A large scale proteogenomic atlas for precision oncology

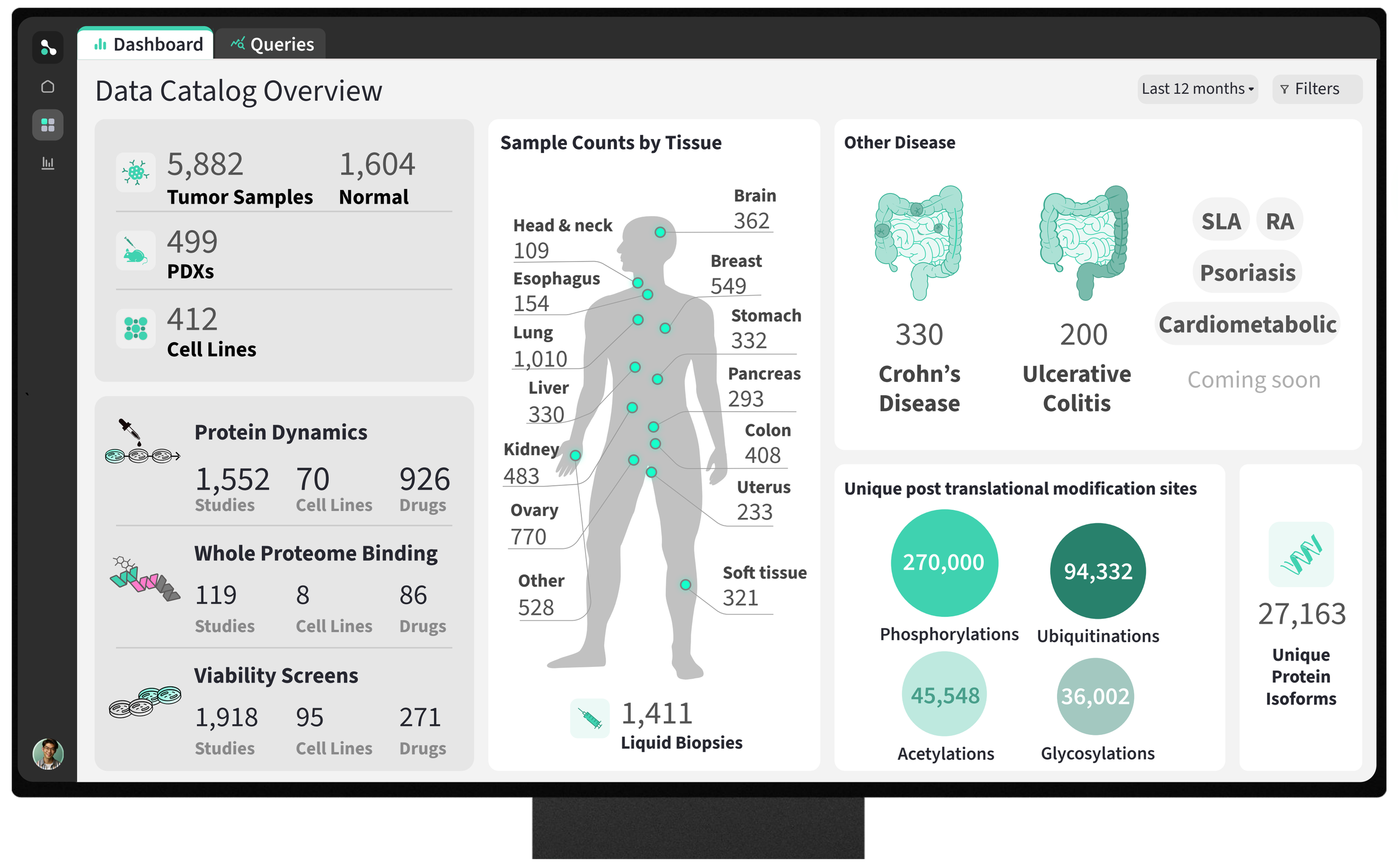
Timothy A. Yap[1], Nitzan Simchi[2], Ofer Givton[2], Dina Daitchman[2], Joseph Rinberg[2], Alon Shtrikman[2], Eran Seger[2], Kirill Pevzner[2]

[1] Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[2] Protai Bio, Ramat Gan, Israel

FPN: 190P

Proteogenomic clinical atlas with diverse cancers and conditions

We introduce a novel atlas containing large-scale and harmonized proteogenomic data and demonstrate its use for biomarker and target discovery. The atlas contains high-quality datasets of proteomics and post translational modifications, derived from clinically annotated patient samples of both disease and normal tissues. It includes samples from cancer patients, as well as Crohn's and Colitis disease, and is continuously expanding to include additional indications. The clinical annotations incorporate treatments, outcomes and general medical data. The atlas offers insights into patient outcomes and drug sensitivities in tumor and normal samples, supporting precision medicine applications.

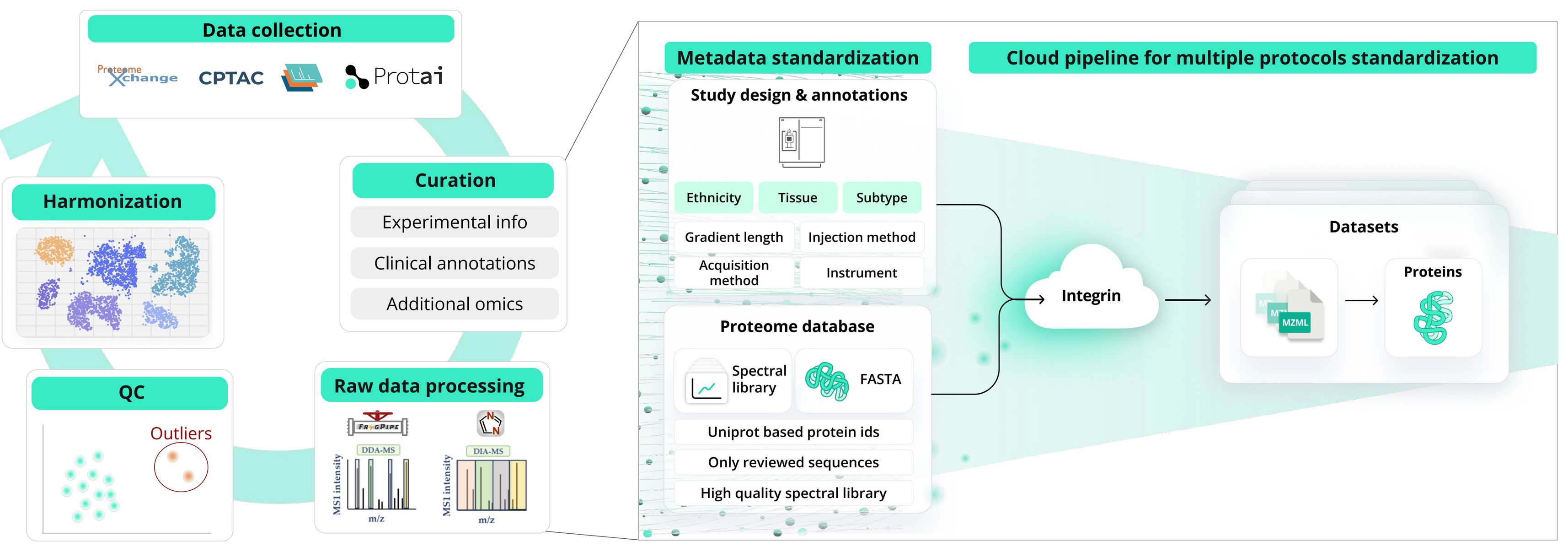

Dashboard
Queries
Data Catalog Overview
Last 12 months
Filters
5,882 Tumor Samples
1,604 Normal
499 PDXs
412 Cell Lines
Protein Dynamics
1,552 Studies
70 Cell Lines
926 Drugs
Whole Proteome Binding
119 Studies
8 Cell Lines
86 Drugs
Viability Screens
1,918 Studies
95 Cell Lines
271 Drugs
Sample Counts by Tissue
Brain 362
Head & neck 109
Breast 549
Esophagus 154
Stomach 332
Lung 1,010
Pancreas 293
Liver 330
Colon 408
Kidney 483
Uterus 233
Ovary 770
Other 528
Soft tissue 321
1,411 Liquid Biopsies
Other Disease
330 Crohn's Disease
200 Ulcerative Colitis
SLA
RA
Psoriasis
Cardiometabolic
Coming soon
Unique post translational modification sites
270,000 Phosphorylations
94,332 Ubiquitinations
45,548 Acetylations
36,002 Glycosylations
27,163 Unique Protein Isoforms

DATA ACCESS

- Python API
- SQL
- Bucket
- Web

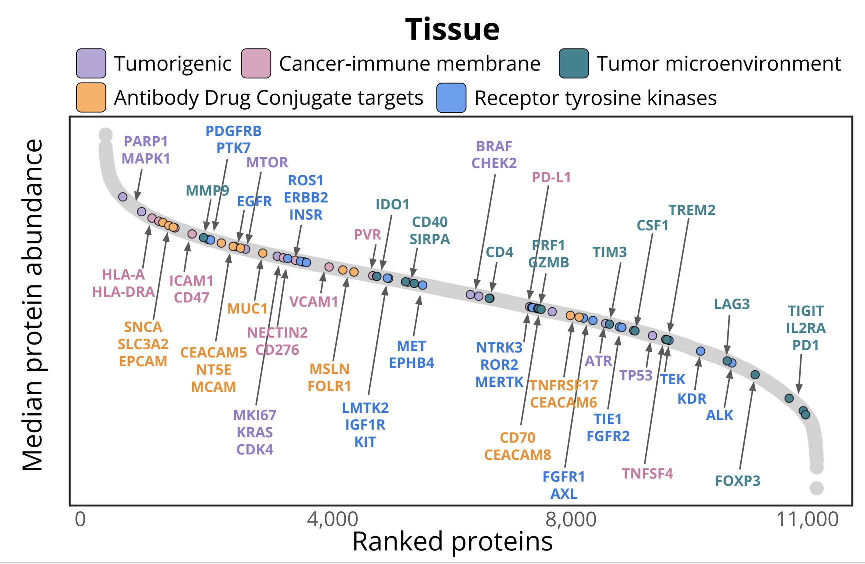
Continuously expanding atlas

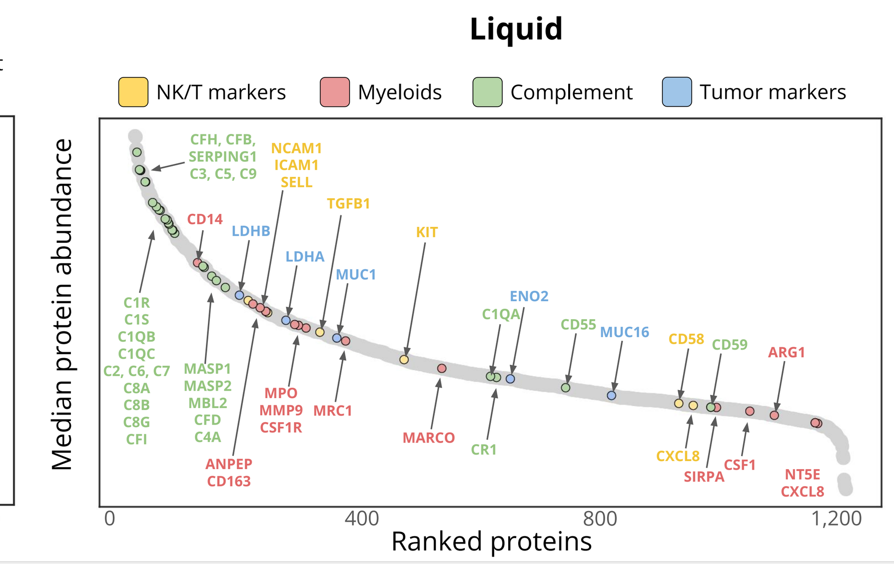
The atlas is continuously expanding by integration of data from several sources. Datasets, study design and clinical annotations are standardised, followed by data processing in a cloud-based pipeline. Quality control and harmonization are subsequently applied.


Data collection
ProteomeXchange
CPTAC
Protai
Harmonization
Curation
Experimental info
Clinical annotations
Additional omics
QC
Outliers
Raw data processing
DDA-MS
DIA-MS
MS1 intensity
m/z
Metadata standardization
Cloud pipeline for multiple protocols standardization
Study design & annotations
Ethnicity
Tissue
Subtype
Gradient length
Injection method
Acquisition method
Instrument
Proteome database
Spectral library
FASTA
Uniprot based protein ids
Only reviewed sequences
High quality spectral library
Integrin
Datasets
MZML
Proteins

Diverse protein markers captured in tumor and biofluid samples

Proteins across a range of biological processes and themes are captured. Tissue analysis identifies oncoproteins such as KRAS, DNA damage (ATR), membrane bound immune regulators (PD-L1) and immune related (PD1). Blood analysis identifies NK/T markers (TGFB1), Myeloids CSF1) and diagnostic markers (MUC16).

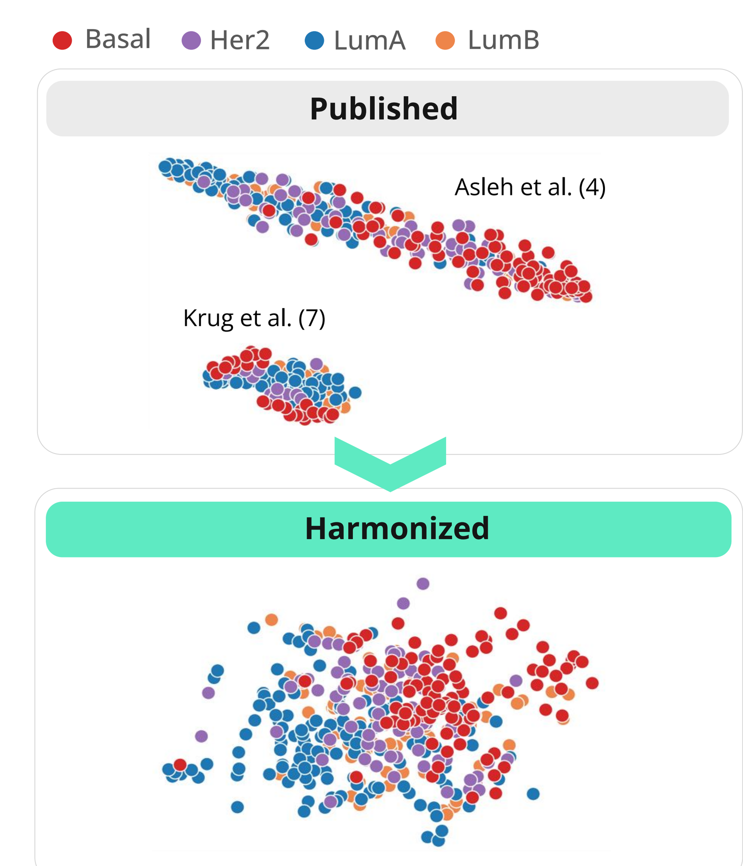

Tissue
Tumorigenic
Cancer-immune membrane
Tumor microenvironment
Antibody Drug Conjugate targets
Receptor tyrosine kinases
Median protein abundance
Ranked proteins
0
4,000
8,000
11,000
PARP1
MAPK1
PDGFRB
PTK7
MTOR
MMP9
EGFR
ROS1
ERBB2
INSR
IDO1
PVR
CD40
SIRPA
BRAF
CHEK2
PD-L1
CD4
PRF1
GZMB
TIM3
CSF1
TREM2
LAG3
TIGIT
IL2RA
PD1
HLA-A
HLA-DRA
ICAM1
CD47
MUC1
VCAM1
SNCA
SLC3A2
EPCAM
CEACAM5
NT5E
MCAM
NECTIN2
CD276
MSLN
FOLR1
MET
EPHB4
NTRK3
ROR2
MERTK
TNFRSF17
CEACAM6
ATR
TP53
TEK
KDR
ALK
MKI67
KRAS
CDK4
LMTK2
IGF1R
KIT
CD70
CEACAM8
TIE1
FGFR2
FGFR1
AXL
TNFSF4
FOXP3


Liquid
NK/T markers
Myeloids
Complement
Tumor markers
Median protein abundance
Ranked proteins
0
400
800
1,200
CFH, CFB, SERPING1
C3, C5, C9
NCAM1
ICAM1
SELL
CD14
LDHB
TGFB1
LDHA
MUC1
KIT
ENO2
C1QA
CD55
MUC16
CD58
CD59
ARG1
C1R
C1S
C1QB
C1QC
C2, C6, C7
C8A
C8B
C8G
CFI
MASP1
MASP2
MBL2
CFD
C4A
MPO
MMP9
CSF1R
MRC1
ANPEP
CD163
MARCO
CR1
CXCL8
SIRPA
CSF1
NT5E
CXCL8

Harmonization reduces technical noise and enhances biological signal

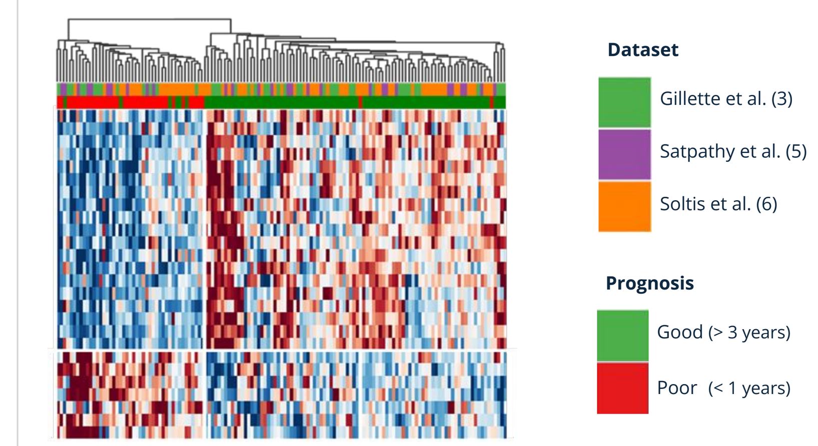
From batch to breast cancer subtypes

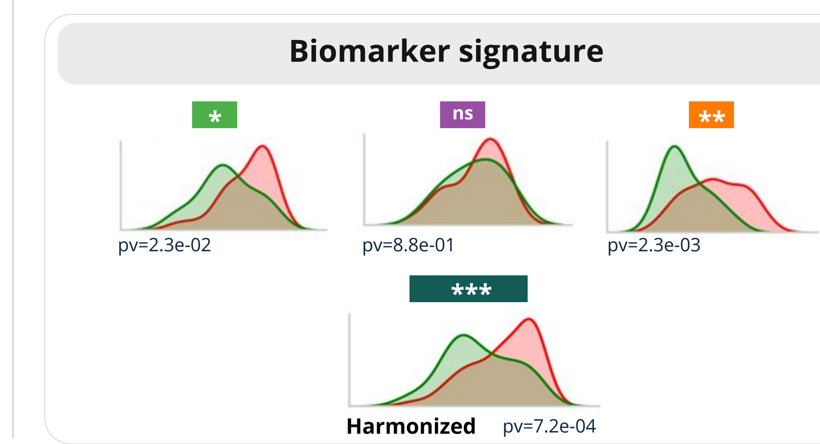
Harmonization overcomes batch effects between different labs, studies, methods and protocols. In this case, breast subtypes from two different studies of different labs are presented, showing that the biological signal is preserved while the dataset batch effect is removed.


Basal
Her2
LumA
LumB
Published
Asleh et al. (4)
Krug et al. (7)
Harmonized

Amplifying prognostic signal in Lung

By leveraging a larger sample size, this approach enables identification of a biomarker that remained undetected when analyzing each dataset separately. This finding highlights the importance of harmonization for revealing previously undiscovered insights, and its applicability extends to other applications.


Dataset
Gillette et al. (3)
Satpathy et al. (5)
Soltis et al. (6)
Prognosis
Good (> 3 years)
Poor (< 1 years)

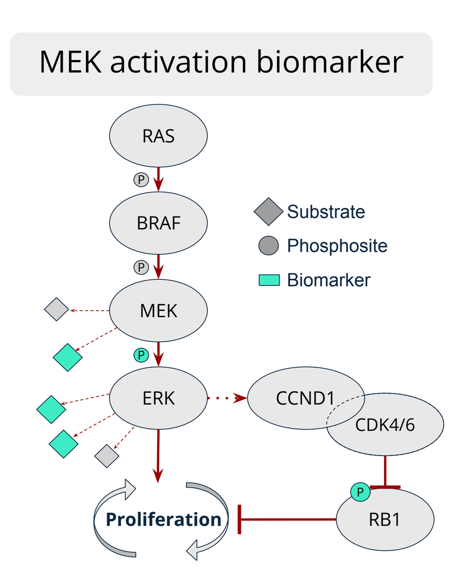

Biomarker signature
*
ns
**
pv=2.3e-02
pv=8.8e-01
pv=2.3e-03

Harmonized
pv=7.2e-04

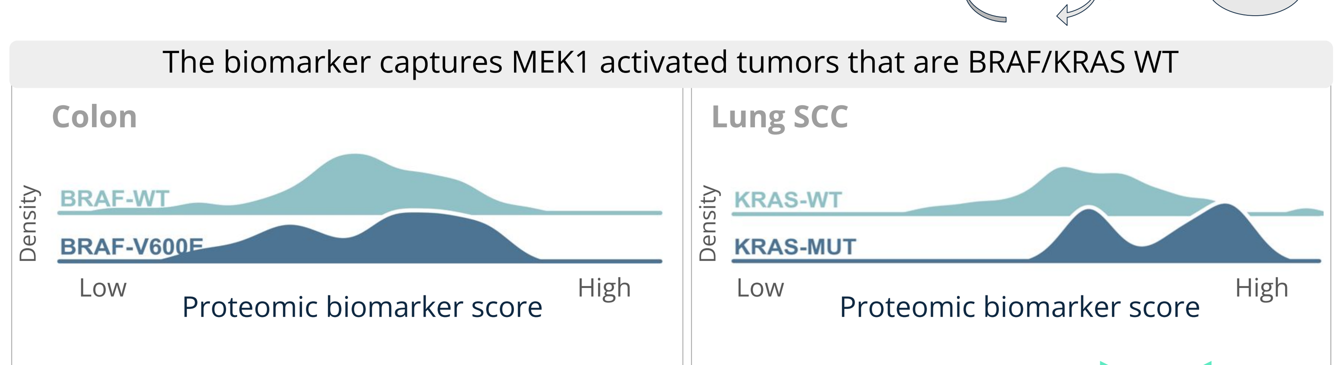
Enhancing patient selection for MEKi with phospho-proteomic signature of MEK activation

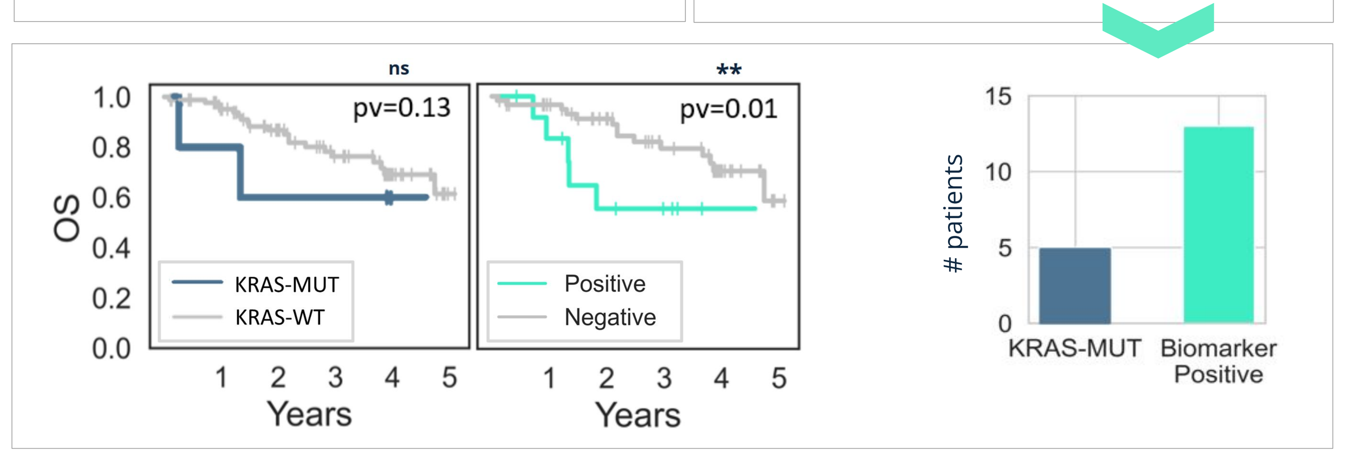
From mutations to function

Using the atlas, we found a phosphoproteomic biomarker for MEK1 inhibitors, rooted in the downstream signaling of BRAF. This biomarker offers a broader perspective by encompassing additional mechanisms of pathway activation beyond the conventional BRAF or KRAS mutations. The analysis reveals that mutations alone may not fully account for the downstream signaling intricacies. The biomarker achieves a more pronounced survival curve, illustrating its effectiveness in capturing the complex dynamics of pathway activation.


MEK activation biomarker
RAS
BRAF
MEK
ERK
CCND1
CDK4/6
RB1
Proliferation
Substrate
Phosphosite
Biomarker

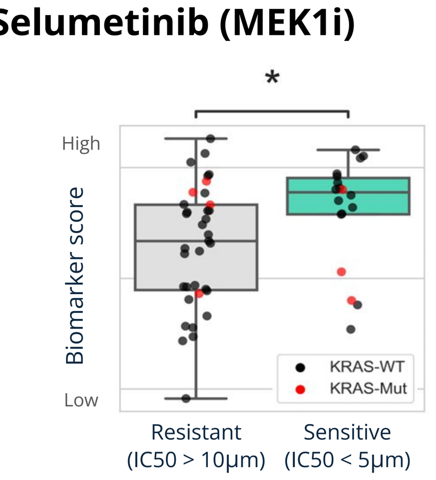
The biomarker captures MEK1 activated tumors that are BRAF/KRAS WT


Colon
BRAF-WT
BRAF-V600E
Lung SCC
KRAS-WT
KRAS-MUT
Density
Low
High
Proteomic biomarker score


ns
**
pv=0.13
pv=0.01
OS
Years
KRAS-MUT
KRAS-WT
Positive
Negative
patients
KRAS-MUT
Biomarker Positive

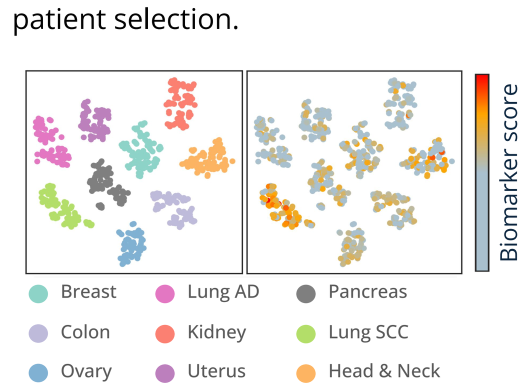
Improved biomarker for Selumetinib (MEK1i)

Validation of the novel proteomic biomarker is shown within a pan cancer cell line dataset, utilizing the MEK1 inhibitor Selumetinib. This validation underscores the biomarkers ability in distinguishing between Selumetinib sensitive and resistant cell lines, and shows potential for treatment in other indications.

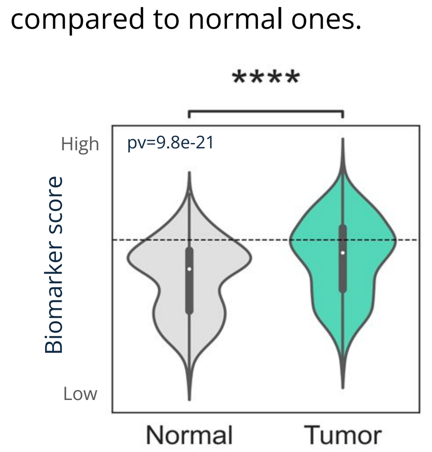

*
High
Low
Biomarker score
KRAS-WT
KRAS-Mut
Resistant (IC50 > 10µm)
Sensitive (IC50 < 5µm)

Indication expansion potential

Proteomic biomarker exhibits considerable potential in identifying LSCC patients who may benefit from MEKi therapy. Additionally, the analysis sheds light on patients from other indications where our biomarker suggests a potential for therapeutic indication expansion and improved patient selection.


Biomarker score
Breast
Lung AD
Pancreas
Colon
Kidney
Lung SCC
Ovary
Uterus
Head & Neck

Leveraging the extensive proteomic clinical atlas of tumor and normal samples, our phospho-proteomic biomarker demonstrates significantly higher activity in tumor samples compared to normal ones.

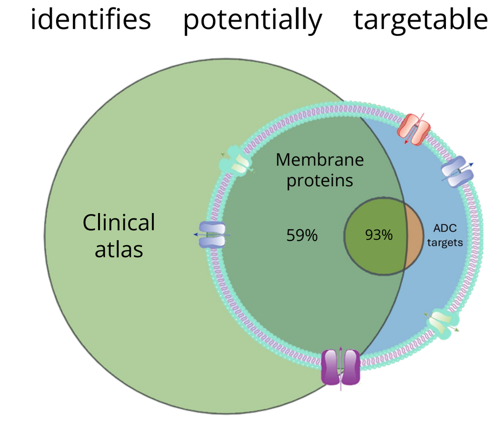


pv=9.8e-21
High
Low
Biomarker score
Normal
Tumor

Novel ADC targets discovery enabled by protein modification analysis

Membrane proteins are widely detected

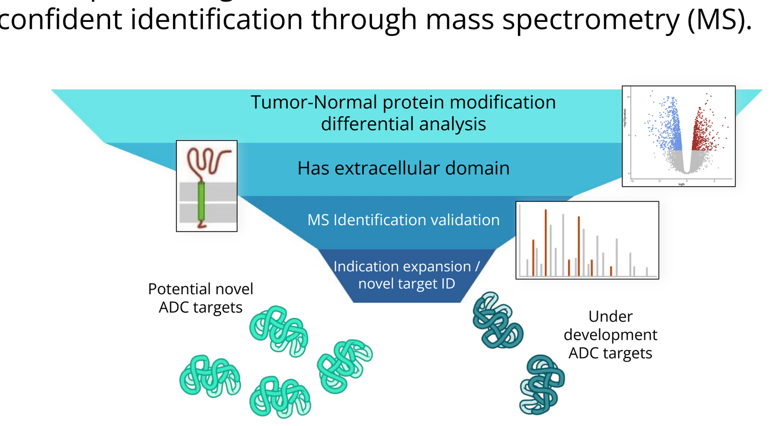
Clinical proteomic atlas includes most of clinical phase ADC targets and identifies potentially targetable membrane proteins.


Clinical atlas
Membrane proteins
59%
93%
ADC targets

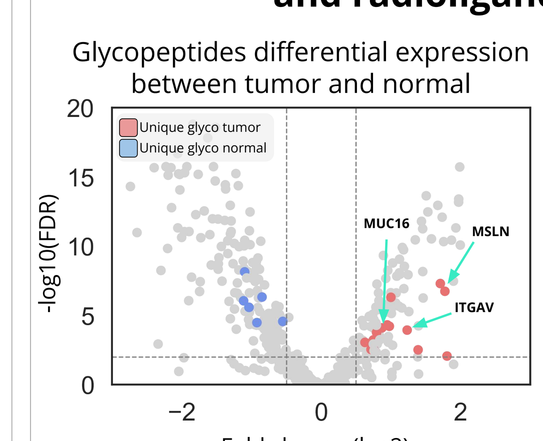
The membrane proteins selected were present in at least two of three datasets: the Cell Surface Protein Atlas, Uniprot, and COMPARTMENTS database (9).

ADC target discovery pipeline

Membrane proteins undergo several filtration steps, which include differential expression in tumors compared to normal tissues at the post-translational modification (PTM) level, possessing an extracellular domain, and having confident identification through mass spectrometry (MS).

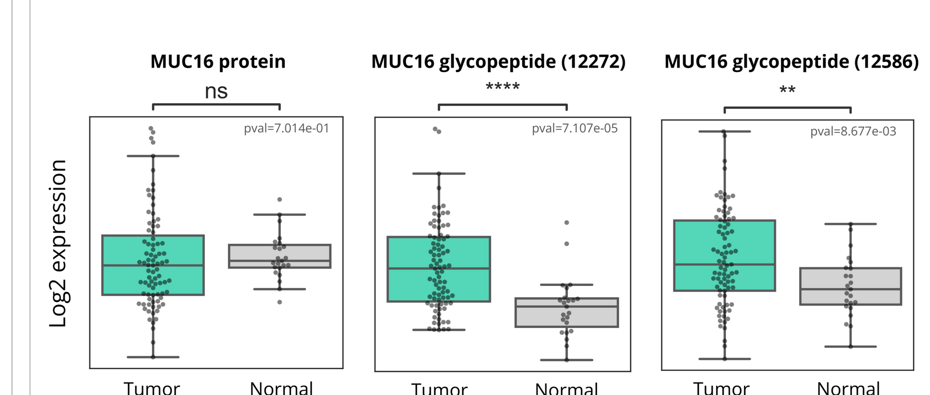

Tumor-Normal protein modification differential analysis
Has extracellular domain
MS Identification validation
Indication expansion / novel target ID
Potential novel ADC targets
Under development ADC targets

Glycoproteins unlock potential for new ADC and radioligand targets


Glycopeptides differential expression between tumor and normal
Unique glyco tumor
Unique glyco normal
-log10(FDR)
Fold change (log2)
MUC16
MSLN
ITGAV

Here, we present protein glycoforms with differences between tumor and normal samples in ovarian cancer patient samples. The highlighted red and blue dots represent glycopeptides from membrane proteins, where the distinction between tumor and normal is not evident at the protein level and could therefore be missed.

MUC16 (CA-125) is a membrane protein utilized as a prognostic marker in ovarian cancer. The expression of MUC16 by itself does not differentiate between tumor and normal samples. However, MUC16 is highly glycosylated in tumor compared to normal (8), revealing a potential tumor-normal expression window, needed for ADC therapy.


MUC16 protein
ns
pval=7.014e-01
MUC16 glycopeptide (12272)

pval=7.107e-05
MUC16 glycopeptide (12586)
**
pval=8.677e-03
Log2 expression
Tumor
Normal

Summary

- A high quality, routinely updated, clinical atlas of cancer proteomics and PTMs is presented.
- Increased sample size translates into qualitative improvement, offering unique precision oncology insights:
 - Predictive and prognostic biomarker discovery.
 - Indication expansion.
 - Discovery of new therapeutic targets and combinations.
- Novel ADC target discovery pipeline based on proteomics and protein modifications suggest new targets and indication expansion for ADC targets under development.

References

1. Shenoy et al. Integrative Proteomic Profiling in High-Grade Serous Ovarian Carcinoma: Unraveling Biomarkers and Therapeutic Targets. ESMO; 2024. FPN # 1957P
2. Perez-Riverol et al. Nucleic Acids Res. (2022)
3. Gillette et al. Cell (2020)
4. Asleh et al. Nature Communications (2022)
5. Satpathy et al. Cell (2021)
6. Soltis et al. Cell (2022)
7. Krug et al. Cell (2020)
8. Hu et al. Cell Reports (2020)
9. Bausch-Fluck et al. PNAS (2018)

contact@protai.bio

Disclosures: T.A.Y. is a scientific advisory board member of Protai Bio. N.S., O.G., J. R., A. S., E. S., K. P. are employees of Protai Bio and own stock in Protai Bio.