PEELing: an integrated and user-centric platform for spatially-resolved proteomics data analysis

Xi Peng[1], Jody Clements[1], Zuzhi Jiang[1,2], Shuo Han[3,*], Stephan Preibisch[1], Jiefu Li[1,*]

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

[2]Yuanpei College, Peking University, Beijing 100871, China

[3]Key Laboratory of RNA Science and Engineering, Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

*Corresponding authors: shuohan@sibcb.ac.cn, lij6@janelia.hhmi.org

Abstract

Summary

Molecular compartmentalization is vital for cellular physiology. Spatially-resolved proteomics allows biologists to survey protein composition and dynamics with subcellular resolution. Here we present PEELing, an integrated package and user-friendly web service for analyzing spatially-resolved proteomics data. PEELing assesses data quality using curated or user-defined references, performs cutoff analysis to remove contaminants, connects to databases for functional annotation, and generates data visualizations—providing a streamlined and reproducible workflow to explore spatially-resolved proteomics data.

Availability and Implementation

PEELing and its tutorial are publicly available at https://peeling.janelia.org/. A Python package of PEELing is available at https://github.com/JaneliaSciComp/peeling/.

Contact

Technical support for PEELing: peeling@janelia.hhmi.org.

Introduction

Localization and function of proteins are always coupled. For instance, proteins for intercellular adhesion and communication are localized to the cell surface while many energy-producing enzymes stay in the mitochondrion. High-resolution, proteome-wide mapping of protein localization is of core importance for understanding cellular organization and processes. Emerging technologies for spatially-resolved proteomics, particularly by proximity labeling, make this possible and broadly applicable in cell biology (Roux *et al.*, 2012; Rhee *et al.*, 2013; Branon *et al.*, 2018; Loh *et al.*, 2016; Geri *et al.*, 2020; Han *et al.*, 2018; Qin *et al.*, 2021). Like other enrichment-based profiling methods, spatially-resolved proteomics can be interfered by ineffective enrichment, non-specific contamination, and other factors. Rigorous assessment of data quality and proper data processing is crucial for interpreting the results and for designing subsequent studies. However, this can be complex and overwhelming, particularly for biologists with limited proteomics experience or systems biology background. To address this, we built PEELing (*p*roteome *e*xtraction from *e*nzymatic *l*abel*ing* data), a platform integrating data quality checks, contaminant removal, functional annotation, and visualization into an automated workflow (**Figure 1a**).

Implementation, Benchmarks, and Results

In **Figure 1**, we used a published cell-surface proteome of mouse developing Purkinje cells (Shuster *et al.*, 2022) (**Figure 1b** and **Supplementary File 1**) to demonstrate the functionalities of PEELing. In this study (Shuster *et al.*, 2022), cell-surface proteins were chemically labelled and isolated ("labelled" groups, hereafter) using horseradish peroxidase (HRP) mediated cell-surface biotinylation (Loh *et al.*, 2016; Li *et al.*, 2020; Shuster *et al.*, 2022). Non-labelled control groups ("control") were included to capture contaminants such as endogenously biotinylated proteins and non-specific binders to isolation reagents. PEELing uses a ratiometric strategy as previously described (Hung *et al.*, 2014) 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.

To assess data quality, PEELing started with pairwise correlation analysis to check whether biological replicates exhibited consistency (**Figure 1c**). To determine whether cell-surface proteins were enriched, PEELing ranked all detected proteins in descending order based on the labelled-to-control ratio and scanned through them to mark curated cell-surface proteins (true positives, TPs; **Supplementary File 10**) and intracellular contaminants (false positives, FPs; **Supplementary File 12**). As shown in **Figure 1d**, true positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) were calculated and plotted against ratio-based ranking. TPR

increased quickly while FPR rose slowly, which led to a single peak of TPR–FPR, revealing a high ranking and an effective enrichment of cell-surface proteins. This is further illustrated by a receiver operating characteristic (ROC) curve bending towards the left-upper corner and exhibiting an area under the curve (AUC) of 0.79 (**Figure 1e**).

To test whether PEELing detects data of poor quality, we generated a pseudo dataset with random values mimicking a failed enrichment (**Supplementary File 2** and **Supplementary Figure 1**). TPR, FPR, and ROC curves all followed the diagonal line (**Supplementary Figure 1b,c,e,f**), showing a complete mix of cell-surface and intracellular proteins without any enrichment (**Supplementary Figure 1d,g**). We then tested reference specificity by analyzing the cell-surface proteome (Shuster *et al.*, 2022) (**Supplementary File 1**) with nuclear and mitochondrial references (**Supplementary File 14–17**). No enrichment of proteins of unintended cellular compartments (**Supplementary Figure 2**) showed the quality of the example data as well as the necessity of data-reference matching in PEELing analysis.

All analyses above used the labelled-to-control ratio as the enrichment indicator. Protein abundance is widely used in proteomic data quantification; however, it is not a reliable enrichment indicator in spatially-resolved profiling compared with the labelled-to-control ratio (**Supplementary Figure 3**, compared with **Figure 1**). From the same cell-surface proteome dataset (Shuster *et al.*, 2022), we obtained protein abundance values of the labelled groups (**Supplementary File 3**). Despite comparable correlations based on abundance (**Supplementary Figure 3a**) and labelled-to-control ratio (**Figure 1c**), cell-surface protein enrichment was not observed when ranking the proteome by abundance (**Supplementary Figure 3b,c,e,f**). Intracellular proteins were highly enriched instead (**Supplementary Figure 3d,g**). Notably, a contaminant (e.g., a non-specific binder to the isolation reagent) can 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. Therefore, labelled-to-control ratio is the preferred input data for PEELing.

Following data quality checks, PEELing performed cutoff analysis to remove contaminants. For each labelled-to-control ratio, PEELing found the ranking position where TPR–FPR was maximal, as indicated by the peak of the green line in **Figure 1d** and the red dot in **Figure 1e**, and retained 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, allowing for unbiased assessment of the proteomic results. 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 databases will not impact the analysis, as long as they are largely accurate and comprehensive. As illustrated in **Supplementary Figure 4**, 10-fold reduction of reference coverage, in either TP/FP or both, did not impair the analysis of the cell-surface proteome of Purkinje cells (compared with **Figure 1**). Additionally, the *Drosophila* proteome has less complete annotations than those of mouse and human, leading to jagged TPR, FPR, and ROC curves (**Supplementary Figure 5c,d**). Nevertheless, it did not interfere with the analysis of a *Drosophila* neuronal surface proteome (Li *et al.*, 2020) (**Supplementary Figure 5** and **Supplementary File 4**). Importantly, in certain physiological or pathological contexts, intracellular proteins can be transported to the cell surface—for instance, RNA helicase U5 snRNP200 in acute myeloid leukemia (Knorr *et al.*, 2023). PEELing retains these proteins if they are ranked highly, potentially enabling researchers to discover novel biomarkers and cellular processes. Despite its robustness to varied reference coverages, we note that PEELing relies on high-quality references (see Supplementary Materials – Method Details on how to create references). When references are not possible to obtain, commonly used statistical analysis is an alternative approach.

PEELing conducted cutoff analysis on all submitted labelled-to-control ratios individually and, for the final proteome, retained only those proteins that passed the cutoff of all ratios, which further removed contaminants. As shown in **Figure 1f**, PEELing displayed the post-cutoff proteome and provided information of the top 100 most enriched proteins for each labelled-to-control ratio (**Figure 1g**). On the website, each UniProt accession number is a clickable link to the corresponding UniProt protein page (Consortium *et al.*, 2023). PEELing then transmitted the post-cutoff proteome to the Panther server (Mi *et al.*, 2019; Thomas *et al.*, 2022) for over-representation analyses on protein localization (**Figure 1h**), function (**Figure 1i**), and pathway (**Figure 1j**) and revealed an enrichment of cell-surface proteins related to cell adhesion and neuronal development, perfectly matching the dataset—a cell-surface proteome of developing Purkinje cells.

Simply plugging in corresponding references, PEELing is ready for analyzing spatially-resolved proteomics data of any cellular compartments, such as the nucleus (**Supplementary Figure 6**) by TurboID (Branon *et al.*, 2018) (**Supplementary File 5**) and miniTurbo (Branon *et al.*, 2018) (**Supplementary File 6**), as well as the mitochondrion (**Supplementary Figure 7** and **8**) by APEX2 (Lam *et al.*, 2014; Han *et al.*, 2017) (**Supplementary File 7**) and BioID (Roux *et al.*, 2012; Branon *et al.*, 2018; Antonicka *et al.*, 2020) (**Supplementary File 8** and **9**). Together, we demonstrated that PEELing is widely applicable to: 1) various organisms, such as human (**Supplementary Figure 6**, **7**, and **8**), mouse (**Figure 1**), and fruit fly (**Supplementary Figure 5**); 2) diverse subcellular compartments, including membrane-enclosed organelles (nucleus and

mitochondrion) and open space (cell surface); 3) all commonly used proximity labeling tools, including peroxidases (APEX2 and HRP) and biotin ligases (BioID, TurboID, and miniTurbo); 4) different mass spectrometry quantification methods including label free (**Supplementary Figure 8**) and isobaric labeling (**Figure 1** and **Supplementary Figure 5**, **6**, and **7**). Moreover, PEELing features a plug-n'-play web service for a complete workflow (**Figure 1a**) including data visualization—all data panels shown here were directly from PEELing—as well as a Python package enabling advanced customization and integration with other bioinformatics tools. It thus provides a user-friendly yet versatile platform for exploring subcellular organization of the proteome.

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 and J. Lohmeyer for administrative assistance.

Author Contributions

X.P., Data curation, Formal analysis, Software, Visualization, Methodology, Writing—review and editing; J.C., Data curation, Formal analysis, Software, Visualization, Methodology, Writing—review and editing; Z.J., Formal analysis, Writing—review and editing; S.H., Data curation, Methodology, Writing—review and editing, Resources, Funding acquisition; S.P., Software, Writing—review and editing, Supervision, Resources, Funding acquisition; J.L., Conceptualization, Data curation, Formal analysis, Visualization, Methodology, Writing—original draft, Supervision, Resources, Funding acquisition.

Conflict of Interest

None declared.

Funding

This work was supported by the Howard Hughes Medical Institute (to S.P. and J.L.), the National Key R&D Program of China (2023YFA1801300 and 2022YFA1304500 to S.H.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0570000 to S.H.), the National Natural Science Foundation of China (92368102 and 22377126 to S.H.), and the Shanghai Municipal Science and Technology Major Project (to S.H.).

Data Availability

All datasets used for benchmarking are included as **Supplementary File 1–17**.

References

Antonicka,H. *et al.* (2020) A High-Density Human Mitochondrial Proximity Interaction Network. *Cell Metab*, **32**, 479-497.e9.

Branon,T.C. *et al.* (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology 2018 36:9*, **36**, 880–887.

Cho,K.F. *et al.* (2020) Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nature Protocols 2020 15:12*, **15**, 3971–3999.

Consortium,T.U. *et al.* (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*, **51**, D523–D531.

Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 2008 26:12*, **26**, 1367–1372.

Geri,J.B. *et al.* (2020) Microenvironment mapping via Dexter energy transfer on immune cells. *Science (1979)*, **367**, 1091–1097.

Han,S. *et al.* (2017) Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem Biol*, **24**, 404–414.

Han,S. *et al.* (2018) Proximity labeling: spatially resolved proteomic mapping for neurobiology. *Curr Opin Neurobiol*, **50**, 17–23.

Hung,V. *et al.* (2014) Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Mol Cell*, **55**, 332–341.

Hung,V. *et al.* (2016) Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nature Protocols 2016 11:3*, **11**, 456–475.

Knorr,K. *et al.* (2023) Systematic evaluation of AML-associated antigens identifies anti-U5 SNRNP200 therapeutic antibodies for the treatment of acute myeloid leukemia. *Nature Cancer 2023 4:12*, **4**, 1675–1692.

Kong,A.T. *et al.* (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature Methods 2017 14:5*, **14**, 513–520.

Lam,S.S. *et al.* (2014) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods 2014 12:1*, **12**, 51–54.

Li,J. *et al.* (2020) Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators. *Cell*, **180**, 373-386.e15.

Loh,K.H. *et al.* (2016) Proteomic Analysis of Unbounded Cellular Compartments: Synaptic Clefts. *Cell*, **166**, 1295-1307.e21.

Mi,H. *et al.* (2019) 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.

Qin,W. *et al.* (2021) Deciphering molecular interactions by proximity labeling. *Nature Methods 2021 18:2*, **18**, 133–143.

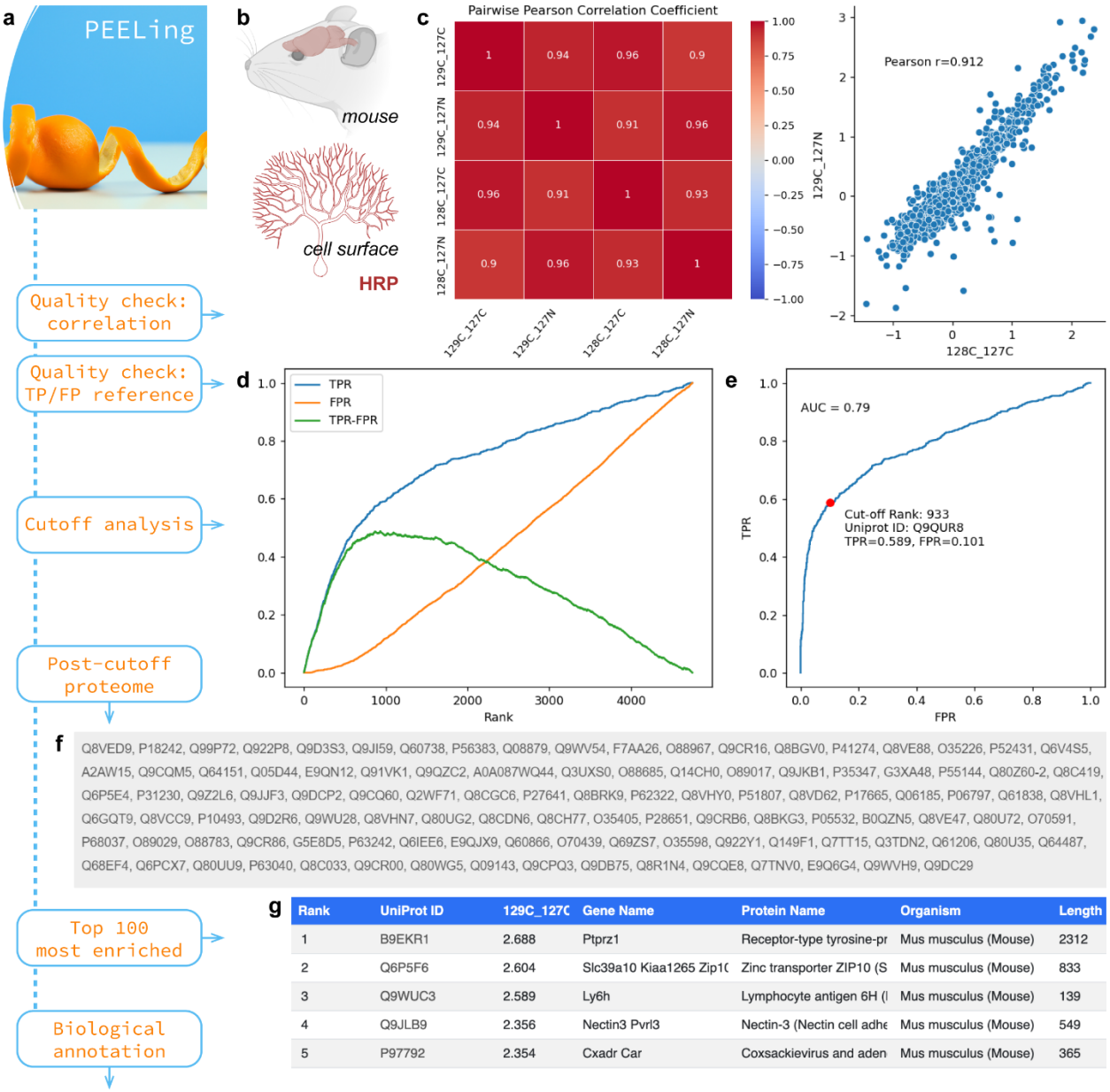
Rhee,H.W. *et al.* (2013) Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science (1979)*, **339**, 1328–1331.

Roux,K.J. *et al.* (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology*, **196**, 801–810.

Shuster,S.A. *et al.* (2022) In situ cell-type-specific cell-surface proteomic profiling in mice. *Neuron*, **110**, 3882-3896.e9.

Thomas,P.D. *et al.* (2022) PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science*, **31**, 8–22.

Thompson,A. *et al.* (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*, **75**, 1895–1904.



f

Q8VED9, P18242, Q99P72, Q922P8, Q9D3S3, Q9JI59, Q60738, P56383, Q08879, Q9WV54, F7AA26, O88967, Q9CR16, Q8BGV0, P41274, Q8VE88, O35226, P52431, Q6V4S5, A2AW15, Q9CQM5, Q64151, Q05D44, E9QN12, Q91VK1, Q9QZC2, A0A087WQ44, Q3UXS0, O88685, Q14CH0, O89017, Q9JKB1, P35347, G3XA48, P55144, Q80Z60-2, Q8C419, Q6P5E4, P31230, Q9Z2L6, Q9JJF3, Q9DCP2, Q9CQ60, Q2WF71, Q8CGC6, P27641, Q8BRK9, P62322, Q8VHY0, P51807, Q8VD62, P17665, Q06185, P06797, Q61838, Q8VHL1, Q6GQT9, Q8VCC9, P10493, Q9D2R6, Q9WU28, Q8VHN7, Q80UG2, Q8CDN6, Q8CH77, O35405, P28651, Q9CRB6, Q8BKG3, P05532, B0QZN5, Q8VE47, Q80U72, O70591, P68037, O89029, O88783, Q9CR86, G5E8D5, P63242, Q6IEE6, E9QJX9, Q60866, O70439, Q69ZS7, O35598, Q922Y1, Q149F1, Q7TT15, Q3TDN2, Q61206, Q80U35, Q64487, Q68EF4, Q6PCX7, Q80UU9, P63040, Q8C033, Q9CR00, Q80WG5, Q09143, Q9CPQ3, Q9DB75, Q8R1N4, Q9CQE8, Q7TNV0, E9Q6G4, Q9WVH9, Q9DC29

g

Top 100 most enriched

Rank	UniProt ID	129C_127C	Gene Name	Protein Name	Organism	Length
1	B9EKR1	2.688	Ptprz1	Receptor-type tyrosine-pr	Mus musculus (Mouse)	2312
2	Q6P5F6	2.604	Slc39a10 Kiaa1265 Zip1(	Zinc transporter ZIP10 (S	Mus musculus (Mouse)	833
3	Q9WUC3	2.589	Ly6h	Lymphocyte antigen 6H (I	Mus musculus (Mouse)	139
4	Q9JLB9	2.356	Nectin3 Pvrl3	Nectin-3 (Nectin cell adhe	Mus musculus (Mouse)	549
5	P97792	2.354	Cxadr Car	Coxsackievirus and aden	Mus musculus (Mouse)	365

Biological annotation

h Location: Panther GO Slim Cellular Component

Term	FDR
plasma membrane	9.857e-25
cell periphery	1.817e-24
plasma membrane protein complex	1.684e-14
cell junction	1.787e-13
receptor complex	4.554e-13
intrinsic component of membrane	9.120e-13
integral component of membrane	1.660e-12
integral component of plasma membrane	4.795e-11
intrinsic component of plasma membrane	9.463e-11
membrane protein complex	1.237e-10

i Function: Panther GO Slim Biological Process

Term	FDR
biological adhesion	2.001e-18
cell adhesion	3.627e-18
cell morphogenesis	5.446e-17
cell-cell adhesion	2.638e-16
nervous system development	5.010e-16
multicellular organism development	5.463e-16
axonogenesis	2.362e-14
axon development	3.389e-14
cell morphogenesis involved in neuron differentiation	3.055e-13
cell junction assembly	3.515e-13

j Pathway: Reactome

Term	FDR
Adherens junctions interactions	1.715e-8
Cell-cell junction organization	2.758e-8
Cell junction organization	2.816e-8
Cell-Cell communication	3.442e-8
Protein-protein interactions at synapses	9.394e-8
Extracellular matrix organization	6.852e-7
Receptor-type tyrosine-protein phosphatases	2.230e-6
ECM proteoglycans	3.094e-6
Post-translational protein phosphorylation	5.581e-5
Neuronal System	8.309e-5

Figure 1. PEELing workflow and the analysis of a mouse cell-surface proteome.

(**a**) PEELing workflow. (**b**) Shuster *et al.*(Shuster *et al.*, 2022) used horseradish peroxidase (HRP) mediated cell-surface biotinylation to capture the cell-surface proteome of mouse Purkinje cells at postnatal day 15. The authors used tandem mass tag (TMT) (Thompson *et al.*, 2003) based quantitative mass spectrometry. 129C and 128C TMT tags were used for cell-surface-labelled samples while 127C and 127N tags were used for non-labelled controls, yielding four labelled-to-control ratios (**Supplementary File 1**). (**c**) Correlation plots and coefficients. (**d**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against 129C:127N ratio-based ranking (x-axis). (**e**) Receiver operating characteristic (ROC) curve, based on ranking by 129C:127N. Red dot, cutoff position. AUC, area under the curve. (**f**) Post-cutoff proteome. (**g**) Top 5 most enriched proteins based on 129C:127C. On the website, this list extends to the top 100. Length unit: amino acids. (**h–j**) Protein ontology analyses for localization (**h**), function (**i**), and pathway (**j**). FDR, false discovery rate. The PEELing icon was produced by DALL-E of OpenAI.

Supplementary Materials

Method Details

Data input

We note that PEELing does not handle raw mass spectral data or provide custom searches (e.g., labeling site identification or non-canonical translation analysis), which needs to be processed by peptide identification and quantification software such as MaxQuant (Cox and Mann, 2008), FragPipe (Kong *et al.*, 2017), Spectrum Mill (Broad Institute), or Proteome Discoverer (Thermo Fisher). The PEELing website and command line program accept tab-separated value (.tsv) files, which contain UniProt accession numbers in their first columns and enrichment indexes in the remaining columns (e.g., **Supplementary File 1–9**). PEELing uses UniProt accession numbers as protein identifiers and updates user input to the current UniProt version to help database searches. As benchmarked and discussed above, labelled-to-control ratio is the preferred enrichment index; however, users may choose other indexes according to their experimental design.

True positive and false positive references

We constructed true positive (TP) and false positive (FP) references using SwissProt-reviewed annotations. The PEELing web service automatically updates these references from the UniProt database (Consortium *et al.*, 2023), ensuring that they remain current and accurate.

Cell-surface TP reference (**Supplementary File 10**) 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. Cell-surface FP reference (**Supplementary File 12**) 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.

Nuclear TP reference (**Supplementary File 14**) is specified by the UniProt term: *(cc_scl_term:SL-0191) AND (reviewed:true)*, which includes SwissProt-reviewed nuclear proteins

(SL-0191). Nuclear FP reference (**Supplementary File 15**) is specified by the UniProt term: *(((cc_scl_term:SL-0091) OR (cc_scl_term:SL-0173) OR (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)) NOT ((cc_scl_term:SL-0191) AND (reviewed:true))*, including SwissProt-reviewed cytosolic (SL-0091), mitochondrial (SL-0173), and cell-surface (cell-surface TP term) proteins that do not express in the nucleus.

Mitochondrial TP reference (**Supplementary File 16**) is specified by the UniProt term: *(cc_scl_term:SL-0173) AND (reviewed:true)*, which includes SwissProt-reviewed mitochondrial proteins (SL-0173). Mitochondrial FP reference (**Supplementary File 17**) is specified by the UniProt term: *(((cc_scl_term:SL-0091) OR (cc_scl_term:SL-0191) OR (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)) NOT ((cc_scl_term:SL-0173) AND (reviewed:true))*, including SwissProt-reviewed cytosolic (SL-0091), nuclear (SL-0191), and cell-surface (cell-surface TP term) proteins that do not express in the mitochondrion.

For custom TP and FP references, PEELing requests two tab-separated value (.tsv) files from the user, each containing one column of UniProt accession numbers (e.g., **Supplementary File 10–17**). Despite the robustness of this algorithm and its tolerance to imperfect references (**Supplementary Figure 4**), constructing high-quality references is essential for proper data filtering and interpretation. The TP reference should contain proteins known to be in the designated cellular compartment while the FP reference should contain proteins that are, to one's best knowledge, not in this compartment. Curated Swiss-Prot/UniProt and Gene Ontology Cellular Component (GOCC) databases, as well as relevant literature, provide resources for creating the TP and FP references. More TP and FP examples and their designing rules can be found in (Hung *et al.*, 2016) and (Cho *et al.*, 2020).

Reference-based data quality check and cutoff analysis

For each enrichment index (e.g., each labelled-to-control ratio), PEELing first ranks all proteins in descending order according to this index. For each protein on the ranked list, accumulated true positive count and false positive count above this ranking position are calculated to obtain true positive rate (TPR), false positive rate (FPR), and their difference (TPR–FPR) at this ranking position (visualized in **Figure 1d** and other plots of TPR, FPR, and TPR–FPR). A receiver operating characteristic (ROC) curve is produced accordingly (**Figure 1e** and other ROC curves).

For each enrichment index, the cutoff is set where TPR–FPR maximizes, representing the largest segregation of intended signal (TP) and unintended noise (FP). PEELing conducts cutoff

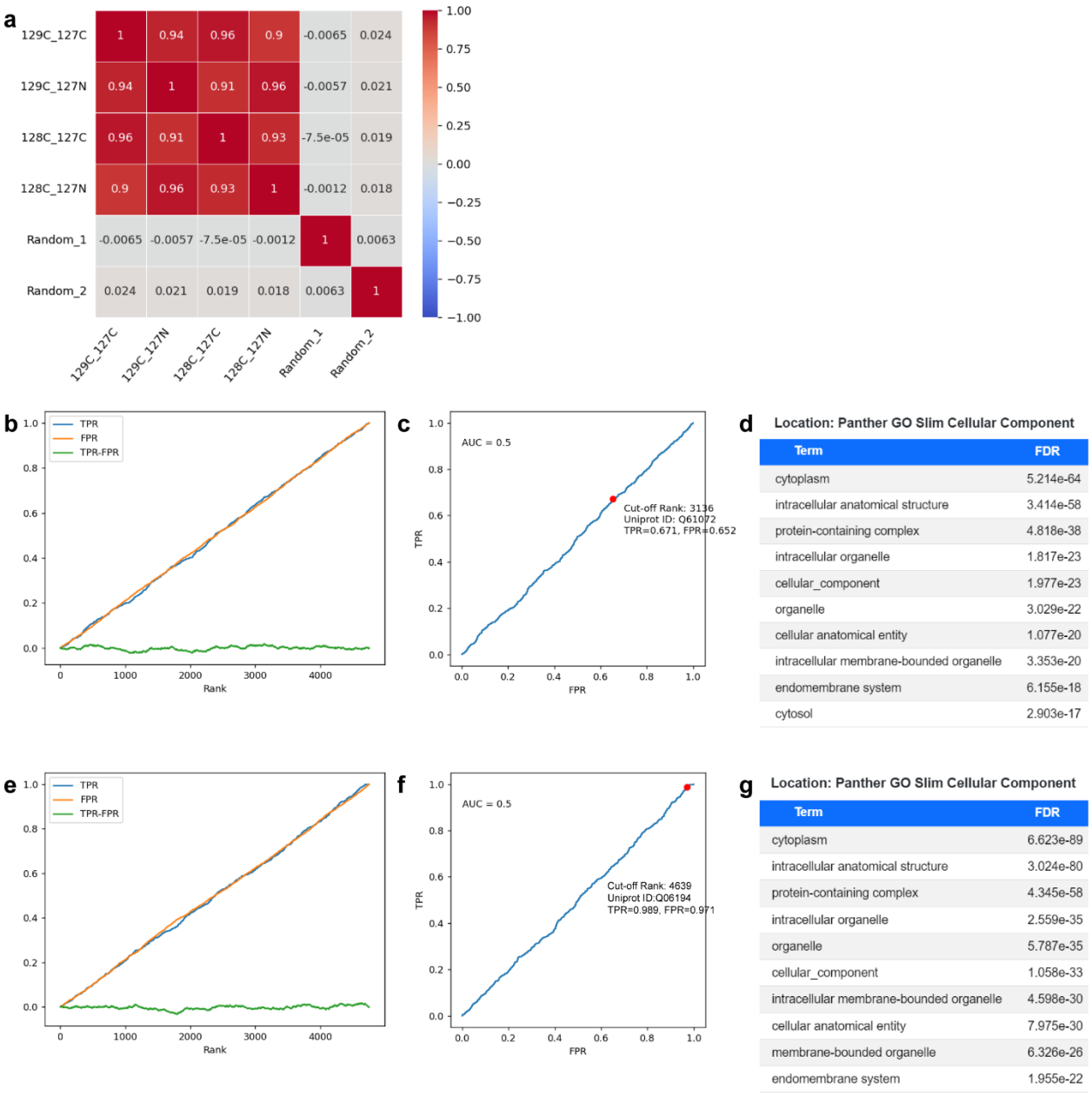
analysis on all submitted enrichment indexes individually and, for the final proteome, retains only those proteins passing the cutoff of all or multiple indexes. Both the PEELing website and command line program offer the optional "Tolerance" setting, enabling users to control the stringency of the cutoff. By default, it is set to 0, meaning that a protein must pass the cutoff of all indexes to be included in the final proteome. If it is set to n, a protein can fail the cutoff in up to n indexes and still be included in the final proteome. Despite the flexibility, we recommend setting the tolerance value to a small number to better filter out contaminants.

Protein annotation

From the post-cutoff proteome, PEELing retrieves from UniProt and displays gene names, protein names, organisms, and protein lengths of the top 100 enriched proteins based on each index. For protein ontology, PEELing sends the post-cutoff proteome to the Panther server (Thomas *et al.*, 2022; Mi *et al.*, 2019) for over-representation analysis. The annotation datasets used include: GO slim cellular component, for protein localization; GO slim biological process, for protein function; and reactome pathway, for signaling pathway. Results are ranked in ascending order by false discovery rate (FDR). Top 10 terms are listed along with their FDRs.

Supplementary Figures

Random values: simulating failed enrichment of cell-surface proteins



Term	FDR
cytoplasm	5.214e-64
intracellular anatomical structure	3.414e-58
protein-containing complex	4.818e-38
intracellular organelle	1.817e-23
cellular_component	1.977e-23
organelle	3.029e-22
cellular anatomical entity	1.077e-20
intracellular membrane-bounded organelle	3.353e-20
endomembrane system	6.155e-18
cytosol	2.903e-17

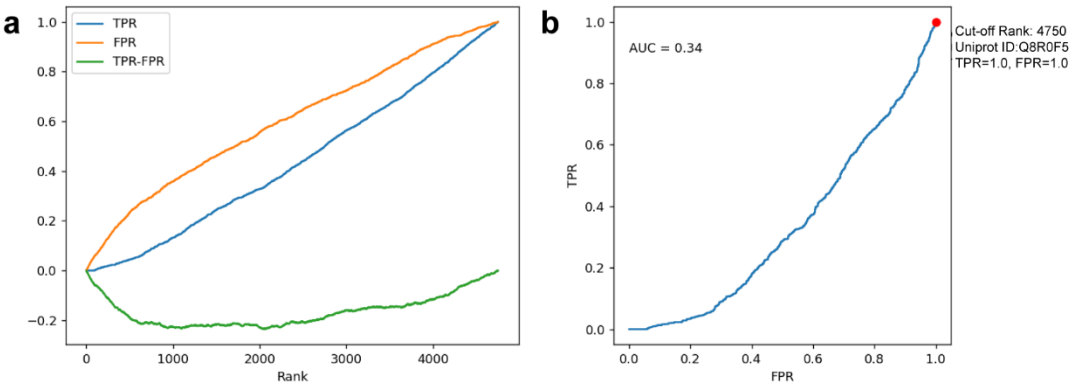
Term	FDR
cytoplasm	6.623e-89
intracellular anatomical structure	3.024e-80
protein-containing complex	4.345e-58
intracellular organelle	2.559e-35
organelle	5.787e-35
cellular_component	1.058e-33
intracellular membrane-bounded organelle	4.598e-30
cellular anatomical entity	7.975e-30
membrane-bounded organelle	6.326e-26
endomembrane system	1.955e-22

Supplementary Figure 1. Testing PEELing with randomized data.

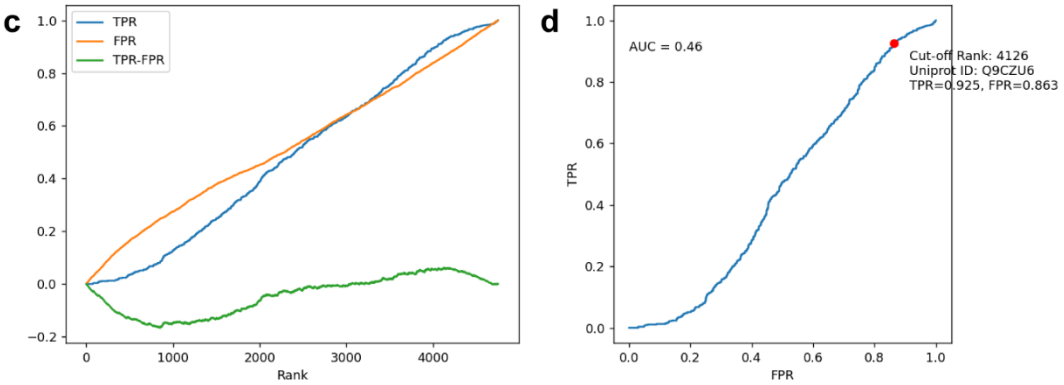
To simulate failed enrichment, two sets of random values Random_1 and Random_2 were generated and added to the data of Purkinje cell cell-surface proteome (**Supplementary File 2**). (**a**) Correlation coefficients showing no correlation between the real and randomized data. (**b**,**e**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR,

green) plotted against Random_1 (**b**) and Random_2 (**e**) based ranking (x-axis). (**c**,**f**) Receiver operating characteristic (ROC) curves, based on Random_1 (**c**) and Random_2 (**f**), respectively. AUC, area under the curve. (**d**,**g**) Subcellular localization annotation did not enrich any cell surface related terms. **d**, Random_1; **g**, Random_2. FDR, false discovery rate.

Mismatched references: analyzing the cell-surface proteome data with nuclear references



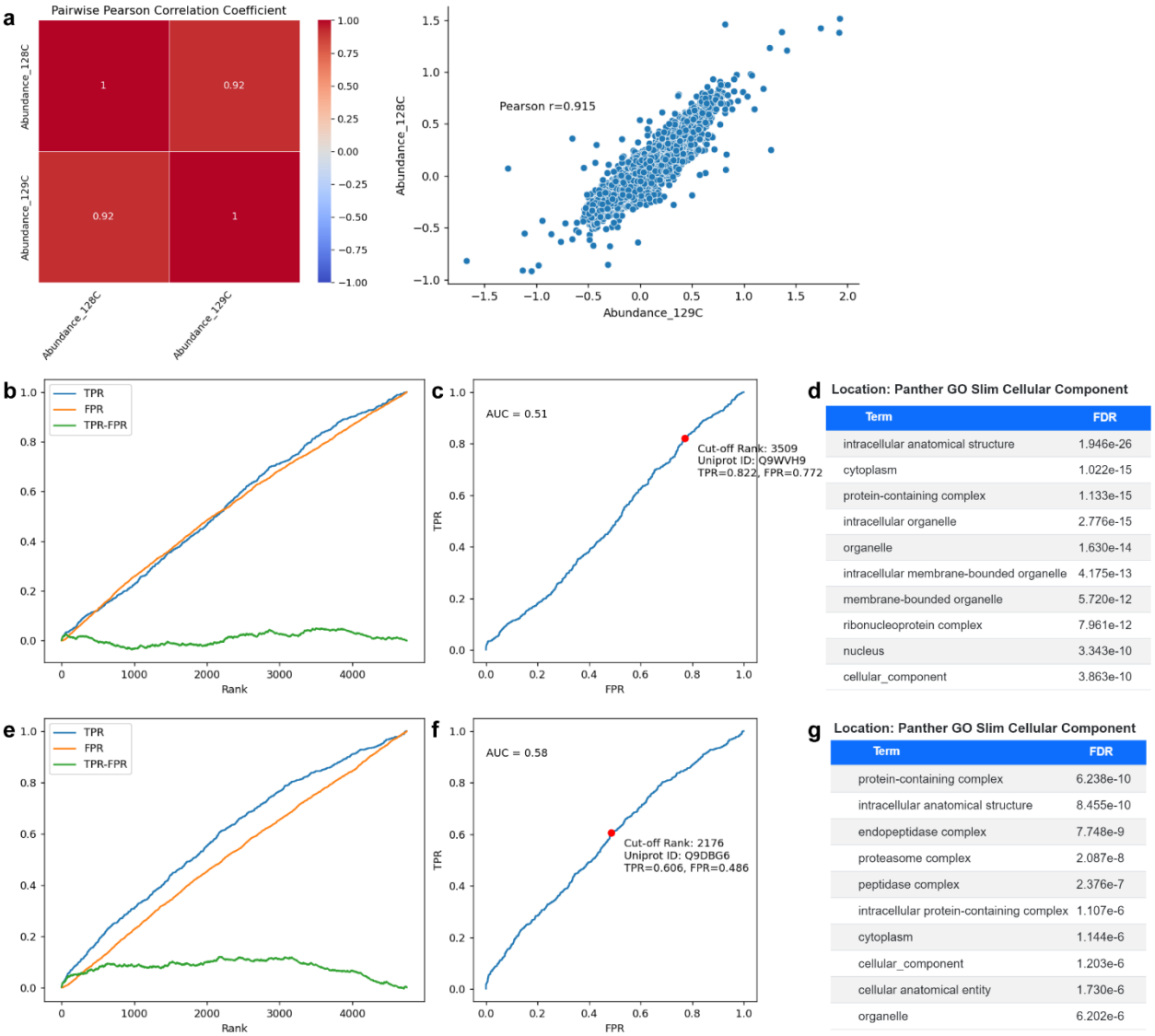
Mismatched references: analyzing the cell-surface proteome data with mitochondrial references



Supplementary Figure 2. Testing PEELing with mismatched references.

The cell-surface proteome of Purkinje cells (**Supplementary File 1**) was analyzed using mismatched references: nucleus (**a**,**b**; **Supplementary File 14**,**15**) and mitochondrion (**c**,**d**; **Supplementary File 16,17**). (**a**,**c**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against 129C:127N ratio-based ranking (x-axis). (**b**,**d**) Receiver operating characteristic (ROC) curves, based on ranking by 129C:127N. AUC, area under the curve.

Analysis based on protein abundances of the labelled groups, instead of labelled-to-control ratios



d Location: Panther GO Slim Cellular Component

Term	FDR
intracellular anatomical structure	1.946e-26
cytoplasm	1.022e-15
protein-containing complex	1.133e-15
intracellular organelle	2.776e-15
organelle	1.630e-14
intracellular membrane-bounded organelle	4.175e-13
membrane-bounded organelle	5.720e-12
ribonucleoprotein complex	7.961e-12
nucleus	3.343e-10
cellular_component	3.863e-10

g Location: Panther GO Slim Cellular Component

Term	FDR
protein-containing complex	6.238e-10
intracellular anatomical structure	8.455e-10
endopeptidase complex	7.748e-9
proteasome complex	2.087e-8
peptidase complex	2.376e-7
intracellular protein-containing complex	1.107e-6
cytoplasm	1.144e-6
cellular_component	1.203e-6
cellular anatomical entity	1.730e-6
organelle	6.202e-6

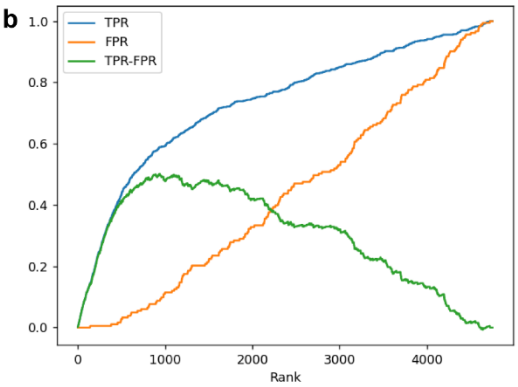
Supplementary Figure 3. Abundance is not a reliable indicator of enrichment.

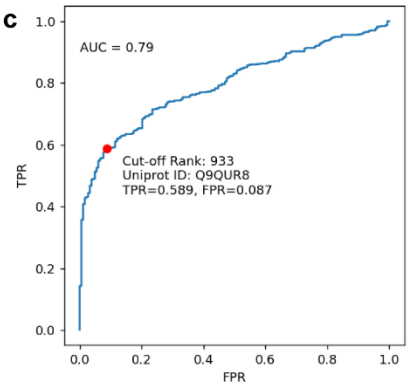
Protein abundance data of cell-surface-labelled samples (TMT tags 128C and 129C) was obtained from the cell-surface proteome of Purkinje cells (Shuster *et al.*, 2022) (**Supplementary File 3**). (**a**) Correlation plots and coefficients. (**b**,**e**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against abundance-based ranking (x-axis). **b**, 128C; **e**, 129C. (**c**,**f**) Receiver operating characteristic (ROC) curves, based on ranking by abundance. **c**, 128C; **f**, 129C. AUC, area under the curve. (**d**,**g**) Subcellular localization of the abundance-ranked top 500 proteins. **d**, 128C; **g**, 129C. FDR, false discovery rate.

a

TP reference: full

FP reference: 10%





d Location: Panther GO Slim Cellular Component

Term	FDR
plasma membrane protein complex	5.095e-15
cell junction	5.383e-15
receptor complex	4.129e-13
membrane	7.830e-11
membrane protein complex	1.746e-10
endomembrane system	6.249e-10
cell periphery	2.823e-9
endoplasmic reticulum	3.042e-9
adherens junction	3.729e-9
plasma membrane	3.912e-9

e Function: Panther GO Slim Biological Process

Term	FDR
cell-cell adhesion	1.920e-24
cell adhesion	2.366e-24
cell morphogenesis	4.625e-23
multicellular organism development	1.509e-22
nervous system development	8.775e-20
axonogenesis	1.322e-17
cell morphogenesis involved in neuron differentiation	3.008e-17
cell morphogenesis involved in differentiation	5.597e-17
neurogenesis	6.325e-17
system development	7.247e-17

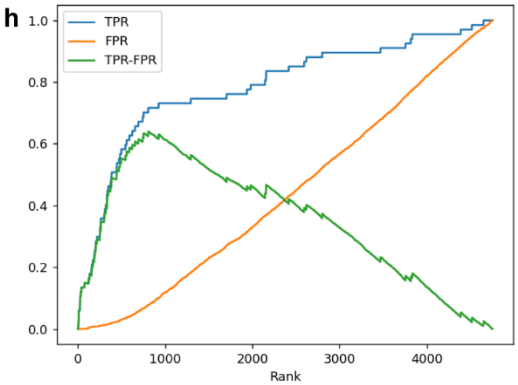
f Pathway: Reactome

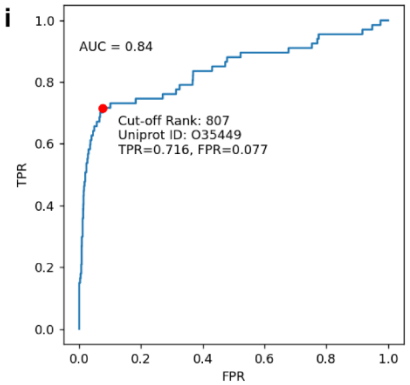
Term	FDR
Cell junction organization	1.188e-7
Adherens junctions interactions	1.422e-7
Cell-Cell communication	1.633e-7
Protein-protein interactions at synapses	2.707e-7
Cell-cell junction organization	2.756e-7
Extracellular matrix organization	5.671e-7
Receptor-type tyrosine-protein phosphatases	2.766e-6
ECM proteoglycans	2.388e-5
Post-translational protein phosphorylation	6.873e-5
Neuronal System	8.284e-5

g

TP reference: 10%

FP reference: full





j Location: Panther GO Slim Cellular Component

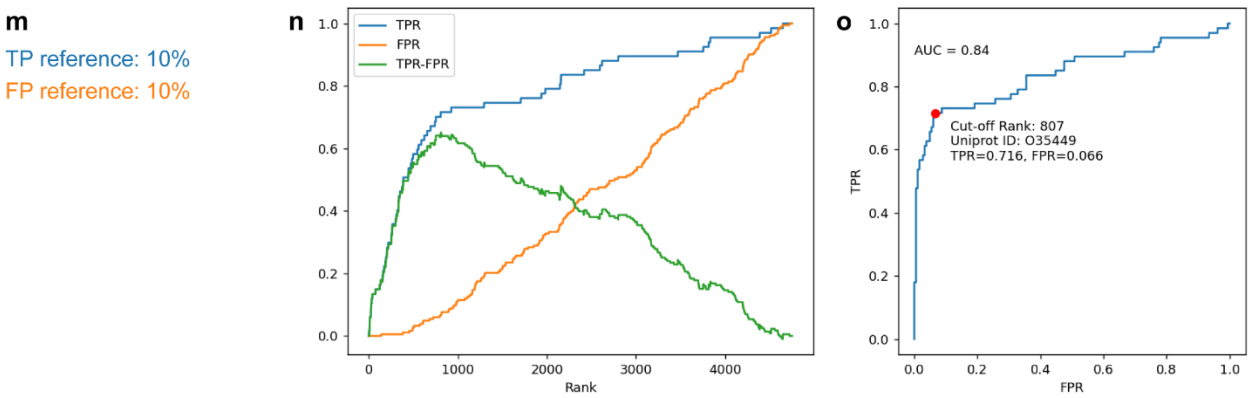
Term	FDR
plasma membrane protein complex	6.527e-17
cell junction	1.388e-14
receptor complex	5.483e-14
cell periphery	6.311e-11
membrane	6.701e-11
membrane protein complex	1.232e-10
plasma membrane	2.100e-10
plasma membrane signaling receptor complex	9.665e-10
endoplasmic reticulum	1.001e-8
adherens junction	1.045e-8

k Function: Panther GO Slim Biological Process

Term	FDR
multicellular organism development	3.405e-25
cell adhesion	3.715e-25
cell morphogenesis	4.147e-25
cell-cell adhesion	5.105e-25
nervous system development	7.785e-23
axonogenesis	1.243e-19
neurogenesis	1.290e-19
system development	1.466e-19
cell morphogenesis involved in neuron differentiation	3.051e-19
cell morphogenesis involved in differentiation	6.023e-19

l Pathway: Reactome

Term	FDR
Adherens junctions interactions	1.776e-7
Cell junction organization	3.533e-7
Cell-cell junction organization	3.788e-7
Cell-Cell communication	3.840e-7
Protein-protein interactions at synapses	8.710e-7
Extracellular matrix organization	2.892e-6
ECM proteoglycans	1.496e-5
Receptor-type tyrosine-protein phosphatases	2.171e-5
Neuronal System	2.660e-4
Post-translational protein phosphorylation	7.147e-4



p Location: Panther GO Slim Cellular Component

Term	FDR
plasma membrane protein complex	6.527e-17
cell junction	1.388e-14
receptor complex	5.483e-14
cell periphery	6.311e-11
membrane	6.701e-11
membrane protein complex	1.232e-10
plasma membrane	2.100e-10
plasma membrane signaling receptor complex	9.665e-10
endoplasmic reticulum	1.001e-8
adherens junction	1.045e-8

q Function: Panther GO Slim Biological Process

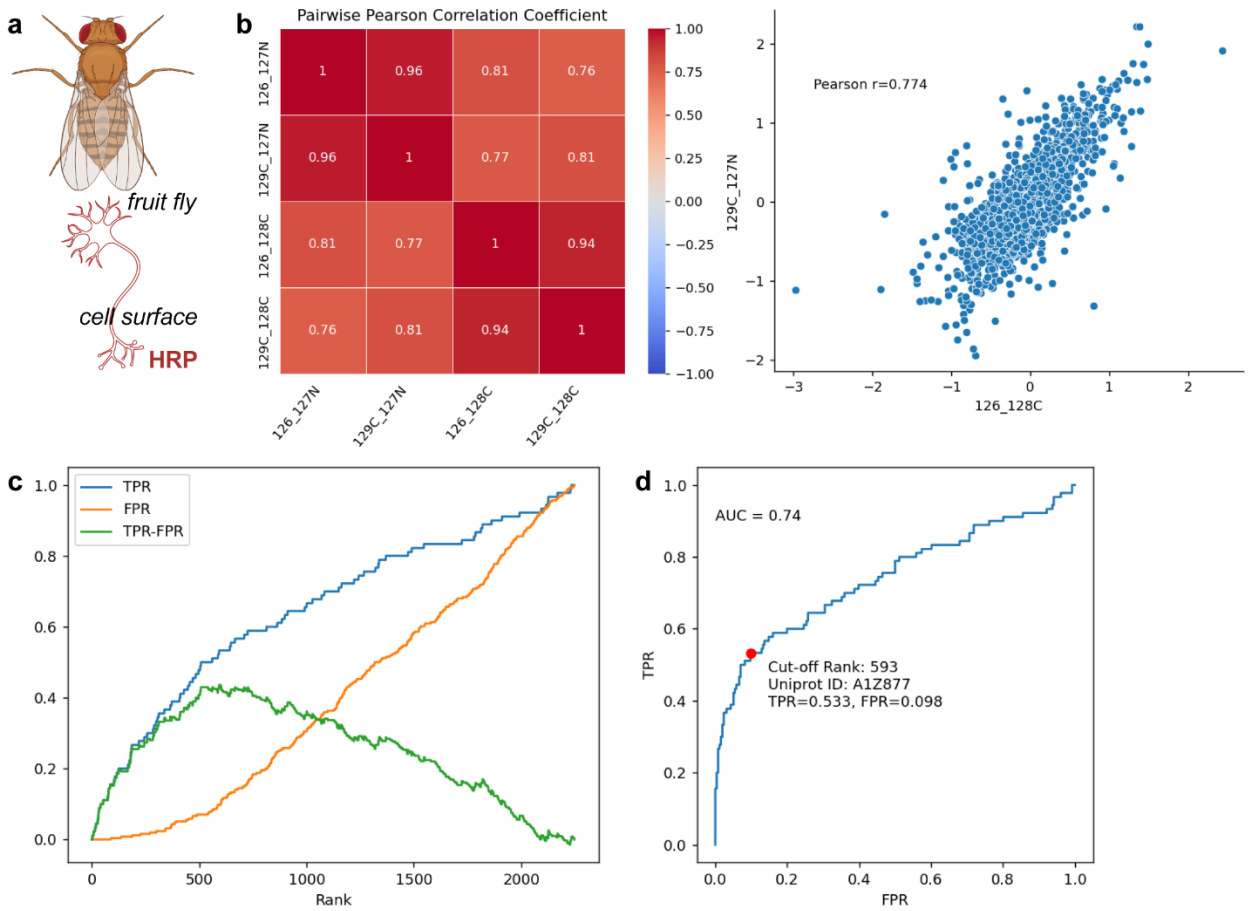
Term	FDR
multicellular organism development	3.405e-25
cell adhesion	3.715e-25
cell morphogenesis	4.147e-25
cell-cell adhesion	5.105e-25
nervous system development	7.785e-23
axonogenesis	1.243e-19
neurogenesis	1.290e-19
system development	1.466e-19
cell morphogenesis involved in neuron differentiation	3.051e-19
cell morphogenesis involved in differentiation	6.023e-19

r Pathway: Reactome

Term	FDR
Adherens junctions interactions	1.776e-7
Cell junction organization	3.533e-7
Cell-cell junction organization	3.788e-7
Cell-Cell communication	3.840e-7
Protein-protein interactions at synapses	8.710e-7
Extracellular matrix organization	2.892e-6
ECM proteoglycans	1.496e-5
Receptor-type tyrosine-protein phosphatases	2.171e-5
Neuronal System	2.660e-4
Post-translational protein phosphorylation	7.147e-4

Supplementary Figure 4. Testing PEELing with coverage-reduced references.

We randomly removed 90% of genes from the cell-surface true-positive (TP) and false-positive (FP) references (**Supplementary File 10,12**) to produce truncated references with only 10% of the original coverage (**Supplementary File 11,13**) and then analyzed the cell-surface proteome of Purkinje cells (**Supplementary File 1**) using the truncated references (**m–r**) or pairing a full reference with a truncated one (**a–l**). (**b**,**h**,**n**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against 129C:127N ratio-based ranking (x-axis). (**c**,**i**,**o**) Receiver operating characteristic (ROC) curve, based on ranking by 129C:127N. Red dot, cutoff position. AUC, area under the curve. (**d–f**, **j–l**, **p–r**) Protein ontology analyses of post-cutoff proteomes for localization (**d**,**j**,**p**), function (**e**,**k**,**q**), and pathway (**f**,**l**,**r**). FDR, false discovery rate.



e Location: Panther GO Slim Cellular Component

Term	FDR
plasma membrane	5.591e-26
cell periphery	1.132e-24
integral component of membrane	8.571e-14
intrinsic component of membrane	1.134e-13
integral component of plasma membrane	1.499e-13
intrinsic component of plasma membrane	1.622e-13
membrane	4.537e-13
neuron projection	1.631e-11
cellular anatomical entity	5.007e-11
cellular_component	1.164e-10

f Function: Panther GO Slim Biological Process

Term	FDR
localization	5.470e-10
transport	1.107e-9
establishment of localization	1.195e-9
ion transport	4.469e-8
cell-cell signaling	1.933e-7
signaling	1.032e-6
cell communication	1.489e-6
anterograde trans-synaptic signaling	2.133e-6
trans-synaptic signaling	2.327e-6
chemical synaptic transmission	2.559e-6

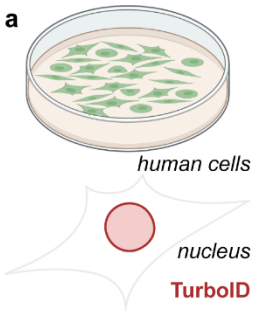
g Pathway: Reactome

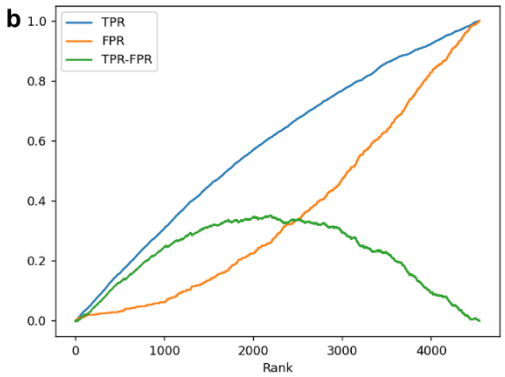
Term	FDR
Neuronal System	5.126e-10
Transmission across Chemical Synapses	2.383e-9
Transport of small molecules	9.498e-8
SLC-mediated transmembrane transport	1.430e-7
Axon guidance	1.351e-4
Nervous system development	1.394e-4
Neurotransmitter release cycle	1.560e-4
Metabolism	3.356e-4
Signal Transduction	4.724e-4
Developmental Biology	5.022e-4

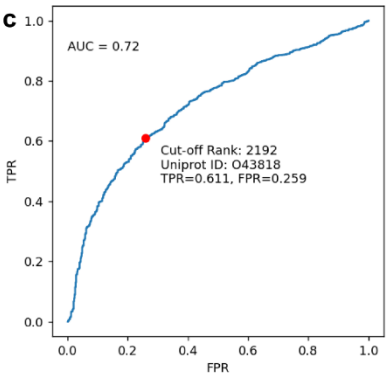
Supplementary Figure 5. PEELing analysis of a *Drosophila* cell-surface proteome.

(**a**) The cell-surface proteome of *Drosophila* mature olfactory projection neurons was profiled by horseradish peroxidase (HRP) based cell-surface proteomics (Li *et al.*, 2020) (**Supplementary File 4**). (**b**) Correlation plots and coefficients. (**c**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against 129C:127N ratio-based ranking (x-axis). (**d**) Receiver operating characteristic (ROC) curve, based on ranking by

129C:127N. Red dot, cutoff position. AUC, area under the curve. (**e–g**) Protein ontology analyses for localization (**e**), function (**f**), and pathway (**g**). FDR, false discovery rate.







d Location: Panther GO Slim Cellular Component

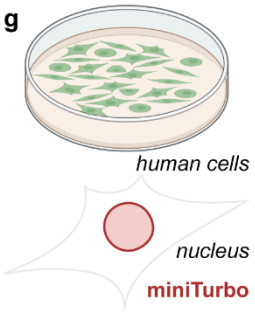
Term	FDR
nucleus	3.848e-157
intracellular organelle	3.825e-101
organelle	1.882e-96
intracellular anatomical structure	4.842e-92
intracellular membrane-bounded organelle	1.846e-90
membrane-bounded organelle	1.183e-85
nuclear protein-containing complex	2.830e-84
nuclear lumen	6.668e-63
organelle lumen	1.155e-57
intracellular organelle lumen	1.271e-57

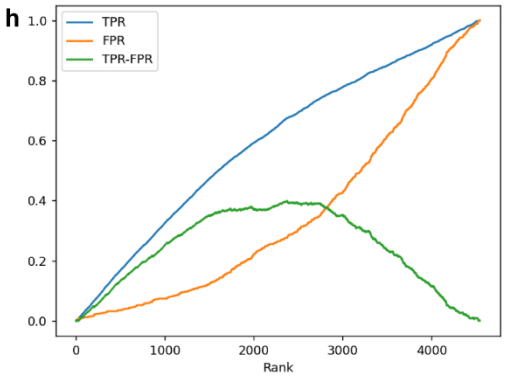
e Function: Panther GO Slim Biological Process

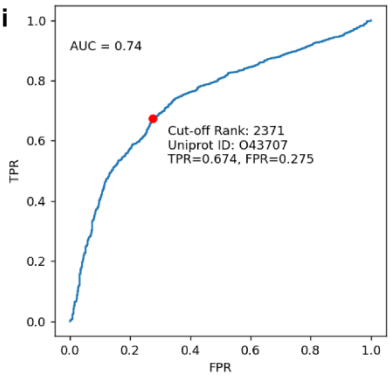
Term	FDR
nucleic acid metabolic process	7.936e-98
nucleobase-containing compound metabolic process	9.954e-88
cellular nitrogen compound metabolic process	9.833e-87
heterocycle metabolic process	5.512e-86
cellular aromatic compound metabolic process	1.942e-85
macromolecule metabolic process	1.068e-83
organic cyclic compound metabolic process	1.314e-83
RNA metabolic process	5.522e-77
gene expression	1.692e-73
nitrogen compound metabolic process	5.120e-70

f Pathway: Reactome

Term	FDR
Metabolism of RNA	1.534e-66
Gene expression (Transcription)	6.176e-45
Processing of Capped Intron-Containing Pre-mRNA	4.920e-40
mRNA Splicing	2.672e-35
Chromatin organization	3.015e-35
Chromatin modifying enzymes	3.768e-35
RNA Polymerase II Transcription	1.334e-34
mRNA Splicing - Major Pathway	1.948e-34
Cell Cycle	3.265e-26
DNA Repair	5.763e-23







j Location: Panther GO Slim Cellular Component

Term	FDR
nucleus	8.559e-144
intracellular organelle	1.783e-98
intracellular membrane-bounded organelle	1.952e-95
organelle	3.908e-93
intracellular anatomical structure	2.237e-92
membrane-bounded organelle	4.483e-90
nuclear protein-containing complex	7.807e-80
nuclear lumen	1.449e-63
organelle lumen	3.772e-58
intracellular organelle lumen	4.149e-58

k Function: Panther GO Slim Biological Process

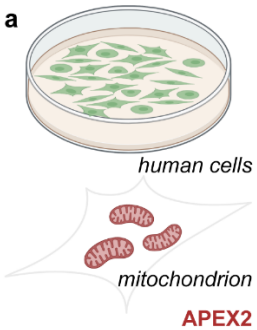
Term	FDR
nucleic acid metabolic process	1.864e-81
cellular nitrogen compound metabolic process	8.242e-76
nucleobase-containing compound metabolic process	2.061e-74
heterocycle metabolic process	2.204e-72
cellular aromatic compound metabolic process	3.869e-72
organic cyclic compound metabolic process	3.271e-71
macromolecule metabolic process	5.281e-67
gene expression	1.996e-64
RNA metabolic process	2.566e-64
nitrogen compound metabolic process	5.271e-58

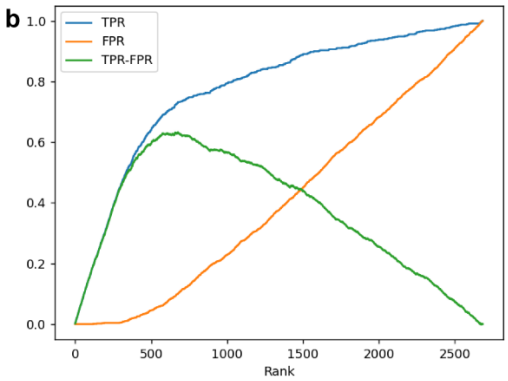
l Pathway: Reactome

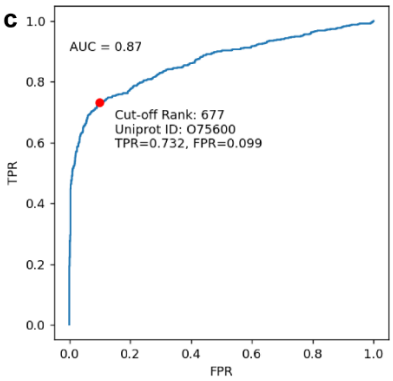
Term	FDR
Metabolism of RNA	2.280e-66
Gene expression (Transcription)	1.687e-34
Processing of Capped Intron-Containing Pre-mRNA	3.843e-34
rRNA processing	3.154e-30
rRNA processing in the nucleus and cytosol	3.807e-30
Chromatin organization	6.901e-29
Chromatin modifying enzymes	8.051e-29
mRNA Splicing	4.832e-28
Cell Cycle	7.489e-28
mRNA Splicing - Major Pathway	5.100e-27

Supplementary Figure 6. PEELing analysis of human nuclear proteomes.

(**a–f**) Analysis of a nuclear proteome of human embryonic kidney 293 cells, profiled by (Branon *et al.*, 2018) using the biotin ligase TurboID (**Supplementary File 5**). (**g–l**) Analysis of a nuclear proteome of human embryonic kidney 293 cells, profiled by (Branon *et al.*, 2018) using the biotin ligase miniTurbo (**Supplementary File 6**). (**b**,**h**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against ratio-based ranking (x-axis). Ratio used: **b**, 129:126; **h**, 130:131. (**c**,**i**) Receiver operating characteristic (ROC) curve. Red dot, cutoff position. AUC, area under the curve. Ratio used: **c**, 129:126; **i**, 130:131. (**d–f**, **j–l**) Protein ontology analyses of post-cutoff proteomes for localization (**d**,**j**), function (**e**,**k**), and pathway (**f**,**l**). FDR, false discovery rate.







d Location: Panther GO Slim Cellular Component

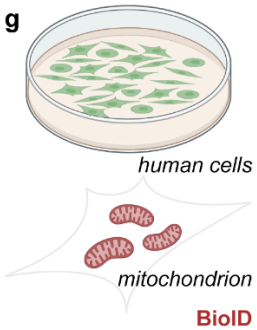
Term	FDR
mitochondrion	3.962e-114
mitochondrial protein-containing complex	6.747e-31
mitochondrial matrix	9.464e-31
cytoplasm	1.678e-29
mitochondrial ribosome	2.793e-19
organellar ribosome	3.192e-19
mitochondrial inner membrane	5.763e-17
organelle inner membrane	6.703e-16
mitochondrial envelope	4.599e-15
ribosome	1.132e-14

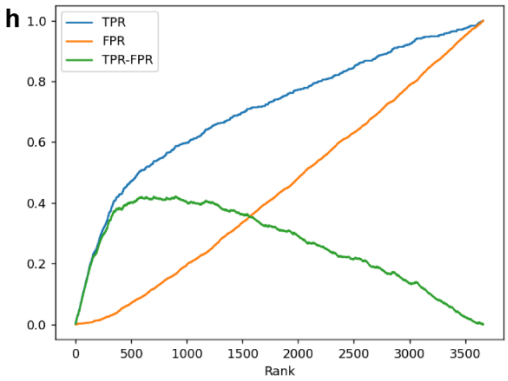
e Function: Panther GO Slim Biological Process

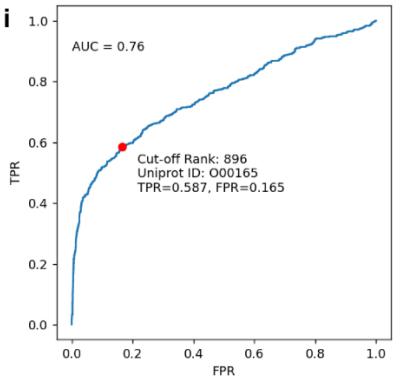
Term	FDR
small molecule metabolic process	2.617e-48
carboxylic acid metabolic process	1.978e-35
oxoacid metabolic process	2.556e-35
organic acid metabolic process	2.130e-34
tRNA metabolic process	3.723e-27
cellular amino acid metabolic process	3.131e-25
ncRNA metabolic process	1.138e-23
organonitrogen compound biosynthetic process	1.133e-20
cellular amide metabolic process	3.687e-18
peptide metabolic process	6.093e-17

f Pathway: Reactome

Term	FDR
Mitochondrial translation	6.667e-59
Mitochondrial translation elongation	1.142e-56
Mitochondrial translation termination	1.713e-56
Mitochondrial translation initiation	1.815e-55
Translation	6.547e-55
Metabolism	4.840e-53
The citric acid (TCA) cycle and respiratory electron transport	1.738e-52
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	3.205e-30
Respiratory electron transport	2.834e-24
Pyruvate metabolism and Citric Acid (TCA) cycle	8.616e-23







j Location: Panther GO Slim Cellular Component

Term	FDR
mitochondrion	1.634e-52
mitochondrial matrix	4.968e-23
mitochondrial protein-containing complex	2.993e-19
mitochondrial ribosome	1.735e-16
organellar ribosome	2.169e-16
cytoplasm	3.561e-12
ribosome	3.840e-12
intracellular anatomical structure	1.389e-10
ribosomal subunit	4.493e-10
mitochondrial large ribosomal subunit	4.962e-10

k Function: Panther GO Slim Biological Process

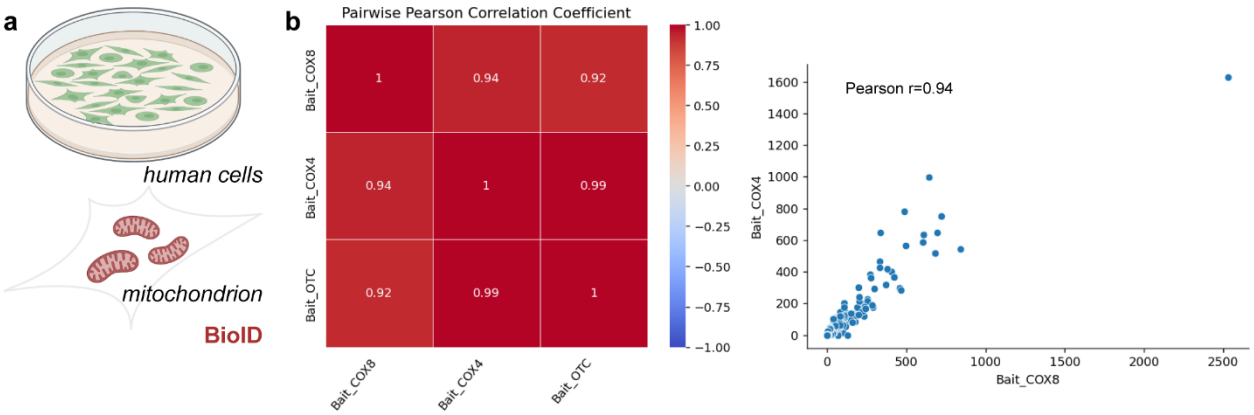
Term	FDR
small molecule metabolic process	2.949e-14
mitochondrial gene expression	1.708e-11
organonitrogen compound biosynthetic process	2.628e-11
cellular amide metabolic process	2.697e-11
translation	2.208e-10
peptide biosynthetic process	2.456e-10
translational elongation	2.650e-10
amide biosynthetic process	8.621e-10
carboxylic acid metabolic process	3.123e-9
peptide metabolic process	3.391e-9

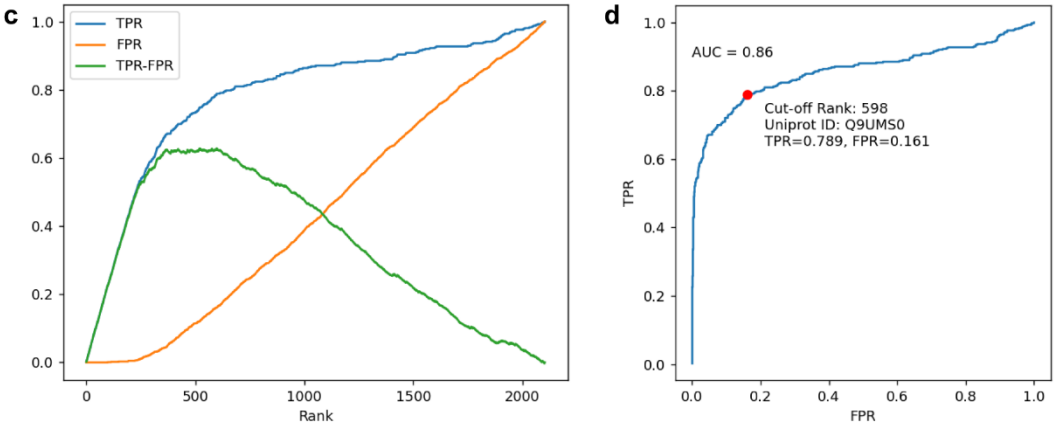
l Pathway: Reactome

Term	FDR
Mitochondrial translation initiation	5.571e-51
Mitochondrial translation elongation	8.357e-51
Mitochondrial translation	1.016e-50
Mitochondrial translation termination	9.547e-50
Translation	1.862e-41
The citric acid (TCA) cycle and respiratory electron transport	6.021e-30
Respiratory electron transport	2.000e-18
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	2.205e-18
Complex I biogenesis	3.401e-14
Metabolism	1.587e-13

Supplementary Figure 7. PEELing analysis of human mitochondrial proteomes.

(**a–f**) Analysis of a mitochondrial nucleoid proteome of human embryonic kidney 293 cells, profiled by (Han *et al.*, 2017) using the peroxidase APEX2 (**Supplementary File 7**). (**g–l**) Analysis of a mitochondrial matrix proteome of human embryonic kidney 293 cells, profiled by (Branon *et al.*, 2018) using the biotin ligase BioID (**Supplementary File 8**). (**b**,**h**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against ratio-based ranking (x-axis). Ratio used: **b**, 126:128; **h**, 126:127. (**c**,**i**) Receiver operating characteristic (ROC) curve. Red dot, cutoff position. AUC, area under the curve. Ratio used: **c**, 126:128; **i**, 126:127. (**d–f**, **j–l**) Protein ontology analyses of post-cutoff proteomes for localization (**d**,**j**), function (**e**,**k**), and pathway (**f**,**l**). FDR, false discovery rate.





e Location: Panther GO Slim Cellular Component

Term	FDR
mitochondrion	1.736e-104
mitochondrial protein-containing complex	2.425e-47
mitochondrial matrix	4.550e-40
mitochondrial ribosome	4.503e-33
organellar ribosome	5.629e-33
ribosome	2.803e-26
cytoplasm	3.443e-21
mitochondrial inner membrane	5.613e-21
ribosomal subunit	6.314e-21
organelle inner membrane	2.011e-20

f Function: Panther GO Slim Biological Process

Term	FDR
cellular metabolic process	1.091e-34
translation	2.811e-30
amide metabolic process	3.434e-30
peptide biosynthetic process	4.012e-30
amide biosynthetic process	4.098e-30
peptide metabolic process	4.430e-28
cellular nitrogen compound biosynthetic process	4.795e-28
cellular biosynthetic process	2.003e-25
cellular nitrogen compound metabolic process	1.880e-24
organonitrogen compound biosynthetic process	2.223e-24

g Pathway: Reactome

Term	FDR
Mitochondrial translation	2.216e-93
Mitochondrial translation elongation	2.932e-92
Mitochondrial translation initiation	4.858e-91
Mitochondrial translation termination	6.477e-91
Translation	3.702e-82
The citric acid (TCA) cycle and respiratory electron transport	8.256e-63
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	1.858e-49
Respiratory electron transport	1.597e-41
Metabolism	1.162e-35
Complex I biogenesis	1.408e-29

Supplementary Figure 8. PEELing analysis of a label-free quantification mass spectrometry dataset.

(**a**) Analysis of a mitochondrial matrix proteome of human embryonic kidney 293 cells, profiled by (Antonicka *et al.*, 2020) using the biotin ligase BioID fused to three different bait proteins COX4, COX8, and OTC and label-free quantification mass spectrometry (**Supplementary File 9**). (**b**) Correlation plots and coefficients. (**c**) True positive rate (TPR, blue), false positive rate (FPR, orange), and their difference (TPR–FPR, green) plotted against the labelled-to-control ratio-based

ranking of the COX4 bait group (x-axis). (**d**) Receiver operating characteristic (ROC) curve, based on the labelled-to-control ratio-based ranking of the COX4 bait group. Red dot, cutoff position. AUC, area under the curve. (**e–g**) Protein ontology analyses for localization (**e**), function (**f**), and pathway (**g**). FDR, false discovery rate.

Supplementary Files

Supplementary File 1. A cell-surface proteome of mouse developing Purkinje cells (Shuster *et al.*, 2022). The authors used horseradish peroxidase (HRP) based cell-surface biotinylation and tandem mass tag (TMT) (Thompson *et al.*, 2003) based quantitative mass spectrometry. TMT tags 129C and 128C were used for labelled samples while 127C and 127N were used for non-labelled controls, producing four labelled-to-control ratios (129C:127C, 129C:127N, 128C:127C, and 128C:127N). Ratios are normalized and $\log_2$-transformed.

Supplementary File 2. Two sets of random values, Random_1 (6th column) and Random_2 (7th column), were generated and added to the data of mouse Purkinje cell cell-surface proteome (Shuster *et al.*, 2022) (1st to 5th columns; from **Supplementary File 1**).

Supplementary File 3. Protein abundance data of the cell-surface proteome of mouse developing Purkinje cells (Shuster *et al.*, 2022), derived from the same mass spectrometry data as **Supplementary File 1**. The authors used horseradish peroxidase (HRP) based cell-surface biotinylation and tandem mass tag (TMT) (Thompson *et al.*, 2003) based quantitative mass spectrometry. TMT tags 129C and 128C were used for labelled samples. Abundance values are normalized and $\log_2$-transformed.

Supplementary File 4. A cell-surface proteome of *Drosophila* mature olfactory projection neurons (Li *et al.*, 2020). The authors used horseradish peroxidase (HRP) based cell-surface biotinylation and tandem mass tag (TMT) (Thompson *et al.*, 2003) based quantitative mass spectrometry. TMT tags 126 and 129C were used for labelled samples while 127N and 128C were used for non-labelled controls, producing four labelled-to-control ratios (126:127N, 129C:127N, 126:128C, and 129C:128C). Ratios are normalized and $\log_2$-transformed.

Supplementary File 5. A nuclear proteome of human embryonic kidney 293 cells (Branon *et al.*, 2018). The authors used the biotin ligase TurboID for spatially-resolved biotinylation in the nucleus and tandem mass tag (TMT) (Thompson *et al.*, 2003) for quantitative mass spectrometry. Two labelled-to-control ratios were provided: TMT tag 128:129 from Replicate 1 and 129:126 from Replicate 2. Ratios are normalized and $\log_2$-transformed.

Supplementary File 6. A nuclear proteome of human embryonic kidney 293 cells (Branon *et al.*, 2018). The authors used the biotin ligase miniTurbo for spatially-resolved biotinylation in the nucleus and tandem mass tag (TMT) (Thompson *et al.*, 2003) for quantitative mass spectrometry.

Two labelled-to-control ratios were provided: TMT tag 130:131 from Replicate 1 and 127:126 from Replicate 2. Ratios are normalized and $\log_2$-transformed.

Supplementary File 7. A mitochondrial nucleoid proteome of human embryonic kidney 293 cells (Han *et al.*, 2017). The authors used the peroxidase APEX2 for spatially-resolved biotinylation in the mitochondrion and tandem mass tag (TMT) (Thompson *et al.*, 2003) for quantitative mass spectrometry. Two labelled-to-control ratios were provided: TMT tag 126:128 from Replicate 1 and 129:131 from Replicate 2. Ratios are normalized and $\log_2$-transformed.

Supplementary File 8. A mitochondrial matrix proteome of human embryonic kidney 293 cells (Branon *et al.*, 2018). The authors used the biotin ligase BioID for spatially-resolved biotinylation in the mitochondrion and tandem mass tag (TMT) (Thompson *et al.*, 2003) for quantitative mass spectrometry. Two labelled-to-control ratios were provided: TMT tag 126:127 from Replicate 1 and 130:131 from Replicate 2. Ratios are normalized and $\log_2$-transformed.

Supplementary File 9. A mitochondrial matrix proteome of human embryonic kidney 293 cells (Antonicka *et al.*, 2020). The authors used the biotin ligase BioID fused to three different bait proteins COX4, COX8, and OTC for spatially-resolved biotinylation in the mitochondrion and label-free quantification mass spectrometry. One labelled-to-control ratio was provided for each bait.

Supplementary File 10. True positive (TP) reference for the cell surface, as of July 2, 2023.

Supplementary File 11. 10%-coverage TP reference for the cell surface, produced by randomly removing 90% of genes from **Supplementary File 10**.

Supplementary File 12. False positive (FP) reference for the cell surface, as of July 2, 2023.

Supplementary File 13. 10%-coverage FP reference for the cell surface, produced by randomly removing 90% of genes from **Supplementary File 12**.

Supplementary File 14. True positive (TP) reference for the nucleus, as of July 2, 2023.

Supplementary File 15. False positive (FP) reference for the nucleus, as of July 2, 2023.

Supplementary File 16. True positive (TP) reference for the mitochondrion, as of July 2, 2023.

Supplementary File 17. False positive (FP) reference for the mitochondrion, as of July 2, 2023.