

Master thesis project plan

Title: The housekeeping proteome: identifying proteins essential for human cells

Student

Thi Huyen Mai Nguyen (maihuyen.nha@gmail.com)

Primary supervisor

Dr Evelina Sjöstedt (evelina.sjostedt@scilifelab.se)

Co-supervisors

Mathias Uhlen (mathias.uhlen@scilifelab.se)

Åsa Sivertsson (asa.sivertsson@scilifelab.se)

Maria Bueno Alvez (maria.bueno@scilifelab.se)

Starting date: January 21

Background

Housekeeping proteins are a well-known concept commonly defined as essential for cell survival (1, 2) and consistently expressed across tissues, cell types or context (1). However, these criteria have been challenged by several research groups (3, 4), highlighting the need for an update on the definition and the list of human housekeeping proteins. The research is critical due to the function of housekeeping proteins as normalization references (for instance, in qualitative Western blotting, (5), a better overview of their function in cell biology or their impact on human disease. One such example is GAPDH, a well-known housekeeping protein involved in glycolysis. Beyond metabolism, GAPDH plays a role in cell death and has been linked to cancer progression, making it a potential target in tumor biology (6).

Project goal

To characterize the housekeeping proteome, a subset of proteins essential for human cell survival and function, as part of the Alpha Cell program by the Human Protein Atlas. The aim is to define the proteomic landscape of proteins ubiquitously expressed across human cells, while distinguishing essential proteins from those that are ubiquitously expressed but non-essential due to functional redundancy (i.e., the cell can survive their knockout).

Aims, objectives and methods

The project consists of four main phases:

1. Defining criteria for housekeeping proteins.

Establish a robust framework to identify housekeeping proteins, focusing on proteins expressed across all human cells. This phase includes literature review and data analysis to define different strategies and rules used for definition.

2. Identification of essential proteins.

The main analysis will focus on data from the Human Protein Atlas (HPA) Alpha Cell program to identify proteins that meet the criteria for being housekeeping. Additionally, extracted data from the DepMap portal will help define the essential proteins, including addressing cases of functional redundancy where knockout (KO) of a protein does not affect survival due to compensation by other proteins.

3. **Validation and refinement of the housekeeping proteome.**
Develop and apply filtering strategies to validate the list of housekeeping proteins. This involves analyzing data robustness and testing the hypothesis across diverse datasets to confirm essentiality and ubiquity, explore different criteria and point of views.
4. **Integration and communication of findings.**
Visualize and present the finalized housekeeping proteome in an accessible format, ensuring the results can be utilized effectively by users of the HPA resource. Create clear documentation to distinguish between truly essential proteins and ubiquitously expressed, non-essential proteins based of different cut off criteria and functional restrictions.

Time plan

- **Project planning:** Now – end of January
- **Phase 1 – Literature review and criteria definition:** February
- **Phase 2 – Identification of gene lists and essential proteins:** February – March
- **Phase 3 – Validation and refinement:** March – April
- **Phase 4 – Visualization and reporting:** April – May
- **Writing:** May

Tools

- R and RStudio
- Human Protein Atlas (HPA) resource
- UniProt and related protein databases

Reference

- (1) Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends in Genetics [Internet]. 2013 Jun 27;29(10):569–74. Available from: <https://pubmed.ncbi.nlm.nih.gov/23810203/>
- (2) Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiological Genomics [Internet]. 2000 Apr 27;2(3):143–7. Available from: <https://journals.physiology.org/doi/epdf/10.1152/physiolgenomics.2000.2.3.143>
- (3) Joshi CJ, Ke W, Drangowska-Way A, O'Rourke EJ, Lewis NE. What are housekeeping genes? PLoS Computational Biology [Internet]. 2022 Jul 13;18(7):e1010295. Available from: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010295>
- (4) Hounkpe BW, Chenou F, De Lima F, De Paula EV. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Research [Internet]. 2020 Jul 9;49(D1):D947–55. Available from: <https://academic.oup.com/nar/article/49/D1/D947/5871367>
- (5) Lee H, Jo J, Hong H, Kim KK, Park J, Cho S, et al. State-of-the-art housekeeping proteins for quantitative western blotting: Revisiting the first draft of the human proteome. PROTEOMICS [Internet]. 2016 Apr 30;16(13):1863–7. Available from: <https://pure.ewha.ac.kr/en/publications/state-of-the-art-housekeeping-proteins-for-quantitative-western-b>
- (6) Colell A, Green DR, Ricci JE. Novel roles for GAPDH in cell death and carcinogenesis. Cell Death and Differentiation [Internet]. 2009 Sep 25;16(12):1573–81. Available from: <https://www.nature.com/articles/cdd2009137>