

PEELing: an integrated and user-centric platform for cell-surface proteomics analysis

Xi Peng¹, Jody Clements¹, Zuzhi Jiang^{1,2}, Stephan Preibisch¹, Jiefu Li^{1,*}

¹Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA

²Yuanpei College, Peking University, Beijing 100871, China

*To whom correspondence should be addressed: lij6@janelia.hhmi.org

Abstract

Proteins localized at the cellular interface mediate cell-cell communication and thus control many aspects of physiology in multicellular organisms. Cell-surface proteomics allows biologists to comprehensively identify proteins on the cell surface and survey their dynamics in physiological and pathological conditions. PEELing provides an integrated package and user-centric web service for analyzing cell-surface proteomics data. With a streamlined and automated workflow, PEELing evaluates data quality using curated references, performs cutoff analysis to remove contaminants, connects to databases for functional annotation, and generates data visualizations. Together with chemical and transgenic tools, PEELing completes a pipeline making cell-surface proteomics analysis handy for every lab.

Main

The human body is made up of about 37 trillion cells¹, while the alga *Tetrrabaena socialis* has only 4 cells². Multicellular organisms, no matter how complex or simple, rely on coordinated behavior of individual cells to carry out physiological functions for survival and reproduction. Cell-surface proteins mediate communication and interaction between cells and often play core roles in the development, physiology, and pathogenesis of multicellular organisms. Chemical biology methods and genetic tools have enabled biologists to profile the cell-surface proteome in cultured cells³⁻⁹ and native tissues¹⁰⁻¹², providing powerful ways to examine the composition and dynamics of cell-surface proteomes in diverse biological contexts. However, evaluating data quality and processing data can be complex, especially for users with limited proteomics experience or programming skills. To address this, we present PEELing, a user-centric platform for cell-surface proteomics analysis that integrates data quality checks, contaminant removal, functional annotation, and visualization into a streamlined workflow.

Ratiometric ranking

In a cell-surface proteomics experiment, cell-surface proteins are often chemically labelled and isolated (“labelled” groups, hereafter) using methods such as cell-surface-restricted biotinylation and streptavidin-based protein isolation. Non-labelled control groups (“control”) are included to capture contaminants such as endogenously biotinylated proteins and non-specific binders to isolation reagents. PEELing uses a ratiometric strategy as previously described^{10,13} in which the labelled-to-control ratio of each protein reflects whether this protein is cell-surface enriched or not. A *bona fide* cell-surface protein should exhibit a high ratio because it should be enriched in the labelled group relative to the control group. A contaminant should have a low ratio since it should be captured similarly in both labelled and control groups. PEELing ranks all detected proteins in descending order based on the labelled-to-control ratio, resulting in high-ranking of *bona fide* cell-surface proteins. This labelled-to-control ratio is a more reliable indicator of cell-surface enrichment than protein abundance of the labelled group because a contaminant could be abundant while a cell-surface protein may have a low expression level. As long as the control group captures the contaminant but not the cell-surface protein, these proteins will be ranked correctly by the labelled-to-control ratio instead of being ranked reversely by abundance.

As described in the tutorial online, labelled-to-control ratios should be uploaded along with protein UniProt accession numbers for PEELing analysis. These ratios can be derived from any mass spectrometry quantification method such as SILAC, TMT, iTRAQ, or label-free. In cases that such a labelled-to-control ratio is not available, protein abundance can be used as a substitute although it is not recommended for the reason discussed above.

Data quality checks

PEELing performs pairwise correlation analysis on labelled-to-control ratios to check whether biological replicates show consistency (**Fig. 1a**). To test whether cell-surface proteins are enriched, PEELing scans through all detected proteins and marks curated cell-surface proteins (true positives, TPs) and intracellular contaminants (false positives, FPs). The reference for TPs includes secreted and membrane-anchored proteins (**Supplementary Data 1**). The FP reference includes nuclear, cytosolic, and mitochondrial proteins that do not localize on the cell surface (**Supplementary Data 2**). As illustrated in **Fig. 1b**, true positive rate (TPR), false positive rate (FPR), and their difference (TPR–FPR) are calculated and plotted against the ratio-based ranking. In a successful cell-surface proteomics experiment, TPR (blue in **Fig. 1b**) increases quickly while FPR (orange in **Fig. 1b**) rises slowly, leading to a single peak of TPR–FPR (green in **Fig. 1b**) and a receiver operating characteristic (ROC) curve bending towards the left-upper corner (**Fig. 1c**).

If the TPR–FPR value fluctuates up and down without forming a single peak or the ROC curve follows the diagonal line without bending towards the left-upper corner, it suggests suboptimal or failed cell-surface proteome capturing. This could be due to an abundance of

intracellular contaminants being enriched. In such cases, it is not recommended to use PEELing for further analysis. Instead, improved sample preparation or other filtering methods should be considered to address the issue.

Cutoff analysis

For each labelled-to-control ratio, PEELing finds the ranking position where TPR–FPR is maximal, as indicated by the peak of the green line in **Fig. 1b** and the red dot in **Fig. 1c**, and retains proteins ranked above this position. The "TPR–FPR maximum" cutoff provides two key benefits: 1) The cutoff position is determined by data quality rather than an arbitrary value. If the data is of high quality with sparse false positives, TPR–FPR will peak later in the ranking, resulting in the retention of more proteins. If the data is heavily contaminated with false positives, TPR–FPR will peak earlier, leading to the retention of fewer proteins. 2) Any protein ranked above the cutoff position is retained, regardless of how it is annotated by a database. Therefore, missing annotations or occasional incorrect annotations in the databases would not impact the analysis, as long as they are largely accurate and comprehensive. Additionally, in certain physiological or pathological contexts, intracellular proteins can be transported to the cell surface. PEELing retains these proteins if they are ranked highly, potentially enabling researchers to discover novel biomarkers and cellular processes.

PEELing conducts cutoff analysis on all labelled-to-control ratios individually and, for the final proteome, retains only those proteins that pass the cutoff of all or multiple ratios, which further eliminates contaminants. The "Tolerance" setting is optional and enables users to control the stringency of the cutoff. By default, it is set to 0, meaning that a protein must pass the cutoff of all ratios to be included in the final proteome. If Tolerance is set to n , a protein can fail the cutoff in up to n ratios and still be included in the final proteome.

Proteome and annotation

As shown in **Fig. 1d**, PEELing displays the post-cutoff proteome and provides basic information on the top 100 most enriched proteins for each labelled-to-control ratio. Each UniProt accession number is a clickable link to the corresponding UniProt protein page.

PEELing submits the post-cutoff proteome to the PANTHER server^{14,15} for over-representation analyses on protein localization (PANTHER GO slim cellular component), function (PANTHER GO slim biological process), and pathway (Reactome). The results show the top 10 terms based on the false discovery rate, as listed in **Fig. 1e**.

Discussion

PEELing offers a user-friendly and automated solution for the analysis of cell-surface proteomics data, designed to be accessible to biologists from all expertise levels. For advanced users, the source code is available for customization or repurposing. Notably, users can customize the TP and FP references and thus extend PEELing for analyzing any spatially-resolved proteomics data; thus, it is not limited to cell-surface proteomics. For example, with a simple change of the TP and FP reference to curated nuclear and non-nuclear proteins, respectively, the program is ready to handle nucleus-targeting proteomics data and provides data quality checks, cutoff analysis, and functional annotation, making PEELing a versatile tool for spatially-resolved proteomics research.

Availability

PEELing is available at <https://peeling.janelia.org/>. A Python package of PEELing is available at <https://pypi.org/project/peeling/> and <https://github.com/JaneliaSciComp/peeling/>. For technical support, contact: peeling@janelia.hhmi.org.

References

1. Bianconi, E. *et al.* An estimation of the number of cells in the human body. *Ann Hum Biol* **40**, 463–471 (2013).
2. Arakaki, Y. *et al.* The Simplest Integrated Multicellular Organism Unveiled. *PLoS One* **8**, e81641 (2013).
3. Jang, J. H. & Hanash, S. Profiling of the cell surface proteome. *Proteomics* **3**, 1947–1954 (2003).
4. Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology* *2009 27:4* **27**, 378–386 (2009).
5. Gundry, R. L. *et al.* A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Mol Cell Proteomics* **11**, 303–316 (2012).
6. Bausch-Fluck, D. *et al.* A Mass Spectrometric-Derived Cell Surface Protein Atlas. *PLoS One* **10**, e0121314 (2015).
7. Loh, K. H. *et al.* Proteomic Analysis of Unbounded Cellular Compartments: Synaptic Clefts. *Cell* **166**, 1295-1307.e21 (2016).
8. Geri, J. B. *et al.* Microenvironment mapping via Dexter energy transfer on immune cells. *Science* (1979) **367**, 1091–1097 (2020).
9. Kirkemo, L. L. *et al.* Cell-surface tethered promiscuous biotinylators enable comparative small-scale surface proteomic analysis of human extracellular vesicles and cells. *Elife* **11**, e73982 (2022).
10. Li, J. *et al.* Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators. *Cell* **180**, 373-386.e15 (2020).

11. Xie, Q. *et al.* Transcription factor Acj6 controls dendrite targeting via a combinatorial cell-surface code. *Neuron* **110**, 2299-2314.e8 (2022).
12. Shuster, S. A. *et al.* In situ cell-type-specific cell-surface proteomic profiling in mice. *Neuron* **110**, 3882-3896.e9 (2022).
13. Hung, V. *et al.* Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Mol Cell* **55**, 332–341 (2014).
14. Mi, H. *et al.* Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols* 2019 14:3 **14**, 703–721 (2019).
15. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* **31**, 8–22 (2022).

Acknowledgements

We thank members of the Scientific Computing Software team, the Open Science Software Initiative, and the Li laboratory at the Janelia Research Campus for support and advice on this project; L. Ding, M. Liu, L. Luo, C. McLaughlin, Y. Peng, M. Wang, Y. Wang, and Z. Zhang for feedback and comments on PEELing; and B. Smith for administrative assistance.

Funding

This work was supported by the Howard Hughes Medical Institute.

Author information

Contributions

X.P. and J.C. conducted code development and implementation. Z.J. benchmarked the program and web service. S.P. and J.L. supervised the project. J.L. wrote the manuscript with input from all authors.

Corresponding author

Correspondence to lij6@janelia.hhmi.org (J.L.).

Ethics declarations

Competing interests

The authors declare no competing interests.

Figure

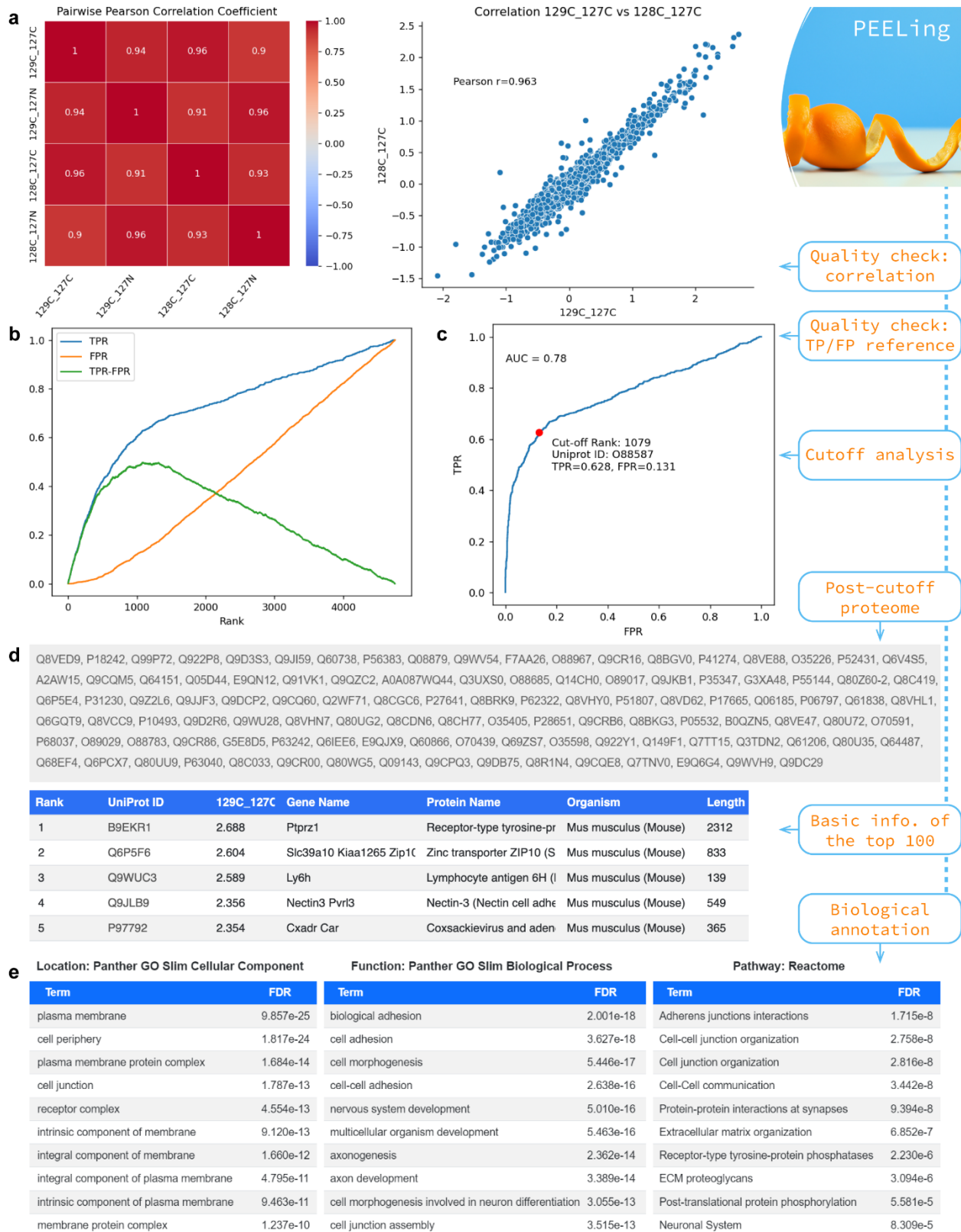


Fig. 1. PEELing workflow and example analyses using a published dataset¹²: the cell-surface proteome of mouse Purkinje cells at postnatal day 15 (**Supplementary Data 3**). **(a)** Correlation plots and coefficients. **(b)** True positive rate (TPR), false positive rate (FPR), and their difference (TPR–FPR) plotted against the ratio-based ranking (x-axis). **(c)** Receiver operating characteristic (ROC) curve. Red dot, cutoff position. AUC, area under the ROC curve. **(d)** Post-cutoff proteome and the top 100 most enriched proteins based on the labelled-to-control ratio. **(e)** PANTHER over-representation analyses for protein localization, function, and pathway. PEELing is named after the iPEEL method¹² for cell-surface proteome extraction by extracellular labeling. The PEELing icon was produced by DALL-E of OpenAI.

Supplementary File

Supplementary Data 1. The true positive (TP) reference, as of April 6, 2023. This list is specified by the UniProt term: *((cc_scl_term:SL-0112) OR (cc_scl_term:SL-0243) OR (keyword:KW-0732) OR (cc_scl_term:SL-9906) OR (cc_scl_term:SL-9907)) AND (reviewed:true)*, which includes SwissProt-reviewed extracellular (SL-0112), secreted (SL-0243), signal peptide-containing (KW-0732), type II transmembrane (SL-9906), or type III transmembrane (SL-9907) proteins. PEELing automatically updates this list from the UniProt database, ensuring that the TP reference remains current and accurate.

Supplementary Data 2. The false positive (FP) reference, as of April 6, 2023. This list is specified by the UniProt term: *((cc_scl_term:SL-0091) OR (cc_scl_term:SL-0173) OR (cc_scl_term:SL-0191)) AND (reviewed:true) NOT (((cc_scl_term:SL-0112) OR (cc_scl_term:SL-0243) OR (keyword:KW-0732) OR (cc_scl_term:SL-9906) OR (cc_scl_term:SL-9907)) AND (reviewed:true))*, including SwissProt-reviewed cytosolic (SL-0091), mitochondrial (SL-0173), and nuclear (SL-0191) proteins that do not express on the cell surface. Some cell-surface proteins, such as the Notch family proteins, are also localized in intracellular compartments and are not considered false positives, and are thus removed from the FP reference. PEELing automatically updates this list from the UniProt database, ensuring that the FP reference remains current and accurate.

Supplementary Data 3. Example data from a published study¹². The authors profiled the cell-surface proteome of mouse Purkinje cells at postnatal day 15, using 4 TMT channels: 2 for labelled replicates (129C and 128C) and 2 for non-labelled controls (127C and 127N). Ratios are log₂-transformed.