DH 607 Introduction to computational multi-omics

Course Project Proposal

Submission Deadline (hard, no late submissions): 1st October 2024, 11:59 PM (via Gradescope)

Total Word Limit: 1000 words

Note: *Proposals will be evaluated based on clarity, feasibility, alignment with course goals, and originality. Please ensure that your proposal is focused and well-structured. If you are working in groups (>1 member) only one member needs to upload the final proposal.*

Project Title

Cross-Tissue Biomarkers for Hepatitis B Virus-Related Hepatocellular Carcinoma: A Comparative Study of Blood and Liver Transcriptomics

---

Broad theme area

'Re-analysis of published datasets'

Group Members

Member 1

- **Name: Shubham Thakur**
- **Roll #: 24D1622**
- **Email: 24d1622@iitb.ac.in**
- **Department/Program: KCDH/PhD**

Member 2

- **Name: Fenil Parmar**
- **Roll #:24D1624**
- **Email:** 24d1624@iitb.ac.in

- **Department/Program: KCDH/PhD**

Member 3

- **Name: Mayurakshi Mukherji**
- **Roll #: 23D1628**
- **Email:** 23d1628@iitb.ac.in
- **Department/Program:KCDH/PhD**

Member Defacto

- **Name:** Saket Choudhary
- **Roll #:**
- **Email:** saketc@iitb.ac.in
- **Department/Program:** KCDH

Project Overview (150-200 words)

Provide a brief summary of your proposed project, outlining the central idea, objective, and significance of the project. Explain what specific genomic concept or tool your project will address.

The central focus of the project will be to study differential gene expression from both tissue and blood between healthy, Hepatitis B virus (HBV) infected, HBV-related Hepatocellular Carcinoma (HBV-HCC), and non-HBV-HCC samples using both bulk RNA-seq and single cell RNA-seq. HBV is a major cause of HCC worldwide. It is estimated that around 50-60% of HCC cases globally are attributed to chronic HBV infection, with this percentage being even higher in regions with high HBV prevalence, such as East Asia and sub-Saharan Africa. There are several other risk factors for HCC such as Cirrhosis, hepatitis C virus infection, high alcohol Consumption, non-alcoholic steatohepatitis, ingestion of the fungal metabolite alflatoxin B1 etc. The project's ultimate goal is to discover biomarkers detectable in blood that are indicative of liver disease, facilitating non-invasive diagnostic tools for early detection and monitoring of HBV-related HCC. Also, to identify biological pathways disrupted during HBV infection and HCC progression. Cross-omics data integration and machine learning models will be employed in future to correlate molecular alterations in the liver with those in circulation, with a special focus on identifying early progression markers.

Research Question and Objectives (200-250 words)

Clearly state the research question your project will explore. Define your project's objectives and explain how these will help in addressing the research question. Include any relevant hypotheses if applicable.

The research questions for our project are as follows –

1) *What are the key genes differentially expressed between HBV patients and HBV-HCC patients in both blood and liver samples?*
   Objective**:** To identify and characterise the specific genes whose expression is altered as HBV progresses into HCC using bulk RNA-seq data and single-cell RNA-seq data.
   Hypothesis**:** Specific sets of genes are consistently dysregulated during the progression of HBV to HCC, with these changes being observable in both blood and liver samples.

2) *How do the gene expression profiles of HBV-HCC differ from those of non-HBV-HCC in both blood and liver?*
   Objective**:** To compare the gene expression patterns between HBV-HCC and non-HBV-HCC, to potentially reveal HBV-specific cancer markers.
   Hypothesis**:** This will highlight biomarkers that are specific to HBV-driven HCC and differentiate it from HCC driven by other causes (e.g., HCV, alcohol). Such markers could be highly valuable in HBV screening programs.

3) *How gene expression profiles of HBV patients differ from healthy individuals in both blood and liver?*
   Objective: Detect early HBV infection markers before HCC develops.
   Rationale: By comparing healthy individuals with HBV-infected patients, you can identify early-stage infection biomarkers, especially in blood, which could be used for early detection and monitoring.

4) *Is there a correlation between the differentially expressed genes in blood and liver samples for HBV-HCC?*
   Objective: To investigate whether gene expression changes in the liver tissue are mirrored in the blood, thus exploring the feasibility of blood-based RNA-seq as a non-invasive screening method for HCC.
   Hypothesis: Key genes differentially expressed in the liver tissue during HCC progression are also detectable in blood samples, enabling non-invasive biomarker screening.

5) *How do epigenetic and proteomic changes validate the findings from differential gene expression analysis in HBV-HCC and non-HBV-HCC?*
   Objective: To use multi-omics approaches (epigenetics, proteomics, etc.) to validate the gene expression findings, confirming affected pathways. A multi-modal approach towards cross-tissue correlations for HBV-HCC biomarkers.
   Hypothesis: Epigenetic and proteomic alterations corroborate the gene expression changes, highlighting key regulatory mechanisms in HBV-related and HBV-unrelated HCC.

Methodology and Approach (250-300 words)

Describe the methods you plan to use to conduct the project. This could include any bioinformatics tools, algorithms, genomic datasets, or experimental techniques. Be specific about the steps involved and explain why the chosen methods are suitable for your research question. If you are developing a new method or benchmarking existing methods, provide your general approach that you will take to develop a new method or benchmark existing methods.

We will be using RNA-seq dataset from liver tissue and blood (both bulk and single cell) to understand differential gene expression in the following condition pairs:

1) Healthy vs. HBV
2) HBV vs. HBV-HCC
3) HBV-HCC vs. non-HBV-HCC
4) HBV-HCC (blood) vs. HBV-HCC (liver) and non-HBV-HCC (blood) vs. non-HBV-HCC (liver)

Steps for analysis:

1) Data collection:
- Obtain bulk RNA-seq and single-cell RNA-seq data (both liver and blood samples) from public repositories like , NCBI, GEO, TCGA, specifically looking for liver and blood transcriptomics datasets.

2) Data Preprocessing:
- Perform quality control (QC) on the RNA-seq data using FastQC.
- Use Trimmomatic to trim low-quality bases and remove adapter sequences from the RNA-seq data.
- Align reads to the reference genome using STAR.
- Quantify gene expression using featureCounts.

3) Differential Gene Expression (DGE) Analysis:
- Perform DGE analysis using DESeq2 for bulk RNA-seq and Scanpy for single-cell RNA-seq data.
- Identify DEGs between particular condition pairs in both blood and liver samples.

4) Cross-Group Comparison:
- Use clustering methods like hierarchical clustering or PCA to visually differentiate between various condition pairs.
- Employ Heatmaps and Volcano Plots to visualize DEGs specific to HBV-driven HCC.

5) Cross-Tissue Comparison:
- Perform DGE analysis separately for blood and liver samples.

- Use correlation analysis to explore the relationships between gene expression changes in blood and liver.

6) Biomarker Identification:
- Focus on HBV-specific cancer biomarkers for diagnostic and screening potential.
- Use machine learning techniques to identify top markers.

7) Functional enrichment analysis:
- Pathway Enrichment: Use DAVID or Reactome to identify pathways that are significantly enriched among DEGs and also to categorize DEGs into biological pathways and processes.
- Gene Set Enrichment (GESA) to detect up- or down-regulated pathways.
- Gene Ontology (GO) and KEGG pathway analysis to interpret biological significance.
- Cross-tissue Pathways: Compare pathways from blood and liver samples across the conditions to understand systemic (blood) vs localized (liver) effects.

8) Multiomics Integration (Epigenome, proteome):
- Methods like iCluster or MOFA+ to integrate multiomics data.
- These methods can help identify shared biological patterns and correlations among different omics layers.

Expected Outcomes and Significance (150-200 words)

Outline the anticipated outcomes of your project. Discuss how these outcomes will contribute to the field of basic genomics or provide insights into disease-related (applied) genomic research. Consider any broader implications or potential applications of your findings. In short, what excites you about the idea?

The best outcome for this work will be the following -

1) Identification of genes that are differentially expressed in HBV and HBV-HCC. This will be useful for screening HBV patients at risk of HCC and thus administer specialised treatment.
2) Identification of genes that are differentially expressed in HBV-HCC and non-HBV-HCC. This will provide an understanding of the differential role of HBV in HCC and thus compare the affected pathways.
3) Identification of genetic markers in liver vs. blood to understand if there is a correlation among the differentially expressed genes in blood and liver. If the results from blood and liver validate each other, blood-based RNA-seq will allow a non-invasive screening for HCC.
4) Use of other modalities such as epigenetics and proteomics to validate the results of the differential gene expression and the affected pathways to emphasise cross-tissue biomarkers.

References

Cite any key references or preliminary research you have consulted in preparing this proposal. Use a standard citation format (e.g., APA, MLA, or any scientific style).

Baudi, I., Inoue, T., Tanaka, Y. (2020). Novel Biomarkers of Hepatitis B and Hepatocellular Carcinoma: Clinical Significance of HBcrAg and M2BPGi. International Journal of Molecular Sciences, 21(3), 949. https://doi.org/10.3390/ijms21030949

Hayashi, S., Nagaoka, K., & Tanaka, Y. (2021). Blood-Based Biomarkers in Hepatitis B Virus-Related Hepatocellular Carcinoma, Including the Viral Genome and Glycosylated Proteins. International journal of molecular sciences, 22(20), 11051. https://doi.org/10.3390/ijms222011051

Ho, D. W., Tsui, Y. M., Chan, L. K., Sze, K. M., Zhang, X., Cheu, J. W., Chiu, Y. T., Lee, J. M., Chan, A. C., Cheung, E. T., Yau, D. T., Chia, N. H., Lo, I. L., Sham, P. C., Cheung, T. T., Wong, C. C., & Ng, I. O. (2021). Single-cell RNA sequencing shows the immunosuppressive landscape and tumor heterogeneity of HBV-associated hepatocellular carcinoma. Nature communications, 12(1), 3684. https://doi.org/10.1038/s41467-021-24010-1

Jiang, Y., Han, Q., Zhao, H., & Zhang, J. (2021). The Mechanisms of HBV-Induced Hepatocellular Carcinoma. Journal of hepatocellular carcinoma, 8, 435–450. https://doi.org/10.2147/JHC.S307962

Liu, Y., Al-Adra, D. P., Lan, R., Jung, G., Li, H., Yeh, M. M., & Liu, Y. Z. (2022). RNA sequencing analysis of hepatocellular carcinoma identified oxidative phosphorylation as a major pathologic feature. Hepatology communications, 6(8), 2170–2181. https://doi.org/10.1002/hep4.1945

Paslaru, L., Bindea, G., Nastase, A., Sorop, A., Zimbru, C., Herlea, V., Hrehoret, D., Brasoveanu, V., Zamfir, R., Dima, S., & Popescu, I. (2022). Comparative RNA-Sequencing Analysis Reveals High Complexity and Heterogeneity of Transcriptomic and Immune Profiles in Hepatocellular Carcinoma Tumors of Viral (HBV, HCV) and Non-Viral Etiology. Medicina (Kaunas, Lithuania), 58(12), 1803. https://doi.org/10.3390/medicina58121803

Rizzo, G. E. M., Cabibbo, G., & Craxì, A. (2022). Hepatitis B Virus-Associated Hepatocellular Carcinoma. Viruses, 14(5), 986. https://doi.org/10.3390/v14050986

Yeh, S. H., Li, C. L., Lin, Y. Y., Ho, M. C., Wang, Y. C., Tseng, S. T., & Chen, P. J. (2023). Hepatitis B Virus DNA Integration Drives Carcinogenesis and Provides a New Biomarker for HBV-related HCC. Cellular and molecular gastroenterology and hepatology, 15(4), 921–929. https://doi.org/10.1016/j.jcmgh.2023.01.001

Zheng, Q., Sun, Q., Yao, H., Shi, R., Wang, C., Ma, Z., Xu, H., Zhou, G., Cheng, Z., & Xia, H. (2024). Single-cell landscape identifies the immunophenotypes and microenvironments of

HBV-positive and HBV-negative liver cancer. Hepatology communications, 8(2), e0364. https://doi.org/10.1097/HC9.0000000000000364