Bioinformatics Summer School Report

Name: Aljori Ahmed Alsheraimi

Associated organization: University of Cambridge

Supervisor's name: N/A

Project title: Gene mapping of proteins associated with Recessive Dystrophic Epidermolysis Bullosa in the context of Cutaneous Squamous Cell Carcinoma

1. Introduction

Epidermolysis Bullosa (EB) is a rare genetic condition that causes wounds and blisters due to the fragility of the skin. It is classified into four major categories: simplex, junctional, dystrophic, and Kindler syndrome. The recessive dystrophic form is considered severe and carries a risk of life-threatening complications. Cutaneous squamous cell carcinoma is the leading cause of death in patients with (RDEB), the persistent inflammation, along with the non-healing wounds create an environment for tumor development. Although clinically significant, it remains poorly studied on a proteomic level, leaving gaps in understanding the underlying pathways.

2. Objectives

The aim of this analysis is to highlight the key proteins that play a role in RDEB associated with cSCC and map them to their corresponding genes, in order to answer the question of whether these proteins contribute to cancer initiation or mainly support the tumor progression, by integrating proteomics with single-cell RNA sequencing to identify potential biomarkers.

3. Methodology and approach

1) Data collection and Processing

- Proteomics data was obtained from Project PXD006914 (PRIDE repository)
- Single-cell RNA data was accessed through GSE181227 (GEO database)
- Applied QC, normalization, and filtered the data to ensure reliability for further analysis

2) Differential Expression Analysis

- Performed DE where the most 50 altered proteins were chosen based on log-fold change.

3) Protein-to-Gene Mapping

- Linked the identified proteins to their genes from the GSE181227 dataset.

4) Data Visualisation and Exploration

- Top proteins were plotted using a heatmap to examine expression patterns, variation among the control and target samples, and their possible biological relevance.

4. Findings and Discussion

By cross-referencing proteins with their associated genes, and investigating the gene expression across the single-cell RNA dataset, which included samples from RDEB with native, wound, non-wound skin, the analysis emphasises potential molecular indicators that could shed light on the relationship between RDEB and cSCC.

5. Challenges and obstacles encountered

I encountered a few limitations which required me to be flexible in my approach. I had one week to work on the project, and since I chose not to use the provided preprocessed datasets and instead source and curate my own, more than half of the week was spent on data preparation, taking into account the amount of (NA) values in the proteomics dataset due to Label-free quantification (LFQ). Regardless, I'm glad I was able to carry on with the analysis successfully.

6. Future Work

What to do next would be to validate the analysis through laboratory experiments, to determine the functional roles of candidate biomarkers.

7. Acknowledgments

I would like to express my gratitude to each member of Kaust Academy for making this opportunity possible, and to the University of Cambridge, particularly the Department of Genetics, for all the guidance and support they provided throughout this journey.

8. Tools

- **Bash**
- **RStudio**

9. Source code

- **https://github.com/jori-as/Kaust-Camabridge-Bio-Project**

10. References

- EMBL-EBI, “PRIDE - PRoteomics IDEntifications Database,” *Ebi.ac.uk*, 2025. https://www.ebi.ac.uk/pride/archive/projects/PXD006914 (accessed Aug. 27, 2025).

- “GEO Accession viewer,” *Nih.gov*, 2024. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181227 (accessed Aug. 27, 2025).