blue line) and Gaps-statistics (red line) in glioblastoma cohort. (B) Quantification of sample similarity using silhouette score based on consensus ensembles result. (C) Consensus heatmap based on results from 10 multi-omics integrative clustering algorithms with cluster number of 3. ....                                                                                                                                                                                                                                                    | 47 |
| Figure 3.2. Comprehensive heatmap based on consensus across 10 algorithms with clinicopathological annotation.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 48 |
| Figure 3.3. Comparison of clinical features of GBM subtypes. (A) Kaplan-Meier survival analysis of the overall survival identified subtypes of GBM. (B) Agreement of identified subtypes of GBM with traditional classification and pathological stage in GBM. (C) The distribution of the clusters with GBM original and transcript subtypes. ....                                                                                                                                                                                                                                                                                                              | 49 |
| Figure 3.4. Comparison of genomic alterations among identified subtypes of GBM. (A) Mutational OncoPrint of identified subtypes of GBM. (B) Comparison of TMB and TiTv among identified subtypes of GBM. (C) Bar plot of fraction genome altered among identified subtypes of GBM. ....                                                                                                                                                                                                                                                                                                                                                                          | 51 |
| Figure 3.5. Box violins for estimated IC50 of Temozolomide among identified subtypes of GBM.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 53 |
| Figure 3.6. PCA plots showing the individual differences in the gene expression profiles among normal and tumor samples. (A) PCA plots created with the expression data of genes obtained from the data set of GSE23400. (B) PCA plots created with the expression data of genes obtained from the data set of GSE13898. Red dots represent healthy samples; blue dots represent tumor samples. PCA, principal component analyses.....                                                                                                                                                                                                                           | 65 |
| Figure 3.7. Kaplan–Meier Plots estimating patients’ survival for ESCC indicating p-value of the log-rank test and hazard ratio for each curve belonging to each module. Censoring samples are shown as “+” marks. Horizontal axis represents time to event. Green and red curves denote low- and high-risk groups, respectively. The green and red numbers below the horizontal axis represent the number of individuals not presenting the event of the corresponding risk group along time. The number of individuals, the number of censored, and the CI of each risk group are shown in the top-right insets. ESCC, esophageal squamous cell carcinoma. .... | 66 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 3.8. Bubble plots indicating protein expression levels of the proteins in five diagnostic modules in different body fluid and tissues. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 67 |
| Figure 3.9. PCA plots showing the individual differences in the gene expression profiles among normal and tumor samples in the modules having significant protein expression levels. Red dots represent healthy samples; blue dots represent tumor samples.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 68 |
| Figure 3.10. Networks indicating drugs associated with proteins in significant prognostic modules resulted in drug repositioning. Drugs are represented in <i>circles</i> and their color varies by ATC groups of the drugs. Modules are represented in <i>yellow hexagons</i> . ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 70 |
| Figure 3.11. <i>In vitro</i> study of risperidone and clozapine on ESCC cell lines and their mechanism(s). (A) Effects of risperidone and clozapine on proliferation of different ESCC cell lines. MTT assay was performed 24 h after treatment with increasing doses of both drugs. Letters indicate the significant difference compared with controls (0 IM). a: $p < 0.05$ , b: $p < 0.01$ ; c: $p < 0.001$ ; d: $p < 0.0001$ ; e: $p < 0.00001$ , f: $p < 0.000001$ . (B) The mechanism(s) indicating the pathways of two antipsychotic drugs, risperidone and clozapine, in cancer. MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide.....                                                                                                                                                                                                                                                                                                                                                         | 74 |
| Figure 3.12. Common DEGs in papillary thyroid cancer and significant overrepresentation analysis of DEGs. (A) Venn diagram representing the comparison of upregulated DEGs among four PTC microarray datasets. (B) Venn diagram representing the comparison of downregulated DEGs among four PTC microarray datasets. (C) Bar graph representing overrepresentation analysis results of common upregulated DEGs among four PTC microarray datasets based on $\log_2(p\text{-value})$ of pathways. (D) Bar graph representing overrepresentation analysis results of common downregulated DEGs among based on $\log_2(p\text{-value})$ of pathways. (E) Plots indicating the clustering of proteins encoded by common upregulated DEGs based on their molecular functions. (F) Plots indicating the clustering of proteins encoded by common downregulated DEGs based on their molecular functions.....                                                                                                                  | 79 |
| Figure 3.13. Hub proteins in PTC datasets. (A) Protein-protein interaction network indicating dataset-specific and common hub proteins among four PTC microarray datasets. Dataset-specific hub proteins are shown in blue, green, yellow, and pink colors. Common hub proteins were shown in orange color and different sizes based on how many datasets they are involved in as significant; small circles for two datasets, medium circles for three datasets, and big circles for four datasets. (B) Kaplan-Meier plots indicating the prognostic power of all hub proteins, hub protein E2F1, and hub protein ERBB2. Each plot has a p-value and hazard ratio to demonstrate its statistical significance. The bottom-right insets present the number of individuals, the number of censored, and the CI of each risk group, and censoring samples are shown as '+' marks. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.) ..... | 82 |
| Figure 3.14. Reporter metabolites associated with thyroid carcinoma. (A) Overrepresentation analysis results for significant reporter metabolites. (B) Heatmap                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |    |

representing significant metabolites, which were observed in at least three datasets, in each dataset based on the p-value..... 84

Figure 3.15. The prognostic power of common reporter molecules related to thyroid carcinoma. (A) Kaplan-Meier plots indicating the prognostic power of common transcription factors having p-value and hazard ratio to demonstrate its statistical significance. The bottom-right insets present the number of individuals, the number of censored, and the CI of each risk group, and censoring samples are shown as ‘+’ marks. (B) Kaplan-Meier plots indicating the prognostic power of common receptors having p-value and hazard ratio to demonstrate its statistical significance. The bottom-right insets present the number of individuals, the number of censored, and the CI of each risk group, and censoring samples are shown as ‘+’ marks. (C) Kaplan-Meier plots indicating the prognostic power of common three miRNAs having p-value to demonstrate its statistical significance. .... 86

Figure 3.16. Bubble plot indicating the binding affinities of repurposed drugs to targeted proteins estimated through the molecular docking simulations. The red lines show the binding affinities between targeted proteins and their inhibitor ligands. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.) ..... 88

**LIST OF TABLES**

Table 1.1. Multi-omics data integration tools and the methods they use. .... 14

Table 3.1. Comparison of clinical features among identified subtypes of GBM. .... 50

Table 3.2. Comparison of mutational frequency among identified subtypes of GBM. . 51

Table 3.3. Differentially interacting proteins with number of differential protein-protein interactions  $\geq 10$  ..... 58

Table 3.4. Differentially interacting proteins-centered modules ..... 58

Table 3.5. Molecular pathways and biological processes significantly associated with differentially interacting protein-centered modules (adjusted p value  $\leq 0.01$ )..... 62

Table 3.6. Differentially interacting protein-centered modules exhibiting high diagnostic potential in several body fluids and tissues. .... 68



## 1 INTRODUCTION

The discovery of the double-helix structure of DNA in 1953, the development of the Sanger DNA sequencing method in 1977, the successful amplification of DNA using polymerase chain reaction (PCR) in 1983, the sequencing of the entire human genome in 2003, and numerous other studies have led to significant advancements in the field of genetics over the past century. These developments have allowed us to understand the complexity of the human genetic structure and how variations in genes can lead to different phenotypes and diseases. Before the Human Genome Project, the genetic basis of around 50-60 diseases was known, but subsequent research following the project's completion has identified the genetic basis for nearly 100 times that number of genetic disorders.

The term of “Genomics” which began to be used with the discovery of the DNA map encoding human genetic material, has quickly led to various omics groups derived from the same root. These include “Epigenomics” which studies genetic changes unrelated to alterations in DNA sequence but affecting gene activity and being transmitted to subsequent generations; “Epi/transcriptomics” which examines changes and regulatory roles at the RNA level; “Proteomics” which investigates functional protein units; “Post-translational modification (PTM) - Epiroteomics” analyzing processes after translation characterized by changes such as sugar, phosphate, and acetyl modifications; “Metabolomics” which examines primary and intermediate metabolites resulting from cellular function; “Lipidomics” enabling high-throughput analysis of lipids; and “Metagenomics” which studies the microbiota living in the body. The high-throughput technologies allow for detailed analysis of these omic structures. However, to definitively establish a situation, integrating multi-omics data is necessary.

The integration of these analyses with various information about the patient, such as their phenotype, clinical history, pathology, laboratory results, and images, will enhance our understanding of disease pathogenesis. Especially within the context of personalized medicine, considering individuals’ nutritional characteristics (nutrigenomics), accurate treatment and prevention methods, and exposure to environmental factors (exposomics) can enable more robust disease risk analyses. To comprehensively understand human health and human diseases, accurate explanation of multi-omics data is essential ((Nalbantoglu et al., 2019; Nalbantoglu and Karadag, 2021)). In the present day, biology

become gradually reliant on multi-layered data created at these levels thanks to the concept of multi-omics ((Argelaguet et al., 2018; Olivier et al., 2019; Subramanian et al., 2020). Low-cost and reproducible sequencing technologies and molecular techniques allow for the production of large quantities of high-quality data, providing an unprecedented view of an organism's system-level complexity. The accessibility of multi-omics data has transformed biology and medicine by paving the way for integrated approaches at the system level. Integrating multi-omics data that provides knowledge about biomolecules from different layers holds promise for systematically understanding complex biology. These integrated approaches merge single omics data sequentially or simultaneously to decipher molecular interactions (Subramanian et al., 2020).

The applications of methods and tools that enable the integration of multi-omics datasets in biology and healthcare aim to uncover answers related to the underlying mechanisms of human diseases. Predicting biomarkers for disease diagnosis, prognosis, and treatment, classification and disease subtyping depending on multi-omics profiles, illuminating disease mechanisms, and making risk and clinical assessment predictions are among the primary reasons for utilizing omics data in the field of health. However, our anticipation is that in the very near future, multi-omics data will play a more prominent role in the healthy lifespan of normal individuals rather than just pathological conditions. “Intelligent healthcare system” analyzing personalized biological data at multiple levels through advanced bioinformatics techniques supported by artificial intelligence will form a basis in subjects including identifying potential risks and ways to avoid them from birth to natural death, establishing appropriate nutrition and environmental conditions, and selecting optimal treatment options if disease is inevitable.

Although big data remain a cornerstone of medical research, there is a growing recognition that it is the meaningful use of big data—clinically, environmentally, and socially—that creates value for health and society. New analytics and sense-making tools that can harness big data for various applications have, therefore, recently emerged, epitomized with the expectations such as “digital transformation of medicine through machine learning” or “artificial intelligence redefining healthcare”(Özdemir and Hekim, 2018; Arga, 2020).

A complete understanding of the genome, its functions, and its links to disease is still a long way off. To address this critical bottleneck, artificial intelligence (AI) techniques, particularly machine learning and deep learning, offer great promise by providing novel predictive tools not inherent in traditional techniques for optimizing key problems in genomic medicine and planetary health (Leung et al., 2016a; Lin and Wu, 2021).

Guiding clinical care of individuals and provision of personalized/precision medicine that optimize diagnostic or therapeutic decisions in tandem call for genomic medicine and planetary health (Shendure et al., 2019; Arga and Sinha, 2021). The main goal of high-throughput sequencing efforts, such as whole genome or whole-exome sequencing, often through next-generation sequencing (NGS) technologies, is to identify genetic variations, understand their associations with diseases or potential therapies, and in this way accelerate the annotation process and advance genomic medicine (Shendure et al., 2019). Increasing efforts are being made to curate population-wide genome sequence data sets and elucidate the relationships of genomic data with phenotype and clinical outcomes. Integrative approaches that incorporate other types of multiomics data sets (e.g., transcriptomics, proteomics, lipidomics, and metabolomics) are also emerging (Karczewski and Snyder, 2018a). Such integrative analyses of multiomics and clinical data make the concept of personalized/precision medicine possible in clinics, but also presents complex computational and sociological challenges. In addition, advances in biotechnology are enabling the rapid generation of highly tailored data sets to test new hypotheses. Single-cell resolution NGS and targeted gene editing tools to study specific gene disorders are the most recent examples (Way and Greene, 2018).

As a result of all these increasing concerns and developments, there is a growing need for computational approaches that can handle the analysis of complex, heterogeneous, and high-dimensional data sets and provide accurate solutions in a fast and cost-effective manner. In addition, diverse data, including molecular expression profiles, physiological measurements, medical imaging, and other clinically and environmentally relevant information, can be integrated into data sets, and some data types can also span many dimensions (Leung et al., 2016; Madabhushi and Lee, 2016). These aspects make the process more complex than could be covered by conventional computational approaches that use predefined rules.

The encouraging thing is that AI algorithms train and learn from data, create their own rules, and improve with experience in a continuous learning process. This innate advantage, not inherent in conventional statistical methods, allows AI to enable new discoveries, such as unrecognized patterns or hard-to-detect relationships, in these data sets without requiring explicit rules to perform these tasks. For these reasons, AI techniques are rapidly gaining attention in genomic medicine and planetary health.

### **1.1. Omics Space**

Developments in technology create research areas called omics. A number of additional omics technologies have been developed through the comprehensive study of the roles, interrelationships, and actions of various types of molecules in the cells of an organism (Olivier et al., 2019). The most common type of omics data is genomic data. Genomic approaches are highly accessible for personalized medicine applications, including disease diagnosis, risk prediction for disease development, and evaluating the suitability and success of individualized treatments. Advances in sequencing technologies now allow faster and more cost-effective whole-genome sequencing, enabling the analysis of all genes at once rather than individually. Genomic-scale association studies and data from whole exome and whole genome sequencing provide numerous DNA sequence variants associated with human traits or diseases. For instance, DNA obtained from cancer cells carries a wide range of somatically inherited changes, such as single nucleotide polymorphisms, insertions, deletions, and copy number variations (Olivier et al., 2019).

In addition to the genomic level, types of transcriptomics have emerged, which is the study of the expression of genes in a cell or organism; proteomics, which is the analysis of proteins; metabolomics, which is the comprehensive analysis of small molecules involved in metabolism; and epigenomics, which is the study of the epigenetic regulation of the entire genome. Developments in technology create new research areas and introduce more omics levels such as lipidomics, metagenomics and glycomics (Olivier et al., 2019).

Sequencing and microarray-based approaches are used in disease genetics research to profile and quantify the transcriptome of cells and tissues. These methods help identify

changes in gene transcription underlying the diagnosis and progression of diseases. After gaining insights into alterations in gene expression, research has shifted toward studying epigenetic regulatory mechanisms that impact gene expression in cells and tissues. These mechanisms include changes in DNA methylation, histone modifications, and the expression of miRNAs and lncRNAs. For patients undergoing chemotherapy, specific miRNA and lncRNA expression profiles can assist in identifying drug-resistant tumor cells and determining whether additional treatment is necessary.

As gene expression analysis has advanced, studies have focused on analyzing cellular proteins—the dominant effectors of cellular function—by examining RNA transcripts and their translational products. Properly evaluating the proteome involves considering various aspects of proteins. Post-translational modifications such as phosphorylation, acetylation, and methylation are often necessary for protein activity or signaling. A protein's cellular localization determines whether a constitutively expressed protein is active or inactive. Most proteomic methods focus on identifying and quantifying individual proteins, often using mass spectrometry or affinity-based protein sequencing. Additional analytical techniques, such as NMR and X-ray crystallography, provide information about the structure of proteins and protein complexes in cells and tissues. Proteomic data play a crucial role in cancer classification, identifying cancer-specific biomarkers, predicting drug sensitivity, and uncovering proteins potentially involved in drug resistance (Kim and Tagkopoulos, 2018; Olivier et al., 2019).

Metabolomics is the scientific study of chemical processes that can analyze small molecules present in cells, tissues, or fluids. It is frequently used in biomarker discovery research. Metabolites are often considered the products of cellular processes mediated by proteins. Therefore, changes in metabolites are assumed to reflect alterations in the functions of enzymes and proteins that mediate these cellular processes. The majority of metabolomic analyses focus on data obtained from plasma or serum samples without requiring invasive tumor biopsies. These analyses aim to identify tumor-specific biomarkers for diagnostic purposes (Olivier et al., 2019).

#### **1.1.1. Major databases storing omics data**

Gene Expression Omnibus (GEO), microarray or microarray, managed by the National Center for Biotechnology Information (NCBI) affiliated with the National Institutes of

Health (NIH), is one of the most used databases in obtaining omics data. It is a public database containing high volumes of genomic data derived from RNA-Seq data, storing approximately 1273 gene expression datasets containing more than 1,147,000 cancer-related samples (Barrett et al., 2013).

The Cancer Genome Atlas (TCGA), funded by the National Institutes of Health, is another database that aims to understand the molecular basis of cancer. It utilizes genomic, transcriptomic, and epigenomic data from over 20,000 primary tumor tissues across 33 cancer types to characterize detailed gene expression patterns in DNA, RNA, proteins, and epigenetic levels (Tomczak et al., 2015a).

The ArrayExpress database, one of the core archival resources of the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), stores functional genomic data from high-throughput experiments such as gene expression, methylation profiling and chromatin immunoprecipitation assays on microarray and sequencing platforms. It stores data from approximately 75,000 experiments and more than 2.5 million analyzes (Athar et al., 2019).

In the SRA data repository, genome data obtained from the DNA-Seq platform to determine genetic variants, transcriptome data obtained from the RNA-Seq platform, and interactome data from the ChIP-exo platform for gene regulation network studies can be accessed (Kodama et al., 2012). Databases producing mass spectrometry (MS)-based proteome data for proteome profiling, quantification of post-translational modifications, and new protein discoveries include ProteomeXchange (Vizcaíno et al., 2014), ProteomicsDB (Samaras et al., 2020), PRIDE (Jones et al., 2006) can be counted. MS-based metabolome data can be obtained for metabolome profiling and discovery of new metabolites with MetaboLights (Kale et al., 2016), while databases such as STRING (Szklarczyk et al., 2015) and BioGRID (Chatr-Aryamontri et al., 2017a) provide yeast two-hybrid data. It stores the protein-protein interactions obtained with (yeast two-hybrid). In addition, our self-established AKBASE database is actively used for our multi-omics study on lung cancer (Çibikdiken et al., 2018).

### **1.1.2. Comprehensive multi-omics data projects/initiatives**

The Cancer Genome Atlas (TCGA) is one of the leading multi-omics projects. Within TCGA, one of the prominent multi-omics projects involves analyzing proteomic data

using Reverse Phase Protein Arrays (RPPAs). These arrays quantify the approximately 200 diverse protein expression in individuals having various cancer types, alongside genomic and epigenomic information (Chang et al., 2013; Das et al., 2020). The resulting data is stored in the Cancer Proteome Atlas (TCPA) (Li et al., 2013a). An example of the benefit of TCGA's multi-omics integration is the identification of risk-modulating non-coding RNAs for prostate cancer by combining RNA-seq and SNP sequencing data (Guo et al., 2016).

The Clinical Proteomic Tumor Analysis Consortium (CPTAC) has subjected samples from various cancer types found in TCGA to LC-MS/MS-based mass spectrometry analysis and enable to determine and quantify thousands of proteins. The obtained proteomic data have been expanded by integrating with genomic data to understand the molecular basis of cancer. In summary, CPTAC presents proteome, glycoproteome, and phosphoproteome data from various cancer tissues provided by TCGA.

The International Cancer Genome Consortium (ICGC), initiated in 2008, is a global effort to create a catalog of mutation abnormalities. It currently contains molecular data from 86 cancer projects with more than 24,000 participants worldwide and represents approximately 82 million somatic mutations (Zhang et al., 2019). ICGC datasets include several high-level data types, including donor, gene, mutation, and cancer drug. The project's goal is to identify all genetic errors in numerous cancers and determine which of these defects cause cancer growth and which have no effect on the disease.

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) offers an integrated database specific to breast cancers, containing clinical characteristics, genomic and transcriptomic profiles, CNV profiles, and SNP genotypes from more than 2,000 tumor samples (Curtis et al., 2012).

TARGET is being conducted to identify molecular events that direct childhood cancers and includes sequencing data, copy number variation, gene expression, miRNA expression, and clinical information of molecular cancer types (Mullighan et al., 2009).

## **1.2. Multi-Omics Data Integration**

Multi-omics data integration can be categorized into two main groups: integration of single-type omics data and integration of multi-omics data. Single-type omics data

integration typically combines genome-scale data related to the same omics level within a specific organism. In this context, most studies focus on genomic and transcriptomic levels. Additionally, there are efforts to integrate genome-scale data for multi-omics types in a specific organism, including joint analysis of transcriptome, proteome, or metabolome expression profiles with genomic data. For instance, integrating transcriptomic and proteomic data has led to the discovery of post-transcriptional activities and correlations between two omics types under the same conditions (Kim and Tagkopoulos, 2018).

### **1.2.1. Principle approaches**

For multi-omic data integration, algorithms can generally be classified as Bayesian, fusion-based, correlation-based, network-based, similarity-based, and other multivariate methods (Subramanian et al., 2020). Some of the multi-omic data integration methods include NetICS (network-based), iCluster, multiple dataset integration (MDI), multi-omics factor analysis (MOFA), and joint Bayesian factor (Bayes pattern fusion analysis, PFA) for fusion-based approaches; PINSplus and NEMO for similarity-based approaches; CNAMet for correlation-based approaches; multiple factor analysis (MFA), joint and individual variation explained (JIVE), and nonnegative matrix factorization (iNMF) for multivariate approaches; and similarity network fusion (SNF) for combined network, similarity, and fusion-based methods. The PARADIGM method combines network-based and Bayesian approaches, while patient-specific data fusion (PSDF) combines Bayesian and fusion-based approaches (**Table 1.1**).

### **1.2.2. Featured method and bioinformatics tools for data integration**

The iCluster method, which is one of the prominent tools in this field, targets to create a single cluster task for examples depending on simultaneous inference obtained from various data types. iCluster is an unsupervised approach which uses a common latent variable model for integrative clustering. It flexibly models relationships between diverse types of data and combines the variance-covariance structures inside data types in a unified construction and decreases the data clusters dimensionality simultaneously. Probabilistic inference is attained using the expectation-maximization algorithm (Shen et al., 2009).

The Pathway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM), which utilizes a Bayes factor graph and is a network-based method, is used to predict the activities of genes as well as the activities of cellular biological processes and molecular pathways in which gene products function, by leveraging multi-omic data in a patient-specific manner (Vaske et al., 2010).

Multi-omics factor analysis (MOFA) is an unsupervised method used to integrate multiple types of omic data across the same or partially overlapping samples. This approach can interpret hidden factors across multi-omic data modalities and helps in extracting a low-dimensional representation of the data. It uses a probabilistic Bayesian framework for model formulation that supports the combination of different noise models to integrate multiple data types, including numerical and categorical data (Argelaguet et al., 2018).

Multiple factor analysis (MFA) is alternative technique which aids in the omic datasets integration by projecting the datasets into a variable space with low-dimension. Multiple factor analysis supplies a sensible illustration of individual and common structures during integration of datasets. Principal component analysis (PCA) is carried out for omic data for the determination of the individual pattern. The global analysis for identifying the common structure involves defining the variance-covariance matrix for each dataset. Also, it offers a matrix of variables which grants to visualize individual and common structures (de Tayrac et al., 2009).

### **1.2.3. Featured web-based portals for data interpretation and visualization**

In addition, the above-mentioned tools and methods which assist in the integration of multi-omic data sets, a variety of portals and platforms are available that aid in the analysis, exploration, interpretation, and visualization of multi-omic data. Some of these visualization portals include cBioPortal, Firebrowse, UCSC Xena, LinkedOmics, 3Omics, NetGestalt, OASIS, Paintomics 3, MethHC (Subramanian et al., 2020).

cBioPortal allows for the analysis, exploration, and visualization of cancer data that includes genomic data, gene expression, methylation, miRNA expression, copy number alterations, and protein abundance data. This portal presently includes data from 233 cancer studies representing more than 30 diverse types of cancer. It intuitively presents cancer data summary, access to downloading of data, analysis and visualization of

network, and correlations between data sets and patient-centric queries. The portal objects to integrate numerous types of data, thus allowing for certain query-based results (Cerami et al., 2012).

Hosted by the Broad Institute, the Firebrowse portal enables the analysis, visualization, and downloading of cancer data from TCGA. This portal provides access to data-specific analysis results and visualization of gene profiles by or specific to cancer type. The iCoMut feature displays a comprehensive summary of the top genes in the available multi-omic data sets for a cancer cohort in the form of heat maps. Genomic variables can be associated with clinical variables and other data types (FireBrowse, n.d.).

UCSC Xena is a web-based tool that allows for the visualization and analysis of cancer data sets from TCGA, CCLE, and more than 40 published studies, as well as the analysis of user data. This tool allows for the comparison of studies between different omic data sets and clinical information. The statistical tools here allow for the dynamic measurement of relationships. The interactive user interface enables discovery by grouping samples according to common clinical features (Goldman et al., 2015).

### **1.3. Applications of Multi-Omics Data Integration in Healthcare**

The applications of tools and methods that enable the integration of multi-omics datasets in the field of health primarily aim to answer various clinically relevant questions related to diseases and mechanisms. These biological questions and relevant literature examples can generally be categorized under four main headings:

- Prediction of biomarkers for the diagnosis, monitoring, and treatment of diseases.
- Disease classification and subtyping depending on multi-omics profiles.
- Illumination of disease mechanisms.
- Risk and clinical assessment predictions.

#### **1.3.1. Prediction of biomarkers for the diagnosis, monitoring, and treatment of diseases**

Biomarkers are molecular footprints of specific cellular functions occurring in a particular phenotype. These biomolecules are part of interconnected biological pathways or processes that facilitate intercellular and intracellular information and mass flow,

allowing us to understand the biological mechanism associated with the relevant phenotype. Integrative analysis can determine confident biomarkers depending on data obtained from multiple and diverse molecular events that cannot be captured by single omics data alone (Sevimoglu and Arga, 2015; Gulfidan et al., 2020a). Since validating biomarkers is a labor-intensive, costly, and time-consuming process, an *in-silico* approach for identifying candidate molecules has emerged as a rational strategy. As integrative analysis approaches continue to evolve and gain wider application, groundbreaking findings that could revolutionize disease diagnosis, prognosis, and treatment are being obtained (Liang et al., 2016; Gov et al., 2017; Dimitrakopoulos et al., 2018; Kori and Arga, 2018; Aydin et al., 2020; Beklen et al., 2021; Turanli et al., 2021).

### **1.3.2. Subtyping and classification of diseases based on multi-omics profiles**

Complex diseases like cancer exhibit heterogeneous characteristics due to significant molecular differences observed in affected individuals (Argelaguet et al., 2017). Additionally, various other factors such as environment and lifestyle can contribute to disease heterogeneity. Therefore, understanding the etiology of the disease, defining different subtypes of the disease, and categorizing diseases into known subgroups or determining appropriate treatment methods for patients with different subtypes are clinically crucial yet challenging tasks. In this context, methods leveraging multi-omics data can be used to identify subtypes of various diseases, including breast cancer, and offer subtype-specific treatment (Bailey et al., 2016; Liu et al., 2016; Turanli et al., 2017; Aydin et al., 2021).

### **1.3.3. Illumination of disease mechanisms**

Figure out the details of molecular mechanisms of disease biology is key to the diagnosis, monitoring, and development of new treatment strategies for diseases. Integration of multi-omics data enables the acquisition of mechanistic information related to disease biology, potentially leading to the development of next-generation treatment systems and personalized medicine applications (Cacchiarelli et al., 2015; Higdon et al., 2015; Guo et al., 2016; Dimitrakopoulos et al., 2018; Turanli et al., 2018; Gulfidan et al., 2020; Ozer et al., 2020).

#### **1.3.4. Risk and clinical assessment predictions**

Workflows integrating multi-omics data with social, demographic, and other clinical information are being developed to create models predicting therapeutic responses or evaluating treatment effectiveness for diseases with complex phenotypes. This integrated method can assist in presenting novel and highly effective therapeutic interventions for diseases characterized by complex phenotypes (Mankoo et al., 2011; Seoane et al., 2014).

#### **1.4. Challenges and Future Predictions in Multi-omics Data Integration**

The approach of multi-omics data integration offers significant advantages, especially in comprehensively understanding biological processes and disease mechanisms. However, it also brings certain challenges. Factors that make multi-omics data integration and analysis difficult include the heterogeneity of individual omics data and the large data sizes that require intensive computational processing (de Maturana et al., 2019; Subramanian et al., 2020). Evaluating the performance of alternative tools on heterogeneous and large-scale data sets remains a critical gap in comparative analyses. Additionally, selecting the right method for addressing specific biological questions is crucial in integrative analyses, emphasizing the need for comprehensive studies evaluating tool selection within the biological context.

Multi-omics data sets can be generated using different platforms, leading to substantial variations in data formats. Bioinformatics-based integration tools require data to be standardized into a predefined format. Consequently, preprocessing omics data sets for integration involves steps such as systematic normalization, data filtering, quality control, and removal of batch effects (Arga, 2019). Given the significant impact of preprocessing on the integration process, careful execution of these steps is essential. For instance, due to the computational demands of integration methods, reducing the size of input data sets has become a prerequisite. The data filtering stage plays a crucial role in eliminating potential noise and reducing the number of attributes that serve as input for the integrated model. However, the lack of universal standards for this step makes it challenging to decide on the most suitable criteria for filtering. Efficient workflows enabling accurate feature selection from omics data which high dimension are urgently needed.

Another dimension that can enhance multi-omics data interpretation is clinical information of individuals. At present, there is no widely accepted approach specifically designed to integrate clinical metadata with omics data. Recent developments in this area primarily focus on reducing challenges. Advanced studies are necessary to establish a framework that enables the interoperability of clinical and omics data.



**Table 1.1.** Multi-omics data integration tools and the methods they use.

| <b>Tool</b>      | <b>Category</b> | <b>Method</b>                                     | <b>Language</b> | <b>Source</b>                     | <b>Link</b>                                                                             |
|------------------|-----------------|---------------------------------------------------|-----------------|-----------------------------------|-----------------------------------------------------------------------------------------|
| Anduril          | Supervised      | Multi-step analysis                               |                 | (Ovaska et al., 2010)             |                                                                                         |
| ATHENA           | Supervised      | Grammatical<br>Evolution Neural<br>Networks, GENN |                 | (Kim et al., 2013)                |                                                                                         |
| BCC              | Unsupervised    | Probabilistic Bayesian<br>Model                   | R package       | (Lock and Dunson,<br>2013)        | <a href="https://github.com/ttriche/bayesCC">https://github.com/ttriche/bayesCC</a>     |
| CCA sparse group | Unsupervised    | Canonical Correlation<br>Analysis, CCA            |                 | (Lin et al., 2013)                |                                                                                         |
| CNAmet           | Unsupervised    | Multi-step analysis                               | R package       | (Louhimo and<br>Hautaniemi, 2011) | <a href="http://csbi.ltdk.helsinki.fi/CNAmet/">http://csbi.ltdk.helsinki.fi/CNAmet/</a> |
| CONEXIC          | Unsupervised    | Probabilistic Bayesian<br>Model                   |                 | (Akavia et al., 2010)             |                                                                                         |
| CoxPath          | Supervised      | Multi-step analysis                               |                 | (Mankoo et al.,<br>2011)          |                                                                                         |
| FSMKL            | Supervised      | Multiple kernel<br>learning                       | MATLAB          | (Seoane et al., 2014)             | <a href="https://github.com/jseoane/FSMKL">https://github.com/jseoane/FSMKL</a>         |

|                      |                                |                             |           |                         |                                                                                                                                                           |
|----------------------|--------------------------------|-----------------------------|-----------|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| GeneticInterPred     | Semi-supervised                | Graph integration           |           | (You et al., 2010)      |                                                                                                                                                           |
| Graph-based learning | Semi-supervised                | Graph integration           |           | (Kim et al., 2012)      |                                                                                                                                                           |
| iBAG                 | Supervised                     | Multi-step analysis         |           | (Jennings et al., 2013) |                                                                                                                                                           |
| iCluster             | Unsupervised                   | Matrix factorization        | R package | (Shen et al., 2012)     | <a href="https://cran.r-project.org/web/packages/iCluster/">https://cran.r-project.org/web/packages/iCluster/</a>                                         |
| iCluster+            | Unsupervised                   | Matrix factorization        | R package | (Mo et al., 2013)       | <a href="https://bioconductor.org/packages/release/bioc/html/iClusterPlus.html">https://bioconductor.org/packages/release/bioc/html/iClusterPlus.html</a> |
| iNMF                 | Unsupervised                   | Matrix factorization        | Python    | TCGA                    | <a href="https://github.com/yangzi4/iNMF">https://github.com/yangzi4/iNMF</a>                                                                             |
| iPAC                 | Unsupervised                   | Multi-step analysis         |           | (Aure et al., 2013)     |                                                                                                                                                           |
| jActiveModules       | Supervised                     | Network simulated annealing |           | (Ideker et al., 2002)   |                                                                                                                                                           |
| JIVE                 | Unsupervised                   | Matrix factorization        | MATLAB    | (Lock et al., 2013)     | <a href="https://genome.unc.edu/jive/">https://genome.unc.edu/jive/</a>                                                                                   |
| Joint Bayes Factor   | Unsupervised                   | Matrix factorization        | MATLAB    | (Ray et al., 2014)      | <a href="https://sites.google.com/site/jointgenomics/">https://sites.google.com/site/jointgenomics/</a>                                                   |
| Joint NMF            | Unsupervised & Semi-supervised | Matrix factorization        | MATLAB    | (Zhang et al., 2008)    | <a href="https://academic.oup.com/nar/article/40/19/9379/2414808">https://academic.oup.com/nar/article/40/19/9379/2414808</a>                             |

|            |                           |                                             |                     |                         |                                                                                                                                                       |
|------------|---------------------------|---------------------------------------------|---------------------|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Lemon-Tree | Unsupervised              | Module network                              |                     | (Bonnet et al., 2015)   |                                                                                                                                                       |
| LRACluster | Unsupervised              | Probabilistic Bayesian Model                | R package           | TCGA                    | <a href="http://bioinfo.au.tsinghua.edu.cn/software/lracluster/">http://bioinfo.au.tsinghua.edu.cn/software/lracluster/</a>                           |
| MCD        | Supervised                | Multi-step analysis                         |                     | (Chari et al., 2010)    |                                                                                                                                                       |
| MCIA       | Unsupervised              | Multiple co-inertia analysis                | R package           | TCGA                    | <a href="http://bioconductor.org/packages/release/bioc/html/omicade4.html">http://bioconductor.org/packages/release/bioc/html/omicade4.html</a>       |
| MDI        | Unsupervised              | Probabilistic Bayesian Model                | MATLAB              | (Kirk et al., 2012)     | <a href="http://www2.warwick.ac.uk/fac/sci/systemsbiology/research/software/">http://www2.warwick.ac.uk/fac/sci/systemsbiology/research/software/</a> |
| MFA        | Unsupervised              | Multiple factor analysis                    | R package           | (Bredel et al., 2005)   | <a href="http://factominer.free.fr/">http://factominer.free.fr/</a>                                                                                   |
| mixOmics   | Supervised & Unsupervised | Multivariate model analysis                 | R package           | TCGA                    | <a href="http://mixomics.org/">http://mixomics.org/</a>                                                                                               |
| MKGI       | Supervised                | Grammatical Evolution Neural Networks, GENN |                     | (Kim et al., 2017)      |                                                                                                                                                       |
| moCluster  | Unsupervised              | Principle Component Analysis                | R package           | TCGA; CPTAC             | <a href="https://bioconductor.org/packages/release/bioc/html/mogsa.html">https://bioconductor.org/packages/release/bioc/html/mogsa.html</a>           |
| MOFA       | Unsupervised              | Probabilistic Bayesian Model                | R package<br>Python | (Dietrich et al., 2018) | <a href="https://github.com/bioFAM/MOFA">https://github.com/bioFAM/MOFA</a>                                                                           |

|                     |                           |                              |           |                           |                                                                                                                                           |
|---------------------|---------------------------|------------------------------|-----------|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| NEMO                | Unsupervised              | Similarity based clustering  | R package | TCGA                      | <a href="https://github.com/Shamir-Lab/NEMO">https://github.com/Shamir-Lab/NEMO</a>                                                       |
| NetICS              | Unsupervised              | Network diffusion method     | MATLAB    | TCGA                      | <a href="https://github.com/cbg-ethz/netics">https://github.com/cbg-ethz/netics</a>                                                       |
| Network propagation | Supervised                | Network based                |           | (Ruffalo et al., 2015)    |                                                                                                                                           |
| PARADIGM            | Unsupervised              | Pathway networks             | Python    | (Vaske et al., 2010)      | <a href="http://paradigm.five3genomics.com/">http://paradigm.five3genomics.com/</a>                                                       |
| PFA                 | Unsupervised              | Fusion method                | MATLAB    | CCLE                      | <a href="http://www.sysbio.ac.cn/cb/chenlab/images/PFApackage_0.1.rar">http://www.sysbio.ac.cn/cb/chenlab/images / PFApackage_0.1.rar</a> |
| PINSPlus            | Unsupervised              | Similarity based clustering  | R package | TCGA and METABRIC         | <a href="https://cran.r-project.org/web/packages/PINSPlus/index.html">https://cran.r-project.org/web/packages/PINSPlus/index.html</a>     |
| PMA                 | Supervised & Unsupervised | Multi-step analysis          | R package | (Witten et al., 2009)     | <a href="https://cran.r-project.org/web/packages/PMA/index.html">https://cran.r-project.org/web/packages/PMA/index.html</a>               |
| Prob_GBM            | Unsupervised              | Probabilistic Bayesian Model |           | (Cho and Przytycka, 2013) |                                                                                                                                           |
| PSDF                | Unsupervised              | Probabilistic Bayesian Model | MATLAB    | (Yuan et al., 2011)       | <a href="https://sites.google.com/site/patientspecificdatafusion/">https://sites.google.com/site/patientspecificdatafusion/</a>           |

|          |              |                                     |                     |                              |                                                                                                                               |
|----------|--------------|-------------------------------------|---------------------|------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| rMKL-LPP | Unsupervised | Multiple kernel learning            |                     | (Speicher and Pfeifer, 2015) | Executable available on request                                                                                               |
| SDP/SVM  | Supervised   | SDP/SVM                             |                     | (Lanckriet)                  |                                                                                                                               |
| sMBPLS   | Unsupervised | Partial Least Squares, PLS          | MATLAB              | TCGA                         | <a href="http://zhoulab.usc.edu/sMBPLS/smbpls_dl.htm">http://zhoulab.usc.edu/sMBPLS/smbpls_dl.htm</a>                         |
| SNF      | Unsupervised | Similarity network fusion           | R package<br>MATLAB | (Wang et al., 2014a)         | <a href="http://compbio.cs.toronto.edu/SNF/SNF/Software.html">http://compbio.cs.toronto.edu/SNF/SNF/Software.html</a>         |
| SNPLS    | Unsupervised | Partial Least Squares, PLS          |                     | (Zhang et al., 2016)         |                                                                                                                               |
| ssCCA    | Unsupervised | Canonical Correlation Analysis, CCA |                     | (Chen et al., 2013)          |                                                                                                                               |
| T-SVD    | Supervised   | Singular vector decomposition       | R package           | TCGA                         | <a href="https://academic.oup.com/nar/article/40/19/9379/2414808">https://academic.oup.com/nar/article/40/19/9379/2414808</a> |

#### **1.4. Current and Emerging Applications of AI in Genomic Medicine**

AI methods have already been applied to almost every aspect of the genomic data analysis pipeline, from presequencing to interpretation, and recent advances have focused on enabling improvements in existing techniques. These efforts offer a number of potential benefits that can help accelerate discovery in genomic medicine, either by directly facilitating the phases of clinical genomic analysis or by improving understanding of health- and disease-related genomic variation (Leung et al., 2016c; Sundaram et al., 2018a; Ho et al., 2019). In line with developments in recent years, machine learning and deep learning are helping to better identify variants in the genome, but understanding the significance of these variations remains a challenge. Therefore, most emerging activities in machine learning and deep learning within genomics are occurring in the analysis and interpretation phase (Dias and Torkamani, 2019).

Although previous efforts have achieved certain standards in genomic data analysis pipelines (Ko et al., 2020), there are still major challenges and limitations that need to be solved, including identifying hard-to-detect variants, predicting phenotypic effects of variants, handling extremely large amounts of genomic sequence data, and integrating genomic data with other relevant information, such as other molecular or clinical data. In contrast, machine learning and deep learning have provided accurate algorithms for identifying somatic and copy number variants (Ho et al., 2019; Sahraeian et al., 2019), and more precise AI tools have been developed for predicting the effects of genetic variants, such as the resulting effects on proteins or key molecular processes such as transcription, signal transduction, and metabolism (Sundaram et al., 2018b; Dias and Torkamani, 2019).

Facilitated by increased computational power in recent years, the resurgence of deep learning has created new opportunities for analyzing massive data sets such as medical imaging data and electronic health records (EHRs). Significant progress has also been made in developing deep learning-driven tools for extracting clinical data from EHRs and integrating with genomic data sets (Esteva et al., 2019; Norgeot et al., 2019).

AI applications have also accelerated discovery in genomic medicine by improving our understanding of the relationship between genomic variation and diseases or disorders.

These include studies to investigate cancer cell development and survival and to elucidate the genetic changes that drive tumor formation and growth (Sundaram et al., 2018b; Sahraeian et al., 2019), to integrate and analyze genomic data along with other types of molecular and clinical data (Karczewski and Snyder, 2018b; Zitnik et al., 2019), and to improve the efficiency and accuracy of gene editing techniques such as CRISPR (Kim et al., 2018; Shen et al., 2018; Leenay et al., 2019).

The popularity of and the growing investments in machine learning and deep learning methods for integrative multiomics data analysis are noteworthy and remain relevant to the current and future design of innovation ecosystems driven by AI (Figure 1.1). To better elucidate the molecular mechanisms under the diseases or the causal relationships between environmental exposure and planetary health, integrative multiomics approach is a preferred strategy (Gov and Arga, 2016; Hasin et al., 2017; Karczewski and Snyder, 2018b; Turanli et al., 2018b; Arga, 2019; Koh and Hwang, 2019; Crandall et al., 2020; Gulfidan et al., 2021).

### Technology Dimensions for Artificial Intelligence Driven Innovation Ecosystems in Genomic Medicine and Planetary Health

#### Data, algorithm, and infrastructure:

- Curate reliable, high-quality training datasets.
- Address standards for data quality, curation, interoperability, and accessibility.
- Develop efficient pipelines to integrate multi-omics datasets as well as metadata.
- Monitor and eliminate sources of bias in data as well as study and algorithm designs.
- Demonstrate validity, clinical and ecological benefits to humans as well as broader impacts on all life in planetary ecosystems.
- Implement a flexible, scalable data processing infrastructure.
- Develop measures to identify and intervene on possible sources of disinformation from data to analyses to contextualization in the clinic and in planetary health practices.



**Figure 1.1.** Technological dimensions in designing innovation systems that can harness the full potential of AI in genomic medicine and planetary health. AI, artificial intelligence.

For example, the emergence of invasive species and the effects of climate change have particular implications for the planetary health and, therefore, a holistic approach to understanding how and why pathogenesis occurs under changing environmental conditions is critical for effective disease management (Crandall et al., 2020). Given the limitations of single-level omics data sets in comprehensively interpreting specific biological phenomena, studies are combining data from multiple omics technologies with clinical data and environmental monitors to find accurate solutions quickly and cost-effectively, and machine learning techniques are becoming increasingly important in analyzing these heterogeneous data sets (Ozer et al., 2020).

The potential to apply machine learning to all phases of genome-based drug development has motivated many pharmaceutical companies to invest resources in this area. Machine learning is being applied to genomic and other omics data sets for a variety of purposes, including disease subtype definition, disease biomarker identification, target discovery, drug repurposing, and drug response prediction (Zafeiris et al., 2018; Ekins et al., 2019; Koromina et al., 2019; Vamathevan et al., 2019; Turanli et al., 2021). Many large pharmaceutical companies have AI-focused research and development (R&D) initiatives or collaborations underway (Baldoni et al., 2020).

### **1.5. Limitations and Grand Challenges**

The application of AI has not yet led to significantly improved outcomes in genomic medicine, and the potential to explore genomic data sets remains largely untapped (Topol, 2019). To drive progress, several interrelated issues must be addressed, including data quality and accessibility, bias, replicability, infrastructure, and uncertainties in privacy, security, regulation, and clinical governance (**Figure 1.2**).

### **Critical governance of emerging AI science in planetary society:**

- Engage with diverse publics commencing with study design and for AI applications aligned with broader societal values, priorities and needs.
- Consider democratic representation in science to address, for example, historical power asymmetries, climate justice, and gender parity.
- Establish accountability and prevent malicious or discriminatory use of AI and algorithms.
- Consider regulatory implications such as privacy and data security, but also think 'upstream' – who designs AI studies, applications, to what end, and by what means?
- For AI impacts in planetary health, consider that health is co-produced not only by human health but also animal and ecosystem health and the interdependencies therein.

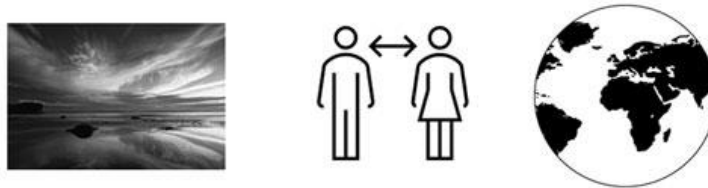


**Figure 1.2.** Beyond the laboratory: societal, critical governance, and policy dimensions for AI-driven innovation ecosystems.

The accuracy of an AI technique is highly dependent on the reliability and quality of the training data. Therefore, data credibility is the primary bottleneck. The performance of AI algorithms is affected by the amount and quality of data used for development or training, so optimized access to highquality genomic and health data is essential. However, big data not only increase the data volume but also of the errors and noise (Arga, 2019). Data sets are noisy, poorly annotated, and generally unstructured. Fundamental to curating reliable and high-quality training data sets are standards (universal protocols) for data quality, curation, interoperability, and accessibility.

In addition, monitoring and eliminating sources of bias within training data sets are also challenges that must be resolved to demonstrate validity and clinical utility. There are biases in the data sets that limit the representation of heterogeneity in society (Demirci et al., 2021). Some populations are underrepresented in data sets used to develop or train AI algorithms. Algorithmic biases may also arise due to the availability of data, the way data are processed and combined, the way questions are formulated, and pre-existing biases in society.

Replicability and reproducibility of results are another problem (Hutson, 2018). Replication and reproduction of the methods and results of AI tools are typically difficult and can be even more challenging in genomic data analysis, given the complexity of the

various steps (i.e., presequencing, sequencing, data processing, analysis, and interpretation) in the data analysis pipeline. Implementation of a flexible scalable data processing infrastructure would streamline this process. It is even more challenging in multiomics R&D, the integration and interpretation of diverse large-scale omics data in a way that provides biological and clinical insights (Arga, 2019; Turanli et al., 2019c). The difficulties often arise from the size and multiple different formats of the data sets generated at different omics levels on different platforms. It is important to ensure that the differences observed in the samples before integration are due to biological variability and are not a technical artifact embedded in the data.

Of course, these issues, in turn, pose challenges for privacy, security, and safety measures (Char et al., 2018; Özdemir, 2019b; Topol, 2019). Security is one of the most controversial aspects of AI applications. Along with the issue of genomics, concerns about security, confidentiality, and ethical use reach the highest levels in society. In addition, there are uncertainties about the regulatory status of AI algorithms used in clinical genomics. How algorithms are certified, how they are used for clinical decision making and patient management, liability for innocuous use, and compliance with transparency and explanation requirements within national regulations are unclear. These concerns need to be effectively managed and addressed to remove the risk of hindering the use and implementation of these technologies. Deciphering legal and regulatory implications and establishing accountability would help curb malicious or discriminatory use of AI.

Interwoven with nearly all of these considerations is the sociological importance of building (patient and health care professional) trust in AI systems (Özdemir, 2020). Given the magnitude of the interrelationships, it is unlikely that these issues can be addressed in isolation and entail the need for efforts to acquire expertise and foster collaboration. AI-driven science is also similar to other forms of science as well in that it requires consideration of democratic representation in technology and innovation to address, for example, historical power asymmetries, climate justice, and gender parity. Although ethical and legal dimensions are important, critical governance of the AI suite of technologies also require thinking about the politics of technology and innovation: who

funds and designs AI studies, applications, to what ends, using which means? (**Figure 1.2**).

Applying AI to genomics requires skills, expertise, data, and resources from multiple disciplines. No single discipline has a monopoly on the subject, but there is a significant need for interdisciplinary collaboration. Moreover, there may be a limit to the extent to which tacit human qualities and contexts can be mimicked or constructed by AI, pointing to the indispensable need for human intelligence, which can be defined as “a collection of contextual tacit knowledge about human values, responsibility, empathy, intuition, or care for another living being that cannot be readily described or performed by algorithms” (Özdemir, 2019a). Considering that it is not cutting-edge technologies but people’s datadriven and value-laden decisions that are transforming society, developing rational ways to integrate AI with human intelligence is a must.

## **1.6. Aims and Scope**

The main purpose of the study is to develop an optimal strategy for multi-omics data integration by enriching multi-omics methods with machine learning approaches. In this context, firstly, it is aimed to investigate currently used analysis methods applied onto multi-omics data, and then to compare the success of these methods in biomarker prediction using various cancer data sets. By developing multi-omics data analysis methods including machine learning techniques, potential biomarkers and therapeutics in cancers will be predicted in the most appropriate way. Thus,

1. Comparative analysis of different methods that interpret multi-omics data will be performed.
2. Comparison of biomarker detection performances of these methods on the specified cancer types will be carried out and the advantages and disadvantages of these methods will be revealed.
3. The molecular mechanisms underlying the specified cancers will be enlightened.
4. Candidate biomarkers at different omic levels for early diagnosis or prognosis of the specified cancers will be determined.
5. Various hypotheses (drug targets, etc.) will be put forward for studies to improve the therapy options of the specified cancers.

In addition, the long-term goal of this study is to enable the production of methods, kits, and devices for the diagnosis and treatment of the specified cancers, besides the establishment of the necessary information infrastructure for experimental and clinical studies to be carried out for the diagnosis, prognosis, and treatment processes in cancer.

### 1.7. How to Read This Thesis?

Even though the main goal of this thesis is to identify potential biomarkers and drug targets for cancer through multi-omics data integration analysis, we carried out different systems biology approaches to accomplish this aim. During the thesis, we published a book chapter, a review article and two research articles. In addition, another research article is currently being prepared for publication. All published or submitted materials are presented under three main sub-topics (**Figure 1.3**) in this thesis due to the format rules for PhD thesis submitted to Marmara University Institute for Graduate Studies in Pure and Applied Sciences.

Book chapter and the review article are mainly stated in Introduction part of the thesis. Research articles are presented within two sub-topics under the Background, Methods and Results sections which include separate sub-headings for each sub-topic. Therefore, following relevant sub-headings is kindly advised for seamless reading pattern.

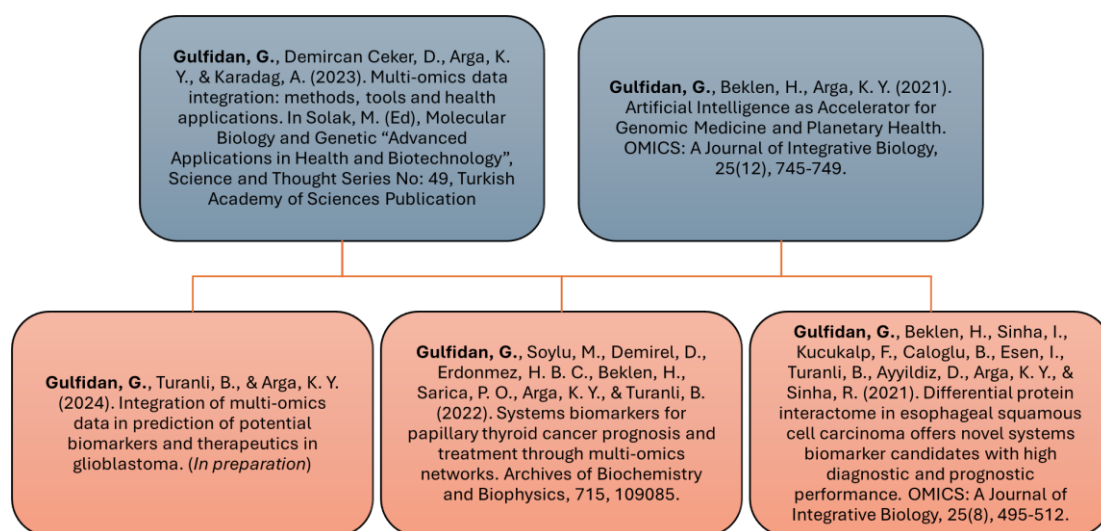


**Figure 1.3.** Sub-topics and publications from this thesis. Book chapter and review (blue), research articles (orange) are presented in the format of the thesis.

## **1.8. Background**

### **1.8.1. Integration of multi-omics data in prediction of potential biomarkers and therapeutics in glioblastoma**

Cancer is a complex disease and the second most common death cause worldwide. Recent statistics in the US depicted that 3% of new estimated deaths in 2024 are by the brain and other nervous system (Siegel et al., 2024). Glioblastoma multiforme (GBM) is one of the most fatal malignancies with a global incidence of approximately 49% of all malignant tumors in all primary adult brain and nervous system tumors. It is estimated that more than 10,000 individuals in the United States will succumb to glioblastoma every year. GBM patients also have lower survival with the average length estimated to be only 8 months (Miller et al., 2021).

It is associated with a dismal prognosis with a short median patient survival of 12– 15 months from diagnosis and a minimal response to therapies (Stupp et al., 2005; Wen and Kesari, 2008). The pathogenesis of this debilitating disease, involving mutations and alterations in cell proliferation, survival, and angiogenesis, is heterogeneous and very complex. The high fatality and recurrence rate of GBM patients compels researchers to pursue novel treatment avenues for a better understanding of GBM pathogenesis (Hanif et al., 2017; Pearson and Regad, 2017).

Over decades, there have only been four drugs and one device ever approved by the FDA specifically for the treatment of glioblastoma. Temozolomide (TMZ) is a key chemotherapeutic agent used in the treatment of GBM despite the challenges of drug resistance. It is an oral alkylating agent that has significantly improved the prognosis for GBM patients when used in combination with radiation therapy. However, drug sensitivity differs for each GBM case (Stupp et al., 2005). Carmustine, and Lomustine which are also alkylating agents and Bevacizumab as monoclonal antibody are other FDA-approved treatment options (Sevastre et al., 2021). Unfortunately, none of these treatments have succeeded in significantly extending patient lives beyond a few extra months.

Molecular evaluation of GBM patients has the utmost importance to understand the molecular mechanism, disease prognosis, and also therapy options to overcome drug

resistance problems. Verhaak et al. (2010) identified four distinct subtypes of samples based on gene expression data comprising TP53, IDH1, IDH2, etc.: Classical, Mesenchymal, Neural and Proneural (Verhaak et al., 2010). According to the 2021 WHO classification, Grade 4 IDH wild-type tumors are considered glioblastoma. Grade 4 IDH mutant tumors are evaluated as high-grade astrocytic tumors, which are now distinguished from glioblastomas (Louis et al., 2021). There are also IDH mutant tumor studies to outline more precise diagnostic, prognostic or therapeutic options (Berzero et al., 2021; Du et al., 2024).

#### **1.8.2. Differential protein interactome in esophageal squamous cell carcinoma offers novel systems biomarker candidates with high diagnostic and prognostic performance**

Esophageal cancer is considered one of the most dangerous cancers worldwide due to its high incidence and mortality rates (Bray et al., 2018). About 70% of esophageal cancer cases occur in men. Esophageal cancer includes two main subtypes: esophageal squamous cell carcinoma (ESCC), which occurs in the squamous cells on the surface of the esophagus and is common in developing countries, and esophageal adenocarcinoma, which occurs in the mucussecreting glandular cells and is commonly observed in developed countries (Griffin and Wahed, 2011; Tungekar et al., 2018). ESCC has a poor prognosis and the long-term survival rate for patients has not improved considerably for the last several decades. Since the precise mechanism of ESCC has not been fully elucidated yet (Cui et al., 2018; Zhang et al., 2020), there is a need for a systems-level understanding of the development and progression of ESCC to improve clinical outcomes.

In recent decades, high-throughput technologies have provided undeniable opportunities to unravel the mechanisms of human disease at multiple molecular levels, enabling the logical shift in cancer treatment toward predictive, preventive, and personalized medicine. The integration and analysis of these multilevel data are essential and play a central role in uncovering disease mechanisms and generating actionable knowledge in the context of precision medicine. The systems biomedicine perspective can shed light on

predictive biomarker discovery by integrating multiomics data and enable predictive diagnostics, targeted prevention, and personalization of current medicine.

Many physiological processes are mediated through cellular pathways influenced by the physical interactions among proteins (Sevimoglu and Arga, 2014). For that reason, decoding the alterations in the protein interactome is critical to reveal a systems-level understanding of tumorigenesis. As it is known, protein-protein interactions (PPIs) are often context specific, which depends on the tissue type, cellular environment, and other dynamically altering phenotypical conditions (Uhlén et al., 2015a; Yeager-Lotem and Sharan, 2015). To date, it has been discovered that there are over 450,000 PPIs between >20,000 human proteins by using various high-throughput proteomics approaches (Oughtred et al., 2019) and various human interactome networks at proteome scale are available (Rolland et al., 2014; Karagoz et al., 2016b; Basha et al., 2018).

The recent studies focused on cancer genes to pave the way for understanding how interactions among these genes can lead to tumorigenesis (Ayyildiz et al., 2017a; Ivanov et al., 2018; Turanli et al., 2019a). The correlations between mRNA and protein levels substantially increase the value of the transcriptomics, allowing the use of genome-wide transcriptomic data as a proxy for prediction of protein levels (Edfors et al., 2016).

We need to pay close attention to the alterations in the interaction patterns of proteins, which provide a differential view of human protein interactome among the various phenotypes to get better discrimination of the molecular mechanisms in specific phenotypes. The concept revealing PPIs that are relatively repressed or activated in each phenotype is called “differential interactome.” This approach enabled the determination of the uncertainty of whether or not a PPI is encountered in the phenotype of interest (Gulfidan et al., 2020b) and its power in revealing new candidate biomarkers. Besides, this method has been effectively used for cancer target discovery, repurposing subtype-specific drug candidates and subtype-specific drug repositioning in several cancers (Ayyildiz et al., 2017b; Gulfidan et al., 2020b; Caliskan et al., 2021).

Over the last decade, the integration of high-throughput omics datasets with the PPI network has attracted great attention to elucidate biological processes and identify disease-associated candidate biomarkers and drug targets. Besides, the individual studies

focused on the neighborhood of specific proteins such as heat shock protein (HSP) (Sun et al., 2018), LOXL4 (Xie et al., 2019b), and the members of scavenger receptor cysteine-rich family (Du et al., 2019), a population-wide compendium of biomarkers was presented through the ESCC Atlas (Tungekar et al., 2018), a database that integrates previously published genomic, transcriptomic, proteomic, and epigenetic genes related to ESCC. Moreover, the proteomic and metabolic signatures of ESCC were elucidated through the integration of transcriptome data with the human transcriptional regulation, PPI, and metabolic networks in our previous study (Karagoz et al., 2016a).

In contrast to these efforts, which focused on biomolecules and the changes in their expression levels between phenotypes, this study provides the first comprehensive view on the PPIs that differ during the transition from a normal to an ESCC tumor phenotype. In this study, we examined the gene expression profiles of 102 individuals with ESCC, used transcriptomic signatures as a proxy for protein levels, and used a systems-level approach to integrate these data with the human protein interactome network (Chatr-Aryamontri et al., 2017b) to analyze the differential interactome of ESCC. These analyses led to the identification of differentially interacting proteins (DIPs) that represent significant alterations in their interaction patterns during the transition from the normal (healthy) to the tumor phenotype, thus generating the differential relevance in the phenotype of interest.

The result of our analysis indicated that a certain number of proteins exhibit differential interactions in tumor cases and have an impact on overall patient survival. Besides, candidate protein modules showing high performance in terms of diagnosis and/or prognosis were identified so that it can be useful to understand the mechanism of tumorigenesis, to improve more precise diagnostic tools, and to develop personalized treatment strategies for patients with ESCC.

### **1.8.3. Systems biomarkers for papillary thyroid cancer prognosis and treatment through multi-omics networks**

Thyroid cancer is a rare type of cancer that affects the thyroid gland, but the estimated number of new cases is rapidly increasing worldwide. Several factors such as age, gender, and being exposed to radiation may escalate the risk of thyroid cancer. Although thyroid

cancer can be diagnosed at any age, women in their 40s or 50s often have a higher risk of cancer than men who are usually in their 60s or 70s. Moreover, thyroid cancer occurred nearly 3 times more often in women (32,130) than in men (12,150) in the United States. Despite the high survival rate in thyroid cancer, the estimated number of deaths in 2021 is still 2200 (Siegel et al., 2021).

There are four fundamental types of thyroid cancer; papillary thyroid carcinoma (PTC, the most common type of thyroid cancer), follicular thyroid carcinoma, medullary thyroid carcinoma, and anaplastic thyroid carcinoma. Papillary and follicular thyroid carcinoma are sometimes called differentiated thyroid cancer. Medullary and anaplastic thyroid carcinoma are referred to as poorly differentiated or undifferentiated thyroid cancer (Shaha, 1998).

There are multiple steps to diagnose the disease which starts with a blood test to check the levels of thyroid hormones. Regarding abnormalities found in blood, ultrasound scan and tissue biopsy are the following steps for the diagnosis. Moreover, computed tomography (CT) and magnetic resonance imaging (MRI) scans can be used to achieve more information about the stage of the cancer (Cabanillas et al., 2016). Early diagnosis is crucial for the treatment of thyroid cancer. The detection of some abnormalities in blood levels such as TSH levels, calcium levels, thyroglobulin levels, T3, and T4 hormone levels has been influenced by the early detection of thyroid cancer.

The recent efforts on molecular characterization of thyroid cancer subtypes influenced the diagnosis and treatment, as well as the molecular markers and analyses, are applied to enhance clinical care. Many attempts have been made to identify thyroid cancer-specific molecular biomarkers including calcitonin, carcinoembryonic antigen, and B-Raf protein encoded by BRAF proto-oncogene, however, they were proven to be not specific enough to distinguish only thyroid carcinomas (Cabanillas et al., 2016).

In parallel with advances in high-throughput technologies and increased omics data, the increasing use of systems biology approaches is becoming essential for integrating heterogeneous data and understanding disease at the system-level. Mapping mRNA expression data to different biological networks including protein-protein interaction networks, transcriptional regulatory networks, or metabolic networks is the quantitative

framework for identifying common disease-state signature (Aydin and Arga, 2019; Kori et al., 2019; Turanli et al., 2019b, 2019d; Caliskan et al., 2020). Alternatively, multi-omics data such as mRNA, miRNA expression or DNA methylation can be integrated to detect potential promising biomarkers or drug target arousing the progress in precision medicine (Xu et al., 2019).

However, the systematic identification of genes associated with PTC is still missing. The earliest effort was only covered 10 differentially expressed genes (DEGs) in PTC because of the technology they utilized (Lee et al., 2009). There have been few systems-based efforts to uncover the molecular mechanisms through the genes associated with the diagnosis and progression of PTC. The gene modules associated with the progression of PTC were found by the integration of a weighted gene co-expression network analysis (WGCNA) and differential expression analysis (Zhao and Li, 2018). In a recent study, a genetic and epigenetic network of protein-protein and protein-DNA interactions was constructed to investigate the pathogenicity and progression mechanisms of PTC and also identified essential biomarkers with specific gene expression signatures (Yeh et al., 2019). In another study, the gene modules were determined by applying WGCNA and differential expression analysis to the genes associated with PTC progression based on TCGA PTC transcriptome sequencing data (Tang et al., 2019).

Meta-analysis of publicly available microarray datasets offers the opportunity to increase the reliability and generalizability of statistical results to provide more robust and accurate findings to determine (Ramasamy et al., 2008). In the present study, an integrative analysis was performed to identify key genes associated with thyroid cancer through a meta-analysis of four publicly available PTC microarray datasets. To achieve significant information on the mechanisms of PTC, we focused on biomolecules from different omics levels. Consequently, this study reports candidate biomolecules that can be considered as prognostic biomarkers or potential therapeutic targets for further experimental and clinical trials for PTC.

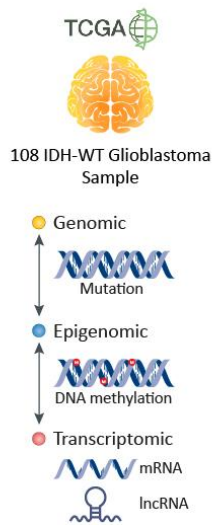
## **2. MATERIALS AND METHODS**

### **2.1. Integration of multi-omics data in prediction of potential biomarkers and therapeutics in glioblastoma**

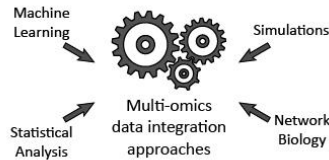
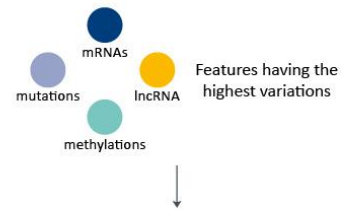
#### **2.1.1. Data collection and preprocessing**

Publicly available data for transcriptome profiling (RNA-Seq raw count and TPM matrix for mRNA and lncRNA), DNA methylation (beta values per CpG), simple nucleotide variation (MAF data), copy number variation, and clinical information were collected for patients suffered from GBM through the NCI's Genomic Data Commons (GDC) portal (Tomczak et al., 2015b) by using "TCGAbiolinks" package in R. The initial data involves 7035 somatic mutations belonging to 4811 genes in 373 tumor samples, DNA copy number variation (180563 segments in 596 tumor samples), DNA methylation (18441 CpGs in 295 samples for 27k array; 335123 CpGs in 155 samples for 450k array) and gene expression (60660 probes in 155 tumor samples). After obtaining all type of data, preprocessing step was performed by creating several matrices for each type of omics data in a way that features on rows and samples in columns. Only 108 GBM patients, consisting of 43 females and 64 males with grade 4, that were common to all data types and carried IDH-wt mutations were included in the study for further analyses which all were performed through the "MOVICS" package in R (Lu et al., 2020) (**Figure 2.1**).

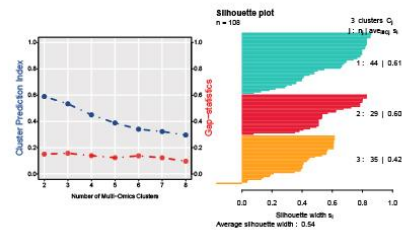
## Obtaining Omics Data



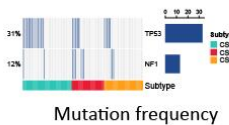
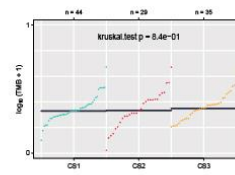
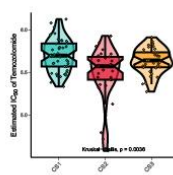
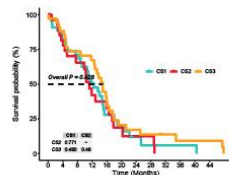
## Multi-omics Data Integration



### Identification of subtypes through clustering analysis



## Comparative analyses of subtypes



Biomarkers

Drug targets

Candidate drugs

**Figure 2.1.** Graphical abstract indicating the multi-omics data integration method.

### 2.1.2. Reducing data dimension and clustering analysis

Features were filtered out that meet some stringent requirements, and that features called elites. They were determined using median absolute deviation for mRNA, lncRNA, and methylation data and frequency for binary omics data of mutation. Also missing values coded in NA were handled by imputing them by k nearest neighbors using a Euclidean metric. All these elites are used in the following clustering analyses.

The data of different omics layers (mRNA, lncRNA, methylation and mutation) of 108 GBM samples were integrated for identification of distinct molecular subtypes for GBM. To ensure the detection of optimal number of molecular subtypes, seven different scenarios of true number of clusters,  $k = 2:8$ , were simulated by using the elites and Cluster Prediction Index (CPI) (Chalise and Fridley, 2017a) and Gaps-statistics (Tibshirani et al., 2001) were calculated for clustering analysis of GBM samples.

### **2.1.3. Multi-omics integration of clusters**

Ten separate multi-omics integrative clustering algorithms (iClusterBayes (Mo et al., 2018), moCluster (Meng et al., 2016), CIMLR (Ramazzotti et al., 2018), IntNMF (Chalise and Fridley, 2017b), ConsensusClustering (Monti et al., 2003), COCA (Hoadley et al., 2014), NEMO (Rappoport and Shamir, 2019), PINSPlus (Nguyen et al., 2019), SNF (Wang et al., 2014b), and LRA (Wu et al., 2015a)) with parameters by default were carried out for the optimal number of clusters and a consensus matrix, which is a probability matrix signifying how many times samples belonging to the same subtype can be clustered together by different multi-omics clustering methods and displays the clustering robustness across different clustering algorithms. Then, the quantification of sample similarity within clusters were assessed by the silhouette score approach.

Multi-omics data were normalized to create a comprehensive multi-omics heatmap. The methylation  $\beta$ -values were converted to M value with the formula of  $M = \log_2(\beta/(1-\beta))$  for stronger visualization, gene expressions and methylation data were then scaled and centered to normalize, and finally a comprehensive multi-omics heatmap based on clustering results and normalized data were generated.

### **2.1.4. Comparisons of glioblastoma subtypes**

Once cancer subtypes have been identified, it's crucial to delve deeper into each subtype to understand their unique characteristics from various perspectives. To this end, subtypes were compared firstly according to clinical and prognosis. Kaplan-Meier curves containing the overall nominal P value calculated by log-rank test and adjusted-p values resulted from pairwise comparison were generated using clusters results together with the information of survival time (unit of day) and survival outcome (0 = censor; 1 = event).

Also, the clinical features were compared among different subtypes by performing appropriate statistical tests.

Mutational frequencies were statistically compared among different subtypes based on binary mutational data (0 = wild; 1 = mutated). Chi-square ( $\chi^2$ ) test was applied for each mutation to identify differentially mutated genes, and Benjamini-Hochberg approach as the correction method was used to determine adjusted p-value. In addition, the subtypes were evaluated in terms of total mutation burden (TMB) using MAF data and fraction genome altered (FGA) using segmented copy number to quantify the genomic alterations.

The subtypes were compared to estimated drug sensitivity (IC50) values through an R package “pRRophetic” developed by Paul Geeleher (Geeleher et al., 2014) by creating statistical models from the gene expression and drug sensitivity data from cell lines in the Cancer Genome Project (CGP) (Garnett et al., 2012) in order to estimate the response of patients to drugs.

At present, glioblastoma have traditional classifications as classical, mesenchymal, neural, preneural (Verhaak et al., 2010). To determine potential relation of subtypes with traditional classifications, the similarity between the current subtypes and other pre-existed classifications was calculated by four statistical methods ranging from 0 to 1 (Rand Index (RI) (Rand, 1971), Adjusted Mutual Information (AMI) (Xuan Vinh et al., 2010), Jaccard Index (JI) (Hancock, 2004), and Fowlkes-Mallows (FM) (Fowlkes and Mallows, 1983). An alluvial diagram was created to visualize the similarities between the current subtypes and pre-existed subtypes.

#### **2.1.5. Differential analysis based on mRNA, lncRNA and methylation**

Differentially expressed analysis (DEA) was conducted to compare transcriptome profiling among different subtypes to identify their potential predictive biomarkers. To perform DEA of each subtype compared to others, DESeq2 (Love et al., 2014a) was applied on RNA-Seq count data to determine differential expressed mRNAs (DEmRNAs), and differential expressed lncRNAs (DElncRNAs). To eliminate false positive DEmRNAs, and DElncRNAs, p-value correction was made with the Benjamini-Hochberg correction method and the cutoff value for adjusted p-values was taken 0.05 as

statistical significance. Down-regulated and up-regulated DEmRNAs, and DElncRNAs not overlapping with any biomarkers identified for other subtypes were determined as candidate biomarkers by taking the fold change threshold value as 1-fold.

For the differential analysis of the methylation data,  $\beta$ -values were firstly converted to M-values due to its more statistically valid for the differential analysis of methylation levels of CpG probes (Xie et al., 2019a). To minimize the unwanted variation within and between samples, quantile data normalization was applied for samples obtained from HumanMethylation27 and HumanMethylation450 array data, separately. To make pairwise comparison of methylation profiles among the subtypes, the limma method (Ritchie et al., 2015) was used on the matrix of M-values to call significant differentially methylated CpG probes (DMPs). The threshold of adjusted p-value corrected by Benjamini-Hochberg method was selected 0.05 and hypomethylated and hypermethylated CpG probes in subtypes were identified according to the threshold of  $|\log_{2}FC| > 1$ . The annotation for methylation of CpG sites to genes was performed based on the information provided by Illumina.

#### **2.1.6. Gene set enrichment analysis of mRNA, lncRNA and methylation**

For each subtype, Gene Set Enrichment Analysis (GSEA) was performed using the corresponding DEA results to identify functional pathways and biological processes specific to that subtype. To this end, gene sets related to GO biological processes (c5.bp.v7.1.symbols.gmt) and molecular pathways (c2.cp.v2023.2.Hs.symbols.gmt) were obtained from The Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). The canonical pathway databases of KEGG, Reactome, PID (Pathway Interaction Database), WikiPathways were used to identify molecular pathways where gene sets involve in. The significance threshold was selected as adjusted P-value  $< 0.05$  and subtype-specific pathways and processes were determined. Genes which were inside these pathways and processes were identified and single sample enrichment scores and subtype-specific enrichment score represented by the mean value within the subtype were calculated by using GSVA R package.

## **2.2. Differential Protein Interactome in Esophageal Squamous Cell Carcinoma Offers Novel Systems Biomarker Candidates with High Diagnostic and Prognostic Performance**

### **2.2.1. Gene expression data**

To analyze the gene expression profiles of transcriptome data in ESCC, the microarray dataset (GSE23400) (Su et al., 2011), including expression profiles of 38,729 probes from 51 ESCC tumor samples and 51 match normal samples, was obtained from GEO database (Barrett et al., 2013).

### **2.2.2. PPIs data**

Protein interactome data, which were physically and experimentally detected in humans and consisted of 35,688 PPIs among 8570 human proteins, were acquired from the BioGRID database (v.3.5.167) (Chatr-Aryamontri et al., 2017a). A network comprising 33,736 PPIs among 8138 proteins was obtained after the filtration of human protein interactome data with proteins that were encoded by the genes denoted in the transcriptome dataset.

### **2.2.3. Identification of differential interactome**

The differential interactome algorithm was executed using R (version 3.5.2) on the gene expression profiles in ESCC to evaluate the relative observation frequency (q-value) for each PPI, as stated previously (Ayyildiz et al., 2017c; Gulfidan et al., 2020a). Briefly, genes were categorized into three levels according to their expression levels within each sample: the expression level of a gene was assigned as 1 or -1, if its expression was significantly higher or lower than the average expression level of the genes in that sample, otherwise it was assigned as 0. Considering this formulation, nine possible states were defined for each interacting protein pair, that is, [0 0], [0 1], [0 - 1], [1 0], [1 1], [1 - 1], [-1 0], [-1 1], and [-1 - 1]. For each state, counts were made of how many times ( $N_0$ ) the state occurred in normal samples and how many times ( $N_1$ ) it occurred in tumor samples. Counts were normalized by the total sizes of normal ( $N_C$ ) and tumor ( $N_T$ ) samples. The relative observation frequency for each PPI for each state is estimated by q-value:

$$q = \frac{\frac{N_1}{N_T}}{\frac{N_0}{N_C} + \frac{N_1}{N_T}}$$

For PPIs to be considered differential PPIs (dPPIs), they should meet the following criteria: (1) q-value greater than 0.90 (significantly activated in tumor phenotype) or <0.10 (significantly repressed in tumor phenotype) and (2) normalized observation frequency in either healthy (normal) or tumor phenotype greater than 20%. The PPI networks were visualized using Cytoscape 3.4.0 software (Shannon et al., 2003).

#### **2.2.4. Categorization of DIPs**

The interaction patterns were determined based on whether interactions between proteins were repressed or activated under the tumorous condition. DIPs were categorized into two groups according to their interaction patterns: Class I includes the proteins that have repressed interactions under tumor condition, while Class II represents the proteins that have activated interactions under tumor condition. Both DIPs and their interacting proteins were named as DIP-centered modules. Each module was statistically evaluated using Kruskal–Wallis test, which compares the observed frequencies of dPPIs among control (healthy) and tumor states, and significant modules were determined based on p-value <0.05.

#### **2.2.5. Functional overrepresentation analysis**

Overrepresentation analyses were carried out using ConsensusPathDB (Kamburov et al., 2013) to recognize functional annotations such as metabolic and signaling pathways, biological processes, and molecular functions significantly related to the gene products. In the analyses of the pathways, the databases of Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al., 2017) and Reactome (Croft et al., 2011) were preferably employed, whereas Gene Ontology (GO) terminology (Ashburner et al., 2000) was used as the source to annotate biological processes. p-values were obtained to determine significant biological processes and pathways through Fisher's exact test followed by Benjamini-Hochberg's correction as the multiple testing correction method. The enrichment results having adjusted  $p < 0.05$  were considered statistically significant.

### 2.2.6. Analysis of diagnostic performance

To evaluate the diagnostic performance of DIPs in each DIP-centered module, principal component analyses (PCA) were executed on the expression values of genes encoding these DIPs. The principal components, which describe at least 80% of total variance, were considered to determine sensitivity, specificity, and diagnostic odds ratio (DOR) metrics. Simulations were performed using an independent validation dataset (GSE13898 with 70 samples) (Kim et al., 2010) as well as the test dataset (i.e., GSE23400 with 102 samples) (Su et al., 2011). The validation dataset (GSE13898) included 28 healthy samples; to keep the sample sizes of healthy and ESCC states the same in the analyses, the randomly selected 28 samples out of the 42 ESCC samples were employed in each simulation, and simulations were repeated several times until the stability was ensured in the performance metrics.

### 2.2.7. Analysis of prognostic power

To discover the prognostic power of each DIP-centered module, survival analyses were performed using Surv-Express tool (Aguirre-Gamboa et al., 2013) using two independent datasets, including clinical data; RNA-Seq originated The Cancer Genome Atlas (TCGA) dataset containing 184 ESCC tumor samples (Tomczak et al., 2015a) and the microarray dataset from GEO (GSE19417) containing 70 tumor samples (Peters et al., 2010). The subjects were divided into two groups (low and high risk) based on their prognostic index (PI), which is the linear component of the Cox model ( $PI = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ , where  $\beta_i$  is the coefficient obtained from the Cox fitting and  $x_i$  is the expression value of each gene). The signatures of survival in each group were estimated by Kaplan–Meier plots and hazard ratios (HR). Statistical significance of each plot was evaluated by the cutoff for log-rank p-value <0.05. HR ( $HR = O_1/E_1/O_2/E_2$ ) was calculated to discover the significance of survival plots according to the rate between the relative death rate in group 1 ( $O_1/E_1$ ) and the relative death rate in group 2 ( $O_2/E_2$ ), where O represents the observed number of deaths and E represents the expected number of deaths.

#### **2.2.8. Evaluation of the secretion levels of candidate biomarkers in body fluids and tissues**

For the proteins represented in diagnostic modules, the secretion levels (ppm) in serum, plasma, saliva, urine, oral epithelium, and hair follicle were investigated through protein expression data presented in the proteomics databases, namely ProteomicsDB (Samaras et al., 2020), MaxQB (Schaab et al., 2012), MOPED (Montague et al., 2014), and GeneCards (Stelzer et al., 2016). A protein secretion level of at least 1 ppm was considered to be included in the data analysis.

#### **2.2.9. Identification of candidate drugs through drug repositioning**

To identify potential therapeutic agents for ESCC, we used geneXpharma tool (Turanli et al., 2017), which reflects disease-gene-drug associations and provides statistically evaluated gene expression profiles from 118 disease-specific datasets and their drug interactions by using comprehensive drug-gene interaction data for 48 diseases under 7 different disease categories. The tool was designed to generate hypotheses according to p-values of the differential expression levels of mRNAs and drug-gene interactions in a specific dataset. The calculation for differential expression analysis was carried out as described in previous (Calimlioglu et al., 2015; Karagoz et al., 2015; Aydin and Arga, 2019; Kori et al., 2019). In the prediction of drug–dataset associations, the statistical linkage method (Amelio et al., 2014) was adopted to gene expression data through a statistical test employing the hypergeometric probability density function (Turanli et al., 2017).

#### **2.2.10. Molecular docking analysis for repurposed drugs**

The 3D structures of target proteins and drug candidates were retrieved from Protein Data Bank (PDB) (Berman et al., 2002) and PubChem (Kim et al., 2019), respectively. The binding sites of the target proteins were taken from PDBe (Velankar et al., 2016). Molecular docking analysis was performed for each target protein and drug candidate, with a total of 160 molecular dockings by Autodock Vina (ver.1.1.2) (Trott and Olson, 2009). In the target protein preparation step, all water molecules were removed, and all polar hydrogens were added to the macromolecule. Also, exhaustiveness was set to eight

for each simulation. Binding affinities were calculated to determine the significance of binding between protein targets and drug candidates. The highest scoring binding affinities (lower than -8 kcal/ mol) were selected for further analysis.

#### **2.2.11. Evaluating repurposed drugs in human ESCC cell lines**

#### **2.2.12. Cell lines and treatments**

All the cell lines were cultivated in well-defined optimum conditions for growth at 37C and 5% CO<sub>2</sub>. TE-5, TE-11, and TE-15 cells are established human ESCC cell lines (obtained from Dr. Douglas Stairs at Penn State College of Medicine, Hershey, PA) maintained in Dulbecco's modified Eagle's medium (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Gemini Bioproducts, West Sacramento, CA) and 1% streptomycinpenicillin (Invitrogen). For treatments, two repurposed drugs based on their binding affinity for proteins analyzed in molecular docking analysis. Risperidone and Clozapine were purchased from Sigma-Aldrich (St. Louis, MO) and incubated with all cell types at 0, 50, 100, and 200  $\mu$ M doses, respectively, for 24 h.

#### **2.2.13. Cell proliferation (MTT Assay)**

Cell proliferation was measured in the three ESCC cell lines following treatments. Briefly, these cell types ( $1 \times 10^4$ /well) were plated on 96-well plates for 24 h and treated with varying doses of Risperidone and Clozapine and methanol as a control. Following the treatments, cells were incubated with 3-(4,5-dimethylthiazol- 2-yl)-2,5-diphenyltetrazolium bromide (MTT; Sigma, St. Louis, MO) (50  $\mu$ g/well) for 3 h in dark at 37C. MTT solution was eliminated from the wells and was replaced by 100  $\mu$ L of DMSO/well to dissolve the purple-blue formazan particles and in Molecular Devices plate reader, the plate was read at 570 nm with correction at 630 nm. MTT assay was executed in triplicate for each treatment dose and the results for the assay were presented as % of control. The MTT assay was compared between control and two drugs at each of the doses by Student's t-test and  $p < 0.05$  were considered significant.

## **2.3. Systems Biomarkers for Papillary Thyroid Cancer Prognosis and Treatment through Multi-Omics Networks**

### **2.3.1. Gene expression datasets of papillary thyroid carcinoma**

Four microarray datasets (GSE3467, GSE3678, GSE6004, GSE29265) from independent studies were obtained from the Gene Expression Omnibus (NCBI-GEO) database [15] to meta-analyze the gene expression profiles in PTC. The datasets were chosen according to including samples from two different phenotypes (normal vs. tumor) having at least 3 samples from each one without any treatment or perturbation. A total of 63 (30 normal, and 33 tumor) samples were examined, including 9 normal and 9 tumor samples from GSE3467 [16], 7 normal and 7 tumor samples from GSE3678, 4 normal and 7 tumor samples from GSE6004 [17], and 10 normal and 10 tumor samples from GSE29265. All datasets included in the study were conducted under the platform of Affymetrix Human Genome U133 Plus 2.0 Array. While GEO datasets used as discovery datasets, PTC from TCGA used as validation dataset.

### **2.3.2. Identification of differentially expressed genes**

DEGs were determined by statistically analyzing the microarray datasets according to the previously published pipeline (Calimlioglu et al., 2015; Kori et al., 2016) that employs Robust MultiArray Average (RMA) normalization (Irizarry et al., 2003) and linear models for microarray data (LIMMA) method (Smyth, 2005) under R/Bioconductor platform (Gentleman et al., 2004). The false discovery rate (FDR) was controlled by Benjamini-Hochberg's correction and the results with adjusted-  $p < 0.01$  were considered significant. To determine regulatory patterns of DEGs, fold-change cut-offs were used as 1 fold (i.e. fold change  $> 1$  for upregulation or fold change  $< 1$  for downregulation). Each dataset was analyzed independently, and results were comparatively analyzed to declare common signatures from these independent studies. Common DEGs determined in all data sets were validated with DEGs determined by differentially expressed analysis of the TCGA-THCA dataset having 552 samples (494 tumor versus 58 normal) of patients suffered from papillary thyroid carcinoma through DESeq2 (Love et al., 2014b).

### **2.3.3. Identification of reporter receptors, transcription factors, and miRNAs**

The reporter molecules were identified using the reporter features algorithm (Patil and Nielsen, 2005), which was previously adapted for potential receptors, transcription factors (TFs), and miRNAs (Kori and Arga, 2018). TF-target gene and miRNA-target gene networks were generated based on the information from the combinatorial human transcriptional regulatory interaction network (Gov and Arga, 2016), Human Transcriptional Regulation Interactions database (HTRIdb) (Bovolenta et al., 2012) and miRTarbase (release 6.0) (Chou et al., 2016). The protein-protein interactions for receptors were obtained from the high confidence physical protein-protein interaction (PPI) data for humans (with confidence score  $\geq 0.8$ ) (Karagoz et al., 2016c). P-values of the calculated reporter molecules were controlled with FDR and the reporter molecules with adjusted-  $p < 0.05$  were considered significant.

### **2.3.4. Identification of reporter metabolites associated with thyroid cancer**

Reporter metabolites were identified through the integration of differential gene expression levels with the tissue-specific genome-scale metabolic network for the thyroid gland (Uhlén et al., 2015b) via BioMet Toolbox (v2.0) (Garcia-Albornoz et al., 2014), and reporter metabolites with  $p < 0.05$  were determined as statistically significant.

### **2.3.5. Gene set overrepresentation analyses**

Overrepresentation analysis of DEGs was carried out through ConsensusPathDB (Kamburov et al., 2009) to find out functional annotations (i.e., biological processes, molecular functions, and pathways) significantly associated with DEGs. In the analyses, KEGG (Kanehisa et al., 2017), Biocarta (Nishimura, 2001), Reactome (Joshi-Tope et al., 2005), and PharmGKB (Whirl-Carrillo et al., 2012) were employed as the data sources for pathways; Gene Ontology (GO) terminology (Ashburner et al., 2000) was used as the data source of biological processes and molecular functions. The overrepresentation analysis of reporter metabolites was performed via MBROLE (v2.0) database (Chagoyen and Pazos, 2011) to find out KEGG metabolic pathways.  $p < 0.05$  was taken into consideration to describe the statistical significance in all analyses.

### **2.3.6. Reconstruction of protein-protein interaction networks and topological analysis**

The high confidence physical protein-protein interaction (PPI) data for humans (with confidence score  $\geq 0.8$ ) (Karagoz et al., 2016c) was used to determine PPIs among proteins encoded by DEGs. The visualization and topological analysis of the PPI network were carried out through Cytoscape (v3.7.0) (Shannon et al., 2003) and the dual-metric approach (i.e. degree as a local metric, and betweenness centrality as a global metric) was applied to the PPI networks of each dataset to determine hub proteins using Cytohubba plugin (Chin et al., 2014). The top 20 hub proteins obtained for each metric in each dataset were merged for further analysis.

### **2.3.7. Prognostic features of reporter molecules and hub proteins**

The common TFs, receptors, miRNAs, and hub proteins among all datasets were investigated in terms of their prognostic potential via survival analyses. Survival analyses were performed for each set of common reporter molecules utilizing from the samples of thyroid carcinoma project (THCA) from the TCGA database via the SurvExpress tool and database (Aguirre-Gamboa et al., 2013) having 489 samples from thyroid carcinoma patients for the common TFs, receptors, and hub proteins and via the OncomiR database (Wong et al., 2018) having 505 samples from thyroid carcinoma patients for the common miRNAs. To discover the prognostic power of each molecule sets, survival analyses were performed by dividing the subjects into two groups (high- and low-risk) according to their prognostic index (PI) which is the linear component of the Cox model and illustrated by Kaplan-Meier plots. Each plot with log-rank p-value  $< 0.05$  was considered as statistically significant and hazard ratio ( $HR = (O_1/E_1)/(O_2/E_2)$ ) was calculated to discover the significance of each survival plot based on the ratio between the relative death rate in group 1 ( $O_1/E_1$ ) and group 2 ( $O_2/E_2$ ) (O: observed number of deaths, E: expected number of deaths).

### **2.3.8. Determination of candidate drugs through drug repositioning**

Potential therapeutic agents were investigated for PTC through the Drug Gene Budger (DGB) (Wang et al., 2019), web-based application, that helps researchers to assess

predicted small molecules, which will downregulate or upregulate the expression of interested genes, with p-value, q-value, fold change and specificity indicating their statistical significance by the LIMMA method. Candidate therapeutic agents for the common hub proteins in all datasets were researched using experimental data from the original Connectivity Map (CMap) dataset (Lamb et al., 2006) which was compiled by the DGB and included systematically profiled transcriptomic response of several human cell lines to treatment with over 20,000 small molecules containing almost all FDA-approved drugs and many preclinical compounds.

### **2.3.9. Molecular docking analysis for repurposed drugs**

Protein Data Bank (PDB) (Berman et al., 2002) and PubChem (Kim et al., 2019) were respectively used to determine the 3-D structures of target proteins (PDB id: 4znx for FYN, and 1s9k for JUN) and repurposed drug candidates, and the binding sites of the targeted proteins were investigated from PDBe (Velankar et al., 2016). Molecular docking analyses were carried out to indicate the interaction between each target protein and drug candidate through Autodock Vina (ver.1.1.2) (Trott and Olson, 2009). All water molecules were removed, and all polar hydrogens were added to the macromolecule in the preparation stage of target protein. In each simulation, exhaustiveness was adjusted to eight. After the analyses, binding affinities were calculated for identification of the significance of interaction between each targeted protein and drug candidates, and the best scoring affinities (lowest binding affinity) (kcal/mol) were selected for further analyses according to the binding affinities of targeted proteins to their inhibitors.

### 3. RESULTS AND DISCUSSION

#### 3.1. Integration of multi-omics data in prediction of potential biomarkers and therapeutics in glioblastoma

##### 3.1.1. Identification of GBM subtypes through clustering

After curation and preprocessing of the data obtained from different omics levels of glioblastoma cohort, features were filtered out for 108 patients having information in all levels through the dimensionality reduction. Top 500 mRNAs, 500 lncRNA, 1,000 promoter CGI probes with high variation were extracted calculating a score based on median absolute deviation, and 12 genes that frequently mutated in at least 10% of the entire cohort were identified (**Table S1**).

The estimation of optimal number for clustering that would reduce noise and retain essential information was made based on elites consisting of filtered features. The optimal subgroup number was decided as 3 according to the calculations of Gaps-statistics and CPI (**Figure 3.1A**). The 10 state-of-the-art multi-omics clustering algorithms were performed to create a consensus clustering by returning a consensus matrix (a probability matrix represents how many times samples belonging to the same subtype can be clustered together by different multi-omics clustering methods). Consequently, 108 GBM patients were divided into three well separated clusters (subtypes) (CS1, n = 44, CS2, n = 29, and CS3, n = 35), and sample similarity was quantified using silhouette score based on consensus ensembles result and average silhouette width were determined as 0.54 based on these three clusters, where the silhouette score of CS1 was 0.61, CS2 was 0.60, and CS3 was 0.42 (**Figure 3.1B**). A corresponding consensus heatmap was generated (**Figure 3.1C**). Three of these ten algorithms (iClusterBayes, CIMLR, and MoCluster) provided feature selection procedure, while remains provided hierarchical clustering of samples.

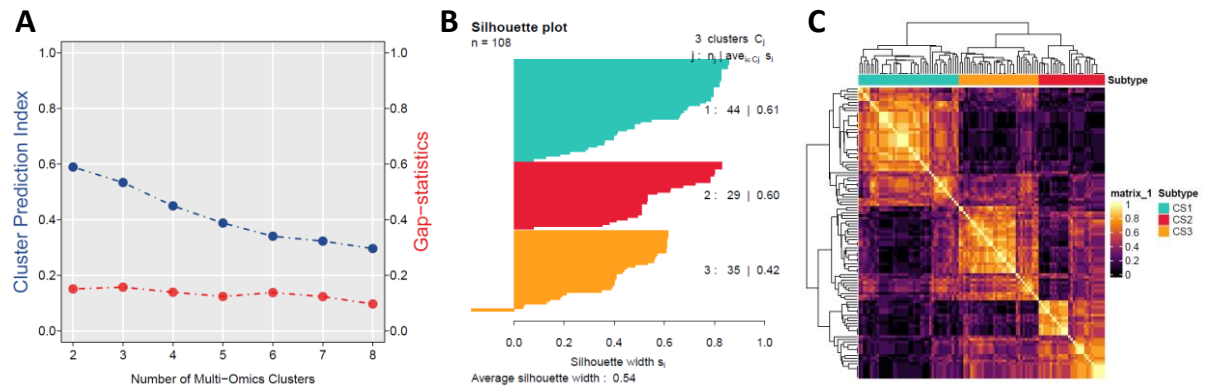


**Figure 3.1.** Identification of GBM subtypes **(A)** Identification of optimal cluster number by calculating CPI (blue line) and Gaps-statistics (red line) in glioblastoma cohort. **(B)** Quantification of sample similarity using silhouette score based on consensus ensembles result. **(C)** Consensus heatmap based on results from 10 multi-omics integrative clustering algorithms with cluster number of 3.

The distributions of mRNA, lncRNA, DNA methylation CpG sites, and mutant genes were visualized in a heat map as shown in Figure 2 after the classification of the multi-omics data. Top 10 features of each omics level are displayed in the right of the plot and the clinical features containing gender, age, and original subtype of GBM are also annotated in the heat map.

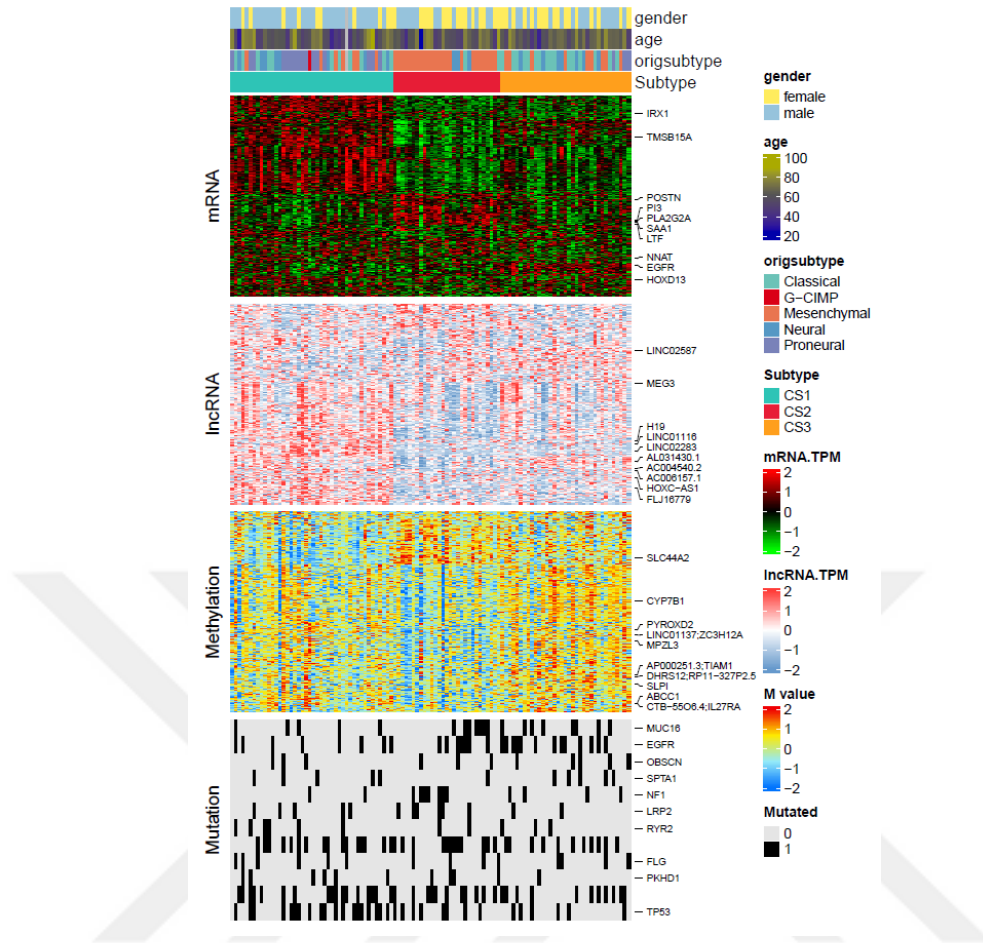


**Figure 3.2.** Comprehensive heatmap based on consensus across 10 algorithms with clinicopathological annotation.

### 3.1.2. Comparisons of identified GBM subtypes

Once GBM subtypes were identified, several characteristics of each subtype should be explored from multiple aspects with downstream analyses provided by MOVICS. First of all, we compared the prognosis potential among different GBM subtypes by generating a Kaplan-Meier Curve using survival time, which referred follow-up time if patients were alive, otherwise referred death time. For the statistical significance, the overall nominal p-value was calculated for all subtypes and also pairwise comparisons were made among different subtypes by log-rank test. The outcome of clinical prognosis indicated that patients in the CS1, CS2 and CS3 groups had no significantly difference in average overall survival time (12.5, 11.2 vs. 14.9 months,  $P = 0.426$ , **Figure 3.3A**). In addition, comparison of the several clinicopathological features performed by appropriate

statistical tests among different subtypes was summarized in the Table 1. CS1 group consisted of a higher number of male patients ( $P = 0.012$ ), while the distribution of age, survival status, survival time, and MGMT promoter showed no significant differences (all  $P > 0.05$ ).

Currently, various cancers have conventional classifications, and it is crucial to assess how well new subtypes align with existing ones to ensure robust clustering analysis and identify novel subtypes. To this end, the similarity between the current subtypes and pre-existed classifications for GBM (original and transcript subtypes) was measured through four statistical methods such as RI, AMI, JI, and FM which indicated that the larger the value is, the more similar the two evaluates are. According to this similarity calculation and the clinical feature comparisons, it was demonstrated that 37.2% of patients in CS1 subtype belonged to the proneural as original subtype; CS2 appeared closer to the mesenchymal as both original (82.8%) and transcription (100%) subtypes, whereas CS3 appeared closer to the classical as both original (45.7%) and transcription (70.6%) subtypes ( $P < 0.001$ ) (**Figure 3.3B-C, Table 3.1**).

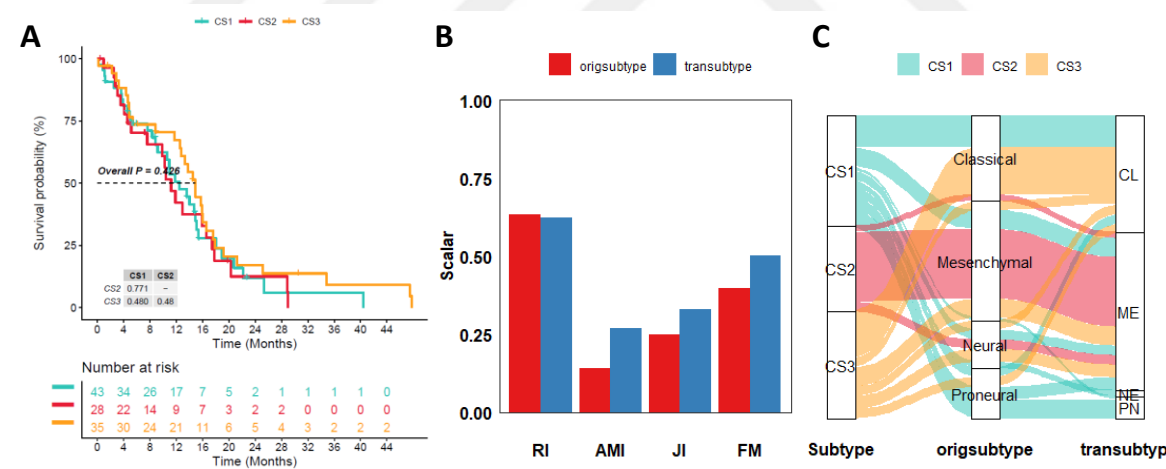


**Figure 3.3.** Comparison of clinical features of GBM subtypes. **(A)** Kaplan-Meier survival analysis of the overall survival identified subtypes of GBM. **(B)** Agreement of identified subtypes of GBM with traditional classification and pathological stage in GBM. **(C)** The distribution of the clusters with GBM original and transcript subtypes.

**Table 3.1.** Comparison of clinical features among identified subtypes of GBM.

|                                 | Level        | CS1                        | CS2                        | CS3                        | P-value | Test    | Significance |
|---------------------------------|--------------|----------------------------|----------------------------|----------------------------|---------|---------|--------------|
| n                               |              | 44                         | 29                         | 35                         |         |         |              |
| Age                             |              | 62.30 ± 13.64              | 61.72 ± 13.01              | 63.46 ± 9.97               | 0.82    |         |              |
| Gender (%)                      | Female       | 10 (23.3)                  | 16 (55.2)                  | 17 (48.6)                  | 0.012   |         |              |
|                                 | Male         | 33 (76.7)                  | 13 (44.8)                  | 18 (51.4)                  |         |         |              |
| Survival status (%)             | 0            | 11 (25.6)                  | 6 (21.4)                   | 5 (14.3)                   | 0.471   |         |              |
|                                 | 1            | 32 (74.4)                  | 22 (78.6)                  | 30 (85.7)                  |         |         |              |
| Survival time<br>(median [IQR]) |              | 279.00<br>[144.50, 453.50] | 300.00<br>[138.00, 505.00] | 419.00<br>[156.00, 521.50] | 0.309   | Nonnorm |              |
| Original subtype (%)            | Classical    | 11 (25.6)                  | 2 (6.9)                    | 16 (45.7)                  | <0.001  | Exact   | **           |
|                                 | G-CIMP       | 1 (2.3)                    | 0 (0.0)                    | 0 (0.0)                    |         |         |              |
|                                 | Mesenchymal  | 7 (16.3)                   | 24 (82.8)                  | 9 (25.7)                   |         |         |              |
|                                 | Neural       | 8 (18.6)                   | 3 (10.3)                   | 7 (20.0)                   |         |         |              |
|                                 | Proneural    | 16 (37.2)                  | 0 (0.0)                    | 3 (8.6)                    |         |         |              |
| Transcription<br>subtype (%)    | Classical    | 13 (36.1)                  | 0 (0.0)                    | 24 (70.6)                  | <0.001  | Exact   | **           |
|                                 | Mesenchymal  | 13 (36.1)                  | 27 (100.0)                 | 10 (29.4)                  |         |         |              |
|                                 | Neural       | 3 (8.3)                    | 0 (0.0)                    | 0 (0.0)                    |         |         |              |
|                                 | Proneural    | 7 (19.4)                   | 0 (0.0)                    | 0 (0.0)                    |         |         |              |
| MGMT promoter (%)               | Methylated   | 14 (31.8)                  | 9 (32.1)                   | 18 (51.4)                  | 0.151   |         |              |
|                                 | Unmethylated | 30 (68.2)                  | 19 (67.9)                  | 17 (48.6)                  |         |         |              |

Gene mutations and copy number alterations significantly impact the early stages and progression of tumorigenesis. Therefore, we compared the genetic alterations in CS1, CS2, and CS3 subtypes to elucidate the genetic mechanism among different subtypes. Subtype-specific mutation might be promising as potential therapeutic target. When we compared the somatic mutation frequency, it was indicated that TP53 and NF1 were differentially mutated genes detected by Chi-square test and the number of mutations were significantly higher in TP53 in CS1 (43.2%,  $P = 0.037$ ), while NF1 had significantly higher mutations in CS2 (24.1%,  $P = 0.043$ ) (**Figure 3.4A, Table 3.2**).

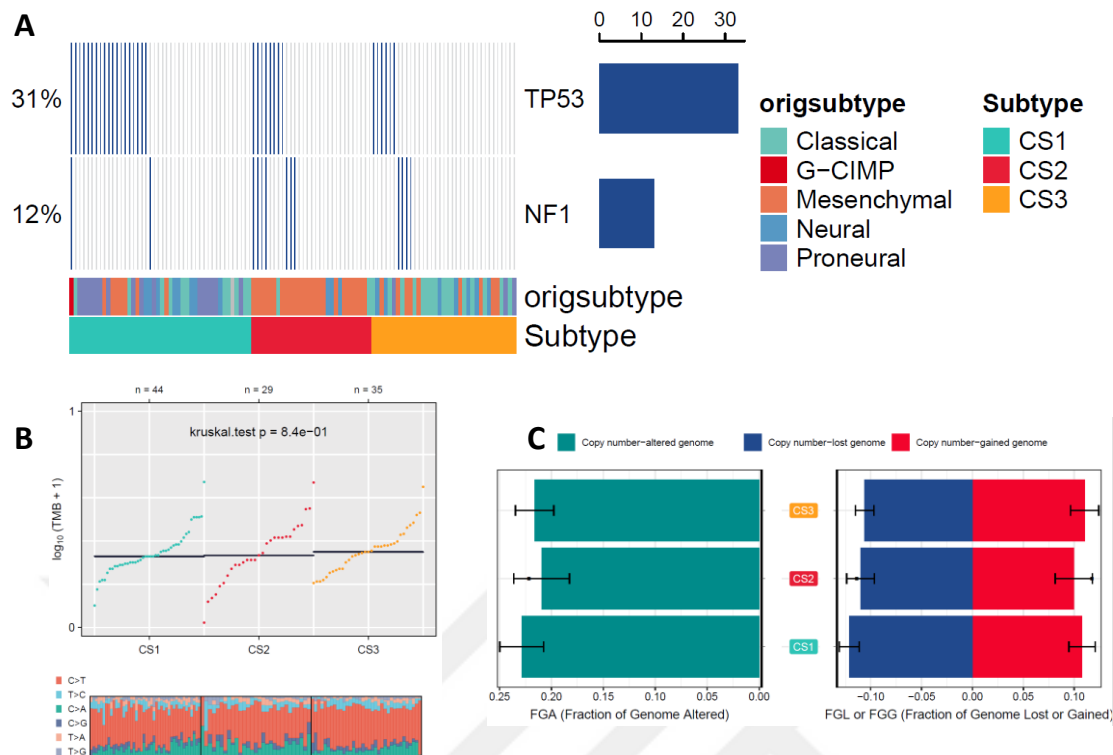


**Figure 3.4.** Comparison of genomic alterations among identified subtypes of GBM. **(A)** Mutational OncoPrint of identified subtypes of GBM. **(B)** Comparison of TMB and TiTv among identified subtypes of GBM. **(C)** Bar plot of fraction genome altered among identified subtypes of GBM.

**Table 3.2.** Comparison of mutational frequency among identified subtypes of GBM.

| Gene (Mutated) | Total mutation  | CS1               | CS2              | CS3              | pvalue       |
|----------------|-----------------|-------------------|------------------|------------------|--------------|
| OBSCN          | 11 (10%)        | 3 (6.8%)          | 3 (10.3%)        | 5 (14.3%)        | 0.567        |
| TTN            | 39 (36%)        | 16 (36.4%)        | 12 (41.4%)       | 11 (31.4%)       | 0.718        |
| MUC16          | 20 (18%)        | 5 (11.4%)         | 7 (24.1%)        | 8 (22.9%)        | 0.263        |
| FLG            | 14 (13%)        | 6 (13.6%)         | 4 (13.8%)        | 4 (11.4%)        | 1.000        |
| LRP2           | 13 (12%)        | 5 (11.4%)         | 6 (20.7%)        | 2 (5.7%)         | 0.171        |
| PTEN           | 39 (36%)        | 17 (38.6%)        | 11 (37.9%)       | 11 (31.4%)       | 0.794        |
| <b>TP53</b>    | <b>33 (31%)</b> | <b>19 (43.2%)</b> | <b>8 (27.6%)</b> | <b>6 (17.1%)</b> | <b>0.037</b> |
| RYR2           | 11 (10%)        | 6 (13.6%)         | 2 (6.9%)         | 3 (8.6%)         | 0.721        |
| EGFR           | 28 (26%)        | 8 (18.2%)         | 7 (24.1%)        | 13 (37.1%)       | 0.163        |
| <b>NF1</b>     | <b>13 (12%)</b> | <b>2 (4.5%)</b>   | <b>7 (24.1%)</b> | <b>4 (11.4%)</b> | <b>0.043</b> |
| SPTA1          | 11 (10%)        | 5 (11.4%)         | 1 (3.4%)         | 5 (14.3%)        | 0.372        |
| PKHD1          | 11 (10%)        | 6 (13.6%)         | 4 (13.8%)        | 1 (2.9%)         | 0.185        |

As is known to all, immunotherapy, which is more effective and has fewer side effects than chemotherapy and targeted therapies, is becoming more important for cancer therapy day by day. Genomic lesions such as tumor mutational burden (TMB) and aneuploidy are related to the activation of immune checkpoint and how patients respond to

immunotherapy in terms of both extent and duration. Therefore, we calculated the TMB which is defined by the number of nonsynonymous mutations per million bases, and fraction genome altered (FGA) which is the percentage of genome that has been affected by copy number gains or losses in order to quantify the genomic alterations being able to beneficial for immunotherapy. After the comparison of TMB among different subtypes, it has been found no significant difference among either three GBM subtypes according to Kruskal-Wallis rank sum test ( $P$ -value = 0.839), or post-hoc Wilcoxon rank sum test with Benjamini-Hochberg adjustment in pairwise comparisons ( $P$ -value > 0.05). The distribution of both TMB and Transitions and Transversions (TiTv) were demonstrated in **Figure 3.4B**. In addition, FGA with also computed specific fraction genome gain (FGG) or fraction genome loss (FGL) was determined per sample within each subtype by using segmented copy number. It could not be observed any differences among three clusters based on either FGG or FGL ( $P$  = 0.966, **Figure 3.4C**).

Predicting the response to medications such as chemotherapeutic agents is crucial because of the highly variable response and side effects. To estimate response of the patients to medications, Gleeleher and colleagues utilized gene expression and in vitro drug sensitivity data obtained from cell lines, along with in vivo tumor gene expression data (Gleeleher et al., 2014). To this end, we explored differences in drug sensitivity among various subtypes by estimating the IC<sub>50</sub> of Temozolomide (TMZ) and it was indicated that CS2 subtype had more drug efficacy for TMZ than other subtypes ( $P$ -value = 0.0036, **Figure 3.5**).

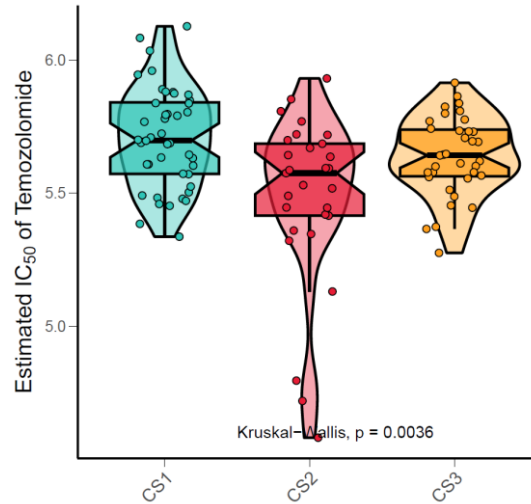


**Figure 3.5.** Box violins for estimated IC<sub>50</sub> of Temozolomide among identified subtypes of GBM.

### 3.1.3. Differential gene expression and methylation probes

Differential expression analyses were performed among different GBM subtypes to identify potential biomarkers passing the significance threshold (adjusted p-value <0.05 and  $|\log FC| > 0$ ) by DESeq2 method. These biomarkers were prioritized according to not overlap with any biomarkers identified for other subtypes. Based on the differential expression results of elite mRNAs among CS1, CS2, and CS3 subtypes, CHI3L1, CA12, NSG2, PTX3, CELF5, PLA2G5, RHOD, PI3, PKIB, and ATCAY were found to be the first 10 mRNAs presenting high discriminative ability in CS1 subtype, while ZDHHC22, LYVE1, CTHRC1, PI3, F13A1, MARCO, CCL2, CYP1B1, MET, and COL5A1 were found in CS2, and ACSBG1, NOS2, CRABP1, COL11A1, KCNF1, TMEFF2, INSYN2A, CSPG5, CPNE4, and SLC6A11 were found in CS3. We detected 150 downregulated mRNA and 120 upregulated mRNA as CS1-specific DEmRNAs, 98 downregulated mRNA and 164 upregulated mRNA as CS2-specific DEmRNAs, and 47 downregulated mRNA and 86 upregulated mRNA as CS3-specific DEmRNAs.

According to the differential expression results of elite lncRNAs among CS1, CS2, and CS3 subtypes, AL118505.1, ELFN2, AL596087.2, MIR124-1HG, AC110285.1, CEROX1, AC093673.1, SILC1, AL359091.1, and MIR9-3HG were found to be the first 10 lncRNAs presenting high discriminative ability in CS1 subtype, while AC007877.1, MIR9-3HG, AL596087.2, AC061992.2, AL133371.2, LINC01614, AC018755.4,

LOXL1-AS1, AL355922.1, and ELFN2 were found in CS2, and LINC01411, AC004816.1, SNTG2-AS1, AL118505.1, AC021242.3, AL008729.2, AUXG01000058.1, HAGLR, SILC1, and AL359091.1 were found in CS3. We identified 119 downregulated lncRNA and 80 upregulated lncRNA as CS1-specific DElncRNAs, 57 downregulated lncRNA and 113 upregulated lncRNA as CS2-specific DElncRNAs, and 32 downregulated lncRNA and 44 upregulated lncRNA as CS3-specific DElncRNAs.

When we examined differential methylation probes with threshold adjusted p-value < 0.05 and  $|\log FC| > 1$ , 55 hypomethylated DMPs, and 106 hypermethylated DMPs in CS2 compared to CS1; 23 hypomethylated DMPs, and 17 hypermethylated DMPs in CS3 compared to CS1; 55 hypomethylated DMPs, and 19 hypermethylated DMPs in CS3 compared to CS2 were detected for data acquired from HumanMethylation27 array. According to HumanMethylation450 array data, 18 hypomethylated DMPs, and 49 hypermethylated DMPs in CS2 compared to CS1; 9 hypomethylated DMPs, and 4 hypermethylated DMPs in CS3 compared to CS1; 84 hypomethylated DMPs, and 29 hypermethylated DMPs in CS3 compared to CS2 were identified. Moreover, in the case of intersection of these two different array data results, 39 CpG sites (cg03852144, cg23090046, and cg24861272 hypomethylated in both; cg16799087, cg19815720, cg09434995, cg18239253, cg00698688, cg25182621, cg21663431, cg03098721, cg13446622, cg23228178, cg06825166, cg26415633, cg13993218, cg27412902, cg11608424, cg24457403, cg21624282, cg14173523, cg23240895, cg12061127, cg25737664, cg12266049, cg02488871, cg01369413, cg07597976, cg03846767, cg05294455, cg20981615, cg16408565, cg25141490, cg23580945, cg25839227, cg18204685, cg02844051, cg10071275, and cg15538820 hypermethylated in both) for the CS2-CS1 comparison, 7 CpG sites (cg18555555, cg08615333, cg06352750, cg12071073, and cg18552413 hypomethylated in both; cg07078114, and cg26186727 hypermethylated in both) for the CS3-CS1 comparison, 27 CpG sites (cg15538820, cg18555555, cg21044104, cg02844051, cg10173075, cg21663431, cg24691255, cg21624282, cg06630567, cg13993218, cg16408565, cg07597976, cg23228178, cg20981615, cg22461018, cg05078019, cg04994456, cg10994126, cg23815000, cg23278885, cg00698688, cg06825166, cg18462653, cg14173523, cg10125195, and cg07484450 hypomethylated in both; cg05221167 hypermethylated in both) for the CS3-CS2 comparison were explored (**Table S2**).

### 3.1.4. Gene Set Enrichment Analysis

Significant molecular pathways and biological processes in which DEmRNAs played roles were investigated for each subtype through GSEA by utilizing gene sets related to GO biological processes, and the canonical pathway databases of KEGG, Reactome, PID, WikiPathways obtained from the MSigDB. CS1-specific downregulated pathways were mainly enriched in immune system related pathways such as the innate immune system, SARS-CoV-2 signaling pathway, neutrophil degranulation, cytokine signaling in immune system, and signaling by interleukins, while upregulated pathways were mainly enriched in neuron related pathways such as synaptic vesicle pathway, transmission across chemical synapses, L1CAM interactions, and neuronal system. CS2-specific downregulated pathways were mainly enriched in neuron related pathways such as neuronal system, synaptic vesicle pathway, and transmission across chemical synapses, while upregulated pathways were mainly enriched in degradation of the extracellular matrix, burn wound healing, syndecan-1 pathway, collagen degradation, and neutrophil degranulation (**Table S3**). The results of GO biological processes presented that response to biotic stimulus, response to bacterium, defense response to other organism, granulocyte migration, humoral immune response, inflammatory response, and leukocyte chemotaxis were enriched in CS1-specific downregulated processes; vesicle mediated transport in synapse, organelle localization, oligodendrocyte differentiation, regulation of synaptic plasticity, regulation of synaptic vesicle cycle, and oligodendrocyte development were enriched in CS1-specific upregulated processes, whereas vesicle mediated transport in synapse, neurotransmitter transport, negative regulation of nervous system development, regulation of synaptic plasticity, neurotransmitter secretion, and synaptic signaling were observed in CS2-specific downregulated processes, and extracellular structure organization, response to bacterium, circulatory system development, cardiovascular system development, response to biotic stimulus, and humoral immune response were observed in CS2-specific upregulated processes (**Table S4**).

Although there was not found any pathways or GO processes which were specific to CS3 subtype, significant downregulated genes appearing in CS3 were annotated for the enriched pathways such as burn wound healing, adaptive immune system, focal adhesion, and cell lineage map for neuronal differentiation, whereas significant upregulated genes

appearing in CS3 were annotated for viral infection pathways, pathways in cancer, spinal cord injury, muscle contraction, and IL-24 signaling pathway. According to GO biological processes which downregulated genes in CS3 involved in, regulation of supramolecular fiber organization, regulation of cytoskeleton organization, oligodendrocyte differentiation, immune response regulating signaling pathway, negative regulation of response to external stimulus, response to antibiotic, and granulocyte migration were mainly observed, while the processes including regulation of neurotransmitter levels, forebrain development, drug metabolic process, macromolecule catabolic process, and sulfur compound metabolic process for upregulated genes in CS3.

### **3.1.5. Discussion**

One of the most fatal primary brain malignancies, GBM, is related to a poor prognosis with a short median patient survival and a slight response to therapies, and has a high recurrence even with radiation therapy, surgery and TMZ treatment (Stupp et al., 2005; Wen and Kesari, 2008; Barzegar Behrooz et al., 2023).

Grade 4 IDH-wild-type tumors (90% of cases) are considered glioblastoma, while grade 4 IDH-mutant tumors are assessed as high-grade astrocytic tumors according to the 2021 WHO Classification of Tumors of the Central Nervous System (Louis et al., 2021). Franceschi et al. reported that IDH1-mutant GBM has overall survival more than three times higher than in IDH-wild-type GBM (Parsons et al., 2008; Franceschi et al., 2021). In this study, we compared the prognosis potential among three identified GBM subtypes through a Kaplan-Meier Curve and observed that there is no statistically significant difference among these subtypes based on average overall survival time ( $P = 0.426$ ). Since we used only IDH-wild-type GBM patients who have the worst prognosis with low survival time to create these subtypes in this study, no significant difference could be observed from the survival analysis.

GBM were classified into four distinct subtypes of samples based on gene expression data; Classical, Mesenchymal, Neural and Proneural (Verhaak et al., 2010). As a result of analysis of similarity between conventional subtypes of GBM and identified subtypes found in this study, we found that CS1 was more similar to proneural subtype, CS2 was more similar to mesenchymal subtype, and CS3 was more similar to classical subtype.

TMZ is one of the FDA approved chemotherapy options for GBM although it causes resistance for some patient groups. For this purpose, we would like to investigate sensitivity of TMZ on identified GBM subtypes in this study. Estimated IC50 values for TMZ were determined significantly lower in CS2, especially compared to CS1 ( $P=0.0036$ ) and we might say that patients belonging to mesenchymal subtype is sensitive to TMZ treatment, whereas patients belonging to proneural subtype is resistant to TMZ treatment. Cai and colleagues were also found that response to TMZ was significantly higher in the mesenchymal subtype by comparing the classical, neural, and proneural subtypes in both TCGA and CGGA cohorts (Cai et al., 2020).

### **3.2. Differential Protein Interactome in Esophageal Squamous Cell Carcinoma Offers Novel Systems Biomarker Candidates with High Diagnostic and Prognostic Performance**

#### **3.2.1. Differential interactome in ESCC**

After applying the differential interactome approach in ESCC, highly probable PPIs differentiated among tumor and healthy states (dPPIs) were identified. As a result, we identified 451 dPPIs among 498 proteins as repressed in tumor state and 1127 dPPIs among 1001 proteins as activated in ESCC.

The differential interactome indicates a network topology with modular organization, in which proteins were bundled around certain proteins (Ayyildiz et al., 2017; Gulfidan et al., 2020). As previously described (Gulfidan et al., 2020), DIPs indicate significant alterations in their interaction patterns during a transition from one phenotype to another phenotype (from healthy normal to tumor in this case) and classified into two discrete categories: proteins presenting (1) an excessive repression pattern (Class I) and (2) an excessive activation pattern (Class II) in their interactome under the phenotype of interest (i.e., ESCC).

Among the 1330 DIPs, 51 DIPs showing a high interaction pattern (representing  $\geq 10$  dPPIs) were categorized into classes. We observed that 14 DIPs presented a repression pattern and were classified as Class I, whereas 37 DIPs presented an activation pattern and were classified as Class II (**Table 3.3**).

**Table 3.3.** Differentially interacting proteins with number of differential protein-protein interactions  $\geq 10$

| <i>DIP</i> | <i>Class</i> | <i>No. of repressed dPPIs</i> | <i>No. of activated dPPIs</i> | <i>No. of total dPPIs</i> | <i>DIP</i> | <i>Class</i> | <i>No. of repressed dPPIs</i> | <i>No. of activated dPPIs</i> | <i>No. of total dPPIs</i> |
|------------|--------------|-------------------------------|-------------------------------|---------------------------|------------|--------------|-------------------------------|-------------------------------|---------------------------|
| MDFI       | I            | 39                            | 3                             | 42                        | AMOT       | II           | 0                             | 19                            | 19                        |
| UBE3A      | I            | 33                            | 4                             | 37                        | TOLLIP     | II           | 0                             | 18                            | 18                        |
| DNMT3B     | I            | 24                            | 0                             | 24                        | CDK7       | II           | 0                             | 16                            | 16                        |
| PSMD10     | I            | 20                            | 0                             | 20                        | NONO       | II           | 0                             | 16                            | 16                        |
| RPGRIP1L   | I            | 20                            | 0                             | 20                        | BAG4       | II           | 0                             | 15                            | 15                        |
| RPL35      | I            | 17                            | 1                             | 18                        | CANX       | II           | 0                             | 15                            | 15                        |
| BCL3       | I            | 15                            | 2                             | 17                        | COPB1      | II           | 2                             | 13                            | 15                        |
| DCTN2      | I            | 16                            | 0                             | 16                        | CDC16      | II           | 0                             | 14                            | 14                        |
| HCK        | I            | 14                            | 0                             | 14                        | EIF3A      | II           | 0                             | 14                            | 14                        |
| KAT7       | I            | 12                            | 2                             | 14                        | CDC34      | II           | 0                             | 13                            | 13                        |
| COPG1      | I            | 10                            | 3                             | 13                        | RGS2       | II           | 0                             | 13                            | 13                        |
| COPE       | I            | 8                             | 3                             | 11                        | EGFR       | II           | 3                             | 9                             | 12                        |
| GNL3       | I            | 9                             | 2                             | 11                        | EIF3M      | II           | 0                             | 12                            | 12                        |
| XRCC4      | I            | 7                             | 3                             | 10                        | POLR3D     | II           | 0                             | 12                            | 12                        |
| HSPA8      | II           | 0                             | 66                            | 66                        | USP10      | II           | 1                             | 11                            | 12                        |
| XRCC6      | II           | 2                             | 50                            | 52                        | EGLN1      | II           | 0                             | 11                            | 11                        |
| CDKN1B     | II           | 3                             | 39                            | 42                        | GNAI2      | II           | 1                             | 10                            | 11                        |
| CHUK       | II           | 1                             | 40                            | 41                        | GNAQ       | II           | 2                             | 9                             | 11                        |
| TRIM27     | II           | 2                             | 29                            | 31                        | QARS       | II           | 0                             | 11                            | 11                        |
| FLNA       | II           | 1                             | 28                            | 29                        | UIMC1      | II           | 0                             | 11                            | 11                        |
| MED19      | II           | 0                             | 27                            | 27                        | EIF3C      | II           | 0                             | 10                            | 10                        |
| CFTR       | II           | 0                             | 26                            | 26                        | ITK        | II           | 0                             | 10                            | 10                        |
| GLP1R      | II           | 0                             | 23                            | 23                        | RPL18      | II           | 1                             | 9                             | 10                        |
| NBN        | II           | 0                             | 21                            | 21                        | UBE2W      | II           | 0                             | 10                            | 10                        |
| PSEN1      | II           | 0                             | 21                            | 21                        | UBLCP1     | II           | 0                             | 10                            | 10                        |
| EZR        | II           | 0                             | 20                            | 20                        |            |              |                               |                               |                           |

DIP, differentially interacting protein; dPPI, differential protein-protein interaction.

We denominated these DIPs together with the protein partners interacting with them as modules, so-called DIPcentered modules (**Table 3.4**), and proposed them as possible systems biomarkers to improve efficient diagnosis, prognosis, and treatment strategies.

**Table 3.4.** Differentially interacting proteins-centered modules

| <b>Module name</b> | <b>Module size</b> | <b>Module proteins</b>                                                                                                                  |
|--------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| <b>mAMOT</b>       | 20                 | AMOT, AMOTL1, AMOTL2, ARHGAP17, DYNLL1, DYNLL2, INADL, ITCH, LATS1, LATS2, LIN7C, MED4, MPDZ, MPP5, NEDD4, NF2, USP9X, WWC1, WWOX, YAP1 |
| <b>mBAG4</b>       | 16                 | BAG4, CD93, DCP1A, DCP1B, DDX6, DNAJB12, DNAJB6, DNAJC7, DNAJC9, EDC3, HSPA1B, HSPA1L, HSPA2, MLF2, STUB1, TNFRSF1A                     |
| <b>mBCL3</b>       | 16                 | BCL3, COPS5, CTBP1, CTBP2, EP300, FOS, HDAC1, HDAC3, KAT5, N4BP2, NCOA1, NFKB1, NFKB2, PIR, RUVBL2, TBL1XR1                             |
| <b>mCANX</b>       | 16                 | APOB, BCAP31, CANX, CFTR, CHST15, EDEM1, EDEM3, EGFR, LMAN1, LPA, MTNR1A, NPHS1, PTPRN, RPN2, TOR1A, VCP                                |

|                |    |                                                                                                                                                                                                                                                                                                                 |
|----------------|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>mCDC16</b>  | 15 | ANAPC1, ANAPC10, ANAPC13, ANAPC15, ANAPC2, ANAPC4, ANAPC5, ANAPC7, BUB1B, CDC16, CDC20, CDC23, MDC1, RB1, TRIM33                                                                                                                                                                                                |
| <b>mCDK7</b>   | 17 | CCNA2, CCNB1, CCNH, CDC37, CDK2, CDK7, CDK9, ERCC2, ERCC3, ERCC5, HNRNPU, HSP90AA1, MNAT1, POLR2A, RPA2, SUPT5H, TCEA1                                                                                                                                                                                          |
| <b>mCDKN1B</b> | 41 | CAMK1, CCNA1, CCNA2, CCND1, CCND2, CCND3, CCNE1, CCNE2, CDK1, CDK2, CDK4, CDK5, CDK6, CDKN1B, CKS1B, COPS5, CUL1, CUL4A, GNL3, GRB2, KPNA1, LYN, MAPK8, MYC, NUP50, PIN1, PSMB1, RCHY1, RFWD2, RHOA, RNF123, RPS6KA1, SIRT6, SKP1, SKP2, SPDYA, SRC, TSC2, UBE3A, XPO1, YWHAQ                                   |
| <b>mCFTR</b>   | 27 | AHSA1, AMFR, CANX, CBL, CFTR, COMMD1, DAB2, DERL1, DNAJA1, DNAJC5, HDAC6, HSP90AA1, HSPA4, HSPA8, HSPB1, MARCH2, NEDD4, RNF5, SLC9A3R1, SLC9A3R2, SNAP23, STUB1, STX1A, UBE2I, UBE3A, USP10, VCP                                                                                                                |
| <b>mCHUK</b>   | 41 | AKT2, ARRB2, CARD11, CD40, CDC37, CHUK, CTNNB1, CUEDC2, EIF2AK2, ERC1, FANCA, FKBP5, FOXO3, HNRNPA1, HSP90AA1, IFIT5, IKBKAP, IKBKG, IRAK1, MAP3K1, MAP3K14, MAP3K7, MAP3K8, MTOR, NCOA3, NCOR2, NFKB1, NFKB2, NFKBIB, PIAS1, PPM1B, PRKCA, PRKCD, PRKCQ, PRKDC, RICTOR, SRC, TNFRSF1A, TP73, TRAF1, TRIM27, 41 |
| <b>mCOPB1</b>  | 14 | ACACA, ARCN1, ARF1, COPA, COPB1, COPB2, COPE, COPG1, COPG2, COPZ1, KCNK3, NUDCD1, SACM1L, UBC                                                                                                                                                                                                                   |
| <b>mDCTN2</b>  | 17 | ACTR10, ACTR1A, ACTR1B, BICD2, CAPZA2, CAPZB, DCTN1, DCTN2, DCTN3, DCTN4, DCTN5, DCTN6, EMILIN1, GADD45A, MAPRE1, MARCKSL1, POLR2A                                                                                                                                                                              |
| <b>mDNMT3B</b> | 25 | CBX3, CBX5, CDC73, CHD4, DNMT1, DNMT3A, DNMT3B, DNMT3L, EED, EZH2, HDAC1, HDAC2, HELLS, HIST1H1A, HIST1H3A, KIF4A, NCAPG, NRIP1, PAF1, PIAS1, SMARCA5, SMC2, TDG, UBE2I, UHRF1                                                                                                                                  |
| <b>mEIF3A</b>  | 15 | EIF3A, EIF3B, EIF3C, EIF3CL, EIF3D, EIF3F, EIF3G, EIF3H, EIF3J, EIF3K, EIF3L, EIF3M, EIF4A1, EIF4B, UPF1                                                                                                                                                                                                        |
| <b>mEZR</b>    | 21 | ARHGDIA, BRCA1, CDH1, DCC, ERBB2, EZR, FASLG, ICAM1, ICAM2, L1CAM, MSN, NF2, PALLD, RDX, S100P, SDC2, SLC9A3R1, SLC9A3R2, SPN, VCAM1, VPS11                                                                                                                                                                     |
| <b>mFLNA</b>   | 29 | ARRB2, ASB2, BRCA1, BRCA2, CASR, CMIP, DBN1, FBLIM1, FLNA, FLNB, GP1BA, HSP90AB1, HSPB7, ICAM1, IGSF8, ITGB1, LMNA, MTNR1A, MTNR1B, MYBL2, MYH9, NPHP1, RALA, SVIL, SYNPO2, TES, TRAF2, TRIO, VHL                                                                                                               |

|                  |    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>mGLP1R</b>    | 24 | ACSF3, AP2M1, APL1A, APLP1, ATP13A1, ATP6AP2, ATP6V0B, ERP29, FDFT1, GCG, GLP1R, GPR37, MFSD5, MRPL22, MYL6, P4HTM, PAQR6, PIGG, SLC15A4, SLC31A2, STMN3, SYNGR3, TMEM147, TSPAN13                                                                                                                                                                                                                                                                                                      |
| <b>mHCK</b>      | 15 | ADAM15, BCR, CBL, CSF3R, EGFR, ELMO1, FRS3, HCK, KHDRBS1, OLIG1, PIK3R3, PTK2, RAPGEF1, RASA1, TNK2                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>mHSPA8</b>    | 67 | ABL1, AHS1, AIPL1, AP2M1, ARRB1, ATP13A2, BAG1, BAG2, BCR, BRCA1, CAPZA1, CAPZB, CCAR2, CCND1, CDC5L, CDH1, CFTR, DNAJA1, DNAJA2, DNAJB6, DNAJC3, DNAJC7, DNAJC9, EDRF1, EGFR, EP300, ERG, ESR1, FANCC, FLYWCH2, GABRB1, HDAC10, HEMGN, HIST3H3, HNRNPD, HSF1, HSP90AA1, HSPA1A, HSPA2, HSPA4, HSPA5, HSPA8, HSPH1, HTT, HYOU1, HYPK, IL16, LAMP2, MDM2, PPP1CA, PPP1CC, PTEN, RAF1, RGS2, RPS4X, SF3A2, SHC1, SNCA, STIP1, STUB1, SUGT1, TRAF3IP1, TRIM25, TUBA1A, UBASH3B, UCHL1, VCP |
| <b>mMDFI</b>     | 40 | AKAP17A, APLN, AQP1, BAHD1, C10orf62, CBFA2T2, CCDC120, DCAF8, ENKD1, FAM124B, FBXL18, GLIDR, GNAI2, HTR1B, MDFI, MYOD1, MYOG, NEU4, NPDC1, NR1H2, OTX1, PGLS, PIN1, PRKAB2, PVRL2, RABL6, SIX1, SLC25A10, SPG7, TRPV6, TYMSOS, VPS72, ZBTB24, ZBTB25, ZNF101, ZNF136, ZNF439, ZNF559, ZNF581, ZNF607                                                                                                                                                                                   |
| <b>mMED19</b>    | 28 | CCNC, CDK19, MED1, MED10, MED11, MED12, MED13, MED13L, MED14, MED15, MED16, MED17, MED18, MED19, MED20, MED23, MED26, MED27, MED28, MED29, MED30, MED31, MED4, MED6, MED7, MED8, MED9, POLR2A                                                                                                                                                                                                                                                                                           |
| <b>mNBN</b>      | 22 | ATF2, ATM, ATR, ATRX, BRCA1, EP300, H2AFX, HIST3H3, MDC1, MDM2, MRE11A, NBN, PRKDC, RAD18, RAD50, RBBP8, RECQL5, RNF8, SIRT1, TERF2, TUBG1, VRK1                                                                                                                                                                                                                                                                                                                                        |
| <b>mNONO</b>     | 17 | CTNNB1, ERG, FUS, H2AFX, HDAC1, HEY1, HNRNPA1, LMNA, MYC, NONO, PARK7, PPP1CA, RNF43, SFPQ, SMARCD1, U2AF2, UBE2A                                                                                                                                                                                                                                                                                                                                                                       |
| <b>mPSEN1</b>    | 22 | APH1A, APP, BCL2, CDC37, CDH1, CTNNB1, DSP, ECSIT, ERLIN2, FLNB, HERPUD1, ICAM5, KCNIP3, NCSTN, PKP4, PSEN1, PSEN2, PSENEN, PSMA5, PSMB1, ST13, TRAF6                                                                                                                                                                                                                                                                                                                                   |
| <b>mPSMD10</b>   | 21 | CLIC1, HSPA4, MDM2, NFKB1, PAAF1, PSMC1, PSMC2, PSMC3, PSMC4, PSMC5, PSMC6, PSMD1, PSMD10, PSMD12, PSMD14, PSMD3, PSMD7, PSMD8, RB1, UBC, UCHL5                                                                                                                                                                                                                                                                                                                                         |
| <b>mRPGRIP1L</b> | 21 | CAPZA2, CCP110, CD2AP, CEP170, CEP192, CSPP1, ECT2, IPO7, KIAA0753, MPRIP, NEK4, NPHP1, NPHP4,                                                                                                                                                                                                                                                                                                                                                                                          |

|                |    |                                                                                                                                                                                                                                                                                                                                                         |
|----------------|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                |    | NXF1, OFD1, PCM1, RAI14, RPS19, RPS27, RPS3A, RPS7, RPS8                                                                                                                                                                                                                                                                                                |
| <b>mRPL35</b>  | 18 | RPL10, RPL10L, RPL11, RPL18, RPL23A, RPL27A, RPL30, RPL35, RPL36, RPL6, RPL7, RPS11, RPS12, RPS19, RPS27, RPS3A, RPS7, RPS8                                                                                                                                                                                                                             |
| <b>mTOLLIP</b> | 19 | DAZAP2, DZIP3, EIF2AK2, IL1R1, IL1RAP, IRAK1, IRAK2, PEG10, RBPMS, RNF26, SMAD7, TLR2, TLR4, TOLLIP, TOM1, TOM1L1, TOM1L2, UBC, WRNIP1                                                                                                                                                                                                                  |
| <b>mTRIM27</b> | 31 | CHD4, CHUK, EIF4E, HDAC1, IKBKE, LNX1, MAGEL2, MBD4, MDM2, NDE1, NDEL1, NFYC, PIAS1, PIAS3, PIAS4, PIK3C2B, PRAM1, RB1, TBK1, TP53, TRIM23, TRIM27, TRIM29, TRIM32, UBE2D1, UBE2D2, UBE2D4, UBE2I, UBE2N, USP2, USP7                                                                                                                                    |
| <b>mUBE3A</b>  | 35 | ANXA1, ARNTL, ASPM, BPY2, CALM1, CDKN1B, CFTR, CTNNB1, E2F1, ESR1, ESR2, HERC2, HSPA4, JADE1, LCK, MAPK6, MCM7, NELFCD, NEURL4, PGR, PML, PRDX1, PSMD14, PSMD4, RAD23A, RAD23B, SHFM1, TSC2, UBE2D2, UBE2E1, UBE2L3, UBE3A, UBQLN2, UCHL5, WBP2                                                                                                         |
| <b>mXRCC6</b>  | 51 | ABL1, ARAP1, BAX, BRCA1, CBX5, CDK2, CHAF1A, CHD1L, CLU, EHF, ELF3, ERG, FMNL1, H2AFX, HDAC6, HIST1H3A, HIST1H3E, HMGA1, HMGA2, HOXB7, HOXC4, HTT, KAT2A, LIG3, MDM2, MRE11A, MSH2, MYC, NCF4, NCOA6, NDRG1, NR1H4, PARP1, PCNA, PDX1, POU2F2, PPID, PTTG1, RNF126, RPA2, SIRT3, SIRT6, SIRT7, TERF2, TERT, TFAP2A, TONSL, VAV1, XRCC4, XRCC6, XRCC6BP1 |

### 3.2.2. Biological characterization of DIP-centered modules

Functional overrepresentation analysis was performed in each DIP-centered module to determine biological processes and molecular pathways associated with the module proteins (**Table 3.5**). The statistically significant enrichments (adjusted  $p < 0.01$ ) showed various biological pathways and processes principally associated with cell cycle, cellular transportation, and cell signaling. Besides, pathways and processes, which were major components of cancer progression, became prominent in the overrepresentation results, such as Hippo, NF- $\kappa$ B, Wnt, and MAPK signaling pathways encountered in many types of cancer, processes related to prostate cancer, transcriptional misregulation of cancer, microRNAs in cancer, and apoptosis. In addition, several pathways and processes related to infections caused by virus (such as viral carcinogenesis, Hepatitis B, HPV, Epstein-Barr virus, and measles), bacteria (such as tuberculosis and Salmonella infection) and

protozoa (like malaria, African trypanosomiasis, and Leishmaniasis) were evident (**Table 3.5**).

**Table 3.5.** Molecular pathways and biological processes significantly associated with differentially interacting protein-centered modules (adjusted p value  $\leq 0.01$ )

| Module  | Associated pathway or biological process                                                                                    | Adjusted p value       |
|---------|-----------------------------------------------------------------------------------------------------------------------------|------------------------|
| mAMOT   | Hippo signaling pathway                                                                                                     | $1.89 \times 10^{-8}$  |
|         | Tight junction                                                                                                              | $1.52 \times 10^{-9}$  |
|         | Regulation of signal transduction                                                                                           | $1.71 \times 10^{-4}$  |
| mCANX   | Cargo recognition for clathrin-mediated endocytosis                                                                         | $8.09 \times 10^{-5}$  |
|         | Protein processing in endoplasmic reticulum                                                                                 | $1.51 \times 10^{-8}$  |
|         | Low density lipoproteins remodeling                                                                                         | $8.39 \times 10^{-5}$  |
| mCDC16  | Inactivation of APC/C via direct inhibition of the APC/C complex                                                            | $3.35 \times 10^{-28}$ |
|         | Inhibition of the proteolytic activity of APC/C required for the onset of anaphase by mitotic spindle checkpoint components | $3.35 \times 10^{-28}$ |
|         | Metaphase/anaphase transition of cell cycle                                                                                 | $2.56 \times 10^{-13}$ |
|         | Mitotic sister chromatid separation                                                                                         | $2.56 \times 10^{-13}$ |
| mCDK7   | Tat-mediated elongation of the HIV-1 transcript                                                                             | $7.53 \times 10^{-17}$ |
|         | Transcriptional Regulation by TP53                                                                                          | $4.52 \times 10^{-16}$ |
|         | DNA repair                                                                                                                  | $7.53 \times 10^{-12}$ |
| mCDKN1B | Viral carcinogenesis                                                                                                        | $1.65 \times 10^{-22}$ |
|         | Hepatitis B                                                                                                                 | $1.47 \times 10^{-16}$ |
|         | Human papillomavirus infection                                                                                              | $6.12 \times 10^{-12}$ |
|         | Measles                                                                                                                     | $9.25 \times 10^{-11}$ |
|         | Epstein-Barr virus infection                                                                                                | $1.93 \times 10^{-10}$ |
|         | Regulation of cell cycle                                                                                                    | $6.63 \times 10^{-20}$ |
| mCFTR   | Protein processing in endoplasmic reticulum                                                                                 | $3.65 \times 10^{-11}$ |
|         | Defective CFTR causes cystic fibrosis                                                                                       | $5.42 \times 10^{-7}$  |
|         | Tight junction                                                                                                              | $1.31 \times 10^{-3}$  |
|         | Protein catabolic process                                                                                                   | $9.46 \times 10^{-12}$ |
|         | Endocytosis                                                                                                                 | $5.68 \times 10^{-4}$  |
| mCHUK   | NF- $\kappa$ B signaling pathway                                                                                            | $1.74 \times 10^{-16}$ |
|         | Apoptotic process                                                                                                           | $4.21 \times 10^{-14}$ |
|         | Epstein-Barr virus infection                                                                                                | $1.96 \times 10^{-12}$ |
|         | MAPK signaling pathway                                                                                                      | $8.61 \times 10^{-12}$ |
|         | Keratinocyte differentiation                                                                                                | $1.02 \times 10^{-10}$ |
| mCOPB1  | ER to Golgi anterograde transport                                                                                           | $5.56 \times 10^{-15}$ |
|         | Golgi-to-ER retrograde transport                                                                                            | $1.59 \times 10^{-15}$ |
|         | ER to Golgi vesicle-mediated transport                                                                                      | $2.03 \times 10^{-13}$ |
| mDCTN2  | Golgi-to-ER retrograde transport                                                                                            | $1.70 \times 10^{-18}$ |
|         | ER-to-Golgi anterograde transport                                                                                           | $8.30 \times 10^{-16}$ |
|         | Vasopressin-regulated water reabsorption                                                                                    | $6.40 \times 10^{-9}$  |

|           |                                                                         |                        |
|-----------|-------------------------------------------------------------------------|------------------------|
|           | Antigen processing and presentation of peptide antigen via MHC class II | $3.44 \times 10^{-20}$ |
| mDNMT3    | Epigenetic regulation of gene expression                                | $1.86 \times 10^{-19}$ |
|           | DNA methylation                                                         | $8.03 \times 10^{-8}$  |
|           | MicroRNAs in cancer                                                     | $8.85 \times 10^{-5}$  |
|           | Transcription, DNA-templated                                            | $3.25 \times 10^{-10}$ |
| mEZR      | Cell adhesion molecules (CAMs)                                          | $7.55 \times 10^{-8}$  |
|           | Tight junction                                                          | $3.25 \times 10^{-6}$  |
|           | Axon guidance                                                           | $8.09 \times 10^{-6}$  |
|           | African trypanosomiasis                                                 | $1.41 \times 10^{-5}$  |
|           | Malaria                                                                 | $3.13 \times 10^{-5}$  |
|           | Cell morphogenesis                                                      | $9.06 \times 10^{-9}$  |
| mFLNA     | Salmonella infection                                                    | $8.20 \times 10^{-3}$  |
|           | Signal transduction                                                     | $2.40 \times 10^{-2}$  |
|           | Cytoskeleton organization                                               | $3.20 \times 10^{-7}$  |
| mGLP1R    | Transport of small molecules                                            | $2.10 \times 10^{-2}$  |
|           | Ion transport                                                           | $1.00 \times 10^{-2}$  |
| mHSPA8    | Protein processing in endoplasmic reticulum                             | $2.30 \times 10^{-12}$ |
|           | Prostate cancer                                                         | $6.00 \times 10^{-7}$  |
|           | Cellular response to heat                                               | $3.40 \times 10^{-18}$ |
| mMDF1     | Generic transcription pathway                                           | $1.05 \times 10^{-2}$  |
|           | Transport of small molecule                                             | $2.87 \times 10^{-2}$  |
| mMED19    | Transcriptional regulation of white adipocyte differentiation           | $4.01 \times 10^{-8}$  |
|           | Thyroid hormone signaling pathway                                       | $6.54 \times 10^{-10}$ |
| mNBN      | Cell cycle                                                              | $2.10 \times 10^{-15}$ |
|           | DNA repair                                                              | $2.20 \times 10^{-15}$ |
|           | Processing of DNA double-strand break ends                              | $2.20 \times 10^{-15}$ |
| mNONO     | Signaling by NOTCH1 in cancer                                           | $2.90 \times 10^{-4}$  |
|           | Transcriptional misregulation in cancer                                 | $4.80 \times 10^{-4}$  |
|           | Thyroid hormone signaling pathway                                       | $1.80 \times 10^{-3}$  |
|           | Wnt signaling pathway                                                   | $1.90 \times 10^{-2}$  |
| mRPGR1P1L | Cilium assembly                                                         | $1.40 \times 10^{-6}$  |
|           | Cytoskeleton organization                                               | $1.80 \times 10^{-8}$  |
| mTOLLIP   | Interleukin-1-mediated signaling pathway                                | $2.72 \times 10^{-8}$  |
|           | NF- $\kappa$ B signaling pathway                                        | $1.53 \times 10^{-5}$  |
|           | Toll-like receptors cascades                                            | $4.51 \times 10^{-5}$  |
|           | Leishmaniasis                                                           | $4.49 \times 10^{-5}$  |
|           | Tuberculosis                                                            | $3.32 \times 10^{-5}$  |
| mTRIM27   | Ubiquitin mediated proteolysis                                          | $7.38 \times 10^{-11}$ |
|           | Epstein-Barr virus infection                                            | $5.86 \times 10^{-6}$  |
|           | SUMOylation of transcription factors                                    | $3.92 \times 10^{-10}$ |
| mUBE3A    | Ubiquitin mediated proteolysis                                          | $5.31 \times 10^{-5}$  |
|           | Generic transcription pathway                                           | $1.74 \times 10^{-3}$  |
|           | Estrogen signaling pathway                                              | $2.21 \times 10^{-3}$  |
| mXRCC6    | DNA repair                                                              | $2.82 \times 10^{-16}$ |
|           | Generic transcription pathway                                           | $9.43 \times 10^{-8}$  |

|  |                                         |                       |
|--|-----------------------------------------|-----------------------|
|  | Transcriptional misregulation in cancer | $6.78 \times 10^{-5}$ |
|--|-----------------------------------------|-----------------------|

CAMs, cell adhesion molecules.

We analyzed the presence of tumor suppressor proteins (TSPs) and oncoproteins among DIPs, considering their crucial roles in genome reprogramming during tumorigenesis. For this purpose, we recruited the list of proteins containing 1217 TSPs and 803 oncoproteins (Lawrence et al., 2014), and observed that 174 TSPs and 125 oncoproteins represented significant changes in their interaction patterns in ESCC.

Furthermore, the involvement of DIPs in biological processes and molecular pathways related to cancer hallmarks was analyzed through functional overrepresentation analyses, which emphasized processes and pathways that enable to acquire or maintain cancer hallmarks consisting of sustaining proliferative signaling, inducing angiogenesis, resisting cell death, deregulating cellular energetics, evading growth suppressors, enabling replicative immortality, activating invasion and metastasis, genome instability, and mutation, and avoiding immune destruction (**Supplementary Figure S1**).

### 3.2.3. Diagnostic performance of DIP-centered modules

The diagnostic potential of each DIP-centered module was analyzed by PCA and performance metrics (i.e., sensitivity, specificity, and DOR) were calculated considering the two most important principal components, which demonstrated at least 80% of the total variance. A total of 8 DIP-centered modules (mAMOT, mCFTR, mCHUK, mDCTN2, mFLNA, mGLP1R, mHSPA8, and mTOLLIP) displayed significantly high diagnostic performance with sensitivity  $\geq 90\%$ , specificity  $\geq 90\%$ , and DOR  $\geq 81$  in the GSE23400 dataset (**Figure 3.6A**). Also, the diagnostic performance of these DIP-centered modules was validated using an independent dataset, that is, GSE13898 (**Figure 3.6B**).

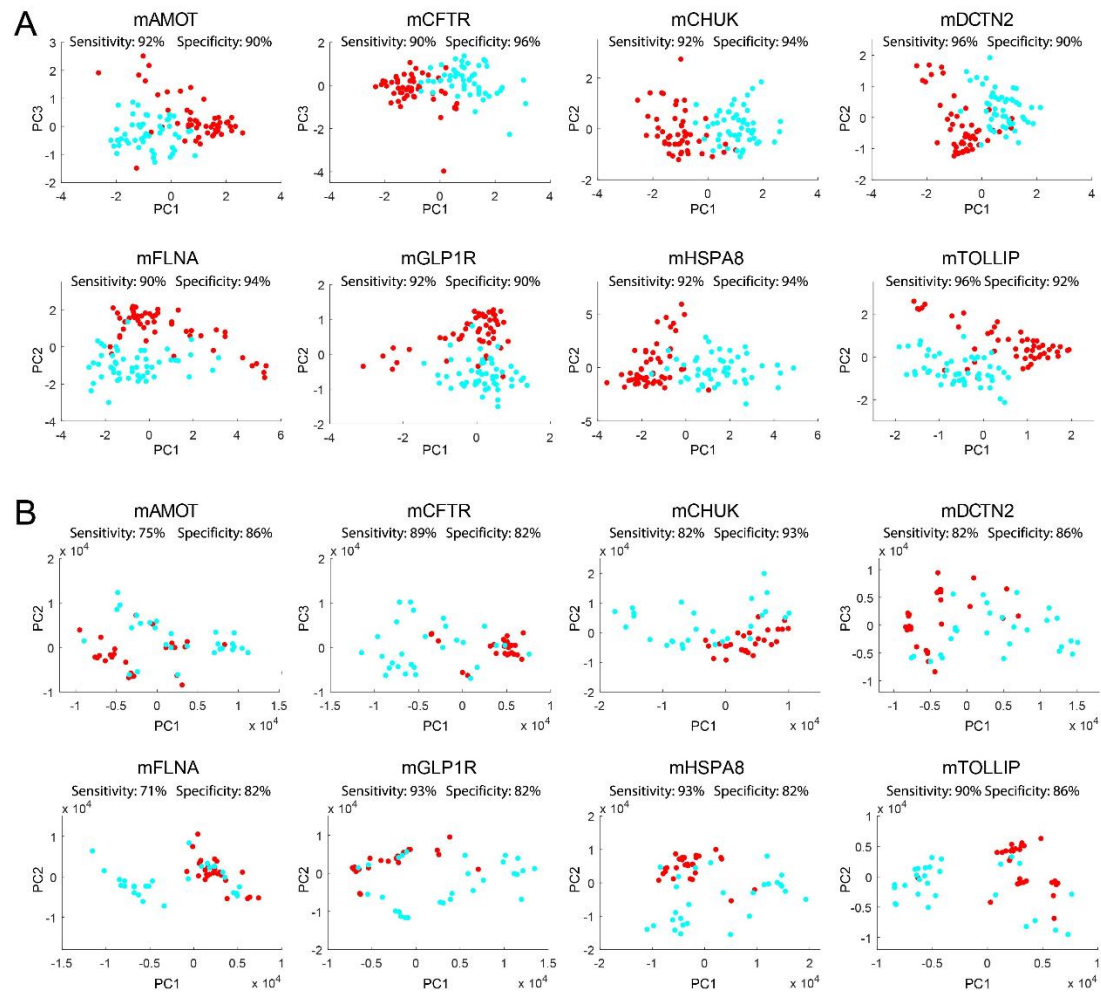


**Figure 3.6.** PCA plots showing the individual differences in the gene expression profiles among normal and tumor samples. (A) PCA plots created with the expression data of genes obtained from the data set of GSE23400. (B) PCA plots created with the expression data of genes obtained from the data set of GSE13898. Red dots represent healthy samples; blue dots represent tumor samples. PCA, principal component analyses.

### 3.2.4. Prognostic performance of DIP-centered modules

For the assessment of the prognostic power of DIPcentered modules as potential systems biomarkers, the patient cohort was separated into low- and high-risk groups based on the expression profiles of genes encoding proteins in each module, and in addition, multivariate survival analyses and risk assessments were achieved. Kaplan–Meier plots were obtained with log-rank test and HRs for quantification of the prognostic capabilities of each module. These analyses were operated utilizing independent datasets with distinct origin (i.e., RNA-seq datasets from TCGA and microarray datasets from NCBI-GEO).

When examining the relationship between the expression levels of DIP-centered modules and prognostic outcome through survival analyses, we monitored that a total of 16 DIP-centered modules (mXRCC6, mHSPA8, mNBN, mCFTR, mMED19, mFLNA, mCDKN1B, mDNMT3B, mUBE3A, mMDFI, mCHUK, mCDC16, mGLP1R, mRPGRIP1L, mTRIM27, and mEZR) indicated high impact on overall survival ( $p < 0.003$  and  $HR > 2.0$ ) of patient with ESCC (**Figure 3.7**).

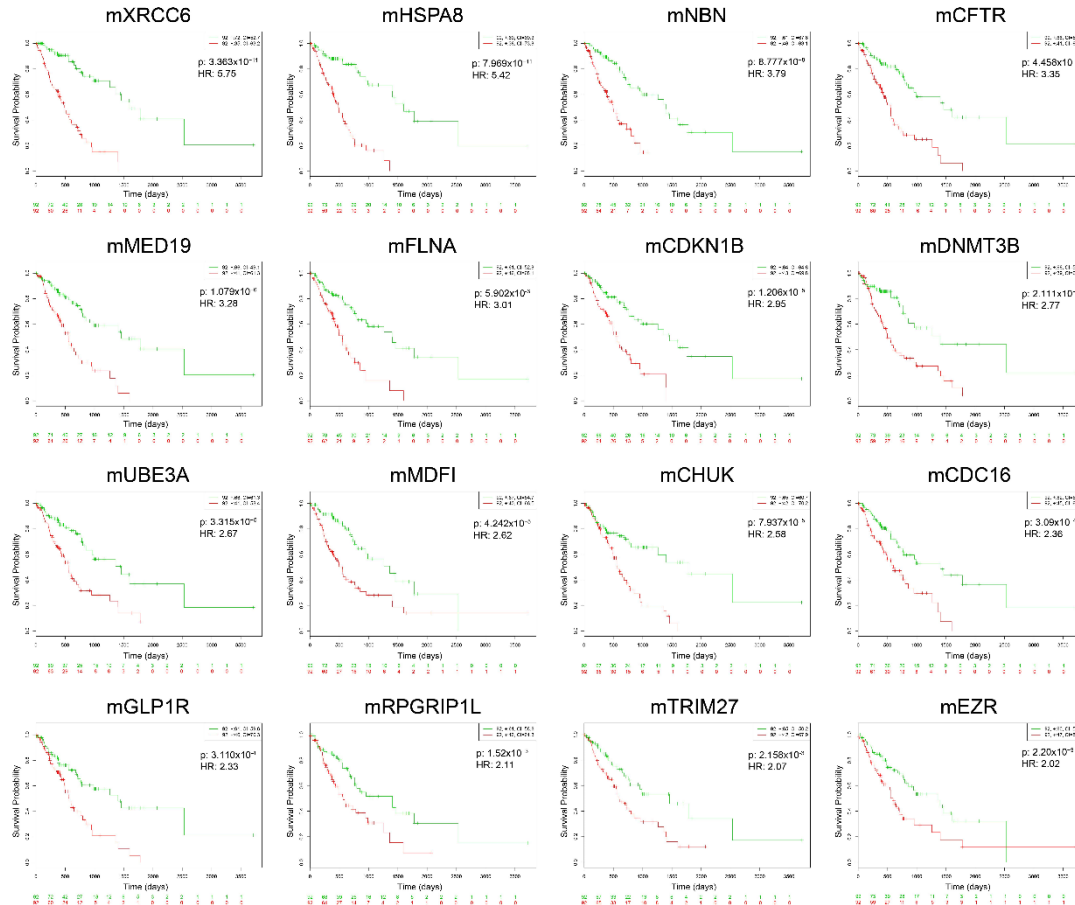


**Figure 3.7.** Kaplan–Meier Plots estimating patients’ survival for ESCC indicating p-value of the log-rank test and hazard ratio for each curve belonging to each module. Censoring samples are shown as “+” marks. Horizontal axis represents time to event. Green and red curves denote low- and high-risk groups, respectively. The green and red numbers below the horizontal axis represent the number of individuals not presenting the event of the corresponding risk group along time. The number of individuals, the number of censored, and the CI of each risk group are shown in the top-right insets. ESCC, esophageal squamous cell carcinoma.

### 3.2.5. Evaluation of biomarker potential of significant modules in body fluids and tissues

For the proteins represented in the DIP-centered modules exhibiting high diagnostic potential, the protein expression levels in several body fluids (i.e., plasma, serum, urine, and saliva) and easy to collect tissues (i.e., oral epithelium and hair follicle) were investigated through data provided in the proteomics databases (Montague et al., 2014; Samaras et al., 2020; Schaab et al., 2012; Stelzer et al., 2016). Relatively high protein expression patterns were reported in various body fluids for five of all DIP-centered modules, which were mCANX, mFLNA, mHSPA8, mTOLLIP, and mXRCC6 (**Figure 3.8**).

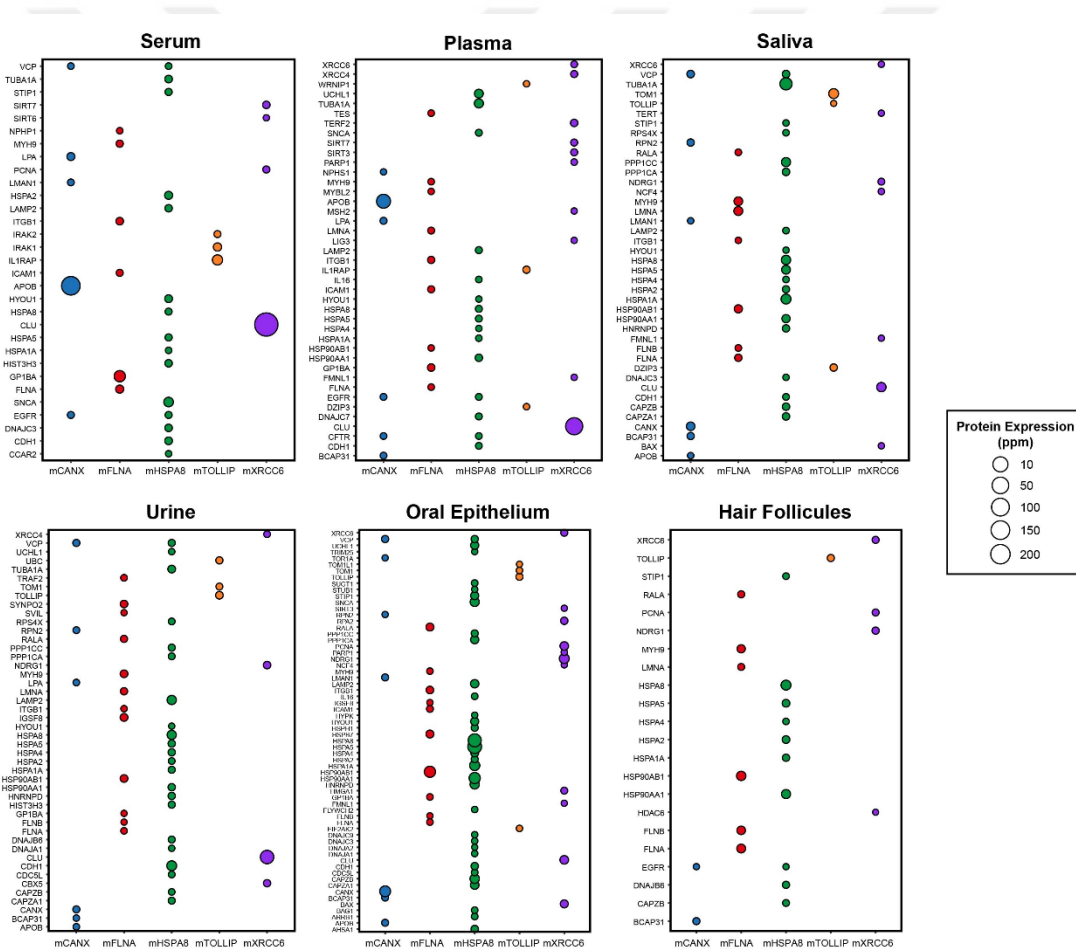


**Figure 3.8.** Bubble plots indicating protein expression levels of the proteins in five diagnostic modules in different body fluid and tissues.

To further evaluate the modules as easily measurable system biomarkers, we filtered them to cover only proteins with high expression patterns (>1 ppm) in the relevant body fluid

or tissue and re-calculated the performance metrics of the modules. We observed that the diagnostic performances of the modules were preserved and even improved for most cases (**Figure 3.9, Table 3.6**).

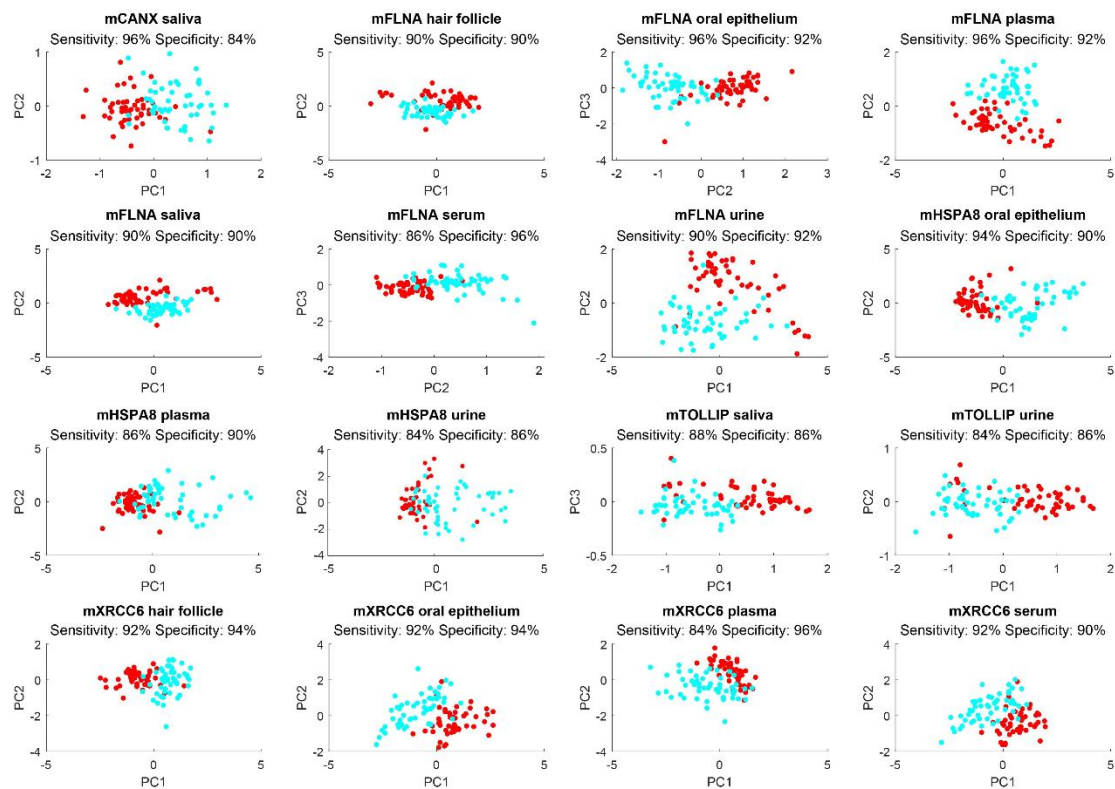


**Figure 3.9.** PCA plots showing the individual differences in the gene expression profiles among normal and tumor samples in the modules having significant protein expression levels. Red dots represent healthy samples; blue dots represent tumor samples.

**Table 3.6.** Differentially interacting protein-centered modules exhibiting high diagnostic potential in several body fluids and tissues.

| Module | Module proteins                                                     | Body fluid/<br>tissue | Sensitivity<br>(%) | Specificity<br>(%) |
|--------|---------------------------------------------------------------------|-----------------------|--------------------|--------------------|
| mCANX  | APOB, BCAP31, CANX, LMAN1, RPN2, VCP                                | Saliva                | 96.08              | 84.31              |
| mFLNA  | FLNA, FLNB, HSP90AB1, LMNA, MYH9, RALA                              | Hair Follicle         | 90.19              | 90.19              |
|        | FLNA, FLNB, GP1BA, HSP90AB1, HSPB7, ICAM1, IGSF8, ITGB1, MYH9, RALA | Oral Epithelium       | 96.08              | 92.16              |

|         |                                                                                                                                                                                                                                               |                 |       |       |
|---------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|-------|-------|
|         | FLNA, GP1BA, HSP90AB1, ICAM1, ITGB1, LMNA, MYBL2, MYH9, TES                                                                                                                                                                                   | Plasma          | 96.08 | 92.16 |
|         | FLNA, FLNB, HSP90AB1, ITGB1, LMNA, MYH9, RALA                                                                                                                                                                                                 | Saliva          | 90.19 | 90.19 |
|         | FLNA, GP1BA, ICAM1, ITGB1, MYH9, NPHP1                                                                                                                                                                                                        | Serum           | 86.28 | 96.08 |
|         | FLNA, FLNB, GP1BA, HSP90AB1, IGSF8, ITGB1, LMNA, MYH9, RALA, SVIL, SYNPO2, TRAF2                                                                                                                                                              | Urine           | 90.19 | 92.16 |
| mHSPA8  | AHSA1, ARRB1, BAG1, CAPZA1, CAPZB, CDC5L, CDH1, DNAJA1, DNAJA2, DNAJC3, DNAJC9, FLYWCH2, HNRNPD, HSP90AA1, HSPA1A, HSPA2, HSPA4, HSPA5, HSPA8, HSPH1, HYOU1, HYPK, IL16, LAMP2, PPP1CA, PPP1CC, SNCA, STIP1, STUB1, SUGT1, TRIM25, UCHL1, VCP | Oral Epithelium | 94.12 | 90.19 |
|         | CDH1, CFTR, DNAJC7, EGFR, HSP90AA1, HSPA1A, HSPA4, HSPA5, HSPA8, HYOU1, IL16, LAMP2, SNCA, TUBA1A, UCHL1                                                                                                                                      | Plasma          | 86.28 | 90.19 |
|         | CAPZA1, CAPZB, CDC5L, CDH1, DNAJA1, DNAJB6, HIST3H3, HNRNPD, HSP90AA1, HSPA1A, HSPA2, HSPA4, HSPA5, HSPA8, HYOU1, LAMP2, PPP1CA, PPP1CC, RPS4X, TUBA1A, UCHL1, VCP                                                                            | Urine           | 84.31 | 86.28 |
| mTOLLIP | DZIP3, TOLLIP, TOM1                                                                                                                                                                                                                           | Saliva          | 88.24 | 86.28 |
|         | TOLLIP, TOM1, UBC                                                                                                                                                                                                                             | Urine           | 84.31 | 86.28 |
| mXRCC6  | HDAC6, NDRG1, PCNA, XRCC6                                                                                                                                                                                                                     | Hair Follicle   | 92.16 | 94.12 |
|         | BAX, CLU, FMNL1, HMGA1, NCF4, NDRG1, PARP1, PCNA, RPA2, SIRT3, XRCC6                                                                                                                                                                          | Oral Epithelium | 92.16 | 94.12 |
|         | CLU, FMNL1, LIG3, MSH2, PARP1, SIRT3, SIRT7, TERF2, XRCC4, XRCC6                                                                                                                                                                              | Plasma          | 84.31 | 96.08 |
|         | CLU, PCNA, SIRT6, SIRT7, XRCC6                                                                                                                                                                                                                | Serum           | 92.16 | 90.19 |

### 3.2.6. Drug repositioning considering prognostic modules as potential drug targets

Considering 16 DIP-centered modules having significant prognostic characteristics as potential drug targets, we identified 53 drug candidates through geneXpharma drug repositioning tool (Turanli et al., 2017) according to p-value ( $<10^{-10}$ ) determined through hypergeometric probability density function (**Figure 3.10**).

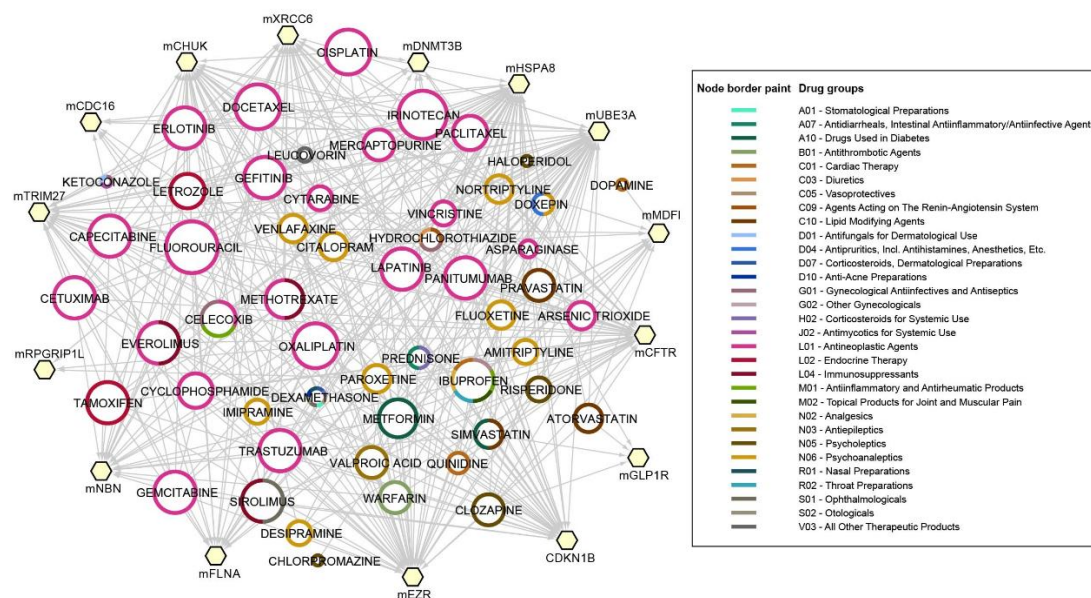


**Figure 3.10.** Networks indicating drugs associated with proteins in significant prognostic modules resulted in drug repositioning. Drugs are represented in *circles* and their color varies by ATC groups of the drugs. Modules are represented in *yellow hexagons*.

Among the drug candidates, fluorouracil was the common drug targeting 13 out of all the modules investigated. It is classified as an antimetabolite and used as an antineoplastic chemotherapy drug. Fluorouracil interferes with DNA synthesis by blocking thymidylate synthetase and slows tumor growth. It is used in the treatment of various cancer types such as colorectal cancer, esophageal cancer, breast cancer, and cervical cancer, and other diseases like actinic keratosis.

On the other hand, each one of the four drugs was proposed as a candidate for a distinct module: chlorpromazine for mTRIM27, dopamine for mMDFI, haloperidol for mHSPA8, and ketoconazole for mCHUK. Chlorpromazine is an antipsychotic medication that

prevents the excess of dopamine in the brain by blocking postsynaptic dopamine receptors in cortical and limbic areas of the brain. It is primarily used to treat psychotic disorders such as schizophrenia and manic depression, bipolar disorder in adults, and severe behavioral problems or hyperactivity in children.

Haloperidol is another antipsychotic medicine that blocks postsynaptic dopamine receptors in the mesolimbic dopaminergic system and increases turnover of brain dopamine. It is used to treat schizophrenia, mania in bipolar disorder, nausea and vomiting, delirium, agitation, and acute psychosis, and to control motor and speech tics in people with Tourette's syndrome.

Dopamine is a medication naturally produced in the body and commonly used in the treatment of very low blood pressure, a slow heart rate, and cardiac arrest when epinephrine is not available. It increases heart rate and cardiac contractility and improves the pumping strength of the heart and the blood flow to the kidneys by exerting an agonist action on beta-adrenoceptors and making the release of norepinephrine from storage sites in sympathetic nerve endings.

Ketoconazole is an antifungal medication used for treatment or prevention of various fungal infections such as tinea, cutaneous candidiasis, blastomycosis, histoplasmosis, pityriasis versicolor, dandruff, and seborrheic dermatitis. It is also used in the treatment of endogenous Cushing's syndrome and excessive hair growth. It increases fluidity of fungal membrane by preventing the synthesis of ergosterol, which has similar function with cholesterol in animals, and prevents growth and spread of the fungus.

Twenty-three of the drug candidates identified were antineoplastic agents. Furthermore, the analysis pointed out several drugs that have been used in diabetes (metformin, simvastatin), cardiac therapy (ibuprofen, quinidine, and dopamine), endocrine therapy (tamoxifen and letrozole), gynecological disorders (celecoxib, hydrochlorothiazide, ketoconazole, and ibuprofen), inflammation treatment (dexamethasone and ibuprofen), and ophthalmological and otological disorders (sirolimus and dexamethasone), besides being used for a variety of conditions such as antithrombotics (warfarin), diuretics (hydrochlorothiazide), vasoprotectives (dexamethasone), antifungals or antimycotics (ketoconazole), antipruritics (doxepin), antidiarrheals (prednisone), corticosteroids

(dexamethasone and prednisone), immunosuppressants (everolimus, sirolimus, and methotrexate), antiinflammatories (ibuprofen and celecoxib), antiepileptics (valproic acid), analgesics (ibuprofen), psycholeptics (clozapine, risperidone, chlorpromazine, and haloperidol), and psychoanaleptics (citalopram, fluoxetine, nortriptyline, paroxetine, venlafaxine, amitriptyline, desipramine, imipramine, and doxepin).

Among all the repurposed drug candidates, 28 drugs, which are mostly antineoplastic agents, have been studied in esophageal cancer previously; on the other hand, 25 drugs (ibuprofen, cyclophosphamide, clozapine, letrozole, mercaptopurine, citalopram, fluoxetine, nortriptyline, paroxetine, venlafaxine, warfarin, amitriptyline, cytarabine, desipramine, imipramine, risperidone, doxepin, hydrochlorothiazide, prednisone, quinidine, asparaginase, chlorpromazine, dopamine, haloperidol, and ketoconazole) were identified as novel drug candidates for the treatment of patients with ESCC.

### **3.2.7. Molecular docking simulations of potential drug target interactions**

There were 39 target proteins (GLP1R, PGR, HTR1B, EGFR, ABL1, NCF4, HDAC2, EP300, ARRB1, CDK1, CFTR, ESR1, RGS2, ERBB2, BCR, FKBP5, HSPA8, EZH2, PML, PSMD14, GCG, MDM2, CCND1, PCNA, CALM1, NCOR2, NELFCD, PRKCA, ESR2, MCM7, CCNE1, LIG3, MYC, RB1, TP53, EIF2AK2, PTEN, VAV1, and CTNNB1) and 24 drug candidates (ibuprofen, cyclophosphamide, clozapine, letrozole, mercaptopurine, citalopram, fluoxetine, nortriptyline, paroxetine, venlafaxine, warfarin, amitriptyline, cytarabine, desipramine, imipramine, risperidone, doxepin, hydrochlorothiazide, prednisone, quinidine, chlorpromazine, dopamine, haloperidol, and ketoconazole) available for the molecular docking analysis. The 3D structure of asparaginase was not available in these databases.

According to molecular docking analysis, 15 drug candidates (amitriptyline, citalopram, clozapine, desipramine, dopamine, doxepin, fluoxetine, imipramine, letrozole, nortriptyline, paroxetine, prednisone, quinidine, risperidone, and warfarin) exhibited significant binding affinities ( $< -8$  kcal/mol) to eight protein targets (ABL1, EGFR, EP300, GLP1R, HDAC2, HTR1B, NCF4, and PGR). Considering the drugs, risperidone and clozapine were those that affected the greatest number of proteins.

Consequently, risperidone targeted three proteins (EGFR, HTR1B, and PGR), and clozapine targeted four proteins (EGFR, GLP1R, HTR1B, and PGR). The predicted binding affinities of risperidone targeting EGFR, HTR1B, and PGR were -9, -9.6, and -9.8 kcal/mol, respectively. In addition, the predicted binding affinities of clozapine for EGFR, GLP1R, HTR1B, and PGR were -8.1, -9.9, -8.6, and -8.1 kcal/mol, respectively. As a result of these simulations, risperidone and clozapine were selected for in vitro validation studies.

### **3.2.8. Validation of growth inhibition by risperidone and clozapine in human ESCC cell lines**

Although antipsychotic medications are widely used for treating various psychiatric diseases for decades, their potential anticancer properties have been reported mostly in recent studies. In addition, the long history of clinical use with accumulated knowledge on safety, efficacy, side effects, or toxicology of psychiatric drugs makes them good candidates for drug repositioning (Huang et al., 2018).

The putative mechanisms of repurposed psychiatric drugs for cancer therapy comprise multiple intracellular functions, including cell stress, cell cycle regulation, survival, and/or apoptosis mediated by cellular signaling pathways such as PI3K/Akt/GSK-3 $\beta$ , STAT3, and WNT signaling. Therefore, antipsychotics can inhibit the growth of tumor cells either by modulation of these pathways or by stimulating immune system. Besides, the charged structures and the lipophilicity of antipsychotic medications have utmost importance to inhibit P-glycoprotein pumps, which result in resistance to chemotherapy drugs. Consequently, antipsychotics act as anticancer agent or increase the efficacy of chemotherapies (Hendouei et al., 2019). On the other hand, there are also previous studies concerning antipsychotic drugs' roles in cancer risk and discussing the potential procancer or anticancer activity of antipsychotics (Fond et al., 2012).

Increasing knowledge on repurposing antipsychotics as antineoplastic agents lead us to assess in vitro cytotoxicity of novel antipsychotic candidates on ESCC cell lines. Moreover, antipsychotic drugs such as clozapine and risperidone have long been known to display polypharmacology due to the multiple protein targets, so these drugs are significant candidates in this drug repositioning study (Roth et al., 2004). Therefore, we

focused on two psycholeptic drugs: risperidone and clozapine. They both strikingly reduced the growth of all the three ESCC cell lines in a dose-dependent manner (**Figure 3.11**). However, the effect of clozapine was more remarkable in all the cell lines compared to risperidone at all the treatment doses tested. The IC<sub>50</sub> concentrations for growth inhibition by both the drugs varied between the three ESCC cells lines: TE-5 [IC<sub>50</sub> (clozapine) = 42.4  $\mu$ M; IC<sub>50</sub> (risperidone) = 88.1  $\mu$ M], TE-11 [IC<sub>50</sub> (clozapine) = 35.6  $\mu$ M; IC<sub>50</sub> (risperidone) = 43.5  $\mu$ M], and TE-15 [IC<sub>50</sub> (clozapine) = 40.0  $\mu$ M; IC<sub>50</sub> (risperidone) = 55.2  $\mu$ M].

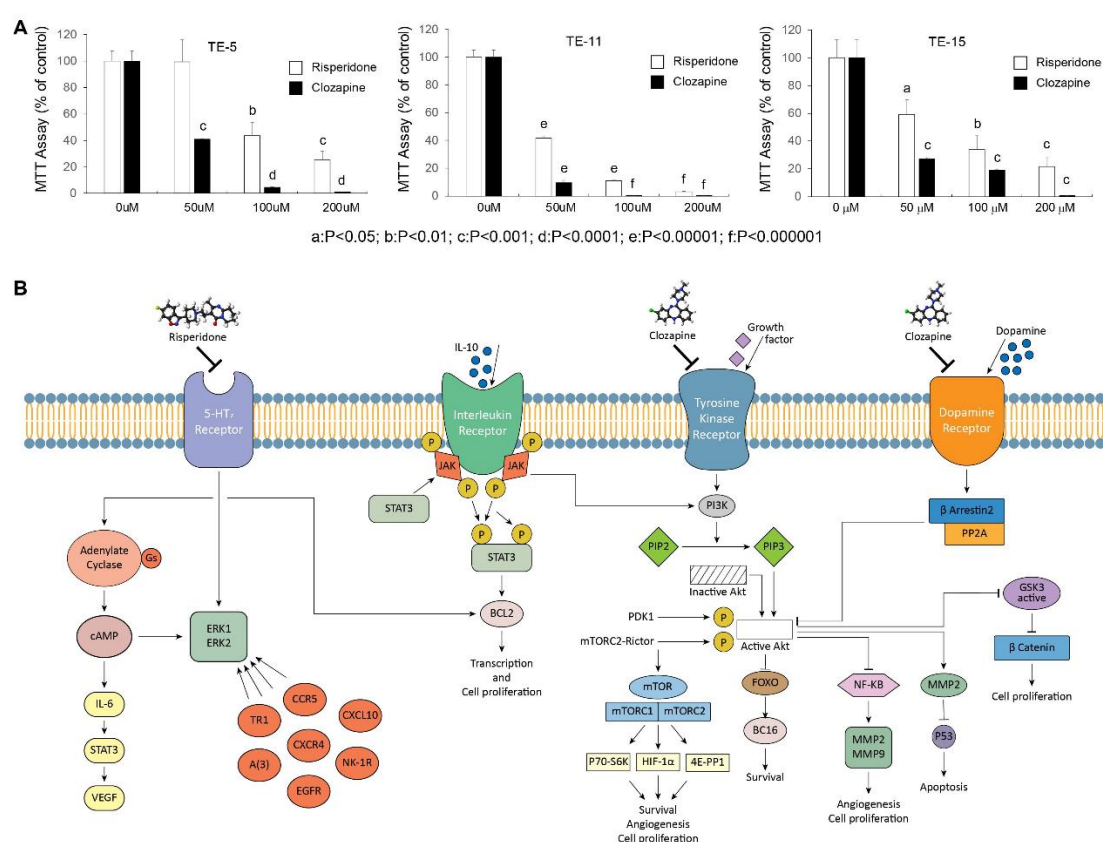


**Figure 3.11.** *In vitro* study of risperidone and clozapine on ESCC cell lines and their mechanism(s). (A) Effects of risperidone and clozapine on proliferation of different ESCC cell lines. MTT assay was performed 24 h after treatment with increasing doses of both drugs. Letters indicate the significant difference compared with controls (0  $\mu$ M). a:  $p < 0.05$ , b:  $p < 0.01$ ; c:  $p < 0.001$ ; d:  $p < 0.0001$ ; e:  $p < 0.00001$ , f:  $p < 0.000001$ . (B) The mechanism(s) indicating the pathways of two antipsychotic drugs, risperidone and clozapine, in cancer. MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide.

### 3.2.9. Discussion

Understanding the molecular mechanism(s) that differentiate one phenotype from other phenotypes requires assessing the active PPIs in each phenotype. As we have described previously (Ayyildiz et al., 2017c), the basis of differential interactome analysis is to comparatively analyze changes in protein interaction patterns and the results lead to differential PPIs (dPPIs) that indicate significant differences between phenotypes. These include “repressed PPIs” that show significantly lower prevalence in the phenotype of interest (e.g., specific cancer) compared to other phenotypes (e.g., healthy state), while “activated PPIs” show significantly higher prevalence in the phenotype of interest.

Network-based approaches not only help to elucidate the mechanisms of cancer and discover therapeutic targets but also help to gain a better understanding of the differential protein interactome in the context of tumorigenesis. In this perspective, the interactome is quite essential to explore genetic alterations in the network, while assessing the system as a whole (Caldera et al., 2017). In this study, we aimed to generate dPPI networks by combining high-throughput transcriptome and protein interactome data for ESCC. The differential interactome methodology was performed on the ESCC microarray dataset to identify the differential protein interactome belonging to the disease.

The network topology of the differential protein interactome in ESCC follows a modular organization in which DIPs were clustered around specific proteins, which we called hub DIPs. We identified a set of hub DIPs that predominantly show an activation pattern in the tumor state, and further analyzed the biological roles of these DIPs and their interacting protein partners. The results of the functional enrichment analysis and the observation that the modules contain molecules (such as TSPs and oncogenes) that are directly associated with cancer features, pathways, and processes strengthened our belief that the modules we identified could be good biomarker candidates. Therefore, special attention should be paid to DIP-centered modules to evaluate their abilities as potential system biomarkers to improve efficient strategies for diagnosis, prognosis, and treatment.

Evaluation of their prognostic performance by survival analysis and their diagnostic performance by PCA analysis confirmed our hypotheses. Furthermore, we showed that a significant proportion of the module proteins were highly expressed in various body fluids

(i.e., plasma, serum, urine, and saliva) and easily obtainable tissues (i.e., oral epithelium and hair follicle).

The development of accurate and robust biomarkers remains a challenging task since the molecular mechanisms of the disease are complex. The conventional way of biomarker discovery has focused on a single gene, protein, or metabolite biomarkers; in contrast, recent studies have focused on the fact that the development and progression of complex diseases such as cancer are mainly caused by the organized activity of a set of biological entities rather than a single molecule that malfunctions (Wang, 2011; Gov and Arga, 2017).

Accordingly, we propose several DIP-centered modules that are tumor specific as potential systems biomarkers and have particular diagnostic and prognostic ability for ESCC. We also considered their diagnostic and prognostic performances assessed by demonstrative stratification and survival analyses. As a treatment strategy, aiming at PPI interfaces has recently arisen (Li et al., 2017a). Various compounds showed promising results as potential inhibitors of PPIs. For instance, Maraviroc, an inhibitor of the CCR5/gp120 interaction, showed effective outcomes on HIV and is present as one of the anti-HIV drugs on the market (Cruz et al., 2014).

Such efforts exhibit the feasibility of targeting PPIs for cancer treatment. When we consider that tumorigenesis is driven by the involvement of a great number of PPIs, we anticipate that crucial survival mechanisms in tumor cells may be inactivated by the interception of critical DIPs (and their modules). Thus, these PPI interfaces were proposed as potential targets for the discovery and development of anticancer therapeutic agents (Turanli and Arga, 2017). Furthermore, these specific dPPIs were evaluated for the drug repositioning to treat ESCC patients (Turanli et al., 2019b, 2019e, 2021). Our drug repositioning analyses not only presented several antineoplastic agents that have been studied in ESCC earlier but also proposed several drugs as novel drug candidates for the treatment of patients with ESCC.

Further experimental validation studies need to be performed to investigate the cytotoxic effects of these agents on available cell lines derived from ESCC patients.

Since biological systems consist of many molecules and complex interactions between molecules instead of pair-wise interactions, the discovery of system biomarkers is very important to predict disease-associated genes and diseasespecific biomarkers and to discover efficient therapeutic strategies and potential drugs. Identification of molecular markers for ESCC is crucial, as there is currently a lack of classical markers for clinical analysis of ESCC. To this end, this study provides several diagnostic and prognostic system biomarkers for ESCC. The DIPs resulting from the differential interactome analyzes were evaluated for their diagnostic and prognostic properties, and four candidate modules (mCFTR, mCHUK, mGLP1R, and mHSPA8) with both high diagnostic and prognostic properties were presented.

In addition, the results of this study suggest that 53 drug candidates, which are already used for various diseases and were found by applying drug repositioning methods to significant prognostic modules, could be used for the treatment of patients with ESCC. In addition to the use of 23 of these drugs in the treatment of other cancers, 30 of them have already been used in the treatment of diseases such as diabetes, cardiac therapy, and psychotic disorders. A total of 25 drugs have not previously been associated with esophageal cancer.

The most observed medication among prognostic modules, fluorouracil, is an antineoplastic antimetabolite chemotherapy drug. It interferes with DNA synthesis and slows tumor growth and is used in the treatment of various cancer types such as colorectal cancer, esophageal cancer, and breast cancer, and other diseases like actinic keratosis.

This study focused on validation of two repurposed antipsychotic medications in three ESCC cell lines. Risperidone acts as a potent dopamine type 2 (D2),  $\alpha_2$  adrenergic receptor, and a 5-HT<sub>2A</sub> antagonist. At a low dose, it may also block the effects of 5-HT<sub>2A</sub> agonists. Risperidone might offer reduced hot flashes in breast cancer patients when used in combination with tamoxifen (Yeh et al., 2014). Risperidone also presented an opportunity to be used for prostate cancer treatment since it reduces proliferation of PC3 cells in vitro and retards tumor growth in vivo. Not only risperidone but also its other formulations are expected to proceed to clinical trials as a repurposed drug for prostate cancer and potentially for other adenocarcinomas (Dilly et al., 2017).

In a recent study, various second-generation antipsychotic drugs were evaluated to be used on colorectal cancer. Clozapine showed some inhibitory effect in human SW480 cells, whereas risperidone (tested at the same doses for ESCC cell lines in vitro in this investigation) exhibited the most notable tumor inhibitory effect in the study, which demonstrates the association between its usage and a lower risk of colorectal cancer (Chen et al., 2020). However, clozapine in ESCC cell lines was more effective in all the three ESCC cell lines compared to its effect in SW480 cells at the same doses (Chen et al., 2020). In another report, clozapine has been reported to be successful based on benefits-to-risk ratio for non-Hodgkins's lymphoma even though it is accompanied with agranulocytosis when given in combination with chemotherapy (Overbeeke et al., 2016).

### **3.3. Systems Biomarkers for Papillary Thyroid Cancer Prognosis and Treatment through Multi-Omics Networks**

#### **3.3.1. The transcriptomic signatures of papillary thyroid cancer**

The number of DEGs in each dataset showed a wide range from 809 to 5138 genes. Among those DEGs, 282 genes showed consistent expression profiles in all datasets, consisting of 133 upregulated genes (**Figure 3.12A**) and 149 downregulated genes (**Figure 3.12B**).

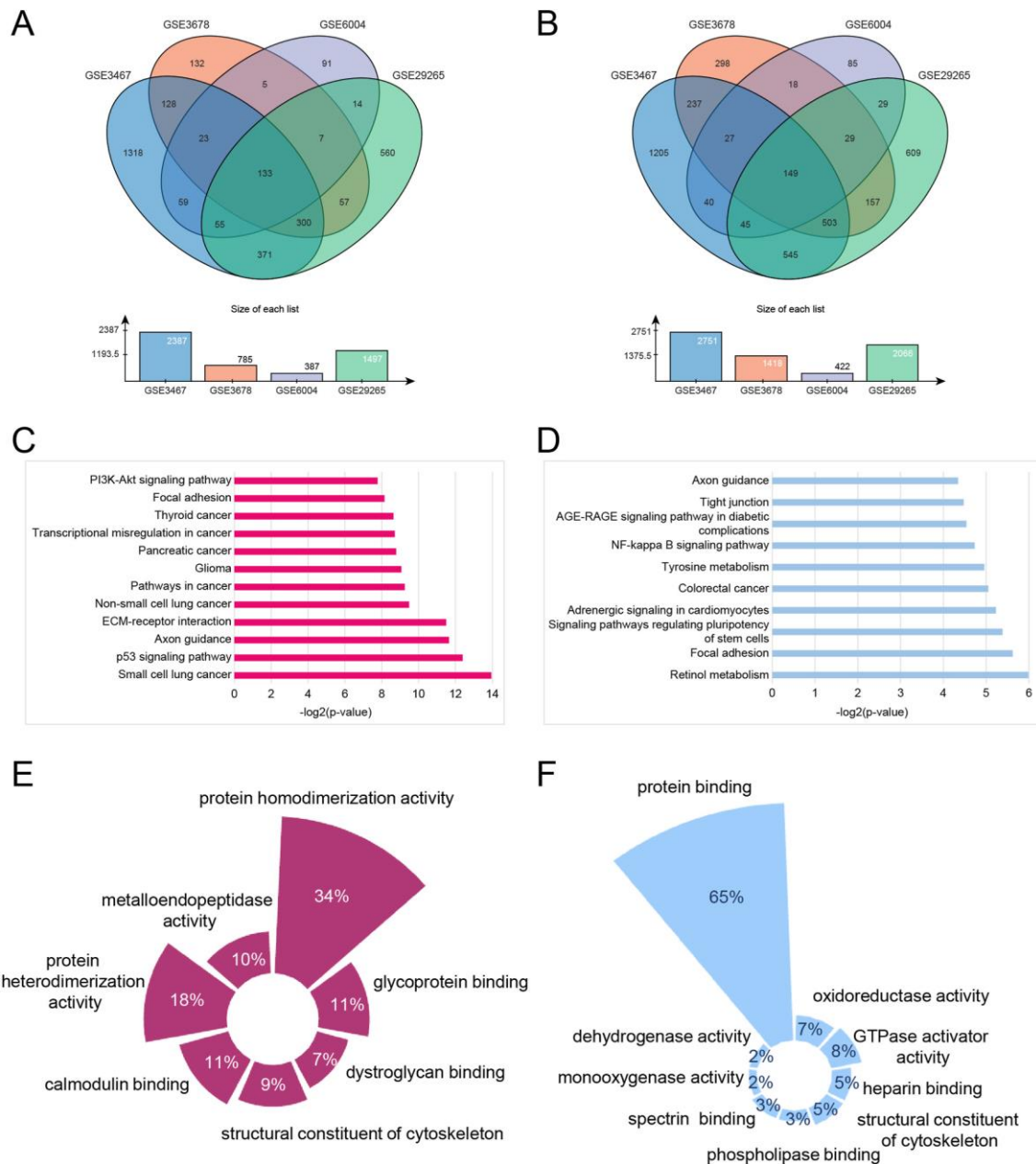


**Figure 3.12.** Common DEGs in papillary thyroid cancer and significant overrepresentation analysis of DEGs. (A) Venn diagram representing the comparison of upregulated DEGs among four PTC microarray datasets. (B) Venn diagram representing the comparison of downregulated DEGs among four PTC microarray datasets. (C) Bar graph representing overrepresentation analysis results of common upregulated DEGs among four PTC microarray datasets based on  $\log_2(p\text{-value})$  of pathways. (D) Bar graph representing overrepresentation analysis results of common downregulated DEGs among four PTC microarray datasets based on  $\log_2(p\text{-value})$  of pathways. (E) Plots indicating the clustering of proteins encoded by common upregulated DEGs based on their molecular functions. (F) Plots indicating the clustering of proteins encoded by common downregulated DEGs based on their molecular functions.

After the analysis of the TCGA-THCA data set having 552 papillary thyroid carcinoma samples via DESeq2, 18,731 genes were found as DEGs based on adjusted p-value<0.01. 278 of these genes were found in common with the genes which were common among the GEO data sets. Further analyses were carried out using with these common DEGs.

Overrepresentation analyses indicated that the upregulated DEGs were significantly associated with several cancer pathways (thyroid cancer, small cell lung cancer, non-small cell lung cancer, glioma, and pancreatic cancer) and cancer-associated molecular pathways such as PI3K-Akt signaling, extracellular matrix (ECM) receptor interaction, and p53 signaling. On the contrary, retinol metabolism, colorectal cancer, tyrosine metabolism, signaling pathways, axon guidance, focal adhesion, and tight junction were downregulated in thyroid cancer (**Figure 3.12C and D**).

Overrepresentation analysis of DEGs indicated the upregulation of glycoprotein binding, protein homodimerization, and heterodimerization activity. On the other hand, protein binding, GTPase activator activity, and oxidoreductase activity were downregulated in thyroid cancer (**Figure 3.12E and F**). In parallel to GO enrichment results, P53 is a tumor suppressor and regulates the cell cycle. There has been found a relationship between p53 and thyroid hormone nuclear receptors and p53 affects growth, development, and differentiation indirectly because of this association. Thereby, the p53 signaling pathway has a significant role in thyroid cancer as previously discussed and parallel to our results [48]. According to focal adhesion kinase, the highest level of expression p125FAK has been observed in patients with distant metastases. On the other hand, p125FAK level has been observed low in patients without distant metastasis. In both cases shows to focal adhesion kinase is a specific pathway and correlated in papillary thyroid cancer (Owens et al., 1996; Golubovskaya et al., 2009). Furthermore, GLI1 becomes major which is a transcription factor in the axon guidance pathway that increases the aggressiveness of the tumor in patients with thyroid cancer (Lee et al., 2015).

According to the results of overrepresentation analysis with long noncoding RNAs (lncRNAs) and mRNAs, 35 pathways related to PTC were detected. It was determined that most of these pathways are related to the p53 signaling pathway, pathways in cancer, PI3K-Akt signaling pathway, and transcriptional misregulation in cancer (Li et al., 2017b). Besides, ECM receptor interaction and tight junction have been determined to

play a critical role in papillary thyroid cancer (Baris et al., 2005; Liu et al., 2013). Retinol metabolism and the clinical effect of retinoids in advanced thyroid cancer were demonstrated in previous studies (Simon et al., 2002; Haugen et al., 2004). Moreover, tyrosine metabolism plays a crucial role in thyroid cancer. It has been predicted that receptor tyrosine kinase inhibitors may be an important treatment strategy; however, it has not yet been determined which inhibitor is more effective (Castellone et al., 2008; Verbeek et al., 2011).

### **3.3.2. The proteomic signatures of papillary thyroid cancer**

To identify hub proteins, a PPI sub-network around proteins encoded by the DEGs was constructed. The PPI network showed a scale-free topology and pointed out the presence of hub proteins, which was designated using degree and betweenness centrality metrics. Seventyfive proteins were determined as hub proteins when PPI networks in all data sets were investigated with these metrics. Among them, six hub proteins (FYN, JUN, LYN, PML, SIN3A, and RARA) were mutually inclusive in at least three datasets (**Figure 3.13A**).

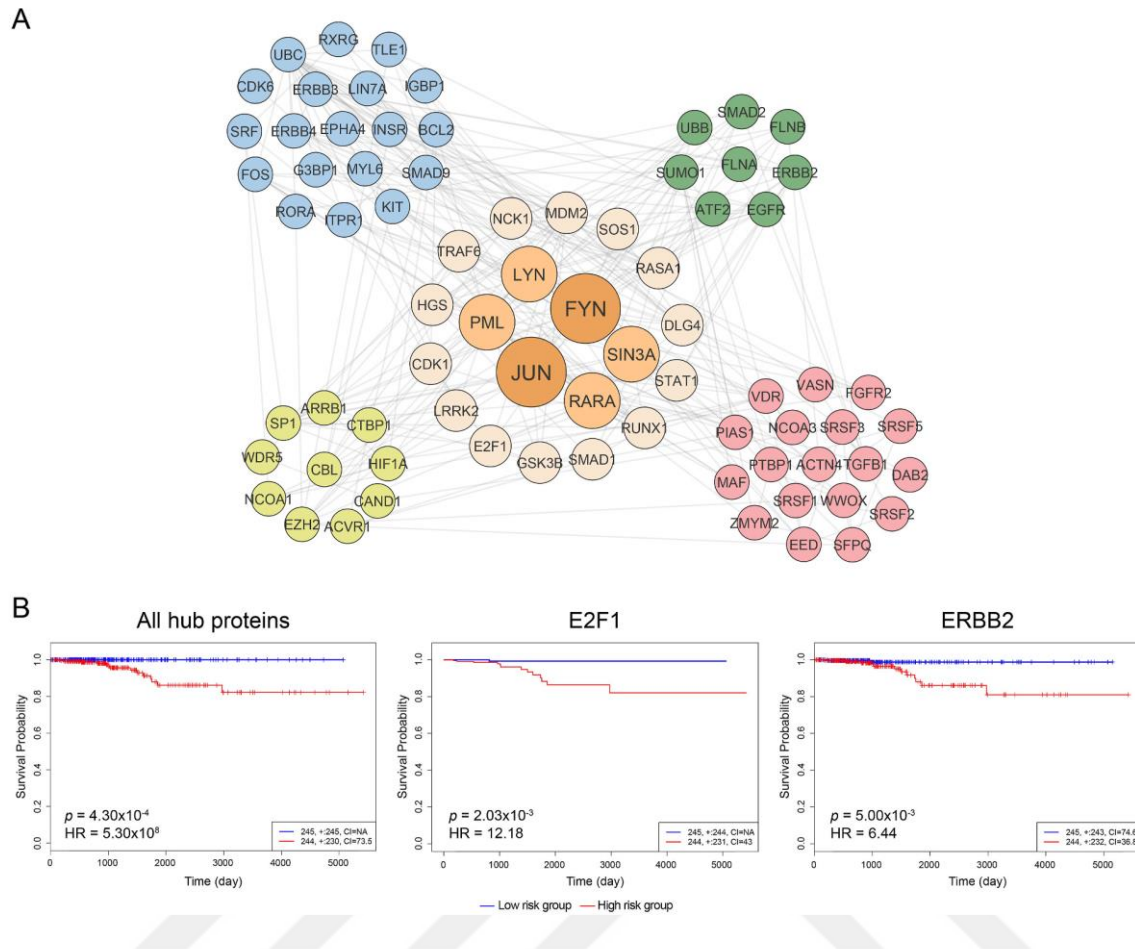


**Figure 3.13.** Hub proteins in PTC datasets. **(A)** Protein-protein interaction network indicating dataset-specific and common hub proteins among four PTC microarray datasets. Dataset-specific hub proteins are shown in blue, green, yellow, and pink colors. Common hub proteins were shown in orange color and different sizes based on how many datasets they are involved in as significant; small circles for two datasets, medium circles for three datasets, and big circles for four datasets. **(B)** Kaplan-Meier plots indicating the prognostic power of all hub proteins, hub protein E2F1, and hub protein ERBB2. Each plot has a p-value and hazard ratio to demonstrate its statistical significance. The bottom-right insets present the number of individuals, the number of censored, and the CI of each risk group, and censoring samples are shown as ‘+’ marks. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Among the hub proteins, tyrosine-protein kinase Fyn proto-oncogene (FYN) is a member of the protein-tyrosine kinase oncogene family and encodes a membrane-associated tyrosine kinase that has been implicated in the control of cell growth. Jun proto-oncogene (JUN, AP-1 transcription factor subunit) has a major role in encoding a specific protein which is highly similar to the viral protein to regulate gene expression interacting DNA sequences. LYN Proto-Oncogene, Src Family Tyrosine Kinase (LYN) is involved in the

regulation of mast cell degranulation, and erythroid differentiation and related to RET signaling and Insulin receptor recycling pathways. PML Nuclear Body Scaffold (PML) is a transcription factor and tumor suppressor and functions as in a wide range of important cellular processes, including tumor suppression, transcriptional regulation, apoptosis, senescence, DNA damage response, and viral defense mechanisms. SIN3 Transcription Regulator Family Member A (SIN3A) is acting as a transcriptional repressor and related to the pathways of androgen receptor signaling and transcriptional misregulation in cancer. Retinoic Acid Receptor Alpha (RARA) involved in the regulation of development, differentiation, apoptosis, granulopoiesis, and transcription of clock genes. The diseases associated with these hub proteins are Wiskott- Aldrich Syndrome for FYN; Sarcoma and Pertussis for JUN; Sarcoma and Hantavirus Pulmonary Syndrome for LYN; Leukemia, Acute Promyelocytic, Somatic and Acute Promyelocytic Leukemia Pml/Rara Type for PML; Witteveen-Kolk Syndrome and Chromosome 15Q24 Deletion Syndrome for SIN3A; and Leukemia, Acute Promyelocytic, Somatic and Myeloid Leukemia for RARA.

Survival analyses were performed for 20 hub proteins which were observed in at least two datasets and illustrated by Kaplan-Meier plots. According to their statistical significance and hazard ratios, the protein module consisting of all hub proteins ( $HR = 5.30 \times 108$ ), the protein of E2F1 ( $HR = 12.18$ ), and the protein of ERBB2 ( $HR = 6.44$ ) were considered as potential prognostic biomarker candidates (**Figure 3.13B**).

### **3.3.3. The metabolic signatures of papillary thyroid cancer**

The number of significant reporter metabolites ( $p$ -values $<0.05$ ) has been determined between 75 and 196. Among all the significant reporter metabolites of each dataset, 31 reporter metabolites were observed in at least three datasets, and two reporter metabolites were found as common, which are lactaldehyde and proline. To obtain more insight into metabolic activities around reporter metabolites, overrepresentation analyses were also performed on 31 significant reporter metabolites through MBROLE (v2.0) database. One of the significant pathways was phenylalanine metabolism which was associated with 5 reporter metabolites such as succinic acid and phenylacetaldehyde. Furthermore, oxidative phosphorylation, purine metabolism, taste transduction, pyrimidine

metabolism, neuroactive ligandreceptor interaction, tyrosine, and pyruvate metabolism were also significantly enriched with the reporter metabolites (**Figure 3.14A**).

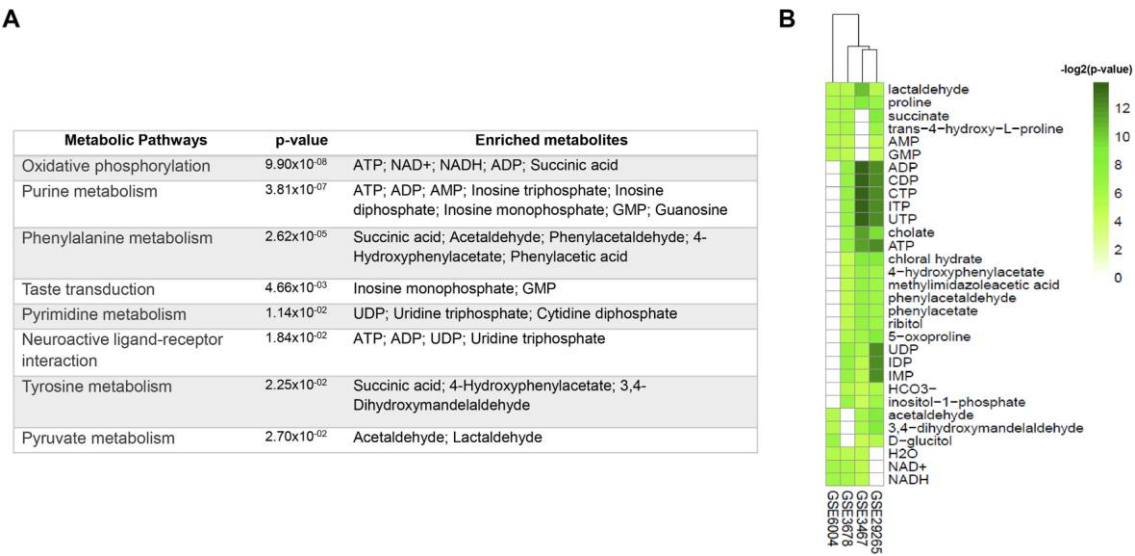


**Figure 3.14.** Reporter metabolites associated with thyroid carcinoma. **(A)** Overrepresentation analysis results for significant reporter metabolites. **(B)** Heatmap representing significant metabolites, which were observed in at least three datasets, in each dataset based on the p-value.

Several metabolites were associated with thyroid carcinoma in advance. Inositol 1-phosphate was a relevant serum metabolic signature of metastatic thyroid carcinoma [59]. Phenylacetate and acetaldehyde were found as biomarkers. In parallel to a systematic review of thyroid cancer biomarkers (Khatami et al., 2019), fatty acids, amino acids, cell membrane metabolites, purine, and pyrimidine metabolites, myo-/scyllo-inositol, and lactate were found common biomarkers.

Lactaldehyde, proline, succinate, chloral hydrate, 4-hydroxyphenylacetate, methylimidazoleacetic acid, and ribitol were not associated with thyroid carcinoma. Thus, they could be potential biomarkers and initiate further investigations (**Figure 3.14B**).

### 3.3.4. The reporter regulatory and signaling elements of papillary thyroid cancer

The significance of receptors was determined through the differential expression patterns of their physically interacting partners. Consequently, 22 proteins were identified as common reporter receptors with a significance level of  $p\text{-value}<0.05$ . In these 22 proteins, two Erb-B2 receptors (ERBB2 and ERBB4), two cyclin-dependent receptors (CDK1 and

CDK2), and three histone deacetylase receptors (HDAC1, HDAC2, and HDAC3) were observed. Furthermore, SRC, EP300, and EGFR were defined as major reporter receptors while considering the statistical significance.

The regulatory elements (TFs or miRNAs) controlling the transcriptional expression of the core genes of thyroid cancer were also identified via the employment of the combinatorial human transcriptional regulatory interaction network. Accordingly, 12 TFs and 110 miRNAs came into prominence with a significance level of  $p\text{-value} < 0.05$  and were identified as the common key regulatory elements in the transcriptional and post-transcriptional control of the core genes in thyroid cancer. TFs were three GATA transcription factors (GATA1, GATA2, and GATA3), two forked-box proteins (FOXA1 and FOXP3), and ETS1, AR, YBX1, PRDM14, E2F4, ESR1, TFAP2C.

The prognostic analysis of potential biomarkers was through the SurvExpress tool that was analyzed as different modules of TFs and receptors, for looking that reporter molecules in thyroid cancer based on risk group and survival days. According to Kaplan-Meier curves, the module of common reporter TFs (HR = 3.67) and the module of common reporter receptors (HR = 12.2) were observed as significant potential prognostic biomarkers (**Figure 3.15A-B**).

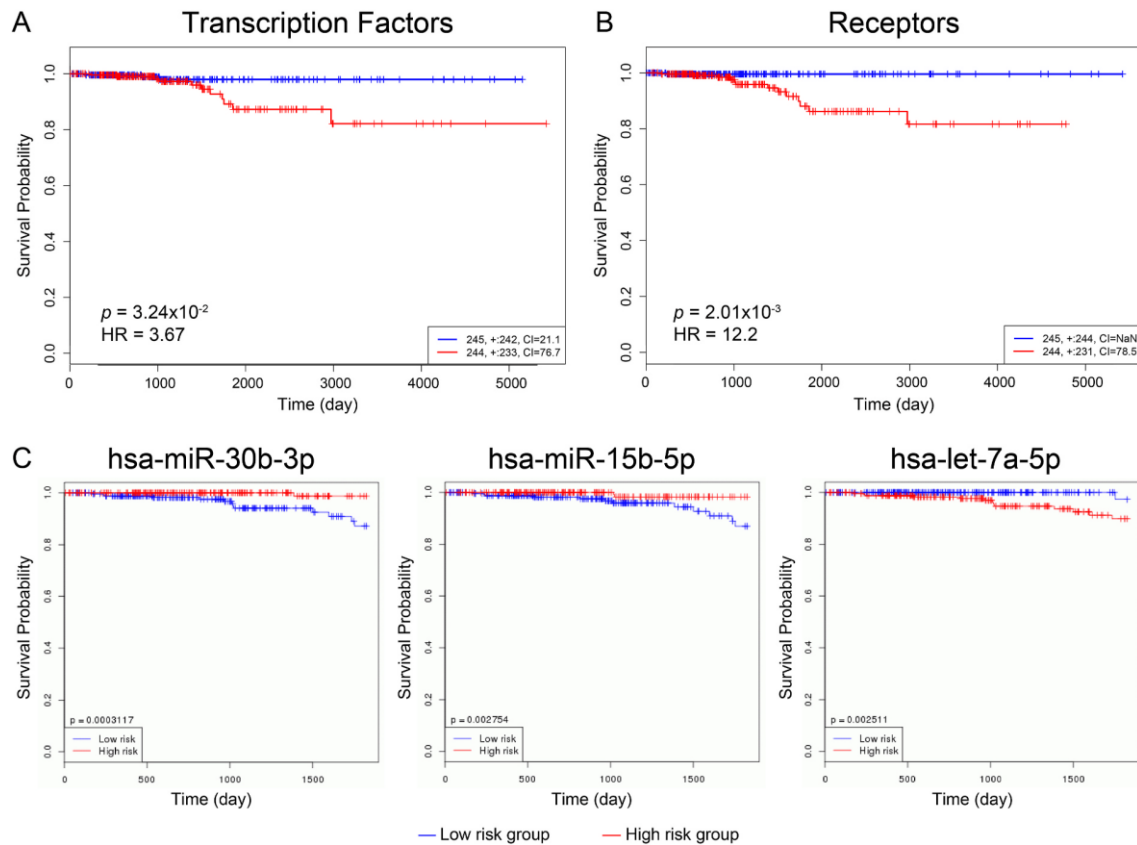


**Figure 3.15.** The prognostic power of common reporter molecules related to thyroid carcinoma. (A) Kaplan-Meier plots indicating the prognostic power of common transcription factors having p-value and hazard ratio to demonstrate its statistical significance. The bottom-right insets present the number of individuals, the number of censored, and the CI of each risk group, and censoring samples are shown as ‘+’ marks. (B) Kaplan-Meier plots indicating the prognostic power of common receptors having p-value and hazard ratio to demonstrate its statistical significance. The bottom-right insets present the number of individuals, the number of censored, and the CI of each risk group, and censoring samples are shown as ‘+’ marks. (C) Kaplan-Meier plots indicating the prognostic power of common three miRNAs having p-value to demonstrate its statistical significance.

In prognostic analysis of miRNAs, 110 miRNAs which were common among four data sets were analyzed using TCGA expression data. Only 11 of reporter miRNAs among them were found as prognostic molecules having statistical significance. Five of these reporter miRNAs (miR-195- 5p, miR-181b-5p, miR-130b-5p, miR-497-5p, and miR-340-5p) were associated with thyroid cancer based on the literature search. Other six miRNAs (miR-30b-3p, miR-15b-5p, let-7a-5p, miR-130b-3p, miR-424- 5p, and miR-193b-3p) could be potential reporter biomolecules. Three of these candidates such as

miR-30b-3p (p-value:  $3.1 \times 10^{-4}$ ), miR-15b- 5p (p-value:  $2.8 \times 10^{-3}$ ) and let-7a-5p (p-value:  $2.5 \times 10^{-3}$ ) showed significant survival probability as seen in **Figure 3.15B**.

### **3.3.5. Identification of candidate repurposed therapeutic agents and molecular docking simulations**

After the determination of hub proteins (FYN and JUN) which are common in all datasets, we employed a drug repositioning methodology using CMap data in order to propose hypothesis with drugs targeting these proteins and currently used for other diseases but tentatively promising in cancer. FYN emerged as having upregulated profile in all healthy-tumor comparisons in the four datasets. To this end, we investigated and found 18 potential drugs which downregulate expression of FYN. Moreover, JUN was found as having downregulated profile in all healthy-tumor comparisons in the four datasets, and 6 potential drugs were found to upregulate the expression of JUN.

Molecular docking simulations were performed for further analysis to evaluate the binding affinities of the repurposed drugs with 3-D structures to their target proteins. The simulations were run three times for each drug-protein binding. In addition, further simulations were run to control the binding affinities between FYN and its known inhibitor PP2, and between JUN and its known inhibitor Tanshinone- IIA. Cut-offs for binding affinities were determined as  $\leq -6$  kcal/mol for FYN, and  $\leq -5.6$  kcal/mol for JUN according to the binding affinities of protein-inhibitor interactions. With reference to this, prochlorperazine (-6 kcal/mol), meclizine (-6.3 kcal/mol), rottlerin ( $-6.8 \pm 0.1$  kcal/mol), and cephaeline ( $-6.7 \pm 0.1$  kcal/mol) targeting FYN, and tretinoin (-5.6 kcal/mol) targeting JUN were proposed as candidate drugs for the treatment of PTC (**Figure 3.16**).

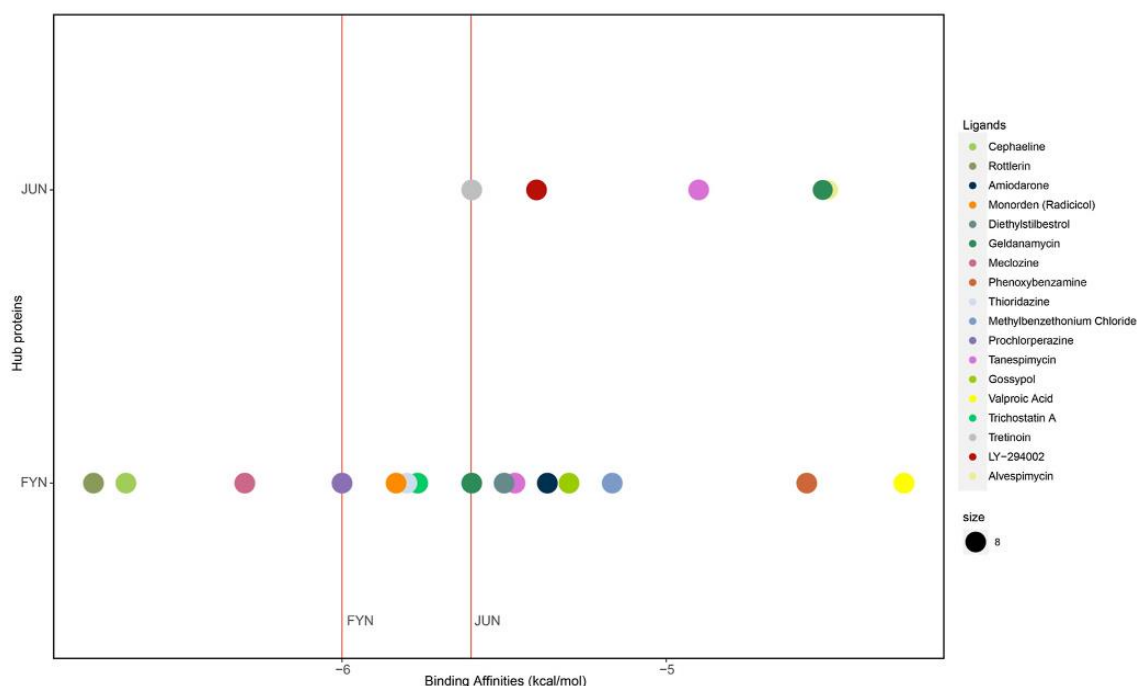


**Figure 3.16.** Bubble plot indicating the binding affinities of repurposed drugs to targeted proteins estimated through the molecular docking simulations. The red lines show the binding affinities between targeted proteins and their inhibitor ligands. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

### 3.3.6. Discussion

Recent cancer statistics showed that thyroid cancer is the most common endocrine cancer. Particularly, PTC accounts for the highest proportion of all thyroid malignant tumors (Siegel et al., 2020). Over the past several decades, a large number of omics data was generated at different molecular levels owing to the rapid development of high-throughput screening technologies and also collected in publicly available data sources. In parallel with the accumulation of data and development of genome-scale biomolecular networks, network analysis at the systems level has been increasingly used to reveal molecular mechanisms behind human diseases (Sevimoglu et al., 2018; Göv, 2019; Turanli and Arga, 2019; Turanli et al., 2019b; Gulfidan et al., 2020a). Although there are recent studies using omics data (Göv, 2019; Tang et al., 2019; Yeh et al., 2019). PTC is not yet supported by system-level biomarkers covering transcripts, miRNAs, TFs, and receptor molecules at once. In our meta-analysis, we analyzed four publicly available PTC datasets to determine DEGs in PTC relative to normal thyroid tissues. As far as we know, this is the first attempt in PTC research. Since we aim to combine information from

multiple microarray datasets for increased reliability and generalizability of results, we only focus on common 282 DEGs including 133 upregulated genes (**Figure 3.12A**) and 149 downregulated genes (**Figure 3.12B**) that found similar PTC gene expression signatures in all datasets.

The overrepresentation analyses revealed significant molecular pathways including p53 signaling pathway, focal adhesion kinase, axon guidance, ECM receptor interaction, tight junction, tyrosine, and retinol metabolism that were previously associated with thyroid cancer (Owens et al., 1996; Bhat et al., 1997; Simon et al., 2002; Haugen et al., 2004; Baris et al., 2005; Golubovskaya et al., 2009; Verbeek et al., 2011; Liu et al., 2013; Lee et al., 2015; Li et al., 2017b). The small cell lung cancer, glioma, pancreatic cancer, and colorectal cancer were the other molecular pathways found by enriched DEGs in PTC relative to normal thyroid tissues, which are demonstrated to contribute to tumor development. Besides, DEGs attached to these pathways might provide us promising biomarkers to uncover the mechanisms for PTC and relations with these cancer types as well.

The PPI networks were also constructed to clarify the interaction among the identified DEGs. Six hub proteins (FYN, JUN, LYN, PML, SIN3A, and RARA) came into prominence as the reporter signaling mediators in thyroid cancer. In a previous study, upregulation of FYN is determined in thyroid carcinogenesis, and this study highlighted the potential role of FYN inhibitors in TC therapy (Zheng et al., 2017). To the best of our knowledge, other hub proteins were not associated with papillary thyroid carcinoma yet.

22 proteins were identified as common reporter receptors and some of them were associated with PTC. For instance, amplification and/or overexpression of ERBB2 which stabilizes ligand binding and enhances kinase-mediated activation of downstream signaling pathways have been reported in numerous cancers, including breast and ovarian tumors. A previous study indicated that the ERBB2 gene is also overexpressed in malignant thyroid tumors (Bièche et al., 2001). ERBB4 is another reporter receptor that has been reported by the upregulation in BRAF wildlife (B-Raf Proto-Oncogene) PTC (Schulten et al., 2015). Upregulation of CDK1 in malignant tumors of thyroid cancer has been reported (Pozo et al., 2015). Moreover, HDACs present a potential target for increasing the sensitivity of thyroid cancer cells to radiotherapy (Wang et al., 2016). SRC,

proto-oncogene tyrosine-protein kinase, is responsible for the regulation of embryonic development and cell growth. Previously, the combined inhibition of SRC and the MAPK pathway showed great promise for improving the overall survival of advanced thyroid cancer patients with BRAF and RAS mutations, and activation of the PI3K pathway and rpS6 phosphorylation represent important biomarkers of response for patients treated with this therapy (Beadnell et al., 2018).

EP300 is indicated as one of the several oncogenic gene mutations that existed in thyroid cancers and a candidate target for the development of future therapies. Further investigation into the role of these genes including EP300 in PTC is required (Pitt et al., 2016). miRNAs are biomolecules with high interactions for cancer types and their deregulation contributes to cancer development. Mainly, the upregulation of some miRNAs leads to the silencing of tumor suppressor genes. On contrary, the downregulation of some miRNAs results in increased oncogene expression (Seema Sethi, 2015). miRNA expression profiles vary not only between tumors and normal tissues but also among different types of tumors or different stages (Li et al., 2013b). Therefore, the differential expression of miRNAs plays an important role in developing a specific strategy for cancer disease. In this study, we analyzed the TCGA data of 110 reporter miRNAs and found prognostic miRNAs for PTC. Concerning statistical significance, eleven of reporter micro-RNAs were found as prognostic. Five of these reporter micro-RNAs (miR-195-5p, miR-181b-5p, miR-130b-5p, miR-497-5p, and miR-340-5p) were associated to thyroid cancer in previous studies, other six miRNAs (miR-30b-3p, miR-15b-5p, let-7a-5p, miR-130b-3p, miR-424-5p, and miR-193b-3p) are potential reporter biomolecules proposed in this study for the first time.

miR-195-5p has previously been detected in many cancer diseases such as breast, colorectal, prostate cancer (Luo et al., 2014; Wu et al., 2015b; Sun et al., 2017). Despite its low expression in tumors, the existence of miR-195-5p in many cancer types makes it critical (Wang et al., 2017b). As well, miR-195-5p plays an active role in thyroid cancer. As a result of the experiment with microarrays, 427 expressed miRNAs were observed and 44 of them including miR-195-5p were found as differentially expressed despite inconsistency on the way of expressions (Swierniak et al., 2013; Hu et al., 2016).

miR-181b-5p defined an upregulated expression profile with an average of 4.75-fold change in 121 tissue samples in PTC while miR-130b-5p showed downregulated expression (Hu et al., 2016).

According to our results, miR-497-5p plays an important role in thyroid cancer in parallel to a previous study. Downregulation of miR-497 in thyroid tissues has been reported and served as a tumor suppressor [81]. It also plays an important role in the regulation of the PI3K/AKT pathway, and suppresses the brain-derived neurotrophic factor, supporting a crucial role in thyroid cancer (Wang et al., 2017a).

miR-340-5p overexpression has proven to cause the progression of thyroid cancer by suppressing BMP4 protein. Similar to our results, keeping this protein and miRNA pair under control might be an important strategy for thyroid cancer treatment (Zhao et al., 2018).

miR-15b-5p is another potential biomolecule. In previous studies, 84 different miRNAs were examined in 113 thyroid tissue samples, and samples were categorized into 3 groups: classic PTC (cPTC), encapsulated PTC (EPTC), and follicular variant of PTC (FVPTC). miR-15b-5p was characterized by downregulated expressions in all groups (Jahanbani et al., 2018).

Circulating miRNA profiles of the let-7 family have been reported as novel and potential noninvasive diagnostic, prognostic, treatment, and surveillance markers for PTC (Perdas et al., 2016). Although Let-7a-5p has been highly expressed with many cancers such as esophageal carcinoma, kidney chromophore, liver hepatocellular carcinoma, stomach adenocarcinoma, uterine corpus endometrial carcinoma (Zhan et al., 2014), it is the first time reported as a reporter miRNA for PTC in the present study.

miR-130b-3p has a downregulated expression profile in PTC, endometrial cancer, ovarian cancer, and colorectal cancer. On the other hand, it has an upregulated profile in gastric cancer, pancreatic cancer, and hepatocellular carcinoma (Chen et al., 2015). Moreover, miR-130b-3p is proved strong downregulated expression in aggressive tumors (Yip et al., 2011).

miR-30b-3p, miR-424-5p, and miR-193b-3p have previously been associated with different types of cancer and have not been directly associated with thyroid cancer until this study.

The transcriptional expression of the core genes of thyroid cancer was controlled by 12 common TFs: ETS1, GATA2, AR, YBX1, FOXP3, GATA1, PRDM14, E2F4, ESR1, GATA3, TFAP2C, and FOXA1. Among them, 7 TFs (ETS1, AR, YBX1, FOXP3, E2F4, ESR1, and FOXA1) were associated with PTC in previous studies. Upregulation of ETS1 was identified as a significant biomarker in differentiated thyroid cancer (Fuhrer et al., 2005); AR in PTC was found to be predominantly methylated at ten sites, compared to matched normal thyroid tissue samples, in the TCGA. Methylation of AR (10 or more sites) was accompanied by a nine-fold reduction in AR expression in male PTC, and a five-fold decrease in female PTC (Gupta et al., 2017), overexpression of YBX1 was observed in some follicular tumors and papillary carcinoma and plays a role in regulating the transcription as well as translation of genes contributing to the anaplastic transformation of thyroid carcinoma (Ito et al., 2003), FOXP3 expression was found in differentiated thyroid carcinoma cells and this expression may exert an important influence on several features of tumor aggressiveness (Cunha et al., 2012), E2F4 loss was stated as it suppresses the development of both pituitary and thyroid tumors in Rb/mice (Lee et al., 2002), ESR1 and ESR ratio can be used as prognostic markers to predict female patient survival and have potential as a therapeutic target was specified in one study (Yi et al., 2017), and FOXA1 was overexpressed in aggressive thyroid cancers and involved in cell cycle progression in an ATC cell line. Therefore, FOXA1 may be an important oncogene in thyroid tumorigenesis and a potential new therapeutic target for the treatment of anaplastic thyroid cancers (Nucera et al., 2009). In addition to these TFs, GATA3 was associated with Braf mutation which affects the PTC (Kim et al., 2010). The other 4 TFs (GATA2, GATA1, PRDM14, and TFAP2C) were proposed in PTC for the first time in the present study.

We applied the drug repositioning strategy to present drug candidates for PTC treatment by looking forward the opposite expression effect of the disease condition. After the validation of interaction between these repurposed drugs and FYN and JUN hub proteins

via molecular docking simulations, five drugs (prochlorperazine, meclizine, rottlerin, cephaeline, and tretinoin) were presented as candidate drugs for the treatment of PTC.

Prochlorperazine is an antipsychotic drug used in the treatment of schizophrenia because of blocking D2 dopamine receptors in the brain (Din and Preuss, 2021), anxiety, acute migraine, and in the relieving nausea and vomiting (Hodgson et al., 2021).

Meclozine (Meclizine) is an H-1 piperazine antihistamine derivative having the feature of antiemetic and antivertigo properties used in the treatment of vertigo and motion sickness (Lee and Jones, 2021). In the study performed by Lin et al., meclozine delivered into human colon cancer cell lines (COLO 205 and HT 29 cells) dose-dependently induced apoptosis by arresting the cell-cycle in the G0/G1 phase, releasing cytochrome C from mitochondria to the cytosol, the translocation of apoptosis-inducing factor (AIF) to the nucleus, and activating caspase 3, 8, and 9 after the upregulation of p53 and p21, downregulation of BCL2 and suppression of CDK2 and CDK4 kinase activity (Lin et al., 2007). Another study demonstrated that meclozine may have efficacy as a repurposed drug in the several cancer treatments such as hepatocellular carcinoma since the combination of meclozine and a glycolysis inhibitor inhibited the growth of human liver cancer cell line in vitro and in a xenotransplantation model and was highly active against acute myeloid leukemia (Guan et al., 2020). The glycolysis inhibitors make cancer cells more susceptible to the toxicity due to increase in phosphoethanolamine based on meclozine by decreasing the energy provided by glycolysis and by making cancer cells dependent on the energy provided by oxidative phosphorylation.

One of the protein kinase inhibitors, rottlerin, has important roles as antitumor, autophagy, anti-proliferative, anti-metastasis, and antiinvasive (Zhao et al., 2017). Rottlerin possessing multiple anti-cancer biological activities could be suggested as a potential agent for treatment of cancers because of its antitumor property with different mechanisms such as inducing of autophagy, inhibiting of cell growth, cell migration and invasion in diverse human cancers like lung, breast, and prostate cancer. In thyroid cancer, rottlerin has several functions including decreasing of adhesion to extracellular matrix, reducing of migration ability, altering of morphology, decreasing of Rho GTPase activity, and reducing of protein levels of integrin  $\beta 1$ , FAK, and focal adhesion complex constituents (Lin et al., 2010; Ma et al., 2018). It was demonstrated that rottlerin is a

potential anti-cancer agent for treatment of thyroid carcinoma since it inhibits the migration of CGTH W-2 human follicular thyroid carcinoma cells by disrupting focal adhesions in a PKC $\delta$ -independent manner (Lin et al., 2010).

Cephaeline is a compound used as an emetic agent and induces vomiting through acting on stomach lining, therefore, it is mostly recommended as an emergency treatment for accidental poisoning (Patel and Patel, 2020). In several studies, cephaeline has been presented as repurposed drugs for the treatment of several cancer types such as kidney renal clear cell carcinoma (Xing et al., 2020), and colorectal cancer (Mastrogamvraki and Zaravinos, 2020).

Tretinoin or all-trans retinoic acid is a vitamin A derivative and it is used in the treatment of acne vulgaris, certain types of promyelocytic leukemia, and fine wrinkles (Bagatin et al., 2018; Dogra et al., 2020; Przybylski et al., 2021). NIS, the plasma membrane glycoprotein sodium/iodide symporter, plays roles in the iodide uptake of thyrocytes and thyroid hormone biosynthesis. It was reported that iodide uptake by thyroid cancer cells during <sup>131</sup>I therapy was decreased by down-regulated expression of the NIS gene and this situation caused resistance to radioiodine therapy. Tretinoin treatment has been shown to elevate gene expression of NIS and stimulates <sup>125</sup>I uptake by ARO human thyroid anaplastic carcinoma cell lines (Jeong et al., 2006). Therefore, it was suggested that tretinoin treatment have potential in the treatment of thyroid cancer besides using as an anticancer drug in the treatment of acute promyelocytic leukemia due to binding to alpha, beta, and gamma retinoic acid receptors (RARs) which have been associated with the development of acute promyelocytic leukemia and squamous cell cancers.

Geldanamycin and tanespimycin were common drugs found for JUN and FYN hub proteins. Geldanamycin, a specific inhibitor of heat shock protein 90 (HSP90) was used in six thyroid cancer cell lines and observed an inhibitory effect on cell proliferation and EGF-induced invasion (Park et al., 2003). Besides, geldanamycin was shown in other studies to inhibit cancer cells growth such as multiple myeloma (Jurczynszyn et al., 2014), oral cancer (Lee, 2018) prostate cancer (Wang et al., 2015; Turanli et al., 2019e). On the other hand, tanespimycin, another HSP90 inhibitor, was also studied to inhibit various cancer cells such as thyroid, breast cancer (Yu et al., 2010; Modi et al., 2011).

#### 4. CONCLUSION

The integrative method for multi-omics data is an impressive strategy for uncovering the underlying mechanisms governing information flow within a cell. A diverse array of publicly accessible tools and techniques exists for the integration of multi-omics data, which facilitates the extraction of significant knowledge. In particular, comprehensive integrated analyzes will be able to reveal the complex mechanisms underlying cancer development and progression. As tools and methods are largely isolated, a standardized structure designed for the efficient handling of multi-omics data end-to-end with easy visualization and interpretation is needed.

As a conclusion, this study indicated that TP53 and NF1 genes have several mutations significantly differentiated among subtypes. There is no statistically significant difference among subtypes based on average overall survival time. Based on the GBM classification, it was found that CS1 was more similar to proneural subtype, CS2 was more similar to mesenchymal subtype, CS3 was more similar to classical subtype. TMZ causes resistance for some GBM patient groups. Estimated IC50 values for TMZ were determined significantly lower in CS2, especially compared to CS1 ( $P = 0.0036$ ). Patients belonging to mesenchymal subtype was sensitive to TMZ treatment more than patients belonging to proneural subtype.

In addition, the present study systematically identified gene expression signatures of PTC relative to normal thyroid tissues from four publicly available datasets. In addition to a meta-analysis of PTC, we also determined reporter biomolecules consisting of hub proteins, receptor proteins, TFs, and miRNAs for the first time. It is worth noting that essential biomarkers are found and their prognostic values are generated efficiently by using systematic approaches. These findings provide not only insights into the mechanisms underlying PTC progression and some potential targets toward PTC. Despite the tremendous significance of the findings of this present study, more efforts are required to achieve the experimental and clinical validation of the findings obtained here.

Also in this study, we validated for the first time the growth inhibitory effects of risperidone and clozapine in three ESCC cell lines. In addition to the recently published data on the cytotoxic properties of antipsychotics on cancer cells, there are also concerns

raised because of side effects, including the potential genotoxic or carcinogenic effects of antipsychotics based on their long-term administrations. Therefore, oncologist and psychiatrists could work closely together to prescribe antipsychotics based on the benefits-to-risk ratio. We further recommend to not only explore the antipsychotics, but also other drugs short-listed in our study as novel drug candidates as an effective strategy for precision and personalized medicine in treating ESCC patients.



## REFERENCES

- Aguirre-Gamboa R, Gomez-Rueda H, Martí'nez-Ledesma E, et al. (2013). SurvExpress: An online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One 8, e74250.
- Akavia, U. D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H. C., . . . Pe'er, D. (2010). An integrated approach to uncover drivers of cancer. *Cell*, 143(6), 1005-1017.
- Amelio I, Gostev M, Knight RA, et al. (2014). DRUGSURV: A resource for repositioning of approved and experimental drugs in oncology based on patient survival information. *Cell Death Dis* 5, e1051.
- Ashburner M, Ball CA, Blake JA, et al. (2000). Gene ontology: Tool for the unification of biology. *Nat Genet* 25, 25–29.
- Arga, K. Y. (2019). Interview with Prof. K. Yalçın Arga: A pioneer of multi-omics science and health care innovation. *Omics: a journal of integrative biology*, 23(9), 460-462.
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., . . . Stegle, O. (2017). Multi-Omics factor analysis disentangles heterogeneity in blood cancer. *BioRxiv*, 217554.
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., . . . Stegle, O. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular systems biology*, 14(6), e8124.
- Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., . . . Papatheodorou, I. (2019). ArrayExpress update—from bulk to single-cell expression data. *Nucleic acids research*, 47(D1), D711-D715.
- Aure, M. R., Steinfeld, I., Baumbusch, L. O., Liestøl, K., Lipson, D., Nyberg, S., . . . Børresen-Dale, A.-L. (2013). Identifying in-trans process associated genes in breast cancer by integrated analysis of copy number and expression data. *PloS one*, 8(1), e53014.
- Aydin B, and Arga KY (2019). Co-expression network analysis elucidated a core module

- in association with prognosis of nonfunctioning non-invasive human pituitary adenoma. *Front Endocrinol* 10, 1–15.
- Aydin, B., Arga, K. Y., & Karadag, A. S. (2020). Omics-driven biomarkers of psoriasis: recent insights, current challenges, and future prospects. *Clinical, Cosmetic and Investigational Dermatology*, 13, 611.
- Aydin, B., Caliskan, A., & Arga, K. Y. (2021). Overview of omics biomarkers in pituitary neuroendocrine tumors to design future diagnosis and treatment strategies. *EPMA Journal*, 12(3), 383-401.
- Ayyildiz D, Gov E, Sinha R, and Arga KY (2017). Ovarian cancer differential interactome and network entropy analysis reveal new candidate biomarkers. *OMICS J Integr Biol* 21, 285–294.
- Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A.-M., Gingras, M.-C., . . . Quinn, M. C. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*, 531(7592), 47-52.
- Barrett T, Wilhite SE, Ledoux P, et al. (2013). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res* 41, D991–D995. Basha O, Shpringer R, Argov CM, and Yeger-Lotem E (2018). The DifferentialNet database of differential protein-protein interactions in human tissues. *Nucleic Acids Res* 46, D522– D526.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., . . . Holko, M. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*, 41(D1), D991-D995.
- Beklen, H., Arslan, S., Gulfidan, G., Turanli, B., Ozbek, P., Karademir Yilmaz, B., & Arga, K. Y. (2021). Differential Interactome Based Drug Repositioning Unraveled Abacavir, Exemestane, Nortriptyline Hydrochloride, and Tolcapone as Potential Therapeutics for Colorectal Cancers. *Front. Bioinforma*, 1(10.3389). <https://doi.org/10.3389/fbinf.2021.710591>
- Berman HM, Battistuz T, Bhat TN, et al. (2002). The protein data bank. *Acta Crystallogr*

D Biol Crystallogr 58, 899–907.

Bonnet, E., Calzone, L., & Michoel, T. (2015). Integrative multi-omics module network inference with Lemon-Tree. *PLoS computational biology*, 11(2), e1003983.

Bray F, Ferlay J, Soerjomataram I, et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394–424.

Bredel, M., Bredel, C., Juric, D., Harsh, G. R., Vogel, H., Recht, L. D., & Sikic, B. I. (2005). Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas. *Cancer research*, 65(19), 8679-8689.

Cacchiarelli, D., Trapnell, C., Ziller, M. J., Soumillon, M., Cesana, M., Karnik, R., . . . Zhang, X. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. *Cell*, 162(2), 412-424.

Caldera M, Buphamalai P, Müller F, and Menche J (2017). Interactome-based approaches to human disease. *Curr Opin Syst Biol* 3, 88–94.

Calimlioglu B, Karagoz K, Sevimoglu T, et al. (2015). Tissuespecific molecular biomarker signatures of type 2 diabetes: An integrative analysis of transcriptomics and protein-protein interaction data. *OMICS J Integr Biol* 19, 563–573.

Caliskan A, Gulfidan G, Sinha R, and Arga KY. (2021). Differential interactome proposes subtype-specific biomarkers and potential therapeutics in renal cell carcinomas. *J Pers Med* 11, 158.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer discovery*, 2(5), 401-404. <https://doi.org/10.1158/2159-8290.cd-12-0095>

Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., . . . Biospecimen Core Resource, C. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*, 45(10), 1113-1120.

<https://doi.org/10.1038/ng.2764>

- Chari, R., Coe, B. P., Vucic, E. A., Lockwood, W. W., & Lam, W. L. (2010). An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. *BMC systems biology*, 4(1), 1-14.
- Chatr-Aryamontri A, Oughtred R, Boucher L, et al. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 45, D369–D379.
- Chen VC-H, Hsieh Y-H, Lin T-C, et al. (2020). New use for old drugs: The protective effect of risperidone on colorectal cancer. *Cancers* 12, 1560.
- Chen, J., & Zhang, S. (2016). Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data. *Bioinformatics*, 32(11), 1724-1732.
- Chen, J., Bushman, F. D., Lewis, J. D., Wu, G. D., & Li, H. (2013). Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. *Biostatistics*, 14(2), 244-258.
- Cho, D.-Y., & Przytycka, T. M. (2013). Dissecting cancer heterogeneity with a probabilistic genotype-phenotype model. *Nucleic acids research*, 41(17), 8011-8020.
- Çibikdiken, A. O., Buyukyilmaz, M., Nacar, M. A., Topcu, A. E., Tambag, Y., & Karadag, A. (2018). Development of a New Platform Through Distributed Storage System for the Bioinformatics Analysis. *DEStech Transactions on Computer Science and Engineering(AIIE* 2017). <https://doi.org/10.12783/dtcse/aiie2017/18233>
- Croft D, O’Kelly G, Wu G, et al. (2011). Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res* 39, D691–D697.
- Cruz DL, de Mello AV, Nogueira LR, et al. (2014). Oral condition and nutritional status of patients in oral clinics. *Revista Univap. São Jose’ dos Campos-SP-Brasil* 21, 21–30.
- Cui G, Li C, Xu G, et al. (2018). Tumor-associated fibroblasts and microvessels

- contribute to the expression of immunosuppressive factor indoleamine 2,3-dioxygenase in human esophageal cancers. *Pathol Oncol Res* 24, 269–275.
- Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., . . . Yuan, Y. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403), 346-352.
- Das, T., Andrieux, G., Ahmed, M., & Chakraborty, S. (2020). Integration of online omics-data resources for cancer research. *Frontiers in Genetics*, 11.
- de Tayrac, M., Lê, S., Aubry, M., Mosser, J., & Husson, F. (2009). Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. *BMC Genomics*, 10(1), 32. <https://doi.org/10.1186/1471-2164-10-32>
- Dietrich, S., Oleś, M., Lu, J., Sellner, L., Anders, S., Velten, B., . . . Walther, T. (2018). Drug-perturbation-based stratification of blood cancer. *The Journal of clinical investigation*, 128(1), 427-445.
- Dilly SJ, Clark AJ, Marsh A, et al. (2017). A chemical genomics approach to drug reprofiling in oncology: Antipsychotic drug risperidone as a potential adenocarcinoma treatment. *Cancer Lett* 393, 16–21.
- Dimitrakopoulos, C., Hindupur, S. K., Häfliger, L., Behr, J., Montazeri, H., Hall, M. N., & Beerenwinkel, N. (2018). Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics*, 34(14), 2441-2448.
- Du Z, Xia Q, Wu B, et al. (2019). The analyses of SRCR genes based on protein-protein interaction network in esophageal squamous cell carcinoma. *Am J Transl Res* 11, 2683–2705.
- Edfors, F, Danielsson F, Hallström BM, et al. (2016). Genespecific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* 12, 883.
- Fond G, Macgregor A, Attal J, et al. (2012). Antipsychotic drugs: Pro-cancer or anti-cancer? A systematic review. *Med Hypothes* 79, 38–42.

- Goldman, M., Craft, B., Swatloski, T., Cline, M., Morozova, O., Diekhans, M., . . . Zhu, J. (2015). The UCSC cancer genomics browser: update 2015. *Nucleic acids research*, 43(D1), D812-D817.
- Gov E, and Arga KY (2017). Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer. *Sci Rep* 7, 1–10.
- Gov, E., Kori, M., & Arga, K. Y. (2017). Multiomics analysis of tumor microenvironment reveals Gata2 and miRNA-124-3p as potential novel biomarkers in ovarian cancer. *Omics: a journal of integrative biology*, 21(10), 603-615.
- Griffin SM, and Wahed S (2011). Oesophageal cancer. *Surgery* 29, 557–562.
- Gulfidan, G., Turanli, B., Beklen, H., Sinha, R., & Arga, K. Y. (2020). Pan-cancer mapping of differential protein-protein interactions. *Scientific reports*, 10(1), 1-12.
- Guo, H., Ahmed, M., Zhang, F., Yao, C. Q., Li, S., Liang, Y., . . . Langstein, J. (2016). Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nature genetics*, 48(10), 1142-1150.
- Hendouei N, Saghafi F, Shadfar F, and Hosseinimehr SJ (2019). Molecular mechanisms of anti-psychotic drugs for improvement of cancer treatment. *Eur J Pharmacol* 856, 172402.
- Higdon, R., Earl, R. K., Stanberry, L., Hudac, C. M., Montague, E., Stewart, E., . . . Kolker, N. (2015). The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. *Omics: a journal of integrative biology*, 19(4), 197-208.
- Huang J, Zhao D, Liu Z, and Liu F (2018). Repurposing psychiatric drugs as anti-cancer agents. *Cancer Lett* 419, 257– 265.
- Ideker, T. (2002). Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18, S233-S240.
- Ivanov AA, Revennaugh B, Rusnak L, et al. (2018). The OncoPPi Portal: An integrative

resource to explore and prioritize protein-protein interactions for cancer target discovery. *Bioinformatics* 34, 1183–1191.

Jennings, E. M., Morris, J. S., Carroll, R. J., Manyam, G. C., & Baladandayuthapani, V. (2013). Bayesian methods for expression-based integration of various types of genomics data. *EURASIP Journal on Bioinformatics and Systems Biology*, 2013(1), 1-11.

Jones, P., Côté, R. G., Martens, L., Quinn, A. F., Taylor, C. F., Derache, W., . . . Apweiler, R. (2006). PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic acids research*, 34(suppl\_1), D659-D663.

Kale, N. S., Haug, K., Conesa, P., Jayseelan, K., Moreno, P., Rocca-Serra, P., . . . Li, X. (2016). MetaboLights: an open-access database repository for metabolomics data. *Current protocols in bioinformatics*, 53(1), 14.13. 11-14.13. 18.

Kamburov A, Stelzl U, Lehrach H, and Herwig R (2013). The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Res* 41, D793–D800.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45, D353– D361.

Karagoz K, Lehman HL, B. Stairs D, Sinha R, and Y. Arga K (2016b). Proteomic and metabolic signatures of esophageal squamous cell carcinoma. *Curr Cancer Drug Targets* 16, 721– 736.

Karagoz K, Sevimoglu T, and Arga KY (2016a). Integration of multiple biological features yields high confidence human protein interactome. *J Theor Biol* 403, 85– 96.

Karagoz K, Sinha R, and Arga KY (2015). Triple negative breast cancer: A multi-omics network discovery strategy for candidate targets and driving pathways. *OMICS J Integr Biol* 19, 115–130.

Kim S, Chen J, Cheng T, et al. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res* 47, D1102–D1109.

- Kim SM, Park YY, Park ES, et al. (2010). Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. *PLoS One* 5, e15074.
- Kim, D., Li, R., Dudek, S. M., & Ritchie, M. D. (2013). ATHENA: Identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. *BioData mining*, 6(1), 1-14.
- Kim, D., Li, R., Lucas, A., Verma, S. S., Dudek, S. M., & Ritchie, M. D. (2017). Using knowledge-driven genomic interactions for multi-omics data analysis: metadimensional models for predicting clinical outcomes in ovarian carcinoma. *Journal of the American Medical Informatics Association*, 24(3), 577-587.
- Kim, D., Shin, H., Song, Y. S., & Kim, J. H. (2012). Synergistic effect of different levels of genomic data for cancer clinical outcome prediction. *Journal of biomedical informatics*, 45(6), 1191-1198.
- Kim, M., & Tagkopoulos, I. (2018). Data integration and predictive modeling methods for multi-omics datasets. *Molecular omics*, 14(1), 8-25.
- Kirk, P., Griffin, J. E., Savage, R. S., Ghahramani, Z., & Wild, D. L. (2012). Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics*, 28(24), 3290-3297.
- Kodama, Y., Shumway, M., & Leinonen, R. (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic acids research*, 40(D1), D54-D56.
- Kori M, Gov E, and Arga KY (2019). Novel genomic biomarker candidates for cervical cancer as identified by differential coexpression network analysis. *OMICS J Integr Biol* 23, 261– 273.
- Kori, M., & Yalcin Arga, K. (2018). Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective. *PloS one*, 13(7), e0200717.
- Lanckriet, G. G, Bie, T. De, N. Cristianini, Jordan, MI, & Noble, WS 2004. A statistical framework for genomic data fusion. *Bioinformatics*, 20.

- Lawrence MS, Stojanov P, Mermel CH, et al. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501. Li K, Wang F, and Hu Z-W (2017). Targeting TRIB3 and PMLRARA interaction against APL. *Oncotarget* 8, 52012–52013.
- Li, J., Lu, Y., Akbani, R., Ju, Z., Roebuck, P. L., Liu, W., . . . Kane, D. W. (2013). TCPA: a resource for cancer functional proteomics data. *Nature methods*, 10(11), 1046-1047.
- Liang, L., Song, L., Yang, Y., Tian, L., Li, X., Wu, S., . . . Ding, K. (2016). Validation of a multi-omics strategy for prioritizing personalized candidate driver genes. *Oncotarget*, 7(25), 38440.
- Lin, D., Zhang, J., Li, J., Calhoun, V. D., Deng, H. W., & Wang, Y. P. (2013). Group sparse canonical correlation analysis for genomic data integration. *BMC bioinformatics*, 14(1), 245. <https://doi.org/10.1186/1471-2105-14-245>
- Liu, G., Dong, C., & Liu, L. (2016). Integrated multiple “-omics” data reveal subtypes of hepatocellular carcinoma. *PloS one*, 11(11), e0165457.
- Lock, E. F., & Dunson, D. B. (2013). Bayesian consensus clustering. *Bioinformatics*, 29(20), 2610-2616.
- Lock, E. F., Hoadley, K. A., Marron, J. S., & Nobel, A. B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The annals of applied statistics*, 7(1), 523.
- López de Maturana, E., Alonso, L., Alarcón, P., Martín-Antoniano, I. A., Pineda, S., Piorno, L., . . . Malats, N. (2019). Challenges in the integration of omics and non-omics data. *Genes*, 10(3), 238.
- Louhimo, R., & Hautaniemi, S. (2011). CNAmets: an R package for integration of copy number, expression and methylation data. *Bioinformatics*, 27(887), 8.
- Mankoo, P. K., Shen, R., Schultz, N., Levine, D. A., & Sander, C. (2011). Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles. *PloS one*, 6(11), e24709.

- Mo, Q., Wang, S., Seshan, V. E., Olshen, A. B., Schultz, N., Sander, C., . . . Shen, R. (2013). Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proceedings of the National Academy of Sciences*, 110(11), 4245-4250.
- Montague E, Stanberry L, Higdon R, et al. (2014). MOPED 2.5- an integrated multi-omics resource: Multi-omics profiling expression database now includes transcriptomics data. *OMICS J Integr Biol* 18, 335–343.
- Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., . . . Schulman, B. A. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *New England Journal of Medicine*, 360(5), 470-480.
- Nalbantoglu, S., & Karadag, A. (2019). Introductory chapter: insight into the OMICS technologies and molecular medicine. In *Molecular Medicine*. IntechOpen.
- Nalbantoglu, S., & Karadag, A. (2021). Metabolomics bridging proteomics along metabolites/oncometabolites and protein modifications: paving the way toward integrative multiomics. *Journal of pharmaceutical and biomedical analysis*, 114031.
- Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D., & Cox, L. A. (2019). The need for multi-omics biomarker signatures in precision medicine. *International journal of molecular sciences*, 20(19), 4781.
- Oughtred R, Stark C, Breitkreutz BJ, et al. (2019). The Bio- GRID interaction database: 2019 update. *Nucleic Acids Res* 47, D529–D541.
- Ovaska, K., Laakso, M., Haapa-Paananen, S., Louhimo, R., Chen, P., Aittomäki, V., . . . Karinen, S. (2010). Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. *Genome medicine*, 2(9), 1-12.
- Overbeeke MR, Beex-Oosterhuis MM, Graveland E, Hoed-van Wijk JN den, and Van Gool AR (2016). Successful clozapine continuation during chemotherapy for the treatment of malignancy: A case report. *Int J Clin Pharm* 38, 199–202.
- Ozer, M. E., Sarica, P. O., & Arga, K. Y. (2020). New machine learning applications to accelerate personalized medicine in breast cancer: rise of the support vector

- machines. *Omics: a journal of integrative biology*, 24(5), 241-246.
- Peters CJ, Rees JRE, Hardwick RH, et al. (2010). A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. *Gastroenterology* 139, 1995–2004.e15.
- Ray, P., Zheng, L., Lucas, J., & Carin, L. (2014). Bayesian joint analysis of heterogeneous genomics data. *Bioinformatics*, 30(10), 1370-1376.
- Rolland T, Tasxan M, Charlotiaux B, et al. (2014). A proteomescale map of the human interactome network. *Cell* 159, 1212– 1226.
- Roth BL, Sheffer DJ, and Kroeze WK (2004). Magic shotguns versus magic bullets: Selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* 3, 353–359.
- Ruffalo, M., Koyutürk, M., & Sharan, R. (2015). Network-based integration of disparate omic data to identify” silent players” in cancer. *PLoS computational biology*, 11(12), e1004595.
- Samaras P, Schmidt T, Frejno M, et al. (2020). ProteomicsDB: A multi-omics and multi-organism resource for life science research. *Nucleic Acids Res* 48, D1153–D1163.
- Schaab C, Geiger T, Stoehr G, Cox J, and Mann M (2012). Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol Cell Proteomics* 11, M111.014068.
- Seoane, J. A., Day, I. N., Gaunt, T. R., & Campbell, C. (2014). A pathway-based data integration framework for prediction of disease progression. *Bioinformatics*, 30(6), 838-845.
- Sevimoglu T, and Arga KY (2014). The role of protein interaction networks in systems biomedicine. *Comput Struct Biotechnol J* 11, 22–27.
- Sevimoglu, T., & Arga, K. Y. (2015). Computational systems biology of psoriasis: Are we ready for the age of omics and systems biomarkers? *Omics: a journal of integrative biology*, 19(11), 669-687.

- Shannon P, Markiel A, Ozier O, et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504.
- Shen, R., Mo, Q., Schultz, N., Seshan, V. E., Olshen, A. B., Huse, J., . . . Sander, C. (2012). Integrative subtype discovery in glioblastoma using iCluster. *PloS one*, 7(4), e35236.
- Shen, R., Olshen, A. B., & Ladanyi, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*, 25(22), 2906-2912.
- Speicher, N. K., & Pfeifer, N. (2015). Integrating different data types by regularized unsupervised multiple kernel learning with application to cancer subtype discovery. *Bioinformatics*, 31(12), i268-i275.
- Stelzer G, Rosen N, Plaschkes I, et al. (2016). The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics* 54, 1.30.1–1.30.33.
- Su H, Hu N, Yang HH, et al. (2011). Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin Cancer Res* 17, 2955–2966.
- Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinformatics and biology insights*, 14, 1177932219899051.
- Sun H, Cai X, Zhou H, et al. (2018). The protein–protein interaction network and clinical significance of heat-shock proteins in esophageal squamous cell carcinoma. *Amino Acids* 50, 685–697.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . Tsafou, K. (2014). *Nucleic Acid Res* 43. D447.
- Tomczak K, Czerwin´ska P, and Wiznerowicz M (2015). The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Wspolczesna Onkologia* 1A,

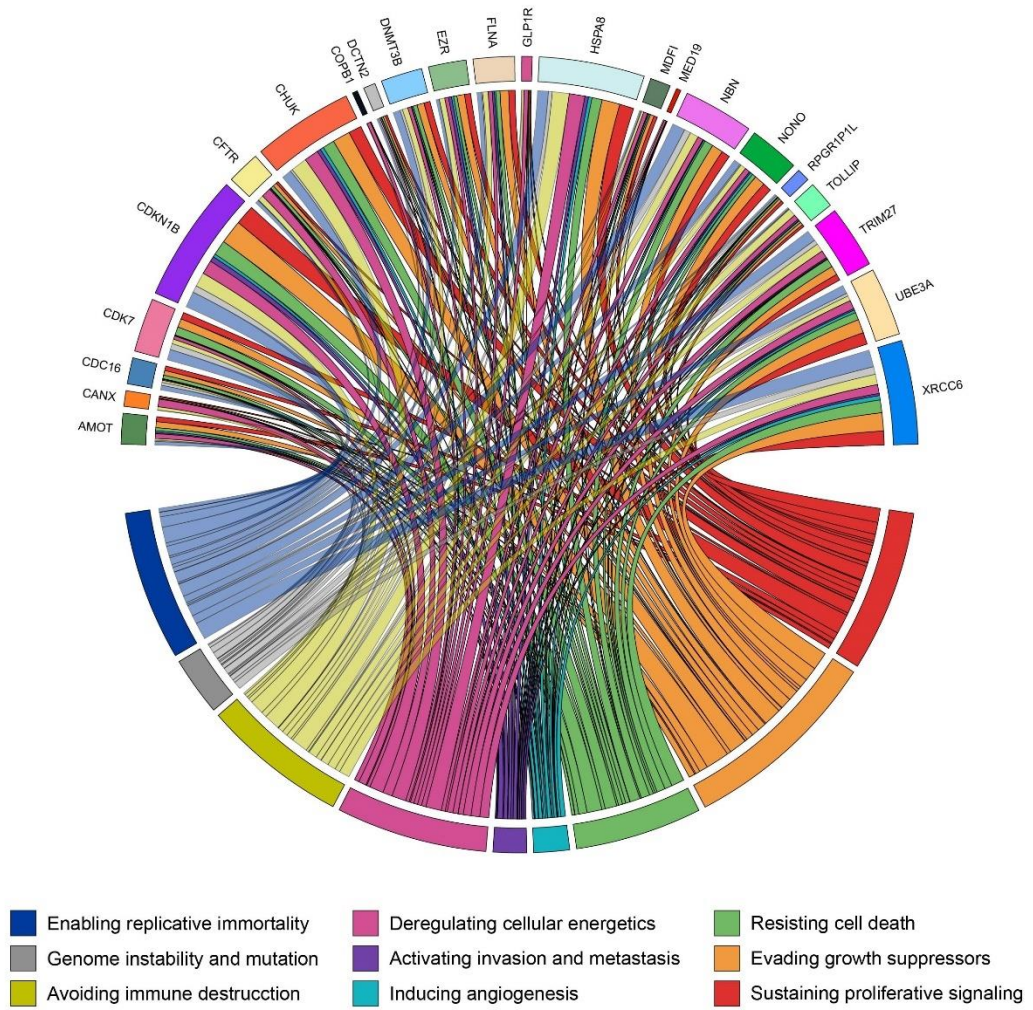
A68–A77.

- Trott O, and Olson AJ (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31, 455–461.
- Tungekar A, Mandarathi S, Mandaviya PR, et al. (2018). ESCC ATLAS: A population wide compendium of biomarkers for Esophageal Squamous Cell Carcinoma. *Sci Rep* 8, 12715.
- Turanli B, Altay O, Bore'n J, et al. (2021). Systems biology based drug repositioning for development of cancer therapy. *Semin Cancer Biol* 68, 47–58.
- Turanli B, and Arga KY (2017). Systems biomedicine acts as a driver for the evolution of pharmacology. *Ann Pharmacol Pharm* 2, 1087.
- Turanli B, Gulfidan G, and Arga KY (2017). Transcriptomicguided drug repositioning supported by a new bioinformatics search tool: GeneXpharma. *OMICS J Integr Biol* 21, 584– 591.
- Turanli B, Karagoz K, Bidkhorji G, et al. (2019a). Multi-omic data interpretation to repurpose subtype specific drug candidates for breast cancer. *Front Genet* 10, 420.
- Turanli B, Zhang C, Kim W, et al. (2019b). Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning. *EBioMedicine* 42, 386–396.
- Turanli, B., Karagoz, K., Gulfidan, G., Sinha, R., Mardinoglu, A., & Arga, K. Y. (2018). A network-based cancer drug discovery: from integrated multi-omics approaches to precision medicine. *Current pharmaceutical design*, 24(32), 3778-3790.
- Uhle'n M, Fagerberg L, Hallstro'm BM, et al. (2015). Tissuebased map of the human proteome. *Science* 347, 1260419.
- Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., . . . Stuart, J. M. (2010). Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*, 26(12), i237-i245.

- Velankar S, Van Ginkel G, Alhroub Y, et al. (2016). PDBe: Improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res* 44, D385–D395.
- Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., . . . Bandeira, N. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology*, 32(3), 223-226.
- Wang X (2011). Role of clinical bioinformatics in the development of network-based Biomarkers. *J Clin Bioinformatics* 1, 28.
- Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., . . . Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*, 11(3), 333-337.
- Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3), 515-534.
- Xie W, Huang P, Wu B, et al. (2019). Clinical significance of LOXL4 expression and features of LOXL4-associated protein– protein interaction network in esophageal squamous cell carcinoma. *Amino Acids* 51, 813–828.
- Yeger-Lotem E, and Sharan R (2015). Human protein interaction networks across tissues and diseases. *Front Genet* 6, 1–5.
- Yeh W-L, Lin H-Y, Wu H-M, and Chen D-R (2014). Combination treatment of tamoxifen with risperidone in breast cancer. *PLoS One* 9, e98805.
- You, Z.-H., Yin, Z., Han, K., Huang, D.-S., & Zhou, X. (2010). A semi-supervised learning approach to predict synthetic genetic interactions by combining functional and topological properties of functional gene network. *BMC bioinformatics*, 11(1), 1-13.
- Yuan, Y., Savage, R. S., & Markowetz, F. (2011). Patient-specific data fusion defines prognostic cancer subtypes. *PLoS computational biology*, 7(10), e1002227.

- Zhang X, Yuan A, Zhao X, Li Z, and Cui G (2020). Tumoral expression of CD166 in human esophageal squamous cell carcinoma: implications for cancer progression and prognosis. *Cancer Biother Radiopharm* 35, 214–222.
- Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., . . . Ferretti, V. (2019). The international cancer genome consortium data portal. *Nature biotechnology*, 37(4), 367-369.
- Zhang, Z.-N., Li, Q., Liu, C., Wang, H.-B., Wang, Q., & Bao, L. (2008). The voltage-gated Na<sup>+</sup> channel Nav1.8 contains an ER-retention/retrieval signal antagonized by the  $\beta 3$  subunit. *Journal of cell science*, 121(19), 3243-3252. <https://doi.org/10.1242/jcs.026856>

APPENDIX



**Supplementary Figure S1.** Circos plot indicating cancer hallmarks by creating the number of molecular pathways/biological processes in each hallmark based on genes taking part in these pathways/processes.

**Supplementary Table S1.** Elite lists of mRNAs, lncRNAs, methylation probes, and genes having mutations

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mRNA | <p> IRX1, SAA1, EGFR, POSTN, PI3, LTF, NNAT, TMSB15A, HOXD13, PLA2G2A, SOX10, F13A1, SLC17A7, SMOC1, MOG, NEFL, DCX, CHGA, HOXB2, LHFPL3, SLN, CHST8, NSG2, NDN, ATCAY, CPLX2, KCNE5, COL20A1, CADM3, MBP, GPR17, PCDHGB7, CHI3L1, MAG, HOXC10, PLPPR1, HLA-DQA2, CD24, TAC1, NNMT, SLPI, LUZP2, NKX2-5, STMN2, GSX1, SEZ6, TNF, CA3, RPRM, HOXB4, UGT8, GALNT13, KCNF1, IBSP, ACKR1, CXCL14, COL6A2, CXCL8, ENPP5, FAM163B, INA, MOBP, CCK, NPNT, SYT1, ADGRB1, VSTM2B, PCDHGB2, S100A8, SLC6A11, HES5, SST, COL11A1, BCAS1, MSTN, VSNL1, DMRTA2, ASIC4, NXPH1, SLC44A5, GALNT17, LY6H, SNCB, DLL3, ENTPD2, CNDP1, NLGN4Y, KCND2, SHD, EYA4, CXCL10, PPP1R1B, MMP9, DPEP1, IGFBPL1, HOXB7, RPH3A, SHISA6, CA9, CARNS1, TCEAL6, CHODL, VAT1L, C6orf15, HOPX, THNSL2, COL28A1, SEZ6L, KLK6, MARCO, PITX1, HAPLN2, COL9A3, IGLON5, PHACTR3, SULF1, LMO1, MOXD1, PEG10, TRH, TMEFF2, STAC, COL8A1, NEUROD1, SLC14A1, CHRNA9, IGF2BP2, SPON1, FA2H, LRRC61, S100A9, CA10, CHI3L2, MYRF, PDYN, AZGP1, COL1A1, NPY, MEOX2, SOX3, RHOD, MAGEE2, C6orf141, HBB, ACTL6B, SNAP91, SH3GL2, GPD1, C5orf49, MYT1, HLA-DRB5, PACSIN1, SPOCD1, CPNE4, ELOVL2, EMILIN3, NKAIN1, NSG1, CCN4, PCDHB2, LGR6, FBLL1, C1QL1, INSYN2A, BTBD17, KDM5D, CLDN11, NTSR2, HP, GABBR2, KCNQ2, CACNG4, GJB2, NEU4, ZNF488, PI16, HOXD10, VGF, TMEM163, ABCC3, SOCS2, TF, PCDHB7, PLP1, CENPV, HOXD9, RCOR2, LGI3, OTX1, GSTM1, CELF3, NEFM, GRID2, SNX22, ERMN, CELF5, HOXC9, GSTM5, SYT13, PTPRN, SNAP25, ERBB3, CFAP300, NPTX1, B3GAT2, ETNPPL, CAV1, METTL7B, PTPN5, SPOCK3, IL13RA2, COL5A1, TMEM132E, RIMS4, NKAIN4, CD163, CRB2, TMEM179, TUBB4A, NETO2, MXRA5, PDLIM4, PDGFRA, SOX11, SYT4, MMD2, PCDHB3, SULT4A1, CNTN2, MYBPC1, HOXA7, LYVE1, TRAM1L1, RPE65, NKX6-2, ZDHHC22, DUSP26, APLNR, DDIT4L, MYBPH, FXYP7, NPTX2, KIF21B, PLA2G5, TMEM132B, TMEM125, CXCL11, NRG1, L1CAM, CREB3L1, MMP19, MCHR1, MAL, C2orf80, CAMK2A, CCER2, LRRC17, CLGN, CPLX1, TPPP, TMEM132C, HPR, CTHRC1, CRTAC1, SFRP1, VEGFA, HOXB3, IFI44L, ELAVL4, CNTFR, NGEF, ATP10B, HIF3A, ESM1, GALR1, FAM110C, AMH, RIT2, ME1, FOLR1, RPL39L, HPCAL4, NTRK2, GRIA2, APOD, MT1A, AQP1, PTGDS, HOXD3, ATP1A2, MTPP, TSPAN18, SLAMF9, SAPCD2, CCL4L2, CCL3L3, CP, TOX3, ZSCAN1, OPCML, RGS6, STMN4, SEMA3A, LYZ, APOL4, EEF1A2, CHRDL1, MDFI, LPL, KCNIP1, ADM, PPP2R2C, CYP1B1, PCDH7, SEMA3E, FCGR2B, FERMT1, MMP7, TMEM229A, BAIAP3, CCL2, ANGPTL1, LIF, CALY, ADAMDEC1, COL6A3, KCNK2, GPR27, GPC2, DCN, DSG2, HOXA5, CACNG8, ANGPTL4, GABRA3, GBP3, PCSK2, SFRP2, SCN4B, PCDHGC4, NOS2, NPIP15, GABRQ, MYH14, HOXA2, RADX, HS3ST2, THBS1, TMEM130, OTP, DES, NELL2, C1QL4, MMP17, PCP4L1, MEX3A, LMO3, RET, SHISA7, FABP7, GAS1, CRYM, GJB1, GAL, ATP1A3, TAGLN3, SLIT1, JPH4, CPNE6, SOX4, FLNC, NCAN, ACY3, NPPA, HBA2, DISP3, MET, PSD2, TWIST1, SYN1, ELFN2, SERPINE1, EFEMP1, ECRG4, UNC13A, CXCL3, LRRC2, CA12, UNC5A, SYT6, SLCO1C1, SLC10A4, KCNJ16, CAMKV, KIF1A, HS3ST3B1, TMSB4Y, ARC, LILRA4, NDRG1, CHGB, ST8SIA5, RPS4Y1, EPHB1, GBX2, </p> |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p>ISG15, FABP5, PCDHB16, ITGA3, PCDHB8, FMOD, TMEM151A, CHST9, FGFR3, SYN2, ENHO, EMX2, SPRY4, PROM1, LRRTM4, PCDHGA9, COL23A1, DDX3Y, ENPP4, PKIB, CNGA3, SEZ6L2, OTOS, CDK5R2, SOX8, TFPI2, SCGB1D2, PCOLCE2, STEAP1, SLC7A3, CRABP1, HLA-DQB1, CSPG5, BST2, TIMP4, EDIL3, INSM1, HAMP, DBX2, HOXA10, CD2, RGS4, PTX3, EPHA3, COL22A1, FAM180A, ARX, SLC32A1, PHF24, DSCAM, FST, BEX5, CALB2, RGS11, EYA1, FUT9, NKAIN3, HOXC11, HOXD11, SNX10, OAS2, MGP, H2AW, DSCAML1, GABRA5, LOX, SCNN1B, CDK5R1, OXTR, GZMA, GPR153, P2RY12, BRSK2, SNORC, HBA1, MRC1, EN1, CACNG3, ACSBG1, RCAN2, CXCL2, BCAN, PRSS35, VCAM1, C7orf57, IP6K3, RALYL, SYT7, CNTN1, IGFBP3, TMIE, GDAP1L1, VSTM2L, RORB, SCRT1</p> |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|



|        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| lncRNA | H19, LINC02283, AC006157.1, LINC01116, LINC02587, HOXC-AS1, AL031430.1, MEG3, AL035446.1, FLJ16779, AC004540.2, LINC00404, AL355922.1, AC107398.3, AL049637.1, AL359091.1, ALDH1L1-AS2, AD000090.1, AUXG01000058.1, AC119673.2, HAGLR, AC006058.1, LINC01831, ELFN2, AL031056.1, CEROX1, CDK6-AS1, HOTAIR, HOXB-AS1, AC110285.2, AC092131.1, AC092958.1, AF131216.3, AL354919.2, SOCS2-AS1, PAX8-AS1, LINC00237, AL512329.2, FOXD3-AS1, KANSL1-AS1, AC025171.5, AL603839.3, AL132711.1, LINC01291, LINC01088, MIR9-3HG, AL928654.2, AC019257.1, NEAT1, AL591686.1, AL162511.1, AC017048.3, MROCK1, MIR124-1HG, AL512785.1, AC091932.1, AC021242.3, AL591686.2, B3GALT5-AS1, LINC00844, AC117464.1, AL450345.2, NFIA-AS2, AL596087.2, AL390755.1, AC090692.1, CPNE8-AS1, AC004233.2, AL355974.1, AC233266.2, AC005696.4, AC090204.1, AL731566.3, AC010378.1, HCP5, IGFBP7-AS1, AC131097.2, AC005197.1, AC100793.4, HOTAIRM1, AC104137.1, LINC02308, AL049871.1, AP001029.1, AC244213.1, AC064875.1, PDE6B-AS1, AL606970.4, LINC02513, AL355916.2, AC008060.4, RAMP2-AS1, AL359881.1, AL645924.1, AL450270.1, AL133325.3, UST-AS2, AL355974.2, AL133465.1, AC013553.3, AL136366.1, ASIC4-AS1, ELOVL2-AS1, MALAT1, LINC02712, AC244517.7, BISPR, AC097713.1, AL049777.1, AL645608.1, LINC00609, HOXA-AS2, AC104407.1, AL121894.2, LINC02732, AL162424.1, LINC01338, AC061992.2, SNHG25, LINC02821, ARMC2-AS1, ASMTL-AS1, ASMTL-AS1.1, AL118505.1, AC083855.2, EMX2OS, C2CD4D-AS1, AC104088.3, LINC01597, CAHM, AC093849.2, AC006435.2, AP006621.2, LINC00685, LINC00685.1, AP001458.1, AC027020.2, MIAT, BNC2-AS1, IDH2-DT, FOXD2-AS1, LINC01315, LINC01351, MIR210HG, AC068535.1, AL356489.2, SOX1-OT, AC026316.5, AL358216.1, AC006058.3, LINC01354, AC009102.2, AC011462.4, AC093895.1, AC018529.1, AC104964.3, SEC61G-DT, LINC01956, AC002074.1, TFAP2A-AS1, AC013391.3, AP001021.1, LINC02743, AC015813.1, LINC00839, AC008735.2, AC016716.1, LINC02604, FAM218A, MACORIS, CCR5AS, AC021739.2, AL512791.1, BAALC-AS2, AC188617.1, LINC01786, AP005432.2, LINC01896, AC123912.4, AC017100.1, Z97200.1, LINC01224, AC018755.4, AC006026.3, AC022034.1, AC132938.1, FP671120.7, LINC00106, LINC00106.1, AL513524.1, LINC01561, LINC01842, AC244517.1, AC092535.5, AC008760.2, LINC01736, AC012467.1, AC093673.2, AC002401.4, MIR325HG, AC103591.3, CTB-178M22.2, AC007344.1, AL645608.7, AL590062.1, AC004540.1, SERTAD4-AS1, LINC01614, AP000757.1, AP001269.2, AC008737.1, AC009549.1, AL031600.1, DDX59-AS1, SERPINB9P1, AC110285.1, AC012405.1, MUC12-AS1, AL365434.2, Z97989.1, AL603839.2, NCAM1-AS1, LINC01503, AC092954.1, AC092924.2, AL365361.1, CHL1-AS2, PTPRD-AS1, AC009902.2, AL008729.2, SLCO4A1-AS1, PPP1R9A-AS1, LINC01546, AC007920.2, PANTR1, AC000067.1, AL354836.1, MIR34AHG, IQCA1-AS1, LINC01894, AL096854.1, AC105052.5, AL512625.2, MIR3659HG, AL353660.1, AC015922.2, AC098479.1, AP001107.5, AC003035.2, AL031709.1, AC015849.3, AC107294.3, AC092111.1, AL031673.1, MAP3K4-AS1, AL355075.2, AL157700.1, AL353150.1, KLHL7-DT, AC007098.1, SILC1, AC133550.3, AC027796.4, AC113189.3, LINC02478, AC092675.1, |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p> MIR9-1HG, ACSL6-AS1, AC034243.1, AC004816.1, AL161729.4, AP000696.1, NBAT1, AP000229.1, AL590666.2, AP002387.1, AL391807.1, LINC01549, AC010889.1, AC004148.1, PTOV1-AS2, AC020907.4, AC083864.5, LINC01871, LINC01480, AC019069.1, RGS5, HLA-DQB1-AS1, AP003032.1, AL022322.1, LINC01235, AC080038.1, AP002840.2, AC009407.1, ANKRD10-IT1, AC093734.1, AL450311.1, MIR600HG, SNHG18, EPHA5-AS1, AC090559.1, AC121493.1, AL645608.6, LINC02762, AC015922.3, AC010531.6, VPS9D1-AS1, AC091435.2, LINC01094, AC009902.3, AC245884.9, AL023803.1, LINC01778, MIR181A2HG, AC093620.1, AC008525.1, AC046168.2, AC126407.1, AC116565.1, AC007130.1, Z70719.1, DBH-AS1, AL513550.1, AC012462.3, AC132872.3, AC008060.1, AC006001.2, NRXN2-AS1, AC009065.2, AC115522.1, AC061708.1, LINC00092, AC073050.1, MCF2L-AS1, AC020558.2, AC025165.4, AC124016.2, AC097462.3, SLC9A3-AS1, LINC01711, AC104695.2, AC024560.3, AC009065.5, AC015813.4, AC011005.4, AC092112.1, AL157871.2, LINC01982, AC084018.1, AC078909.2, AL590235.2, AC122129.1, WNT5A-AS1, AP003392.1, AC108449.2, Z97653.1, AC079610.2, AC147067.2, LLPH-DT, AC006960.3, AC093627.7, LINC01816, AP000439.2, CASC15, AL031651.2, AL139424.2, AL133371.2, AC009120.2, AL590666.1, AC068057.1, AL139287.1, AL121933.2, AC099792.1, HOXC-AS2, AL359715.2, AL021707.6, AC073140.1, AC096645.1, AC110995.1, FEZF1-AS1, AC131097.3, AL162411.1, AL122125.1, LINC02100, AP003170.3, AC133919.2, AC005014.3, AC015726.2, LINC00342, AC011445.1, LINC01141, AC018638.6, AC093673.1, AC074135.1, AC011625.1, PRKAR1B-AS1, LINC01411, AL139349.1, LINC01579, AL157904.1, AC011444.3, LINC00641, AC105206.2, LINC02716, OSTM1-AS1, THBS3-AS1, PHACTR3-AS1, AC084117.1, FAM181A-AS1, LINC02883, AC104532.2, AC107464.2, AC008759.3, AC004863.1, RASSF10-DT, AC009120.3, AC015909.3, AC004253.1, AL021707.8, AC110620.2, MIF-AS1, AC020915.1, AL512625.3, AC245407.2, AL731567.1, AC124045.1, AL031668.2, AC245060.2, NALT1, AC125807.2, AC005519.1, LINC02134, INSYN1-AS1, AC007877.1, RNF207-AS1, SNHG15, AC010729.1, AL133330.1, AC010624.2, USP12-AS2, BX293535.1, AL136295.6, AL049838.1, AC002091.2, AP003065.1, AF131215.5, AC026471.4, MIR222HG, AC004816.2, AC016590.3, PEG13, AL022162.1, TGFB2-AS1, AC104521.1, AL031186.1, AC021087.1, AC097658.2, AC141002.1, AQP4-AS1, CRIM1-DT, AC005330.1, AP001107.4, AC107375.1, AC020928.1, AC242842.1, AC092117.1, AC016168.2, AC015540.1, AC125613.1, AC026316.2, MAFA-AS1, AL356740.1, MIR4453HG, AC048341.2, FP671120.6, AL355338.1, BAALC-AS1, AC106820.3, AC027348.1, SNTG2-AS1, AL049840.2, AL118516.1, AC245884.8, AL353803.5, AL365205.4, AC010719.1, AL365181.4, AP001505.1, FAM66A, AL157823.2, LOXL1-AS1, AC020905.1, AC127502.2, AC060766.4, AC144652.1, MIR22HG, PCAT6, LINC02526, AC233728.1, SLC25A25-AS1 </p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Methylation | cg00015770, cg00027083, cg00030047, cg00041575, cg00066816, cg00117172, cg00128877, cg00221494, cg00233307, cg00242839, cg00263760, cg00278366, cg00318573, cg00334507, cg00396894, cg00466492, cg00489401, cg00503840, cg00563926, cg00565688, cg00616129, cg00626466, cg00643392, cg00698688, cg00708598, cg00739120, cg00755043, cg00767581, cg00772000, cg00795812, cg00891541, cg00910067, cg00962459, cg00969446, cg01009664, cg01081263, cg01155039, cg01172899, cg01186777, cg01193293, cg01200177, cg01242619, cg01255591, cg01280080, cg01294702, cg01295203, cg01305625, cg01335367, cg01353448, cg01366419, cg01369413, cg01377911, cg01414934, cg01418124, cg01420388, cg01441777, cg01454134, cg01543654, cg01556075, cg01561916, cg01565918, cg01566404, cg01580681, cg01610488, cg01669948, cg01683883, cg01757745, cg01765461, cg01770232, cg01777397, cg01817029, cg01830294, cg01851399, cg01883966, cg01966465, cg01985396, cg01989224, cg02008154, cg02016419, cg02019333, cg02029926, cg02055963, cg02064402, cg02082342, cg02128567, cg02161900, cg02164046, cg02248486, cg02250594, cg02255004, cg02260587, cg02276665, cg02282237, cg02293044, cg02332525, cg02347487, cg02409351, cg02488871, cg02489552, cg02497758, cg02545192, cg02554564, cg02608019, cg02611419, cg02630888, cg02672220, cg02675652, cg02732509, cg02748539, cg02786019, cg02787991, cg02844051, cg02844545, cg02849695, cg02859934, cg02882813, cg02919422, cg02983451, cg02989940, cg02994956, cg03014957, cg03085312, cg03098721, cg03214697, cg03224418, cg03240509, cg03317245, cg03330058, cg03365437, cg03440267, cg03464655, cg03469054, cg03506489, cg03532879, cg03565081, cg03589001, cg03611555, cg03630088, cg03642518, cg03662459, cg03664992, cg03699904, cg03702236, cg03742272, cg03770548, cg03819692, cg03826976, cg03846767, cg03852144, cg03870862, cg03874199, cg03876618, cg03883348, cg03929796, cg03941108, cg03945800, cg03962522, cg03963198, cg03964111, cg03975694, cg03980304, cg04001333, cg04037228, cg04086012, cg04265576, cg04317399, cg04329454, cg04376312, cg04377282, cg04418492, cg04425624, cg04431133, cg04438497, cg04456238, cg04461802, cg04481779, cg04513422, cg04528819, cg04587829, cg04587910, cg04595372, cg04598121, cg04600618, cg04618528, cg04619381, cg04623955, cg04632671, cg04637372, cg04720330, cg04721098, cg04726200, cg04765277, cg04784672, cg04797323, cg04809136, cg04897683, cg04898797, cg04994456, cg05026186, cg05078019, cg05158615, cg05184394, cg05215575, cg05221167, cg05241571, cg05248781, cg05250458, cg05294455, cg05330360, cg05335315, cg05358404, cg05373457, cg05386606, cg05403071, cg05436658, cg05477024, cg05501357, cg05546044, cg05589246, cg05600717, cg05671350, cg05714219, cg05774801, cg05839235, cg05958582, cg05982504, cg06038133, cg06092815, cg06220235, cg06263495, cg06268694, cg06274159, cg06277657, cg06291867, cg06327515, cg06346081, cg06352750, cg06392241, cg06424894, cg06462291, cg06491116, cg06529761, cg06630567, cg06637774, cg06638433, cg06736444, cg06738602, cg06759890, cg06793062, cg06821120, cg06822995, cg06825166, cg06834875, cg06836736, cg06873352, cg06896207, cg06911113, cg06918467, cg06974755, cg06980380, cg07007400, cg07017374, cg07027513, cg07038400, cg07039113, cg07059052, cg07059360, cg07078114, cg07103493, cg07123069, cg07141002, cg07158339, cg07168556, cg07177852, cg07237830, cg07294541, cg07307078, cg07354440, cg07376232, cg07376535, cg07404485, cg07405796, cg07455279, cg07484450, cg07597976, cg07621046, cg07654934, cg07664183, |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| cg07671603, cg07685786, cg07693270, cg07703401, cg07715201, cg07749074, cg07753583, cg07758904, cg07778029, cg07823492, cg07846167, cg07879977, cg07914866, cg07922606, cg07935568, cg07950803, cg07952391, cg07965823, cg07979357, cg07986525, cg08089301, cg08097882, cg08101264, cg08109815, cg08137716, cg08191854, cg08216808, cg08254089, cg08303146, cg08307816, cg08315770, cg08317263, cg08397758, cg08411049, cg08460435, cg08527127, cg08548888, cg08558340, cg08573687, cg08575537, cg08594695, cg08611714, cg08615333, cg08649013, cg08668790, cg08768421, cg08806153, cg08832227, cg08924430, cg09022993, cg09059945, cg09068492, cg09095562, cg09191232, cg09191626, cg09201719, cg09242541, cg09243021, cg09254939, cg09259772, cg09299388, cg09381003, cg09419670, cg09434995, cg09448880, cg09486093, cg09492887, cg09527362, cg09541248, cg09577651, cg09587549, cg09588653, cg09595479, cg09666654, cg09697795, cg09721427, cg09737668, cg09744051, cg09816471, cg09871043, cg09871315, cg09893305, cg09937039, cg09949775, cg09952204, cg09954385, cg09971646, cg10003443, cg10037005, cg10046620, cg10071275, cg10078829, cg10106388, cg10125195, cg10135717, cg10173075, cg10175795, cg10189695, cg10235817, cg10245048, cg10303487, cg10319505, cg10453365, cg10515956, cg10521852, cg10523019, cg10525488, cg10528989, cg10549973, cg10644361, cg10694152, cg10784341, cg10798171, cg10883303, cg10922280, cg10964421, cg10970409, cg10994126, cg11004890, cg11091262, cg11120551, cg11164400, cg11191210, cg11195002, cg11204562, cg11260422, cg11260848, cg11263296, cg11265941, cg11271605, cg11277126, cg11285003, cg11319389, cg11377136, cg11438428, cg11525285, cg11530960, cg11536940, cg11540692, cg11608424, cg11611600, cg11657808, cg11667754, cg11706111, cg11706163, cg11719157, cg11719784, cg11757293, cg11897314, cg11990309, cg12045002, cg12061127, cg12061236, cg12071073, cg12078929, cg12119029, cg12144803, cg12163132, cg12188860, cg12266049, cg12277666, cg12293634, cg12324629, cg12374431, cg12387247, cg12437481, cg12456510, cg12513481, cg12571423, cg12671744, cg12758687, cg12768605, cg12770741, cg12792367, cg12819826, cg12902039, cg12936220, cg12946225, cg12955127, cg12958778, cg12966875, cg13021384, cg13042487, cg13055001, cg13062935, cg13081704, cg13099330, cg13119609, cg13125510, cg13191049, cg13266292, cg13302823, cg13315147, cg13337662, cg13384396, cg13397379, cg13415207, cg13446622, cg13447818, cg13547644, cg13562542, cg13656062, cg13682722, cg13739417, cg13744194, cg13878456, cg13897627, cg13921352, cg13929328, cg13968061, cg13969584, cg13993218, cg14088161, cg14141399, cg14153740, cg14173523, cg14184817, cg14204735, cg14221171, cg14262937, cg14269477, cg14417329, cg14498423, cg14561282, cg14563260, cg14566624, cg14592065, cg14696396, cg14759043, cg14802951, cg14851685, cg14859460, cg14886269, cg14893161, cg14911395, cg14959707, cg14991487, cg15001381, cg15014458, cg15032239, cg15043057, cg15107670, cg15146752, cg15149645, cg15164103, cg15239123, cg15250797, cg15309006, cg15309264, cg15375239, cg15387123, cg15425811, cg15452573, cg15517609, cg15518950, cg15536490, cg15537850, cg15538820, cg15539420, cg15540820, cg15552238, cg15576195, cg15660498, cg15686608, cg15711744, cg15748507, cg15760840, cg15835825, cg15853125, cg15873301, cg15887846, cg15916061, cg15928398, cg15949277, cg15984661, cg16005224, cg16072462, cg16080552, cg16092786, cg16154416, cg16158874, cg16175263, cg16176600, cg16179125, cg16191009, cg16232126, cg16254309, cg16270890, |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| cg16272420, cg16293105, cg16306115, cg16313343, cg16319578, cg16393207, cg16408565, cg16418329, cg16426459, cg16427670, cg16428251, cg16431978, cg16474696, cg16483916, cg16488098, cg16516400, cg16536450, cg16592658, cg16604516, cg16639185, cg16652063, cg16708981, cg16722536, cg16743289, cg16799087, cg16812893, cg16853982, cg16864658, cg16884042, cg16893574, cg16907566, cg16944093, cg17001430, cg17162024, cg17173423, cg17181653, cg17183546, cg17188046, cg17192247, cg17207590, cg17241310, cg17264618, cg17272843, cg17274064, cg17301902, cg17357062, cg17387870, cg17465304, cg17484237, cg17496921, cg17534899, cg17558126, cg17560332, cg17586860, cg17612991, cg17641252, cg17651821, cg17685628, cg17699374, cg17711870, cg17727529, cg17729667, cg17741572, cg17783509, cg17838516, cg17906786, cg17923358, cg17940013, cg17959722, cg17992056, cg18138484, cg18151487, cg18180155, cg18204685, cg18219418, cg18236079, cg18236477, cg18239253, cg18239753, cg18265887, cg18267374, cg18345369, cg18392783, cg18462653, cg18463686, cg18530748, cg18552413, cg18555440, cg18555555, cg18565355, cg18618334, cg18636641, cg18665384, cg18691434, cg18702197, cg18705776, cg18750960, cg18793806, cg18809289, cg18815943, cg18838701, cg18839416, cg18847227, cg18854666, cg18881723, cg18901980, cg18956481, cg18967533, cg18972811, cg19000186, cg19006008, cg19017177, cg19042947, cg19096540, cg19168338, cg19172575, cg19220586, cg19241311, cg19264571, cg19325985, cg19370451, cg19372178, cg19404582, cg19438514, cg19456540, cg19502744, cg19535609, cg19576304, cg19594666, cg19596755, cg19636861, cg19668234, cg19718882, cg19740375, cg19756068, cg19766460, cg19795898, cg19797376, cg19798224, cg19815720, cg19841506, cg19868691, cg19896198, cg19904463, cg19917856, cg19948393, cg19988449, cg20006825, cg20011352, cg20028470, cg20050113, cg20050826, cg20052718, cg20059151, cg20070090, cg20125091, cg20134215, cg20158248, cg20189782, cg20191453, cg20253551, cg20264732, cg20322977, cg20339230, cg20339650, cg20356482, cg20366906, cg20377955, cg20406482, cg20420433, cg20451680, cg20507276, cg20537992, cg20582779, cg20587968, cg20615832, cg20616414, cg20622019, cg20644981, cg20645058, cg20673481, cg20789620, cg20801056, cg20807545, cg20822579, cg20839149, cg20870559, cg20959866, cg20973210, cg20977864, cg20979799, cg20981615, cg20994801, cg21044104, cg21053323, cg21057046, cg21091679, cg21115977, cg21116314, cg21120063, cg21129531, cg21163415, cg21184174, cg21252483, cg21274570, cg21296230, cg21303011, cg21351102, cg21410991, cg21428681, cg21478490, cg21513385, cg21542793, cg21554552, cg21577049, cg21589115, cg21591742, cg21602520, cg21624282, cg21663431, cg21665000, cg21685266, cg21696393, cg21771250, cg21815667, cg21835643, cg21852292, cg21870884, cg21975377, cg22061523, cg22130834, cg22134325, cg22189019, cg22215728, cg22244418, cg22268231, cg22289360, cg22289837, cg22294908, cg22325572, cg22334665, cg22338307, cg22341310, cg22402007, cg22418909, cg22461018, cg22461835, cg22534509, cg22546318, cg22584138, cg22598563, cg22740835, cg22752533, cg22777952, cg22797169, cg22815110, cg22836229, cg22892110, cg22892904, cg22919728, cg22956483, cg22995106, cg23003832, cg23090046, cg23131007, cg23189044, cg23228178, cg23239444, cg23240895, cg23244913, cg23273897, cg23278885, cg23282559, cg23283495, cg23303408, cg23323671, cg23326689, cg23413307, cg23418591, cg23432345, cg23495733, cg23539753, cg23566503, cg23571857, cg23580945, cg23606079, cg23642747, cg23651728, |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|          | cg23694248, cg23698287, cg23710218, cg23713520, cg23771603, cg23777956, cg23801057, cg23815000, cg23843505, cg23855989, cg23881725, cg23941599, cg23950724, cg23960723, cg23995753, cg24056567, cg24101578, cg24132694, cg24167037, cg24235037, cg24251035, cg24292612, cg24309555, cg24335149, cg24378421, cg24432073, cg24443250, cg24457403, cg24516901, cg24523000, cg24562819, cg24621042, cg24621354, cg24642468, cg24691255, cg24693053, cg24715735, cg24719601, cg24768561, cg24792360, cg24826867, cg24861272, cg24891133, cg24898753, cg24899750, cg24920358, cg24921089, cg24924779, cg24975222, cg24989962, cg25028542, cg25033144, cg25057743, cg25084878, cg25107791, cg25141490, cg25182621, cg25192419, cg25193494, cg25239996, cg25358289, cg25384157, cg25426302, cg25431974, cg25432696, cg25437385, cg25483003, cg25484904, cg25509184, cg25553916, cg25558099, cg25589890, cg25671438, cg25711779, cg25720804, cg25725843, cg25737664, cg25763788, cg25766046, cg25802093, cg25839227, cg25882366, cg25957124, cg25990647, cg26018901, cg26128092, cg26135325, cg26163537, cg26186727, cg26200585, cg26219051, cg26220985, cg26221631, cg26252167, cg26365553, cg26415633, cg26416466, cg26504021, cg26523005, cg26531804, cg26537639, cg26556719, cg26607785, cg26608332, cg26614073, cg26620157, cg26632776, cg26656113, cg26711820, cg26746469, cg26774312, cg26776069, cg26833602, cg26845300, cg26952662, cg26980692, cg27056119, cg27096144, cg27114120, cg27202708, cg27210390, cg27320127, cg27321931, cg27412902, cg27416067, cg27418851, cg27443224, cg27449489, cg27465566, cg27532722, cg27590397, cg27619475, cg27622610, cg27634151, cg27637521, cg27652350 |
| Mutation | OBSCN, TTN, MUC16, FLG, LRP2, PTEN, TP53, RYR2, EGFR, NF1, SPTA1, PKHD1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

**Supplementary Table S2.** Differential methylated probes (DMPs) indicating logFC in different arrays

|            | CpG Probe  | LogFC_27k array | LogFC_450k array | Direction       |
|------------|------------|-----------------|------------------|-----------------|
| CS2 vs CS1 | cg03852144 | -2.328          | -3.060           | hypomethylated  |
|            | cg23090046 | -1.497          | -1.434           | hypomethylated  |
|            | cg24861272 | -1.159          | -1.543           | hypomethylated  |
|            | cg16799087 | 1.032           | 1.632            | hypermethylated |
|            | cg19815720 | 1.050           | 1.594            | hypermethylated |
|            | cg09434995 | 1.080           | 1.939            | hypermethylated |
|            | cg18239253 | 1.145           | 1.804            | hypermethylated |
|            | cg00698688 | 1.166           | 2.099            | hypermethylated |
|            | cg25182621 | 1.193           | 1.150            | hypermethylated |
|            | cg21663431 | 1.217           | 2.466            | hypermethylated |
|            | cg03098721 | 1.249           | 1.647            | hypermethylated |
|            | cg13446622 | 1.253           | 1.482            | hypermethylated |
|            | cg23228178 | 1.269           | 2.124            | hypermethylated |

|            |            |        |        |                 |
|------------|------------|--------|--------|-----------------|
|            | cg06825166 | 1.332  | 1.443  | hypermethylated |
|            | cg26415633 | 1.346  | 1.859  | hypermethylated |
|            | cg13993218 | 1.434  | 1.630  | hypermethylated |
|            | cg27412902 | 1.436  | 2.282  | hypermethylated |
|            | cg11608424 | 1.438  | 1.696  | hypermethylated |
|            | cg24457403 | 1.480  | 1.576  | hypermethylated |
|            | cg21624282 | 1.488  | 2.072  | hypermethylated |
|            | cg14173523 | 1.495  | 1.914  | hypermethylated |
|            | cg23240895 | 1.509  | 1.808  | hypermethylated |
|            | cg12061127 | 1.581  | 2.040  | hypermethylated |
|            | cg25737664 | 1.626  | 2.002  | hypermethylated |
|            | cg12266049 | 1.638  | 1.951  | hypermethylated |
|            | cg02488871 | 1.662  | 1.685  | hypermethylated |
|            | cg01369413 | 1.679  | 1.951  | hypermethylated |
|            | cg07597976 | 1.691  | 2.183  | hypermethylated |
|            | cg03846767 | 1.730  | 1.736  | hypermethylated |
|            | cg05294455 | 1.735  | 1.693  | hypermethylated |
|            | cg20981615 | 1.797  | 1.795  | hypermethylated |
|            | cg16408565 | 1.817  | 2.048  | hypermethylated |
|            | cg25141490 | 1.818  | 1.999  | hypermethylated |
|            | cg23580945 | 1.982  | 1.963  | hypermethylated |
|            | cg25839227 | 2.186  | 1.835  | hypermethylated |
|            | cg18204685 | 2.262  | 1.671  | hypermethylated |
|            | cg02844051 | 2.438  | 2.006  | hypermethylated |
|            | cg10071275 | 2.456  | 2.430  | hypermethylated |
|            | cg15538820 | 3.199  | 1.924  | hypermethylated |
| CS3 vs CS1 | cg18555555 | -2.918 | -2.508 | hypomethylated  |
|            | cg08615333 | -2.779 | -2.317 | hypomethylated  |
|            | cg06352750 | -2.014 | -1.709 | hypomethylated  |
|            | cg12071073 | -1.531 | -1.888 | hypomethylated  |
|            | cg18552413 | -1.464 | -1.797 | hypomethylated  |
|            | cg07078114 | 1.405  | 2.152  | hypermethylated |
|            | cg26186727 | 2.918  | 2.432  | hypermethylated |
|            | cg15538820 | -3.339 | -2.446 | hypomethylated  |
| CS3 vs CS2 | cg18555555 | -2.838 | -2.716 | hypomethylated  |
|            | cg21044104 | -2.721 | -1.891 | hypomethylated  |
|            | cg02844051 | -2.175 | -2.400 | hypomethylated  |
|            | cg10173075 | -2.042 | -2.696 | hypomethylated  |
|            | cg21663431 | -1.813 | -3.108 | hypomethylated  |
|            | cg24691255 | -1.762 | -2.054 | hypomethylated  |
|            | cg21624282 | -1.658 | -1.981 | hypomethylated  |
|            | cg06630567 | -1.657 | -1.796 | hypomethylated  |
|            | cg13993218 | -1.627 | -3.082 | hypomethylated  |

|            |        |        |                 |
|------------|--------|--------|-----------------|
| cg16408565 | -1.544 | -2.022 | hypomethylated  |
| cg07597976 | -1.542 | -2.013 | hypomethylated  |
| cg23228178 | -1.535 | -2.268 | hypomethylated  |
| cg20981615 | -1.533 | -2.428 | hypomethylated  |
| cg22461018 | -1.480 | -1.521 | hypomethylated  |
| cg05078019 | -1.460 | -1.635 | hypomethylated  |
| cg04994456 | -1.403 | -1.760 | hypomethylated  |
| cg10994126 | -1.361 | -1.754 | hypomethylated  |
| cg23815000 | -1.352 | -1.422 | hypomethylated  |
| cg23278885 | -1.334 | -2.048 | hypomethylated  |
| cg00698688 | -1.331 | -1.857 | hypomethylated  |
| cg06825166 | -1.315 | -1.584 | hypomethylated  |
| cg18462653 | -1.294 | -1.750 | hypomethylated  |
| cg14173523 | -1.266 | -1.685 | hypomethylated  |
| cg10125195 | -1.173 | -1.785 | hypomethylated  |
| cg07484450 | -1.136 | -1.793 | hypomethylated  |
| cg05221167 | 1.899  | 1.686  | hypermethylated |

**Supplementary Table S3.** CS1- and CS2-specific enriched pathway results.

| CS1_specific Pathways                                    | setSize | enrichmentScore | NES    | pvalue | p.adjust |
|----------------------------------------------------------|---------|-----------------|--------|--------|----------|
| REACTOME_INNATE_IMMUNE_SYSTEM                            | 20      | -0.699          | -2.011 | 0.001  | 0.034    |
| WP_NETWORK_MAP_OF_SARS_COV_2_SIGNALING_PATHWAY           | 19      | -0.693          | -1.988 | 0.001  | 0.034    |
| REACTOME_NEUTROPHIL_DEGRANULATION                        | 14      | -0.695          | -1.839 | 0.003  | 0.034    |
| REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM             | 28      | -0.587          | -1.808 | 0.002  | 0.034    |
| REACTOME_SIGNALING_BY_INTERLEUKINS                       | 19      | -0.627          | -1.798 | 0.005  | 0.037    |
| WP_OVERVIEW_OF_PROINFLAMMATORY_AND_PROFIBROTIC_MEDIATORS | 12      | -0.702          | -1.787 | 0.007  | 0.037    |
| REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION               | 30      | -0.571          | -1.779 | 0.002  | 0.034    |
| WP_BURN_WOUND_HEALING                                    | 11      | -0.711          | -1.775 | 0.007  | 0.037    |
| KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION              | 16      | -0.645          | -1.762 | 0.006  | 0.037    |
| REACTOME_GLYCOSAMINOGLYCAN_METABOLISM                    | 11      | -0.698          | -1.744 | 0.007  | 0.037    |
| REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS              | 11      | -0.695          | -1.735 | 0.007  | 0.037    |
| REACTOME_INTERLEUKIN_4_AND_INTERLEUKIN_13_SIGNALING      | 12      | -0.667          | -1.700 | 0.012  | 0.052    |

|                                                                       |                |                        |            |               |                 |
|-----------------------------------------------------------------------|----------------|------------------------|------------|---------------|-----------------|
| WP_NUCLEAR_RECEPTORS_META_PATHWAY                                     | 12             | -0.655                 | -1.669     | 0.016         | 0.058           |
| REACTOME_METABOLISM_OF_CARBOHYDRATES                                  | 12             | -0.650                 | -1.656     | 0.016         | 0.058           |
| WP_SPINAL_CORD_INJURY                                                 | 16             | -0.603                 | -1.648     | 0.013         | 0.053           |
| REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX                      | 15             | -0.612                 | -1.643     | 0.011         | 0.052           |
| REACTOME_ECM_PROTEOGLYCAN                                             | 12             | -0.627                 | -1.596     | 0.023         | 0.070           |
| PID_INTEGRIN1_PATHWAY                                                 | 11             | -0.639                 | -1.595     | 0.021         | 0.070           |
| REACTOME_COLLAGEN_DEGRADATION                                         | 11             | -0.633                 | -1.580     | 0.024         | 0.070           |
| REACTOME_COLLAGEN_FORMATION                                           | 15             | -0.585                 | -1.571     | 0.024         | 0.070           |
| PID_SYNDECAN_1_PATHWAY                                                | 10             | -0.639                 | -1.559     | 0.028         | 0.074           |
| REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES | 10             | -0.639                 | -1.559     | 0.028         | 0.074           |
| KEGG_FOCAL_ADHESION                                                   | 15             | -0.576                 | -1.547     | 0.024         | 0.070           |
| REACTOME_COLLAGEN_BIOSYNTHESIS_AND_MODIFYING_ENZYMES                  | 12             | -0.598                 | -1.522     | 0.037         | 0.091           |
| WP_FOCAL_ADHESION_PI3K_AKT_MTOR_SIGNALING_PATHWAY                     | 17             | -0.547                 | -1.520     | 0.038         | 0.092           |
| REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS                             | 21             | -0.509                 | -1.482     | 0.041         | 0.092           |
| WP_SYNAPTIC_VESICLE_PATHWAY                                           | 10             | 0.674                  | 2.010      | 0.008         | 0.040           |
| REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES                        | 22             | 0.487                  | 1.867      | 0.006         | 0.037           |
| REACTOME_L1CAM_INTERACTIONS                                           | 11             | 0.540                  | 1.673      | 0.041         | 0.092           |
| REACTOME_NEURONAL_SYSTEM                                              | 29             | 0.388                  | 1.616      | 0.029         | 0.074           |
| <b>CS2_specific Pathways</b>                                          | <b>setSize</b> | <b>enrichmentScore</b> | <b>NES</b> | <b>pvalue</b> | <b>p.adjust</b> |
| REACTOME_NEURONAL_SYSTEM                                              | 29             | -0.491                 | -1.884     | 0.008         | 0.022           |
| WP_SYNAPTIC_VESICLE_PATHWAY                                           | 10             | -0.659                 | -1.825     | 0.008         | 0.022           |
| REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES                        | 22             | -0.510                 | -1.812     | 0.007         | 0.022           |
| REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX                      | 15             | 0.778                  | 2.171      | 0.002         | 0.007           |
| WP_BURN_WOUND_HEALING                                                 | 11             | 0.847                  | 2.167      | 0.002         | 0.007           |
| PID_SYNDECAN_1_PATHWAY                                                | 10             | 0.858                  | 2.124      | 0.002         | 0.007           |
| REACTOME_COLLAGEN_DEGRADATION                                         | 11             | 0.826                  | 2.112      | 0.002         | 0.007           |
| WP_NETWORK_MAP_OF_SARS_COV_2_SIGNALING_PATHWAY                        | 19             | 0.707                  | 2.108      | 0.001         | 0.007           |

|                                                                               |    |       |       |       |       |
|-------------------------------------------------------------------------------|----|-------|-------|-------|-------|
| REACTOME_ASSEMBLY_OF_COLL<br>AGEN_FIBRILS_AND_OTHER_MUL<br>TIMERIC_STRUCTURES | 10 | 0.839 | 2.078 | 0.002 | 0.007 |
| REACTOME_NEUTROPHIL_DEGRA<br>NULATION                                         | 14 | 0.731 | 2.027 | 0.002 | 0.007 |
| PID_INTEGRIN1_PATHWAY                                                         | 11 | 0.792 | 2.025 | 0.002 | 0.007 |
| REACTOME_INNATE_IMMUNE_SY<br>STEM                                             | 20 | 0.675 | 2.022 | 0.003 | 0.011 |
| REACTOME_EXTRACELLULAR_M<br>ATRIX_ORGANIZATION                                | 30 | 0.592 | 1.931 | 0.001 | 0.007 |
| REACTOME_INTEGRIN_CELL_SUR<br>FACE_INTERACTIONS                               | 11 | 0.752 | 1.923 | 0.002 | 0.007 |
| REACTOME_COLLAGEN_FORMATI<br>ON                                               | 15 | 0.665 | 1.858 | 0.002 | 0.007 |
| REACTOME_ECM_PROTEOGLYCA<br>NS                                                | 12 | 0.699 | 1.851 | 0.002 | 0.007 |
| REACTOME_INTERLEUKIN_4_AND<br>_INTERLEUKIN_13_SIGNALING                       | 12 | 0.699 | 1.851 | 0.002 | 0.007 |
| KEGG_FOCAL_ADHESION                                                           | 15 | 0.661 | 1.847 | 0.002 | 0.007 |
| WP_OVERVIEW_OF_PROINFLAMM<br>ATORY_AND_PROFIBROTIC_MEDI<br>ATORS              | 12 | 0.696 | 1.844 | 0.002 | 0.007 |
| REACTOME_ADAPTIVE_IMMUNE_<br>SYSTEM                                           | 12 | 0.686 | 1.816 | 0.002 | 0.007 |
| REACTOME_COLLAGEN_CHAIN_T<br>RIMERIZATION                                     | 11 | 0.697 | 1.784 | 0.005 | 0.017 |
| REACTOME_COLLAGEN_BIOSYNT<br>HESIS_AND_MODIFYING_ENZYME<br>S                  | 12 | 0.669 | 1.772 | 0.003 | 0.012 |
| REACTOME_SIGNALING_BY_INTE<br>RLEUKINS                                        | 19 | 0.589 | 1.755 | 0.007 | 0.022 |
| WP_PLEURAL_MESOTHELIOMA                                                       | 14 | 0.630 | 1.746 | 0.008 | 0.022 |
| WP_FOCAL_ADHESION_PI3K_AKT<br>_MTOR_SIGNALING_PATHWAY                         | 17 | 0.599 | 1.736 | 0.009 | 0.022 |
| REACTOME_CYTOKINE_SIGNALIN<br>G_IN_IMMUNE_SYSTEM                              | 28 | 0.534 | 1.714 | 0.008 | 0.022 |
| WP_PI3K_AKT_SIGNALING_PATH<br>WAY                                             | 16 | 0.591 | 1.685 | 0.012 | 0.029 |
| KEGG_CYTOKINE_CYTOKINE_RE<br>CEPTOR_INTERACTION                               | 16 | 0.586 | 1.672 | 0.015 | 0.036 |
| WP_FOCAL_ADHESION                                                             | 12 | 0.587 | 1.555 | 0.043 | 0.097 |
| REACTOME_SIGNALING_BY_RECE<br>PTOR_TYROSINE_KINASES                           | 24 | 0.479 | 1.484 | 0.041 | 0.094 |

**Supplementary Table S4.** CS1- and CS2-specific enriched GO processes results.

| CS1_specific GO Terms | setSize | enrichment<br>Score | NES | pvalue | p.adjust |
|-----------------------|---------|---------------------|-----|--------|----------|
|-----------------------|---------|---------------------|-----|--------|----------|

|                                                         |    |        |        |       |       |
|---------------------------------------------------------|----|--------|--------|-------|-------|
| GO_RESPONSE_TO_BIOTIC_STIMULUS                          | 50 | -0.670 | -2.274 | 0.001 | 0.017 |
| GO_RESPONSE_TO_BACTERIUM                                | 35 | -0.690 | -2.219 | 0.001 | 0.017 |
| GO_DEFENSE_RESPONSE_TO_OTHER_ORGANISM                   | 33 | -0.677 | -2.160 | 0.001 | 0.017 |
| GO_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN             | 21 | -0.746 | -2.158 | 0.001 | 0.017 |
| GO GRANULOCYTE MIGRATION                                | 12 | -0.831 | -2.113 | 0.001 | 0.017 |
| GO_DEFENSE_RESPONSE                                     | 56 | -0.611 | -2.101 | 0.001 | 0.017 |
| GO_HUMORAL_IMMUNE_RESPONSE                              | 16 | -0.766 | -2.092 | 0.001 | 0.017 |
| GO_INFLAMMATORY_RESPONSE                                | 35 | -0.650 | -2.088 | 0.001 | 0.017 |
| GO_LEUKOCYTE_CHEMOTAXIS                                 | 16 | -0.758 | -2.071 | 0.001 | 0.017 |
| GO_ANTIMICROBIAL_HUMORAL_RESPONSE                       | 13 | -0.795 | -2.067 | 0.001 | 0.017 |
| GO_INNATE_IMMUNE_RESPONSE                               | 25 | -0.687 | -2.051 | 0.001 | 0.017 |
| GO_CELLULAR_RESPONSE_TO_BIOTIC_STIMULUS                 | 13 | -0.666 | -2.024 | 0.001 | 0.017 |
| GO_CELL_CHEMOTAXIS                                      | 21 | -0.774 | -2.013 | 0.001 | 0.017 |
| GO_NEUTROPHIL_MIGRATION                                 | 10 | -0.695 | -2.012 | 0.001 | 0.017 |
| GO_RESPONSE_TO_LIPID                                    | 48 | -0.827 | -2.010 | 0.001 | 0.017 |
| GO_RESPONSE_TO_CYTOKINE                                 | 49 | -0.591 | -2.001 | 0.001 | 0.017 |
| GO_LEUKOCYTE_MIGRATION                                  | 27 | -0.587 | -1.986 | 0.001 | 0.017 |
| GO_POSITIVE_REGULATION_OF_MAPK_CASCADE                  | 22 | -0.653 | -1.984 | 0.001 | 0.017 |
| GO_MYELOID_LEUKOCYTE_MIGRATION                          | 16 | -0.676 | -1.983 | 0.001 | 0.017 |
| GO_POSITIVE_REGULATION_OF_RESPONSE_TO_EXTERNAL_STIMULUS | 23 | -0.720 | -1.968 | 0.001 | 0.017 |
| GO_POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS         | 30 | -0.664 | -1.955 | 0.001 | 0.017 |
| GO_REGULATION_OF_VASCULATURE_DEVELOPMENT                | 21 | -0.627 | -1.953 | 0.001 | 0.017 |
| GO_REGULATION_OF_MULTIORGANISM_PROCESS                  | 16 | -0.664 | -1.923 | 0.001 | 0.017 |
| GO_POSITIVE_REGULATION_OF_LOCOMOTION                    | 28 | -0.700 | -1.913 | 0.003 | 0.025 |
| GO_REGULATION_OF_CHEMOTAXIS                             | 14 | -0.616 | -1.884 | 0.001 | 0.017 |
| GO_REGULATION_OF_LEUKOCYTE_MIGRATION                    | 11 | -0.706 | -1.871 | 0.003 | 0.025 |

|                                                             |    |        |        |       |       |
|-------------------------------------------------------------|----|--------|--------|-------|-------|
| GO_POSITIVE_REGULATION_OF_INTRACELLULAR_SIGNAL_TRANSDUCTION | 36 | -0.756 | -1.868 | 0.001 | 0.017 |
| GO_REGULATION_OF_CELLULAR_COMPONENT_MOVE                    | 51 | -0.578 | -1.864 | 0.001 | 0.017 |
| GO_EXTRACELLULAR_STRUCTURE_ORGANIZATION                     | 36 | -0.545 | -1.861 | 0.001 | 0.017 |
| GO_REGULATION_OF_IMMUNE_SYSTEM_PROCESS                      | 43 | -0.576 | -1.859 | 0.001 | 0.017 |
| GO_POSITIVE_REGULATION_OF_CHEMOTAXIS                        | 10 | -0.560 | -1.859 | 0.001 | 0.017 |
| GO_VIRAL_LIFE_CYCLE                                         | 13 | -0.763 | -1.855 | 0.001 | 0.017 |
| GO_CYTOKINE_MEDIATED_SIGNALING_PATHWAY                      | 31 | -0.712 | -1.852 | 0.001 | 0.017 |
| GO_DEFENSE_RESPONSE_TO_BACTERIUM                            | 14 | -0.589 | -1.849 | 0.001 | 0.017 |
| GO_REGULATION_OF_INTRACELLULAR_SIGNAL_TRANSDUCTION          | 59 | -0.695 | -1.840 | 0.003 | 0.025 |
| GO_POSITIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT           | 12 | -0.530 | -1.840 | 0.001 | 0.017 |
| GO_INTERSPECIES_INTERACTION_BETWEEN_ORGANISMS               | 24 | -0.723 | -1.837 | 0.001 | 0.017 |
| GO_REGULATION_OF_RESPONSE_TO_EXTERNAL_STIMULUS              | 38 | -0.620 | -1.833 | 0.001 | 0.017 |
| GO_NEGATIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT           | 10 | -0.561 | -1.827 | 0.001 | 0.017 |
| GO_CARDIOVASCULAR_SYSTEM_DEVELOPMENT                        | 43 | -0.746 | -1.813 | 0.001 | 0.017 |
| GO_ACTIVATION_OF_IMMUNE_RESPONSE                            | 11 | -0.546 | -1.812 | 0.001 | 0.017 |
| GO_REGULATION_OF_CELL_ADHESION                              | 36 | -0.732 | -1.810 | 0.001 | 0.017 |
| GO_IMMUNE_RESPONSE_REGULATING_SIGNALING_PATHWAY             | 12 | -0.561 | -1.808 | 0.001 | 0.017 |
| GO_MYELOID_LEUKOCYTE_ACTIVATION                             | 23 | -0.710 | -1.804 | 0.001 | 0.017 |
| GO_CELLULAR_RESPONSE_TO_LIPID                               | 27 | -0.608 | -1.788 | 0.001 | 0.017 |
| GO_NEGATIVE_REGULATION_OF_CELL_MOTILITY                     | 18 | -0.584 | -1.776 | 0.001 | 0.017 |
| GO_REGULATION_OF_BINDING                                    | 16 | -0.628 | -1.759 | 0.001 | 0.017 |
| GO_REGULATION_OF_APOPTOTIC_SIGNALING_PATHWAY                | 13 | -0.642 | -1.755 | 0.005 | 0.039 |

|                                                        |    |        |        |       |       |
|--------------------------------------------------------|----|--------|--------|-------|-------|
| GO_POSITIVE_REGULATION_OF_IMMUNE_RESPONSE              | 13 | -0.675 | -1.755 | 0.003 | 0.025 |
| GO_CIRCULATORY_SYSTEM_DEVELOPMENT                      | 57 | -0.674 | -1.754 | 0.003 | 0.025 |
| GO_NEGATIVE_REGULATION_OF_LOCOMOTION                   | 19 | -0.509 | -1.754 | 0.001 | 0.017 |
| GO_NEGATIVE_REGULATION_OF_MOLECULAR_FUNCTION           | 48 | -0.611 | -1.738 | 0.001 | 0.017 |
| GO_CELL_ACTIVATION                                     | 42 | -0.513 | -1.736 | 0.001 | 0.017 |
| GO_AMINOGLYCAN_METABOLIC_PROCESS                       | 12 | -0.525 | -1.733 | 0.001 | 0.017 |
| GO_BLOOD_VESSEL_MORPHOGENESIS                          | 39 | -0.680 | -1.727 | 0.004 | 0.036 |
| GO_POSITIVE_REGULATION_OF_DEFENSE_RESPONSE             | 14 | -0.527 | -1.727 | 0.002 | 0.025 |
| GO_WOUND_HEALING                                       | 26 | -0.651 | -1.725 | 0.004 | 0.035 |
| GO_NEGATIVE_REGULATION_OF_HYDROLASE_ACTIVITY           | 25 | -0.570 | -1.724 | 0.002 | 0.025 |
| GO_RESPONSE_TO_INTERFERON_GAMMA                        | 12 | -0.577 | -1.723 | 0.002 | 0.025 |
| GO_POSITIVE_REGULATION_OF_PHOSPHORUS_METABOLIC_PROCESS | 41 | -0.677 | -1.721 | 0.005 | 0.040 |
| GO_POSITIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS    | 43 | -0.520 | -1.717 | 0.001 | 0.017 |
| GO_TAXIS                                               | 40 | -0.517 | -1.714 | 0.001 | 0.017 |
| GO_NEGATIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS    | 45 | -0.521 | -1.710 | 0.003 | 0.031 |
| GO_REGULATION_OF_CATABOLIC_PROCESS                     | 14 | -0.509 | -1.707 | 0.001 | 0.017 |
| GO_CYTOKINE_SECRETION                                  | 12 | -0.643 | -1.705 | 0.005 | 0.040 |
| GO_REGULATION_OF_MAPK_CASCADE                          | 34 | -0.671 | -1.704 | 0.005 | 0.040 |
| GO_REGULATION_OF_DEFENSE_RESPONSE                      | 20 | -0.531 | -1.703 | 0.006 | 0.041 |
| GO_REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS    | 57 | -0.593 | -1.702 | 0.005 | 0.039 |
| GO_POSITIVE_REGULATION_OF_PROTEIN_MODIFICATION_PROCESS | 37 | -0.493 | -1.697 | 0.002 | 0.024 |
| GO_REGULATION_OF_RESPONSE_TO_STRESS                    | 39 | -0.523 | -1.695 | 0.006 | 0.041 |
| GO_REGULATION_OF_EPITHELIAL_CELL_MIGRATION             | 10 | -0.516 | -1.691 | 0.004 | 0.037 |
| GO_REGULATION_OF_PEPTIDASE_ACTIVITY                    | 24 | -0.694 | -1.687 | 0.004 | 0.036 |



|                                                            |    |        |        |       |       |
|------------------------------------------------------------|----|--------|--------|-------|-------|
| GO_REGULATION_OF_PHOSPHORUS_METABOLIC_PROCESS              | 61 | -0.601 | -1.625 | 0.013 | 0.064 |
| GO_RESPONSE_TO_TUMOR_NECROSIS_FACTOR                       | 13 | -0.465 | -1.623 | 0.002 | 0.024 |
| GO_LEUKOCYTE_DIFFERENTIATION                               | 14 | -0.623 | -1.621 | 0.017 | 0.078 |
| GO_CELL_MOTILITY                                           | 87 | -0.610 | -1.617 | 0.015 | 0.072 |
| GO_POSITIVE_REGULATION_OF_CELL_DEATH                       | 30 | -0.449 | -1.616 | 0.003 | 0.030 |
| GO_MUSCLE_STRUCTURE_DEVELOPMENT                            | 36 | -0.519 | -1.615 | 0.008 | 0.050 |
| GO_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND                  | 85 | -0.500 | -1.614 | 0.009 | 0.054 |
| GO_REGULATION_OF_CELL_SUBSTRATE_ADHESION                   | 17 | -0.448 | -1.612 | 0.004 | 0.036 |
| GO_COAGULATION                                             | 12 | -0.576 | -1.607 | 0.020 | 0.085 |
| GO_REGULATION_OF_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY | 11 | -0.631 | -1.604 | 0.013 | 0.066 |
| GO_TUBE_DEVELOPMENT                                        | 62 | -0.646 | -1.598 | 0.018 | 0.079 |
| GO_METAL_ION_HOMEOSTASIS                                   | 29 | -0.457 | -1.598 | 0.004 | 0.036 |
| GO_REGULATION_OF_CELLULAR_CATABOLIC_PROCESSES              | 11 | -0.516 | -1.596 | 0.013 | 0.065 |
| GO_REGULATION_OF_PROTEOLYSIS                               | 28 | -0.644 | -1.592 | 0.018 | 0.079 |
| GO_RESPONSE_TO ABIOTIC_STIMULUS                            | 47 | -0.520 | -1.591 | 0.011 | 0.060 |
| GO_POSITIVE_REGULATION_OF_ERK1_AND_ERK2_CASCADE            | 10 | -0.472 | -1.590 | 0.008 | 0.049 |
| GO_POSITIVE_REGULATION_OF_CYTOKINE_PRODUCTION              | 11 | -0.651 | -1.583 | 0.014 | 0.067 |
| GO_TUBE_MORPHOGENESIS                                      | 53 | -0.639 | -1.580 | 0.019 | 0.083 |
| GO_BIOLOGICAL_ADHESION                                     | 85 | -0.461 | -1.579 | 0.007 | 0.044 |
| GO_ORGANONITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS            | 39 | -0.438 | -1.577 | 0.006 | 0.043 |
| GO_HOMEOSTATIC_PROCESSES                                   | 67 | -0.481 | -1.576 | 0.018 | 0.079 |
| GO_IMMUNE_EFFECTOR_PROCESS                                 | 26 | -0.447 | -1.575 | 0.004 | 0.036 |
| GO_ERK1_AND_ERK2_CASCADE                                   | 19 | -0.520 | -1.575 | 0.021 | 0.087 |
| GO_SKELETAL_MUSCLE_ORGAN_DEVELOPMENT                       | 14 | -0.554 | -1.574 | 0.020 | 0.085 |
| GO_SULFUR_COMPOUND_BIOSYNTHETIC_PROCESS                    | 10 | -0.592 | -1.569 | 0.032 | 0.115 |

|                                                             |    |        |        |       |       |
|-------------------------------------------------------------|----|--------|--------|-------|-------|
| GO_RESPONSE_TO_FIBROBLAST_GROWTH_FACTOR                     | 10 | -0.644 | -1.565 | 0.019 | 0.083 |
| GO_CELLULAR_RESPONSE_TO_DNA_DAMAGE_STIMULUS                 | 10 | -0.642 | -1.561 | 0.020 | 0.086 |
| GO_NEGATIVE_REGULATION_OF_CELL_ADHESION                     | 16 | -0.641 | -1.558 | 0.020 | 0.086 |
| GO_APOPTOTIC_SIGNALING_PATHWAY                              | 20 | -0.569 | -1.555 | 0.035 | 0.124 |
| GO_RESPONSE_TO_NUTRIENT                                     | 12 | -0.541 | -1.555 | 0.022 | 0.089 |
| GO_RESPONSE_TO_OXYGEN_LEVELS                                | 21 | -0.611 | -1.551 | 0.025 | 0.099 |
| GO_REGULATION_OF_HEMOPOIESIS                                | 13 | -0.536 | -1.551 | 0.026 | 0.099 |
| GO_RESPONSE_TO_GROWTH_FACTOR                                | 40 | -0.596 | -1.550 | 0.035 | 0.124 |
| GO_POSITIVE_REGULATION_OF_CATALYTIC_ACTIVITY                | 36 | -0.472 | -1.548 | 0.018 | 0.079 |
| GO_REGULATION_OF_IMMUNE_RESPONSE                            | 19 | -0.480 | -1.547 | 0.014 | 0.067 |
| GO_POSITIVE_REGULATION_OF_MOLECULAR_FUNCTION                | 50 | -0.544 | -1.547 | 0.025 | 0.099 |
| GO_LOCOMOTION                                               | 95 | -0.455 | -1.545 | 0.013 | 0.066 |
| GO_ANATOMICAL_STRUCTURE_FORMATION_INVOLVED_IN_MORPHOGENESIS | 58 | -0.423 | -1.544 | 0.006 | 0.043 |
| GO_REGULATION_OF_PROTEIN_SERINE_THREONINE_KINASE_ACTIVITY   | 18 | -0.446 | -1.542 | 0.007 | 0.044 |
| GO_CONNECTIVE_TISSUE_DEVELOPMENT                            | 21 | -0.550 | -1.542 | 0.033 | 0.116 |
| GO_RESPONSE_TO_ANTI-BIOTIC                                  | 17 | -0.531 | -1.536 | 0.029 | 0.110 |
| GO_REGULATION_OF_CELL_DEATH                                 | 66 | -0.550 | -1.535 | 0.037 | 0.129 |
| GO_CHEMICAL_HOMEOSTASIS                                     | 50 | -0.436 | -1.535 | 0.011 | 0.060 |
| GO_LEUKOCYTE_MEDIATED_IMMUNITY                              | 21 | -0.452 | -1.534 | 0.014 | 0.069 |
| GO_NEGATIVE_REGULATION_OF_SIGNALING                         | 55 | -0.530 | -1.534 | 0.029 | 0.110 |
| GO_DNA_METABOLIC_PROCESS                                    | 11 | -0.446 | -1.533 | 0.012 | 0.064 |
| GO_MACROMOLECULE_CATABOLIC_PROCESS                          | 21 | -0.620 | -1.532 | 0.023 | 0.093 |
| GO_RESPONSE_TO_CHEMOKINE                                    | 10 | -0.529 | -1.531 | 0.030 | 0.111 |
| GO_REGULATION_OF_BODY_FLUID_LEVELS                          | 23 | -0.629 | -1.528 | 0.028 | 0.109 |
| GO_PROTEOLYSIS                                              | 43 | -0.519 | -1.526 | 0.030 | 0.111 |

|                                                                       |    |        |        |       |       |
|-----------------------------------------------------------------------|----|--------|--------|-------|-------|
| GO_CARTILAGE_DEVELOPMENT                                              | 18 | -0.460 | -1.525 | 0.019 | 0.083 |
| GO_SKIN_DEVELOPMENT                                                   | 12 | -0.543 | -1.523 | 0.036 | 0.127 |
| GO_POSITIVE_REGULATION_OF_CELL_ADHESION                               | 19 | -0.599 | -1.523 | 0.037 | 0.129 |
| GO_AGING                                                              | 12 | -0.535 | -1.522 | 0.041 | 0.137 |
| GO_RECEPTOR_SIGNALING_PATHWAY_VIA_STAT                                | 10 | -0.598 | -1.518 | 0.042 | 0.140 |
| GO_NEGATIVE_REGULATION_OF_INTRACELLULAR_SIGNAL_TRANSDUCTION           | 19 | -0.622 | -1.512 | 0.032 | 0.116 |
| GO_CELLULAR_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND                    | 53 | -0.531 | -1.510 | 0.043 | 0.142 |
| GO_NEGATIVE_REGULATION_OF_CELLULAR_RESPONSE_TO_GROWTH_FACTOR_STIMULUS | 10 | -0.438 | -1.500 | 0.021 | 0.086 |
| GO_PROTEIN_KINASE_B_SIGNALING                                         | 14 | -0.613 | -1.490 | 0.039 | 0.134 |
| GO_RESPONSE_TO_TOXIC_SUBSTANCE                                        | 23 | -0.560 | -1.484 | 0.050 | 0.158 |
| GO_COFACTOR_METABOLIC_PROCESS                                         | 12 | -0.504 | -1.483 | 0.038 | 0.130 |
| GO_MYELOID_LEUKOCYTE_DIFFERENTIATION                                  | 10 | -0.583 | -1.481 | 0.048 | 0.153 |
| GO_ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY                   | 49 | -0.608 | -1.478 | 0.041 | 0.136 |
| GO_MONOCARBOXYLIC_ACID_TRANSPORT                                      | 10 | -0.435 | -1.473 | 0.031 | 0.112 |
| GO_REGULATION_OF_PROTEIN_MODIFICATION_PROCESS                         | 57 | -0.606 | -1.472 | 0.042 | 0.139 |
| GO_NEGATIVE_REGULATION_OF_CELL_POPULATION_PROLIFERATION               | 31 | -0.425 | -1.466 | 0.021 | 0.086 |
| GO_EXOCYTOSIS                                                         | 41 | -0.463 | -1.453 | 0.046 | 0.149 |
| GO_CELLULAR_HOMEOSTASIS                                               | 39 | -0.439 | -1.448 | 0.041 | 0.137 |
| GO_CARBOHYDRATE_DERIVATIVE_CATABOLIC_PROCESS                          | 10 | -0.440 | -1.442 | 0.040 | 0.136 |
| GO_POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS                       | 77 | -0.593 | -1.442 | 0.049 | 0.155 |
| GO_SECRETION                                                          | 87 | -0.397 | -1.417 | 0.029 | 0.109 |
| GO_ANIMAL_ORGAN_MORPHOGENESIS                                         | 71 | -0.391 | -1.409 | 0.030 | 0.111 |
| GO_VESICLE_MEDIATED_TRANSPORT_IN_SYNAPSE                              | 24 | -0.396 | -1.400 | 0.046 | 0.149 |

|                                                      |     |       |       |       |       |
|------------------------------------------------------|-----|-------|-------|-------|-------|
| GO_ORGANELLE_LOCALIZATION                            | 18  | 0.658 | 2.611 | 0.006 | 0.041 |
| GO_OLIGODENDROCYTE_DIFFERENTIATION                   | 13  | 0.647 | 2.333 | 0.005 | 0.039 |
| GO_REGULATION_OF_SYNAPTIC_PLASTICITY                 | 19  | 0.687 | 2.232 | 0.004 | 0.036 |
| GO_REGULATION_OF_SYNAPTIC_VESICLE_CYCLE              | 11  | 0.574 | 2.099 | 0.005 | 0.040 |
| GO_OLIGODENDROCYTE_DEVELOPMENT                       | 10  | 0.652 | 1.992 | 0.008 | 0.049 |
| GO_ENSHEATHMENT_OF_NEURONS                           | 17  | 0.682 | 1.984 | 0.008 | 0.049 |
| GO_NEUROTRANSMITTER_SECRETION                        | 16  | 0.525 | 1.859 | 0.015 | 0.071 |
| GO_SYNAPTIC_VESICLE_EXOCYTOSIS                       | 11  | 0.545 | 1.857 | 0.010 | 0.056 |
| GO_NEGATIVE_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT | 20  | 0.591 | 1.806 | 0.023 | 0.091 |
| GO_REGULATION_OF_NERVOUS_SYSTEM_PROCESS              | 13  | 0.485 | 1.785 | 0.011 | 0.060 |
| GO_NEUROTRANSMITTER_TRANSPORT                        | 25  | 0.543 | 1.765 | 0.021 | 0.087 |
| GO_REGULATION_OF_NEUROTRANSMITTER_SECRETION          | 11  | 0.428 | 1.731 | 0.006 | 0.041 |
| GO_REGULATION_OF_EXOCYTOSIS                          | 17  | 0.563 | 1.722 | 0.026 | 0.102 |
| GO_CALCIIUM_ION_REGULATED_EXOCYTOSIS                 | 10  | 0.486 | 1.718 | 0.025 | 0.099 |
| GO_CELLULAR_PROTEIN_COMPLEX_ASSEMBLY                 | 20  | 0.586 | 1.705 | 0.031 | 0.111 |
| GO_DEVELOPMENTAL_MATURATION                          | 24  | 0.460 | 1.697 | 0.011 | 0.060 |
| GO_PERIPHERAL_NERVOUS_SYSTEM_DEVELOPMENT             | 11  | 0.424 | 1.682 | 0.017 | 0.077 |
| GO_ESTABLISHMENT_OF_ORGANELLE_LOCALIZATION           | 10  | 0.549 | 1.678 | 0.030 | 0.111 |
| GO_DICARBOXYLIC_ACID_TRANSPORT                       | 10  | 0.544 | 1.585 | 0.046 | 0.149 |
| GO_PROTEIN_CONTAINING_COMPLEX_ASSEMBLY               | 34  | 0.531 | 1.546 | 0.050 | 0.158 |
| GO_NEUROGENESIS                                      | 114 | 0.359 | 1.532 | 0.008 | 0.051 |
| GO_CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT                | 85  | 0.281 | 1.531 | 0.030 | 0.111 |
| GO_NEURON_DEVELOPMENT                                | 77  | 0.292 | 1.523 | 0.021 | 0.087 |
| GO_SYNAPSE_ORGANIZATION                              | 35  | 0.296 | 1.495 | 0.017 | 0.079 |

|                                                      |                |                         |            |               |                 |
|------------------------------------------------------|----------------|-------------------------|------------|---------------|-----------------|
| GO_REGULATION_OF_VESICLE_MEDIATED_TRANSPORT          | 34             | 0.341                   | 1.458      | 0.016         | 0.077           |
| GO_SYNAPTIC_SIGNALING                                | 66             | 0.329                   | 1.407      | 0.025         | 0.099           |
| GO_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT          | 63             | 0.285                   | 1.395      | 0.015         | 0.071           |
|                                                      |                | 0.262                   | 1.270      | 0.040         | 0.136           |
| <b>CS2_specific GO Terms</b>                         | <b>setSize</b> | <b>enrichment Score</b> | <b>NES</b> | <b>pvalue</b> | <b>p.adjust</b> |
| GO_VESICLE_MEDIATED_TRANSPORT_IN_SYNAPSE             | 24             | -0.660                  | -2.398     | 0.003         | 0.021           |
| GO_NEUROTRANSMITTER_TRANSPORT                        | 25             | -0.587                  | -2.166     | 0.003         | 0.021           |
| GO_NEGATIVE_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT | 20             | -0.614                  | -2.090     | 0.003         | 0.021           |
| GO_REGULATION_OF_SYNAPTIC_PLASTICITY                 | 19             | -0.617                  | -2.046     | 0.003         | 0.021           |
| GO_NEUROTRANSMITTER_SECRETION                        | 16             | -0.630                  | -2.012     | 0.003         | 0.020           |
| GO_SYNAPTIC_SIGNALING                                | 66             | -0.428                  | -2.006     | 0.005         | 0.028           |
| GO_INTRACELLULAR_TRANSPORT                           | 28             | -0.523                  | -1.994     | 0.004         | 0.022           |
| GO_REGULATION_OF_SYNAPTIC_VESICLE_CYCLE              | 11             | -0.688                  | -1.983     | 0.003         | 0.020           |
| GO_OLIGODENDROCYTE_DIFFERENTIATION                   | 13             | -0.645                  | -1.955     | 0.003         | 0.020           |
| GO_SYNAPSE_ORGANIZATION                              | 35             | -0.480                  | -1.953     | 0.004         | 0.024           |
| GO_REGULATION_OF_SYNAPSE_STRUCTURE_OR_ACTIVITY       | 14             | -0.613                  | -1.908     | 0.003         | 0.020           |
| GO_SIGNAL_RELEASE                                    | 36             | -0.449                  | -1.835     | 0.004         | 0.024           |
| GO_FOREBRAIN_DEVELOPMENT                             | 30             | -0.471                  | -1.816     | 0.004         | 0.023           |
| GO_REGULATION_OF_NEUROTRANSMITTER_SECRETION          | 11             | -0.626                  | -1.804     | 0.009         | 0.039           |
| GO_REGULATION_OF_NEUROTRANSMITTER_TRANSPORT          | 15             | -0.573                  | -1.798     | 0.009         | 0.041           |
| GO_DENDRITE_DEVELOPMENT                              | 11             | -0.617                  | -1.778     | 0.011         | 0.047           |
| GO_NEGATIVE_REGULATION_OF_NEURON_DIFFERENTIATION     | 14             | -0.568                  | -1.770     | 0.008         | 0.039           |
| GO_SYNAPTIC_VESICLE_EXOCYTOSIS                       | 11             | -0.613                  | -1.767     | 0.011         | 0.047           |
| GO_ORGANELLE_LOCALIZATION                            | 18             | -0.532                  | -1.755     | 0.007         | 0.033           |

|                                                       |     |        |        |       |       |
|-------------------------------------------------------|-----|--------|--------|-------|-------|
| GO_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT           | 63  | -0.379 | -1.733 | 0.006 | 0.029 |
| GO_CALCIIUM_ION_REGULATED_EXOCYTOSIS                  | 10  | -0.614 | -1.724 | 0.014 | 0.054 |
| GO_TELENCEPHALON_DEVELOPMENT                          | 21  | -0.496 | -1.712 | 0.010 | 0.042 |
| GO_DEVELOPMENTAL_MATURATION                           | 24  | -0.464 | -1.688 | 0.014 | 0.054 |
| GO_CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT                 | 85  | -0.347 | -1.687 | 0.006 | 0.031 |
| GO_METENCEPHALON_DEVELOPMENT                          | 11  | -0.579 | -1.666 | 0.034 | 0.107 |
| GO_SYNAPSE_ASSEMBLY                                   | 11  | -0.578 | -1.664 | 0.034 | 0.107 |
| GO_HEAD_DEVELOPMENT                                   | 56  | -0.361 | -1.638 | 0.005 | 0.027 |
| GO_OLIGODENDROCYTE_DEVELOPMENT                        | 10  | -0.578 | -1.626 | 0.036 | 0.111 |
| GO_REGULATION_OF_TRANS_SYNAPTIC_SIGNALING             | 38  | -0.388 | -1.602 | 0.017 | 0.062 |
| GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ORGANELLE | 11  | -0.549 | -1.582 | 0.048 | 0.137 |
| GO_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC                | 10  | -0.553 | -1.555 | 0.044 | 0.128 |
| GO_MULTICELLULAR_ORGANISMAL_SIGNALING                 | 18  | -0.461 | -1.522 | 0.040 | 0.119 |
| GO_REGULATION_OF_NEURON_DIFFERENTIATION               | 46  | -0.350 | -1.511 | 0.023 | 0.079 |
| GO_REGULATION_OF_NEUROTRANSMITTER_LEVELS              | 27  | -0.401 | -1.504 | 0.042 | 0.124 |
| GO_CELL_PART_MORPHOGENESIS                            | 42  | -0.351 | -1.466 | 0.036 | 0.111 |
| GO_REGULATION_OF_CELL_DEVELOPMENT                     | 63  | -0.313 | -1.434 | 0.028 | 0.091 |
| GO_NEUROGENESIS                                       | 114 | -0.279 | -1.411 | 0.008 | 0.037 |
| GO_NEURON_DIFFERENTIATION                             | 97  | -0.259 | -1.286 | 0.047 | 0.134 |
| GO_EXTRACELLULAR_STRUCTURE_ORGANIZATION               | 36  | 0.667  | 2.250  | 0.001 | 0.014 |
| GO_RESPONSE_TO_BACTERIUM                              | 35  | 0.649  | 2.175  | 0.001 | 0.014 |
| GO_CIRCULATORY_SYSTEM_DEVELOPMENT                     | 57  | 0.588  | 2.155  | 0.001 | 0.014 |
| GO_CARDIOVASCULAR_SYSTEM_DEVELOPMENT                  | 43  | 0.617  | 2.151  | 0.001 | 0.014 |
| GO_RESPONSE_TO_BIOTIC_STIMULUS                        | 50  | 0.598  | 2.147  | 0.001 | 0.014 |
| GO_HUMORAL_IMMUNE_RESPONSE                            | 16  | 0.769  | 2.137  | 0.001 | 0.014 |
| GO_POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS       | 30  | 0.655  | 2.136  | 0.001 | 0.014 |

|                                                   |    |       |       |       |       |
|---------------------------------------------------|----|-------|-------|-------|-------|
| GO_NEGATIVE_REGULATION_OF_PROTEOLYSIS             | 19 | 0.724 | 2.128 | 0.001 | 0.014 |
| GO_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN       | 21 | 0.708 | 2.117 | 0.001 | 0.014 |
| GO_INFLAMMATORY_RESPONSE                          | 35 | 0.629 | 2.108 | 0.001 | 0.014 |
| GO_NEGATIVE_REGULATION_OF_HYDROLASE_ACTIVITY      | 25 | 0.676 | 2.101 | 0.001 | 0.014 |
| GO_CYTOKINE_PRODUCTION                            | 27 | 0.662 | 2.101 | 0.001 | 0.014 |
| GO_REGULATION_OF_PEPTIDASE_ACTIVITY               | 24 | 0.680 | 2.093 | 0.001 | 0.014 |
| GO_RESPONSE_TO_CYTOKINE                           | 49 | 0.586 | 2.091 | 0.001 | 0.014 |
| GO_NEGATIVE_REGULATION_OF_PEPTIDASE_ACTIVITY      | 18 | 0.724 | 2.086 | 0.001 | 0.014 |
| GO_REGULATION_OF_IMMUNE_SYSTEM_PROCESS            | 43 | 0.598 | 2.083 | 0.001 | 0.014 |
| GO_CELL_ACTIVATION                                | 42 | 0.596 | 2.075 | 0.001 | 0.014 |
| GO_DEFENSE_RESPONSE                               | 56 | 0.568 | 2.071 | 0.001 | 0.014 |
| GO_RESPONSE_TO_LIPID                              | 48 | 0.578 | 2.068 | 0.001 | 0.014 |
| GO_DEFENSE_RESPONSE_TO_OTHER_ORGANISM             | 33 | 0.623 | 2.065 | 0.001 | 0.014 |
| GO_POSITIVE_REGULATION_OF_CYTOKINE_PRODUCTION     | 11 | 0.809 | 2.058 | 0.002 | 0.014 |
| GO_WOUND_HEALING                                  | 26 | 0.657 | 2.055 | 0.001 | 0.014 |
| GO_ANTIMICROBIAL_HUMORAL_RESPONSE                 | 13 | 0.773 | 2.052 | 0.002 | 0.014 |
| GO_POSITIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT | 12 | 0.782 | 2.040 | 0.002 | 0.014 |
| GO_CELLULAR_RESPONSE_TO_BIOTIC_STIMULUS           | 13 | 0.768 | 2.038 | 0.002 | 0.014 |
| GO_BLOOD_VESSEL_MORPHOGENESIS                     | 39 | 0.593 | 2.038 | 0.001 | 0.014 |
| GO_REGULATION_OF_PROTEOLYSIS                      | 28 | 0.633 | 2.018 | 0.001 | 0.014 |
| GO_CELL_CHEMOTAXIS                                | 21 | 0.675 | 2.018 | 0.001 | 0.014 |
| GO_RESPONSE_TO_TUMOR_NECROSIS_FACTOR              | 13 | 0.759 | 2.014 | 0.002 | 0.014 |
| GO_RESPONSE_TO_GROWTH_FACTOR                      | 40 | 0.580 | 2.005 | 0.001 | 0.014 |
| GO_LEUKOCYTE_DIFFERENTIATION                      | 14 | 0.741 | 2.001 | 0.002 | 0.014 |
| GO_PROTEOLYSIS                                    | 43 | 0.574 | 2.001 | 0.001 | 0.014 |
| GO_REGULATION_OF_HYDROLASE_ACTIVITY               | 52 | 0.555 | 1.993 | 0.001 | 0.014 |

|                                                             |    |       |       |       |       |
|-------------------------------------------------------------|----|-------|-------|-------|-------|
| GO_NEGATIVE_REGULATION_OF_MOLECULAR_FUNCTION                | 48 | 0.557 | 1.992 | 0.001 | 0.014 |
| GO_GRANULOCYTE_MIGRATION                                    | 12 | 0.763 | 1.989 | 0.002 | 0.014 |
| GO_LEUKOCYTE_CHEMOTAXIS                                     | 16 | 0.715 | 1.988 | 0.001 | 0.014 |
| GO_LEUKOCYTE_MIGRATION                                      | 27 | 0.626 | 1.984 | 0.001 | 0.014 |
| GO_REGULATION_OF_CELL_ADHESION                              | 36 | 0.587 | 1.980 | 0.001 | 0.014 |
| GO_NEGATIVE_REGULATION_OF_LOCOMOTION                        | 19 | 0.672 | 1.976 | 0.001 | 0.014 |
| GO_RESPONSE_TO_WOUNDING                                     | 33 | 0.595 | 1.974 | 0.001 | 0.014 |
| GO_RESPONSE_TO_TRANSFORMING_GROWTH_FACTOR_BETA              | 15 | 0.716 | 1.968 | 0.003 | 0.020 |
| GO_NEGATIVE_REGULATION_OF_CELL_MOTILITY                     | 18 | 0.678 | 1.953 | 0.001 | 0.014 |
| GO_REGULATION_OF_CELLULAR_COMPONENT_MOVEMENT                | 51 | 0.542 | 1.947 | 0.001 | 0.014 |
| GO_REGULATION_OF_VASCULATURE_DEVELOPMENT                    | 21 | 0.646 | 1.932 | 0.001 | 0.014 |
| GO_ACTIVATION_OF_IMMUNE_RESPONSE                            | 11 | 0.758 | 1.928 | 0.002 | 0.014 |
| GO_OSSIFICATION                                             | 25 | 0.620 | 1.926 | 0.001 | 0.014 |
| GO_RESPONSE_TO_ANTIBIOTIC                                   | 17 | 0.676 | 1.924 | 0.003 | 0.020 |
| GO_MYELOID_LEUKOCYTE_MIGRATION                              | 16 | 0.692 | 1.924 | 0.003 | 0.020 |
| GO_BONE_DEVELOPMENT                                         | 13 | 0.720 | 1.912 | 0.002 | 0.014 |
| GO_MYELOID_LEUKOCYTE_ACTIVATION                             | 23 | 0.628 | 1.907 | 0.001 | 0.014 |
| GO_REGULATION_OF_MULTIORGANISM_PROCESS                      | 16 | 0.685 | 1.904 | 0.003 | 0.020 |
| GO_OSTEOBLAST_DIFFERENTIATION                               | 17 | 0.668 | 1.901 | 0.003 | 0.020 |
| GO_POSITIVE_REGULATION_OF_RESPONSE_TO_EXTERNAL_STIMULUS     | 23 | 0.625 | 1.898 | 0.001 | 0.014 |
| GO_POSITIVE_REGULATION_OF_INTRACELLULAR_SIGNAL_TRANSDUCTION | 36 | 0.562 | 1.896 | 0.001 | 0.014 |
| GO_NEGATIVE_REGULATION_OF_CELL_POPULATION_PROLIFERATION     | 31 | 0.575 | 1.892 | 0.003 | 0.020 |
| GO_NEUTROPHIL_MIGRATION                                     | 10 | 0.765 | 1.890 | 0.002 | 0.014 |

|                                                               |    |       |       |       |       |
|---------------------------------------------------------------|----|-------|-------|-------|-------|
| GO_REGULATION_OF_CYST<br>EINE_TYPE_ENDOPEPTIDAS<br>E_ACTIVITY | 10 | 0.764 | 1.887 | 0.002 | 0.014 |
| GO_NEGATIVE_REGULATIO<br>N_OF_CATALYTIC_ACTIVIT<br>Y          | 38 | 0.553 | 1.886 | 0.001 | 0.014 |
| GO_REGULATION_OF_APOP<br>TOTIC_SIGNALING_PATHW<br>AY          | 13 | 0.709 | 1.882 | 0.002 | 0.014 |
| GO_INTERSPECIES_INTERA<br>CTION_BETWEEN_ORGANIS<br>MS         | 24 | 0.611 | 1.881 | 0.003 | 0.020 |
| GO_IMMUNE_RESPONSE_RE<br>GULATING_SIGNALING_PAT<br>HWAY       | 12 | 0.720 | 1.879 | 0.002 | 0.014 |
| GO_MYELOID_CELL_DIFFER<br>ENTIATION                           | 15 | 0.682 | 1.876 | 0.003 | 0.020 |
| GO_REGULATION_OF_BINDI<br>NG                                  | 16 | 0.673 | 1.871 | 0.003 | 0.020 |
| GO_POSITIVE_REGULATION<br>_OF_LOCOMOTION                      | 28 | 0.586 | 1.870 | 0.004 | 0.024 |
| GO_CELLULAR_RESPONSE_<br>TO_LIPID                             | 27 | 0.589 | 1.868 | 0.001 | 0.014 |
| GO_POSITIVE_REGULATION<br>_OF_IMMUNE_RESPONSE                 | 13 | 0.703 | 1.866 | 0.003 | 0.020 |
| GO_RESPONSE_TO_OXYGEN<br>_CONTAINING_COMPOUND                 | 85 | 0.480 | 1.863 | 0.001 | 0.014 |
| GO_CELL_MOTILITY                                              | 87 | 0.478 | 1.857 | 0.001 | 0.014 |
| GO_CELL_JUNCTION_ORGA<br>NIZATION                             | 13 | 0.697 | 1.850 | 0.003 | 0.020 |
| GO_INNATE_IMMUNE_RESP<br>ONSE                                 | 25 | 0.595 | 1.849 | 0.003 | 0.020 |
| GO_NEGATIVE_REGULATIO<br>N_OF_CELL_ADHESION                   | 16 | 0.661 | 1.838 | 0.003 | 0.020 |
| GO_POSITIVE_REGULATION<br>_OF_CELL_DEATH                      | 30 | 0.561 | 1.830 | 0.004 | 0.024 |
| GO_SIGNAL_TRANSDUCTIO<br>N_BY_PROTEIN_PHOSPHOR<br>YLATION     | 39 | 0.531 | 1.826 | 0.001 | 0.014 |
| GO_CELL_MATRIX_ADHESI<br>ON                                   | 17 | 0.640 | 1.819 | 0.004 | 0.024 |
| GO_REGULATION_OF_INTR<br>ACELLULAR_SIGNAL_TRAN<br>SDUCTION    | 59 | 0.493 | 1.813 | 0.001 | 0.014 |
| GO_COAGULATION                                                | 12 | 0.693 | 1.808 | 0.003 | 0.020 |
| GO_DEFENSE_RESPONSE_T<br>O_BACTERIUM                          | 14 | 0.669 | 1.807 | 0.005 | 0.026 |
| GO_TUBE_DEVELOPMENT                                           | 62 | 0.487 | 1.801 | 0.001 | 0.014 |
| GO_CELL_JUNCTION_ASSE<br>MBLY                                 | 11 | 0.704 | 1.793 | 0.005 | 0.026 |
| GO_NEGATIVE_REGULATIO<br>N_OF_CELLULAR_RESPONS                | 10 | 0.724 | 1.790 | 0.003 | 0.020 |



|                                                                             |    |       |       |       |       |
|-----------------------------------------------------------------------------|----|-------|-------|-------|-------|
| GO_SKELETAL_SYSTEM_MORPHOGENESIS                                            | 28 | 0.540 | 1.722 | 0.008 | 0.039 |
| GO_REGULATION_OF_CELL_SUBSTRATE_ADHESION                                    | 17 | 0.605 | 1.722 | 0.009 | 0.039 |
| GO_REGULATION_OF_CELL_MATRIX_ADHESION                                       | 11 | 0.675 | 1.718 | 0.017 | 0.063 |
| GO_REGULATION_OF_CELL_DEATH                                                 | 66 | 0.459 | 1.716 | 0.002 | 0.020 |
| GO_ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY                         | 49 | 0.481 | 1.716 | 0.001 | 0.014 |
| GO_AGING                                                                    | 12 | 0.657 | 1.714 | 0.011 | 0.045 |
| GO_ANATOMICAL_STRUCTURE_FORMATION_INVOLVED_IN_MORPHOGENESIS                 | 58 | 0.466 | 1.707 | 0.001 | 0.014 |
| GO_TRANSMEMBRANE_RECEPTOR_PROTEIN_SERINE_THREONINE_KINASE_SIGNALING_PATHWAY | 20 | 0.577 | 1.705 | 0.010 | 0.042 |
| GO_VIRAL_LIFE_CYCLE                                                         | 13 | 0.642 | 1.702 | 0.009 | 0.041 |
| GO_EPITHELIAL_CELL_PROLIFERATION                                            | 20 | 0.576 | 1.702 | 0.010 | 0.042 |
| GO_TRANSFORMING_GROWTH_FACTOR_BETA_RECEPTOR_SIGNALING_PATHWAY               | 11 | 0.666 | 1.694 | 0.017 | 0.063 |
| GO_RESPONSE_TO_INORGANIC_SUBSTANCE                                          | 30 | 0.519 | 1.694 | 0.008 | 0.038 |
| GO_NEGATIVE_REGULATION_OF_TRANSMEMBRANE_TRANSPORT                           | 10 | 0.683 | 1.687 | 0.009 | 0.041 |
| GO_CELL_ACTIVATION_INVOLVED_IN_IMMUNE_RESPONSE                              | 19 | 0.574 | 1.687 | 0.013 | 0.052 |
| GO_MYELOID_LEUKOCYTE_MEDIATED_IMMUNITY                                      | 19 | 0.574 | 1.687 | 0.013 | 0.052 |
| GO_MULTICELLULAR_ORGANISMAL_HOMEOSTASIS                                     | 21 | 0.562 | 1.681 | 0.010 | 0.042 |
| GO_REGULATION_OF_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY                  | 11 | 0.660 | 1.679 | 0.020 | 0.070 |
| GO_ANIMAL_ORGAN_MORPHOGENESIS                                               | 71 | 0.443 | 1.678 | 0.006 | 0.031 |
| GO_TAXIS                                                                    | 40 | 0.485 | 1.676 | 0.006 | 0.032 |
| GO_NEGATIVE_REGULATION_OF_INTRACELLULAR_SIGNAL_TRANSDUCTION                 | 19 | 0.569 | 1.673 | 0.014 | 0.054 |
| GO_REGULATION_OF_DEFENSE_RESPONSE                                           | 20 | 0.566 | 1.672 | 0.014 | 0.054 |
| GO_REGULATION_OF_CELLULAR_RESPONSE_TO_GROWTH_FACTOR_STIMULUS                | 20 | 0.566 | 1.672 | 0.014 | 0.054 |

|                                                                                           |    |       |       |       |       |
|-------------------------------------------------------------------------------------------|----|-------|-------|-------|-------|
| GO_CARTILAGE_DEVELOPMENT                                                                  | 18 | 0.580 | 1.670 | 0.013 | 0.052 |
| GO_REGULATION_OF_LYMPHOCYTE_ACTIVATION                                                    | 12 | 0.637 | 1.661 | 0.021 | 0.074 |
| GO_SENSORY_PERCEPTION                                                                     | 28 | 0.520 | 1.658 | 0.014 | 0.054 |
| GO_RESPONSE_TO_MECHANICAL_STIMULUS                                                        | 16 | 0.596 | 1.656 | 0.018 | 0.065 |
| GO_GASTRULATION                                                                           | 11 | 0.649 | 1.653 | 0.022 | 0.074 |
| GO_CELLULAR_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND                                        | 53 | 0.458 | 1.650 | 0.005 | 0.027 |
| GO_REGULATION_OF_TRANSMEMBRANE_RECEPTOR_PROTEIN_SERINE_THREONINE_KINASE_SIGNALING_PATHWAY | 17 | 0.580 | 1.650 | 0.019 | 0.067 |
| GO_HEART_DEVELOPMENT                                                                      | 26 | 0.525 | 1.643 | 0.014 | 0.054 |
| GO_PROTEIN_PHOSPHORYLATION                                                                | 60 | 0.445 | 1.642 | 0.005 | 0.026 |
| GO_REGULATION_OF_CHEMOTAXIS                                                               | 14 | 0.607 | 1.642 | 0.023 | 0.079 |
| GO_SUPRAMOLECULAR_FIBER_ORGANIZATION                                                      | 28 | 0.513 | 1.637 | 0.014 | 0.054 |
| GO_NEGATIVE_REGULATION_OF_RESPONSE_TO_STIMULUS                                            | 64 | 0.439 | 1.632 | 0.005 | 0.026 |
| GO_POSITIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS                                       | 43 | 0.468 | 1.631 | 0.009 | 0.041 |
| GO_RESPONSE_TO_NUTRIENT                                                                   | 12 | 0.624 | 1.628 | 0.027 | 0.091 |
| GO_POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS                                           | 77 | 0.423 | 1.626 | 0.005 | 0.026 |
| GO_RESPONSE_TO_OXYGEN_LEVELS                                                              | 21 | 0.543 | 1.624 | 0.020 | 0.070 |
| GO_REPRODUCTION                                                                           | 44 | 0.459 | 1.617 | 0.014 | 0.054 |
| GO_REGULATION_OF_WNT_SIGNALING_PATHWAY                                                    | 17 | 0.568 | 1.617 | 0.026 | 0.087 |
| GO_APOPTOTIC_SIGNALING_PATHWAY                                                            | 20 | 0.546 | 1.612 | 0.018 | 0.067 |
| GO_RESPONSE_TO ABIOTIC_STIMULUS                                                           | 47 | 0.454 | 1.612 | 0.010 | 0.042 |
| GO_TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_SIGNALING_PATHWAY                       | 31 | 0.489 | 1.610 | 0.015 | 0.056 |
| GO_RESPONSE_TO_EXTRACELLULAR_STIMULUS                                                     | 17 | 0.563 | 1.600 | 0.030 | 0.098 |
| GO_POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS                                | 90 | 0.408 | 1.595 | 0.007 | 0.034 |

|                                                        |    |       |       |       |       |
|--------------------------------------------------------|----|-------|-------|-------|-------|
| GO_ION_HOMEOSTASIS                                     | 32 | 0.483 | 1.592 | 0.019 | 0.067 |
| GO_HOMEOSTATIC_PROCESSES                               | 67 | 0.423 | 1.592 | 0.010 | 0.042 |
| GO_REPRODUCTIVE_SYSTEM_DEVELOPMENT                     | 23 | 0.524 | 1.591 | 0.024 | 0.082 |
| GO_RESPONSE_TO_ORGANIC_CYCLIC_COMPOUND                 | 46 | 0.451 | 1.589 | 0.015 | 0.058 |
| GO_NEGATIVE_REGULATION_OF_ION_TRANSPORT                | 10 | 0.641 | 1.584 | 0.031 | 0.099 |
| GO_RESPONSE_TO_REACTIVE_OXYGEN_SPECIES                 | 14 | 0.586 | 1.583 | 0.031 | 0.099 |
| GO_POSITIVE_REGULATION_OF_CHEMOTAXIS                   | 10 | 0.640 | 1.582 | 0.031 | 0.099 |
| GO_REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS    | 57 | 0.431 | 1.581 | 0.018 | 0.067 |
| GO_REGULATION_OF_HEMOPOIESIS                           | 13 | 0.595 | 1.579 | 0.038 | 0.116 |
| GO_POSITIVE_REGULATION_OF_ERK1_AND_ERK2_CASCADE        | 10 | 0.638 | 1.577 | 0.036 | 0.111 |
| GO_NEGATIVE_REGULATION_OF_SIGNALING                    | 55 | 0.434 | 1.576 | 0.019 | 0.067 |
| GO_SKELETAL_SYSTEM_DEVELOPMENT                         | 48 | 0.438 | 1.568 | 0.020 | 0.070 |
| GO_LYMPHOCYTE_ACTIVATION                               | 15 | 0.570 | 1.567 | 0.036 | 0.111 |
| GO_TISSUE_MORPHOGENESIS                                | 43 | 0.449 | 1.563 | 0.026 | 0.087 |
| GO_SENSORY_ORGAN_MORPHOGENESIS                         | 21 | 0.521 | 1.560 | 0.040 | 0.119 |
| GO_POSITIVE_REGULATION_OF_PROTEIN_MODIFICATION_PROCESS | 37 | 0.459 | 1.559 | 0.020 | 0.070 |
| GO_RESPONSE_TO_OXIDATIVE_STRESS                        | 16 | 0.559 | 1.554 | 0.038 | 0.116 |
| GO_MUSCLE_STRUCTURE_DEVELOPMENT                        | 36 | 0.460 | 1.552 | 0.026 | 0.088 |
| GO_MULTI_MULTICELLULAR_ORGANISM_PROCESS                | 10 | 0.627 | 1.548 | 0.042 | 0.124 |
| GO_POSITIVE_REGULATION_OF_CELL_ADHESION                | 19 | 0.523 | 1.537 | 0.041 | 0.121 |
| GO_REGULATION_OF_CELL_ACTIVATION                       | 16 | 0.552 | 1.534 | 0.044 | 0.129 |
| GO_REGULATION_OF_CELLULAR_CATABOLIC_PROCESSES          | 11 | 0.602 | 1.533 | 0.049 | 0.138 |
| GO_NEGATIVE_REGULATION_OF_CELL_DEATH                   | 37 | 0.450 | 1.529 | 0.028 | 0.091 |
| GO_METAL_ION_HOMEOSTASIS                               | 29 | 0.473 | 1.520 | 0.030 | 0.098 |

|                                                   |    |       |       |       |       |
|---------------------------------------------------|----|-------|-------|-------|-------|
| GO_POSITIVE_REGULATION_OF_SIGNALING               | 81 | 0.390 | 1.506 | 0.018 | 0.066 |
| GO_RECEPTOR_MEDIATED_ENDOCYTOSIS                  | 26 | 0.481 | 1.506 | 0.041 | 0.121 |
| GO_RESPONSE_TO_ENDOGENOUS_STIMULUS                | 74 | 0.393 | 1.499 | 0.029 | 0.095 |
| GO_CHEMICAL_HOMEOSTASIS                           | 50 | 0.417 | 1.495 | 0.037 | 0.114 |
| GO_REGULATION_OF_PHOSPHORYLATION                  | 55 | 0.408 | 1.480 | 0.040 | 0.119 |
| GO_EPITHELIUM_DEVELOPMENT                         | 57 | 0.402 | 1.473 | 0.046 | 0.132 |
| GO_DEVELOPMENTAL_PROCESS_INVOLVED_IN_REPRODUCTION | 28 | 0.459 | 1.463 | 0.047 | 0.134 |
| GO_SECRETION                                      | 87 | 0.373 | 1.450 | 0.031 | 0.099 |
| GO_REGULATION_OF_PHOSPHORUS_METABOLIC_PROCESS     | 61 | 0.391 | 1.447 | 0.049 | 0.138 |
| GO_REGULATION_OF_CELL_POPULATION_PROLIFERATION    | 70 | 0.377 | 1.432 | 0.046 | 0.132 |

