Integrative Analysis of Proteomic, Transcriptomic, and FACS-based Surface Marker Data for ADC Target Discovery in Cancer Cell Lines

N. Obier[1], V. Vuaroqueaux[2], J. Krumm[3], A-L Peille[2] , D. Feger[1] ,S. Ulrich[1], H. Weber[1], J. Wallner[3], H. Hahne[3], J. Ehlert[1]

[1]Reaction Biology Europe GmbH, Freiburg, Germany,[2]APEX OncoScience SAS, Mulhouse, France,[3]OmicScouts GmbH, a Momentum Biotechnologies company, Freising, Germany

Introduction

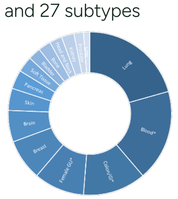
In the preclinical space of cancer drug discovery, multi-omics characterization of cancer models is essential for model selection, understanding mechanisms of action, and early biomarker discovery.

We launched a panel of 160 cell lines (CL Proliferation Panel) to study drug responses across a large variety of cancer types and validated their molecular characteristics at both the transcriptomic and genomic levels.

Recently, we characterized this set of models at the proteomic level using mass spectrometry, identifying a total of more than 14.000 proteins.

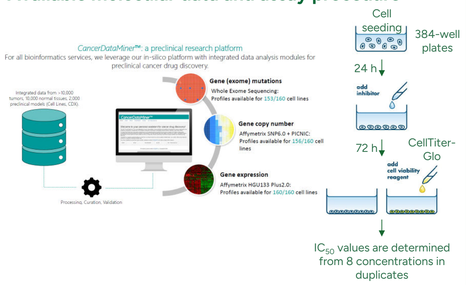
In the present work, we aim to investigate the relevance of this new dataset in the daily practice of cancer model utilization.

ProLiFilerTM cell line in vitro cellular assay platform: The panel currently consists of 160 human tumor cell lines (CLs) derived from 15 tumor types and 27 subtypes



Entities	Number of CLs
Lung (Non-Small and Small Cell Lung tumors)	33
Blood (Leukemia, Lymphoma, Myeloma)	29
Colon/GI	20
Female GU (Ovary, Uterus)	16
Breast	13
Brain (Glioblastoma, Neuroblastoma)	10
Skin (Melanoma)	7
Pancreas	6
Soft Tissue	5
Bladder	4
Bone	4
Head and Neck	4
Kidney	4
Prostate	3
Liver	2

Available molecular data and assay procedure

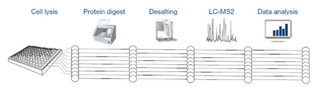


New Proteomic mass spectrometry

ProteomeScout™ – Experimental conditions and workflow

Labelfree quantification – Data independent acquisition (DIA)

- SDS-based cell lysis including sonication
- Determination of protein amount (BCA assay)
- Digestion of 50ug input using a KingFisher magnet handler
- Desalting of peptides using an Agilent Bravo system
- Measurement of desalted peptides with a 120 min gradient (LF-DIA) using a Neo Vanquish LC coupled to an Exploris Orbitrap
- Raw file processing using DIANN
- Data analysis using OmicScouts proprietory data analysis workflow



MS methods: Cell pellets were lysed in lysis buffer consisting of 5% SDS and 50 mM Tris-HCl (pH 8.5)
Total protein was determined using a BCA assay (Thermo Fisher Scientific) according to the manufacturer's instructions
Proteins were further digested into peptides using a magnetic SP3-like protocol [1] with 1:25 trypsin:protein over night
Peptides were desalted by C18 solid-phase extraction using Bravo C18 Cartridges (Agilent) as described in [2] and dried *in vacuo*
Desalted peptides were measured on an Orbitrap Exploris480 equipped with a Neo Vanquish LC system.
Sample were measured in a data-independent acquisition (DIA) mode, and an active gradient of 120 min. Raw data was processed with DIA-NN and subjected to our proprietary data analysis pipeline.

[1] Hughes et al, Nat Protoc. 2019 Jan;14(1):68-85; [2] Rappsilber et al, Nat Protoc. 2007;2(8):1896-906

Results (Protein Assay QC)

Proteomic Profiles Show High Reproducibility Across Duplicate Samples. PCA analysis demonstrates that replicates cluster together showing their similar protein profiles. Euclidean distance analysis confirmed numerically high similarity between duplicates.

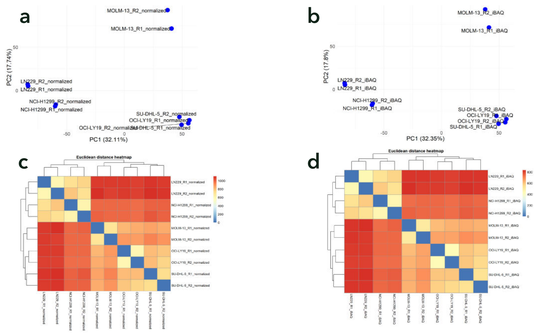


Fig. 1: PCA and Euclidean distance analysis of cell line proteomic profile replicates. Five cell lines were analyzed in duplicate. Principal component analysis (PCA) (a–b) and Euclidean distance (c–d) were used to assess the reproducibility of the mass spectrometry–based proteomic profiles. Analyses were performed using both normalized (a & c) and iBAQ (Intensity-Based Absolute Quantification, (b & d)) data. Prior to analysis, datasets were curated by removing proteins with missing values across all samples, resulting in a total of 1,222 proteins. Data were log2-transformed, and missing or zero values were replaced with 1 for downstream analysis. For PCA, numbers in parentheses indicate the percentage of total variance explained by each principal component.

Results (Exploratory Analysis)

Overall, proteomics profiles, consistent with transcriptomics data, separated hematological from solid tumor cell lines and further distinguished tumor origins within the solid tumor group.

Analysis of the top 10% most variable genes in both mRNA and proteomics datasets revealed that transcriptomics measurements exhibit extensive variability, producing heatmaps with strong contrast; in opposite, proteomics measurements, with many undetected proteins, yield heatmaps with reduced contrast and uniformity.

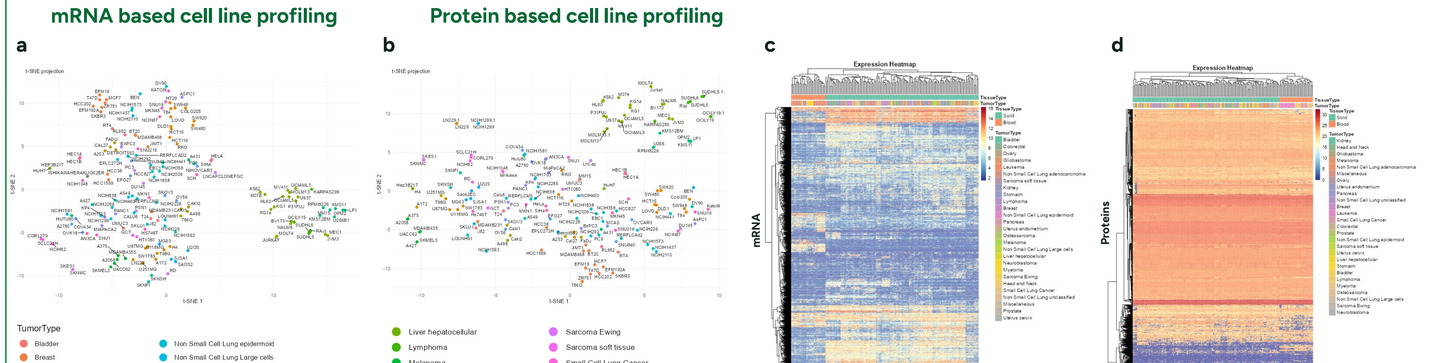


Fig. 2: Exploratory analysis of transcriptomics and proteomics profiles across the 160 ProLiFiler cell lines. t-SNE dimensionality reduction was applied to: **(a)** curated Affymetrix U133 Plus 2.0 transcriptomic data retaining JetSet best* probesets (n = 19,179) and **(b)** mass spectrometry–based proteomics quantification data (n = 14,325). (**c–d**) Heatmaps generated using the top 10% of transcripts and proteins with the highest variance. Data used were log2-transformed.

Results (Correlation mRNA and Protein) – Top 20 ADC Targets

Across the top 20 ADC targets, protein detection correlated variably with mRNA detection with Rho coefficient ranging from 0.8 e.g. for erbb2 to ~ 0 for genes including FOLH1 or TF (mean Rho Value: 0.46; 50% with rho>0.5)

	Mass Spectrometry (IBAQ)					Affymetrix (RSEM normalized)					
	Min-Max	Median	Mean	SD	Variance	Min-Max	Median	Mean	SD	Variance	Rho (RNA affy vs Protein MS)
ERBB2	0-21	13.5	11.5	5.6	31.4	2.7-14.6	7.4	7.0	2.5	6.4	0.80
ERBB3	0-20	0.0	9.1	9.3	85.6	2.1-13.4	6.4	6.7	3.7	14.0	0.77
TACSTD2	0-25.9	0.0	10.3	11.3	127.6	1.9-15.4	2.3	6.7	5.4	29.2	0.77
EGFR	0-28.3	21.9	19.4	7.4	54.1	1.9-15.4	10.1	8.5	4.0	15.7	0.77
CD276	0-22.6	20.1	19.2	4.2	17.7	2.8-10.1	7.9	7.8	1.3	1.7	0.72
MET	0-25.7	20.9	20.6	3.5	12.4	2.3-15.3	11.3	10.0	3.7	13.6	0.71
CD274	0-22	0.0	9.0	9.4	88.2	1.8-13	2.8	4.4	2.8	7.9	0.63
ROR1	0-21.3	18.3	14.4	8.1	65.4	2.7-10.8	6.0	6.0	2.4	5.8	0.62
NECTIN4	0-21	0.0	7.3	9.3	85.7	2-6.7	2.1	2.7	1.2	1.4	0.59
VTCN1	0-22.1	0.0	3.0	6.8	46.5	2.3-12.7	2.4	3.3	2.3	5.3	0.50
FOLR1	0-18.2	12.3	7.4	7.2	52.4	2.3-12.2	3.1	4.5	2.5	6.5	0.49
CEACAM5	0-24.5	0.0	9.1	9.7	95.0	2.2-15.1	2.3	3.6	3.1	9.5	0.38
DLL3	0-20.5	0.0	3.4	7.3	54.0	2-9.4	2.4	3.3	1.5	2.4	0.29
CLDN18	0-20.2	0.0	0.5	2.9	8.3	2.3-10	2.6	2.8	1.0	1.0	0.27
MS4A1	0-23.8	0.0	4.1	8.2	67.6	2.3-13.5	2.4	2.8	1.8	3.2	0.23
CDH17	0-26.3	0.0	9.8	11.7	137.7	2.4-14.5	2.6	3.5	2.6	6.8	0.20
SLC34A2	0-24.4	0.0	4.3	8.2	67.7	2.5-12.3	2.7	3.0	1.4	1.9	0.15
CDH6	0-22.4	20.1	18.9	4.9	24.4	2-12	2.1	3.0	2.3	5.4	0.14
FOLH1	0-21.3	17.3	12.3	8.1	65.8	2.1-12.5	2.2	2.8	1.3	1.8	0.06
TF	0-16.5	12.7	12.8	1.6	2.4	2.6-14.6	2.7	3.1	1.8	3.4	0.03

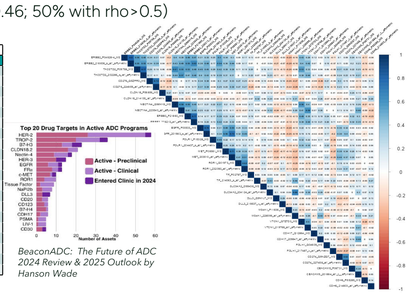


Figure 3. Descriptive statistics and correlation analysis between protein abundance and mRNA expression for the top 20 most widely investigated ADC targets. . Method: Spearman correlation analysis was performed using log2-transformed MS-iBAQ and log(TPM + 1) RNA-seq data. **Left table:** reported descriptive statistics and rho correlation coefficient. **Right figure:** Correlation plot showing the results of correlation analyses for each protein–RNA-seq pair among the 20 ADC targets. Blue indicates correlations with ρ > 0.5, and red indicates correlations with ρ < 0.5. Color intensity is proportional to the ρ value.

Results (Correlation, Individual Analysis)

The correlations between mRNA and MS protein detection is influenced by biological parameters such as gene alterations, dynamic range of expression, and the number of samples expressing the targets. We observed, however, that for low rho values, certain genes with no mRNA detection still showed signals at the protein level, suggesting possible cross-reactivity with other proteins.

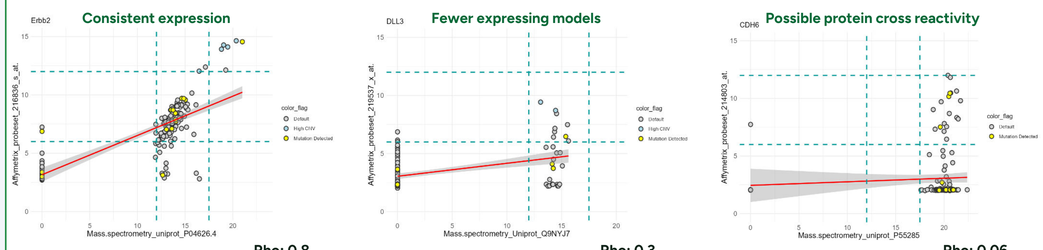


Fig. 4 Individual gene correlation analyses. Correlation plots show concordance between mRNA (Affymetrix) and protein (iBAQ) levels across 160 cell lines. Cell lines are annotated for high CNV (Affymetrix SNP6; PICNIC > 8) and mutation status (whole-exome sequencing; data from CancerDataMiner™). mRNA and protein levels were log2-transformed. Arbitrary cutoffs were applied to define expression categories (low, medium, high): 6 and 12 for mRNA, and 12 and 17.5 for protein.

Targeted Proteins and Sensitivity to ADC

Following validation of MS-derived protein levels using OncoFlow-FACS platform, integrative analysis information provide valuable insights into ADC behavior when evaluated using the 160-cell line proLiFiler panel.

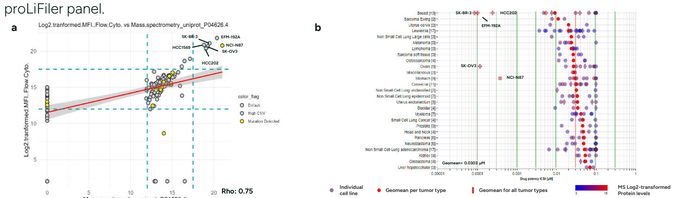


Fig 5. In vitro activity of Kadcyla (ERBB2-targeting ADC) in relation to the molecular features of cell lines (a) Integrative correlation analysis of MS-derived and FACS-derived ERBB2 protein quantification. (b) Scatter plot showing the Kadcyla IC50 values and the level of ERBB2 protein expression for each CL across cancer subtypes of the 160 CL panel. Horizontal-axis: IC50 value per CL, vertical-axis: histological (sub)types sorted from top to bottom by increasing geomean IC50 values. In brackets: the number of CLs per tumor subtype.

Summary

MS-based proteomics reliably quantifies ~14,000 proteins and can classify cell lines by cancer type, similar to RNA-seq.

mRNA–MS-derived protein correlations are influenced by expression range, gene alterations, and detection factors.

For interpreting ProLiFiler data, integrating mRNA, mutation (s), CNV, and protein information is recommended.

Targets showing mRNA–protein discordance should be reassessed individually (e.g., with OncoFlow).

Contact Information

Jan Ehlert

Head of Cellular Pharmacology

Reaction Biology Europe
Engesserstr. 4
79108 Freiburg, Germany

+49-761-769996-1645
Jan.Ehlert@reactionbiology.com
www.reactionbiology.com