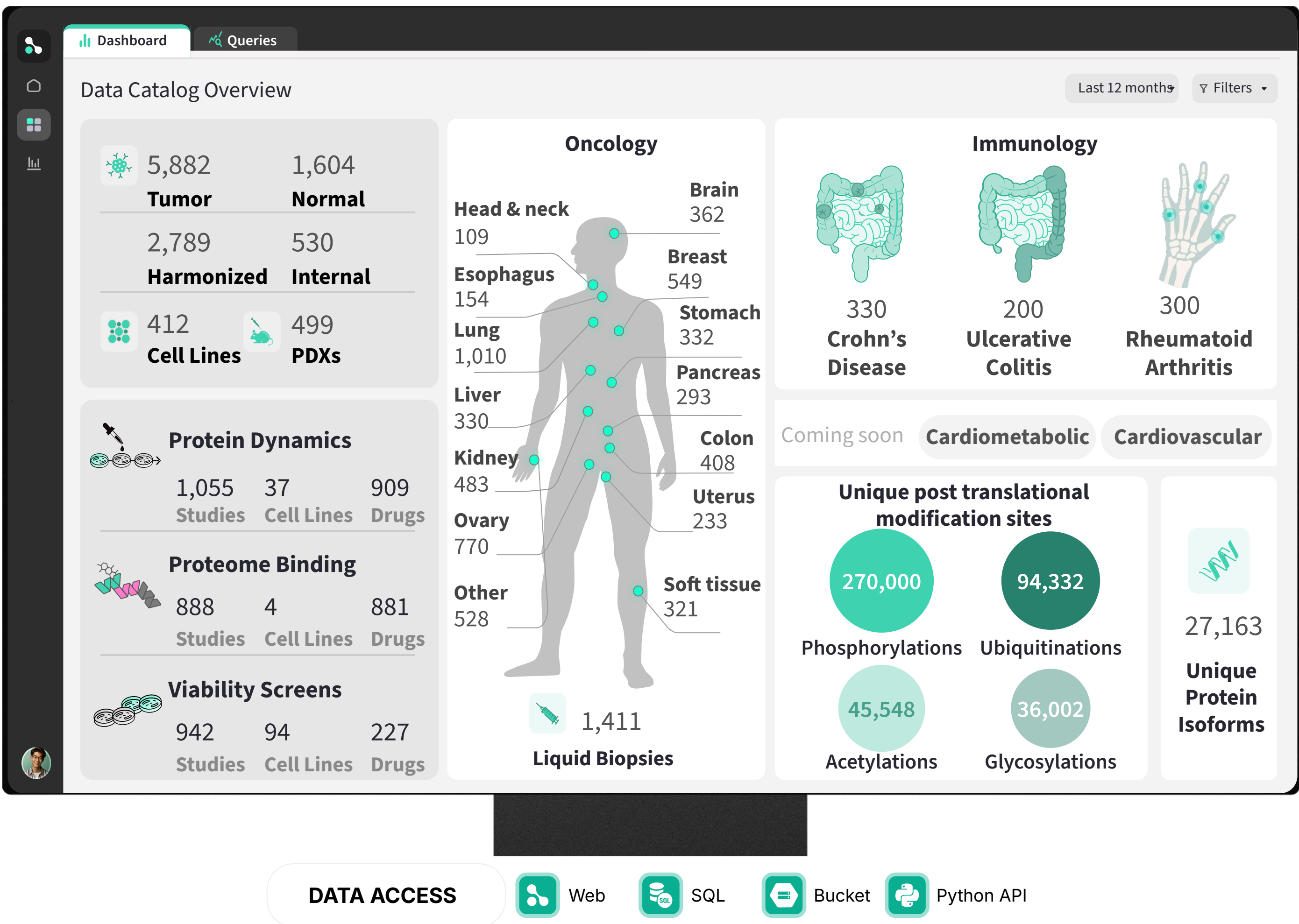


Proteogenomic clinical atlas with diverse datasets

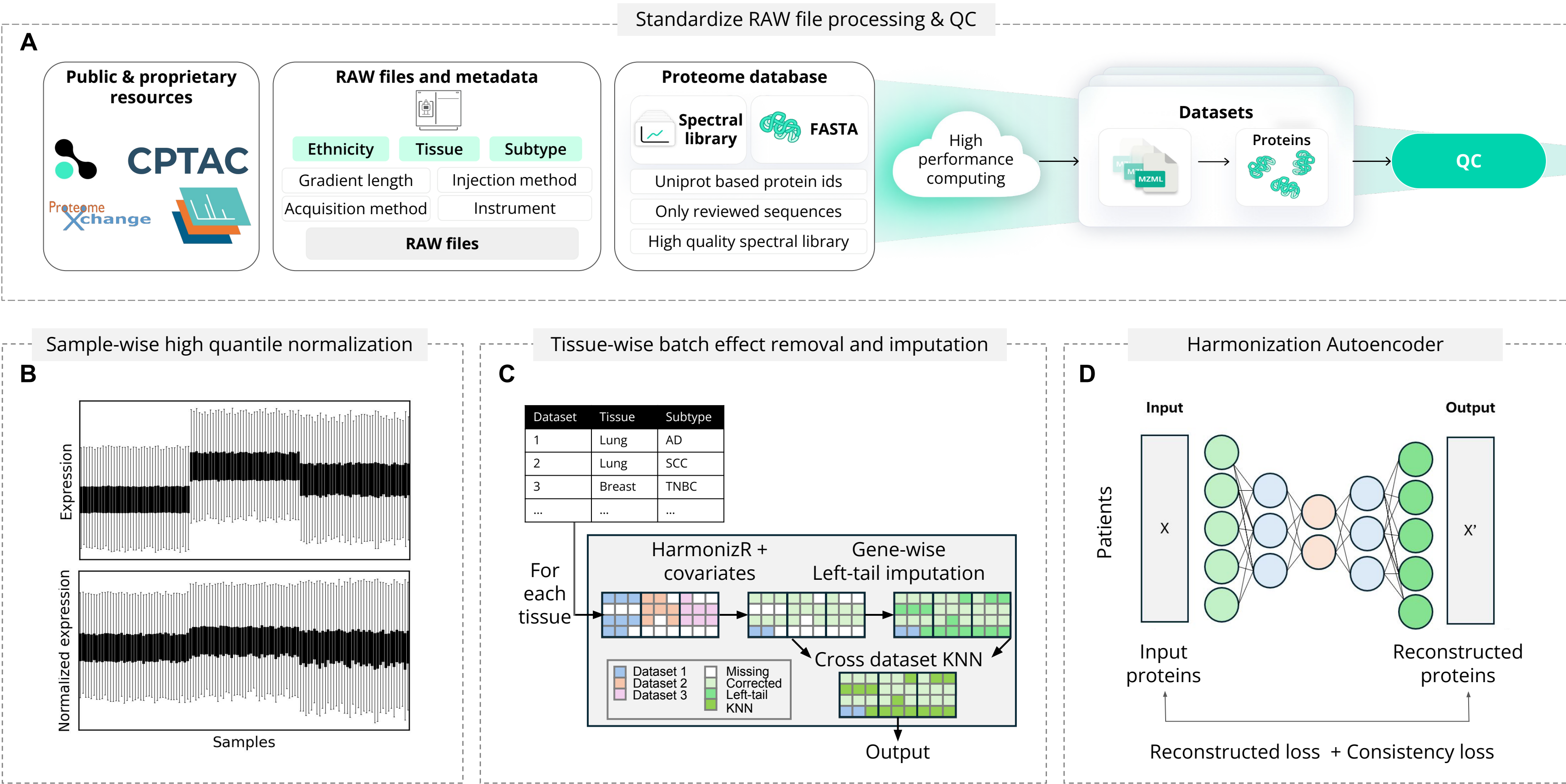


We assembled a novel proteogenomic atlas with high-quality proteomics and PTM data from clinically annotated patient samples across diseases, including cancer, Crohn's, and colitis. Continuously expanding, it integrates treatments, outcomes, and medical data to support precision medicine and drug discovery.

As the growth of proteomics datasets is at an unprecedented high, we developed a harmonization strategy for large-scale integrative analyses. Our algorithm enables the direct comparison of datasets, regardless of differences in method, cohort size, experimental design, or tissue type.

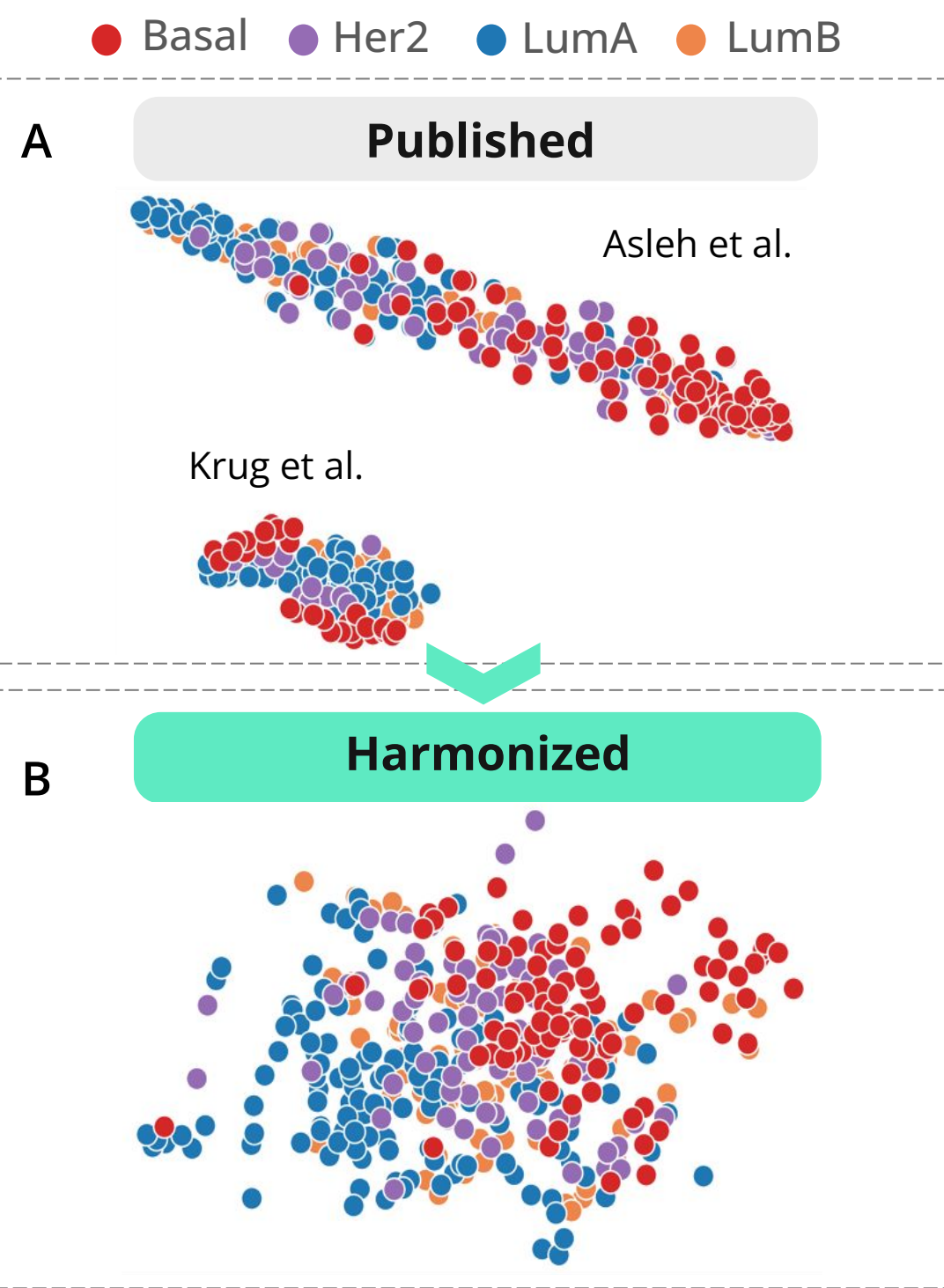
The resulting harmonized pan-cancer proteomics dataset is a robust resource, designed for high-impact applications such as target discovery, biomarker identification, and indication expansion.

Multistep harmonization mitigates proteomics batch effects



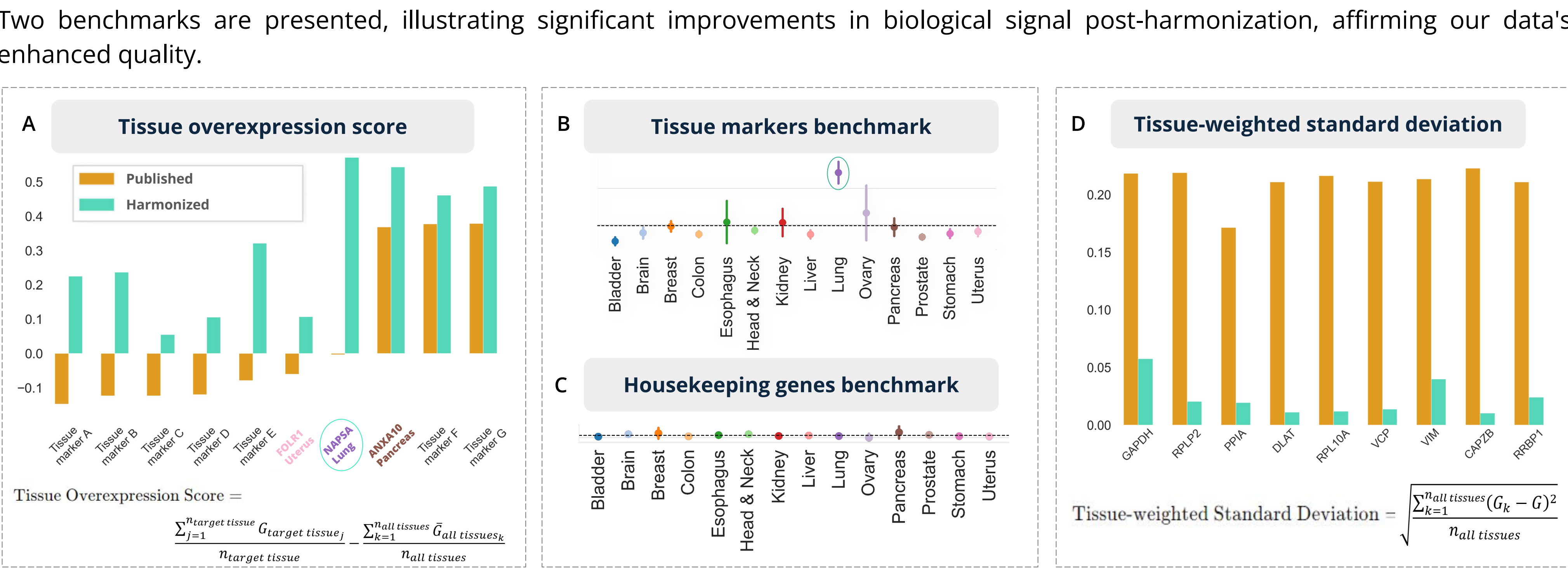
Our approach tackles different aspects of batch effects in proteomics data beginning with (A) standardized reanalysis of raw data using a cloud based pipeline (B) sample-wise quantile normalization (C) indication-wise imputation coupled with batch-effect removal and (D) a pan-cancer autoencoder for biological signal smoothing.

Biological signal enhancement



Efficient batch effect removal and enhancement of biological signal after harmonization (A) Breast subtypes dataset from two different studies and labs (B) Data after harmonization.

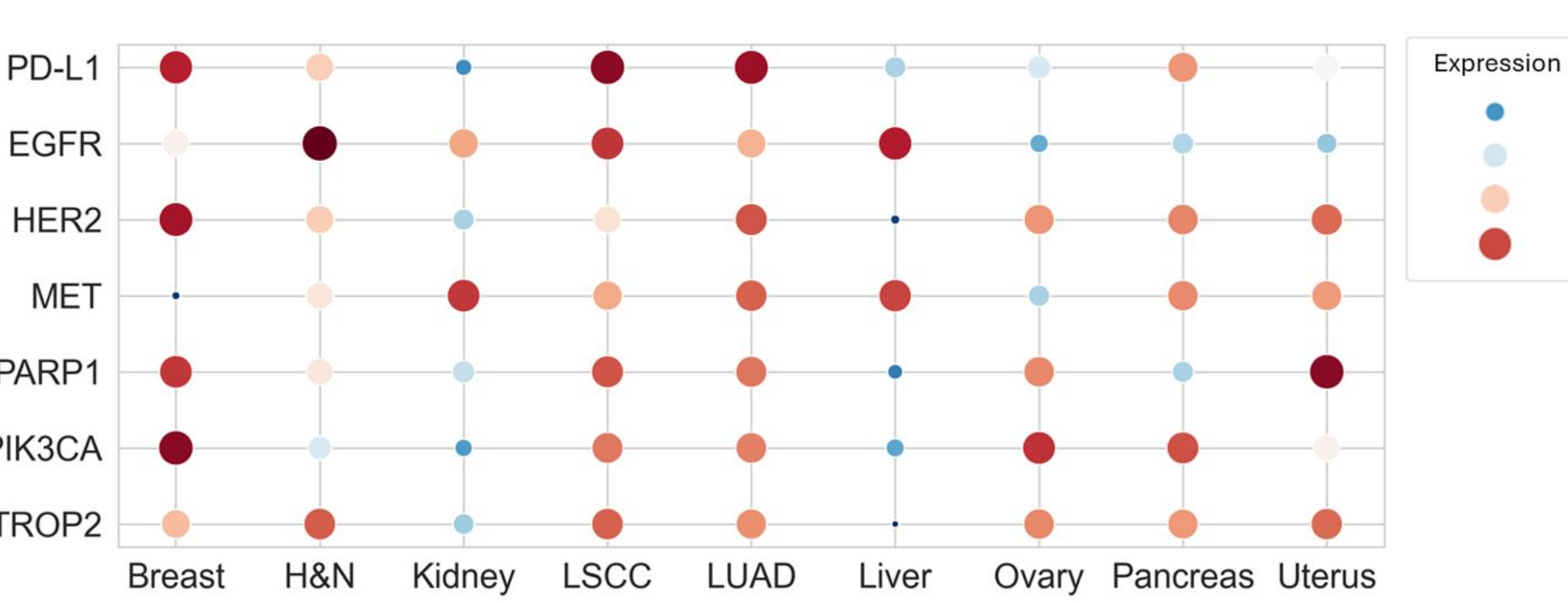
Biologically-driven harmonization benchmark



Tissue markers benchmark
50 literature validated tissue specific cancer markers were collected from 2 pan-cancer datasets. These were used to evaluate the harmonized enhancement of tissue biological signal (A, B).

Housekeeping genes benchmark
To ensure consistent quality and comparability within our datasets, the standard deviation of housekeeping genes across different indications is measured and expected to be minimized (C, D).

Pan-cancer biomarker prevalence

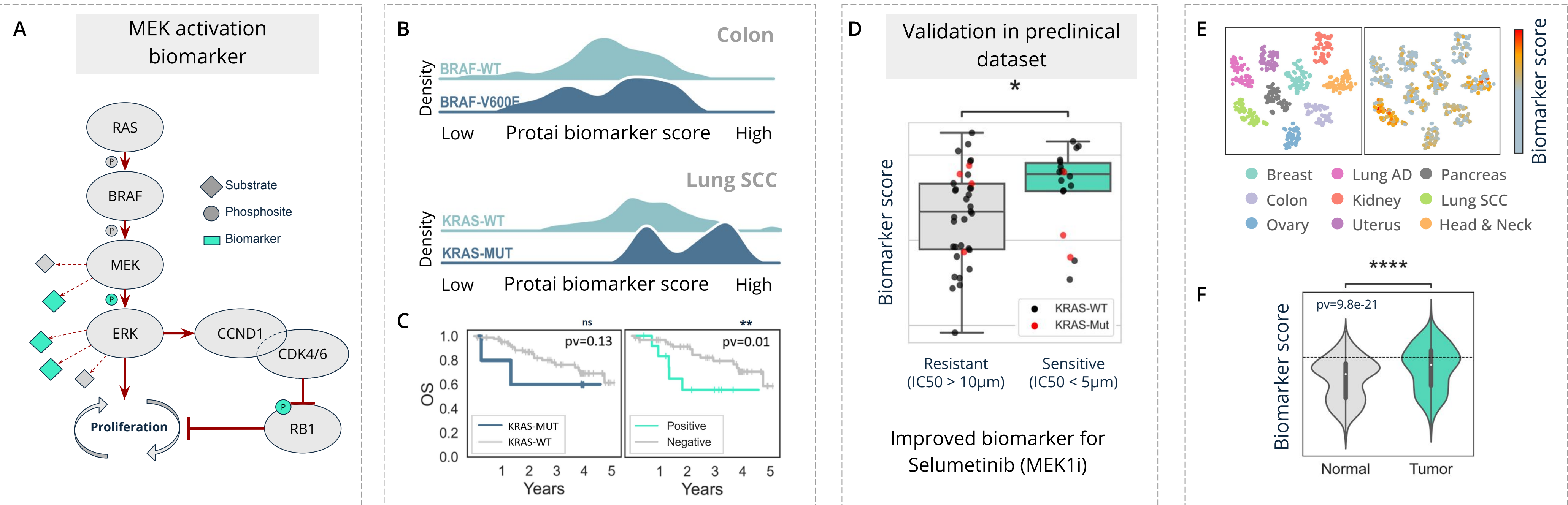


To assess the potential of harmonized atlas for precision medicine, we review protein expression levels of several druggable targets across multiple indications.

Reassuringly, HER2 and PIK3CA levels in breast cancer are high, in agreement with respective drugs being approved therapies in this indication.

Additionally, we observe overexpression of MET and EGFR in hepatocellular carcinoma (HCC), presenting a promising avenue for targeted therapies. Given their role in tumor progression and treatment resistance, these markers could help guide new therapeutic strategies for HCC.

Enhancing patient selection for MEKi with phospho-proteomic signature of MEK activation



A novel phosphoproteomic biomarker for MEK1 inhibitors arising from the atlas. (A) Mutations alone do not fully account for the downstream BRAF signaling intricacies. (B) The biomarker captures MEK1 activated tumors that are BRAF/KRAS WT and (C) achieves a more pronounced survival curve. (D) Validation of the novel biomarker within a pan cancer cell line dataset, utilizing the MEK1 inhibitor Selumetinib, distinguishes between sensitive and resistant cell lines. (E) The biomarker exhibits a potential in identifying LSCC patients who may benefit from MEKi therapy and for therapeutic indication expansion and improved patient selection. (F) Leveraging Protai's extensive clinical atlas of tumor and normal samples, our phospho-proteomic biomarker demonstrates significantly higher activity in tumor samples compared to normal ones.

Summary

- **Proteogenomic atlas** enables leveraging proteomics data growth for precision medicine.
- **AI based harmonization** corrects batch effects while preserving biological signals.
- **Benchmarking ensures quality** by balancing batch correction and biology.
- **Pan-cancer use cases** include
 - Biomarker prevalence
 - Target discovery
 - Indication expansion

References

1. Perez-Riverol et al. Nucleic Acids Res. (2022)
2. Gillette et al. Cell (2020)
3. Asleh et al. Nature Communications (2022)
4. Satpathy et al. Cell (2021)
5. Soltis et al. Cell (2022)
6. Krug et al. Cell (2020)
7. Hu et al. Cell Reports (2020)
8. Bausch-Fluck et al. PNAS (2018)
9. Song et al. Nature Communications (2022)
10. Zhou et al. Journal of Hematology & Oncology (2020)



contact@protai.bio