CAP 5510 – Project Proposal

Sleep Deprivation and Its Impact on Human Gene Expression: A Transcriptomic Analysis

Team Member: Neha Eshwaragari

Abstract

Sleep is one of the most important biological processes since it is necessary for preserving long-term health, immunological response, and cognitive function. The biological effects of the widespread sleep deprivation brought on by modern lifestyles, however, are poorly understood.My goal in this project is to use a publicly available blood transcriptome dataset (**GSE98566**) to investigate the molecular signatures of sleep deprivation. I will use differential gene expression (DGE) analysis and functional enrichment techniques to pinpoint particular genes and pathways that are changed when sleep deprivation occurs as opposed to regular sleep. The immunological, stress-response, and circadian pathways—all of which are probably impacted by sleep disturbances—will receive particular attention. The expected outcome of this work is a comprehensive set of candidate biomarkers and pathways that may explain how disrupted sleep contributes to disease susceptibility and altered physiological states.

Methodology

The proposed methodology is structured as a bioinformatics pipeline, combining data preprocessing, statistical modeling, pathway enrichment, and visualization:

1. **Dataset Retrieval and Preprocessing:** I will acquire the transcriptomic dataset **GSE98566** from the NCBI Gene Expression Omnibus (GEO).This dataset includes the genome-wide expression profiles of individuals who sleep and those who don't. Raw data will undergo a quality assessment to identify batch effects, missing values, and outliers.To ensure that expression values are comparable across samples, normalization will be done using variance-stabilizing transformations.
2. **Differential Gene Expression (DGE) Analysis:** Two popular techniques will be used for DGE: `DESeq2`, which applies negative binomial generalized linear modeling, and `limma`, which applies linear modeling with empirical Bayes shrinkage.If a gene exhibits significant $\log_2$ fold changes and adjusted p-values below a predetermined threshold (e.g., FDR ¡ 0.05), it will be considered differentially expressed.This two-pronged strategy will improve robustness and enable cross-validation of results.
3. **Pathway and Functional Enrichment:** The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) categories will be used to map genes that have been found to be significantly up- or down-regulated to biological pathways.The focus of the analysis will be on processes that are known to be impacted by irregular sleep cycles, including immune response, inflammatory signaling, circadian regulation, and metabolic pathways.
4. **Comparative Analysis of Methods:** To assess overlap and divergence in identified gene sets, I will compare the outcomes from `DESeq2` and `limma`.To measure agreement, Venn diagrams and Jaccard similarity coefficients will be employed.This step will demonstrate how methodological decisions can affect how transcriptomic data are interpreted.
5. **Visualization and Interpretation:** It is necessary to visualize transcriptomic changes.I'll make heatmaps to group the most highly differentially expressed genes, volcano plots to display significant fold-changes, and enrichment barplots to give an overview of the pathway analysis.These figures taken as a whole will form the basis for interpretation, highlighting the biological mechanisms that sleep deprivation alters over time.

Plan of Action

The research plan will be divided into stages that align with the methodology:

1. **First: Data Acquisition and Preprocessing** — Download GSE98566, perform quality control, normalize datasets.
2. **Second: Differential Gene Expression Analysis** — Implement `DESeq2` and `limma`, identify significant genes.
3. **Third: Pathway Enrichment and Comparative Study** — Perform GO/KEGG enrichment, compare DESeq2 vs. limma results.
4. **Fourth: Visualization and Reporting** — Generate figures (heatmaps, volcano plots, enrichment charts) and draft preliminary results.

Workload Statement

Since I am completing this project individually, I will be responsible for all tasks, including data preprocessing, differential expression analysis, enrichment studies, and result visualization.

Expected Outcomes

This project is expected to produce:

- A list of differentially expressed genes (DEGs) associated with sleep deprivation.
- Pathway enrichment results highlighting disrupted biological functions.
- Comparative insights into differences between analysis methods.
- Visualizations that effectively communicate transcriptomic alterations.
- Candidate biomarkers that may be further validated in future studies on sleep and disease risk.

Preliminary Literature

- Möller-Levet et al. (2013). *Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome.* Proceedings of the National Academy of Sciences.
- Archer et al. (2014). *Sleep deprivation influences on immune-related gene expression. Sleep.*
- GEO Dataset Documentation: GSE98566.