Indivumed Therapeutics

APPLICATION NOTE

High-throughput Phosphoproteomics Profiling of Non-small Cell Lung Cancer Tissues

INTRODUCTION

Lung cancer remains the cause of most cancer-related deaths worldwide [1]. Despite considerable advances in new therapies including tyrosine kinase inhibitors and immunotherapies, sustained responses are rare, and prognosis remains poor [2]. Although prior molecular characterization has facilitated the development of effective inhibitors for non-small cell lung cancer (NSCLC) driven by gene changes in EGFR, ALK, ROS1, BRAF, MET, NTRK, or RET, deep scale proteomics and phosphoproteomics profiling is needed to advance precision medicine [3, 4]. Recent developments in data-independent acquisition (DIA) LC-MS/MS and robust chromatographic separation now present the opportunity to make proteomics available to routine analysis [5]. Here, we leverage these technological developments as part of our extensive cancer discovery platform, nRavel®, which combines proteomic and phosphoproteomic data with whole genome sequencing, RNA sequencing, and clinical data for each patient in the database, thus providing an unprecedented insight into tumor biology. The data generated comes from high-quality tissue that we collect by following standard operating procedures to maintain an average ischemia time of approximately 10 minutes along with strictly adhered to preservation protocols to minimize functional alterations in tumor biology (Figure 1).

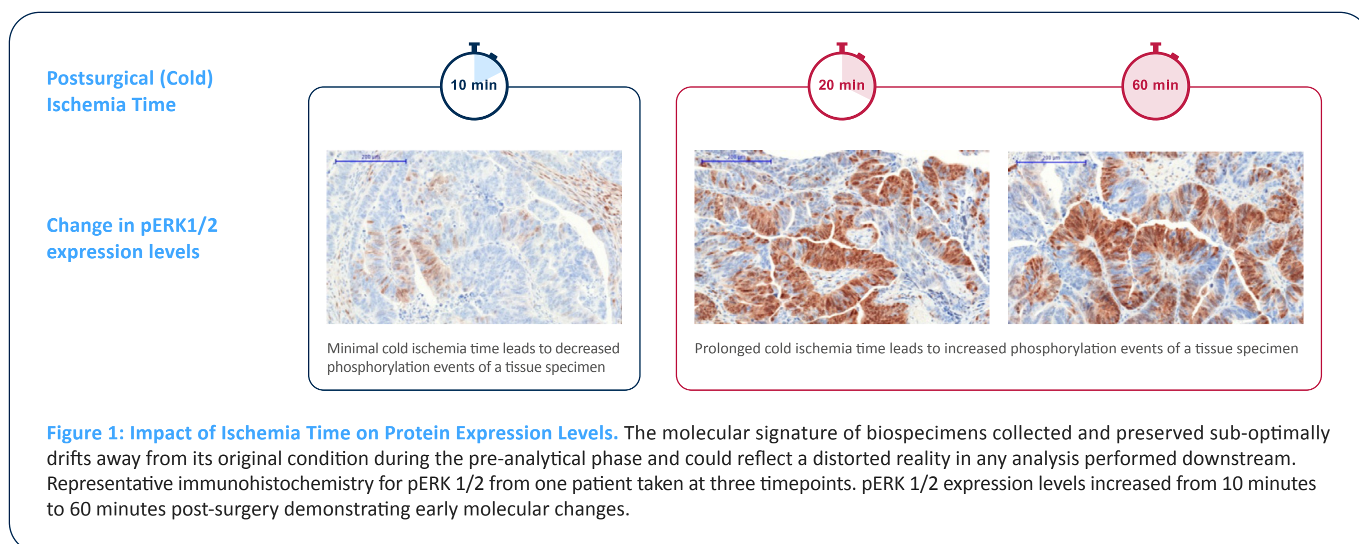

Figure 1: Impact of Ischemia Time on Protein Expression Levels. The molecular signature of biospecimens collected and preserved sub-optimally drifts away from its original condition during the pre-analytical phase and could reflect a distorted reality in any analysis performed downstream. Representative immunohistochemistry for pERK 1/2 from one patient taken at three timepoints. pERK 1/2 expression levels increased from 10 minutes to 60 minutes post-surgery demonstrating early molecular changes.

Through our partnership with Biognosys, a workflow has been established that is capable of routine profiling 850 whole proteome (WP) and 650 phosphoproteome (PP) tumor samples per month with an average depth of 6,000 proteins (WP) and 18,000 phosphopeptides (PP) (Figure 2). In this study, we performed DIA mass spectrometry-based proteomic analysis on tissue samples from 772 NSCLC patients, including a detailed characterization of the EGFR signaling pathway. Our findings provide valuable insights into NSCLC biology and highlight the precision medicine potential of the nRavel® platform.

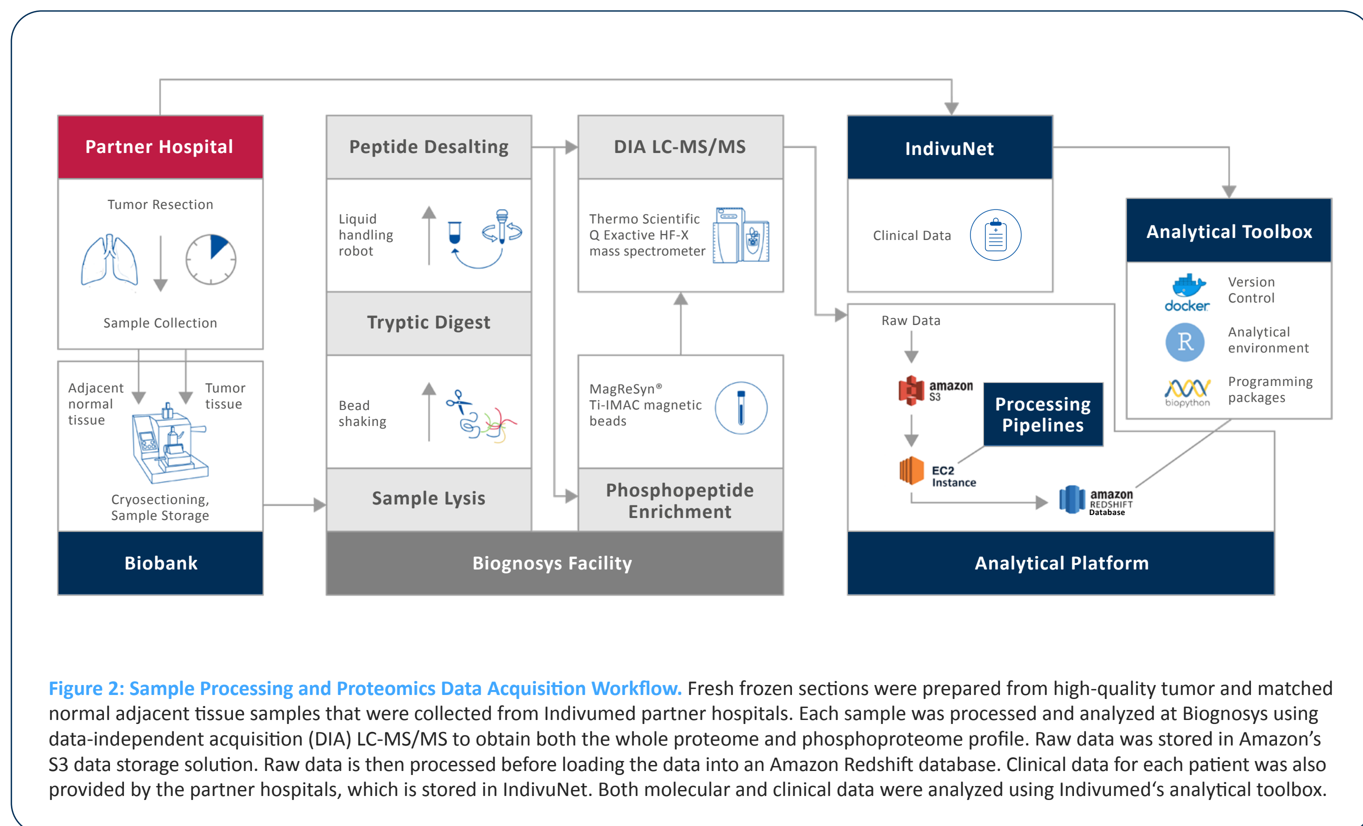

Figure 2: Sample Processing and Proteomics Data Acquisition Workflow. Fresh frozen sections were prepared from high-quality tumor and matched normal adjacent tissue samples that were collected from Indivumed partner hospitals. Each sample was processed and analyzed at Biognosys using data-independent acquisition (DIA) LC-MS/MS to obtain both the whole proteome and phosphoproteome profile. Raw data was stored in Amazon's S3 data storage solution. Raw data is then processed before loading the data into an Amazon Redshift database. Clinical data for each patient was also provided by the partner hospitals, which is stored in IndivuNet. Both molecular and clinical data were analyzed using Indivumed's analytical toolbox.

Indivumed Therapeutics

MATERIALS AND METHODS

Collection of Patient Tumor Tissue

- Patients with NSCLC gave informed consent and institutional review board approval was obtained from partner clinical sites. The use of human tissue was approved, and anonymization of patient data and samples was compliant with data protection regulations.
- Tumor tissue and matched normal adjacent tissue from NSCLC patients was collected according to Indivumed's standard operating procedures within 10 minutes of surgical resection and tissue samples were snap frozen in liquid nitrogen.

Proteomics

- Fresh frozen (FF) tissue from the Indivumed biobank was shipped to Biognosys to be processed and analyzed through Biognosys' proprietary MS-based Hyper Reaction Monitoring (HRM™) DIA discovery proteomics workflow.
- Whole proteome (WP) peptide samples were generated by lysis, trypsin digestion, and desalting of 5–10 mg of fresh frozen tissue. 30 µg of each WP sample was used for proteome profiling.
- Phosphoproteome (PP) samples were generated by phosphopeptide enrichment using 500 µg of the corresponding WP sample using MagReSyn® Ti-IMAC magnetic beads (ReSyn Biosciences).
- Final peptide samples were loaded into an ACQUITY M-Class micro-flow LC system (Waters) and peptides were eluted using a short gradient into a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) for analysis.
- Data acquisition was performed by 50 non-linear DIA segments.

Bioinformatics Analysis

- To map peptides and phosphopeptides, Biognosys' DIA proteomics software, Spectronaut (version 13.8.200403) was run on raw mass spec output files in peptide-centric mode with enabled PTM localization.
- Cohort-based protein and peptide normalization, principal component analysis, and boxplots were undertaken in R3.6.3.
- To ascertain coverage of the EGFR signaling pathway, EGFR components were extracted using the GO term GO:0007173.

Simple Western

- Protein lysates were prepared from fresh frozen sections of tumor and matched normal adjacent tissue obtained from three NSCLC patients using Indivumed's standard protocol.
- Protein concentration was determined using BCA assay.
- Samples were analyzed using Simple Western Size™ assay to detect:
  - Total EGFR – 0.125 mg/ml sample concentration, EGF Receptor (D38B1) XP® Rabbit mAb 1:25 (Cell Signaling Technologies).
  - pEGFR (Tyr1110) – 2 mg/ml sample concentration, EGF Receptor (Tyr1086) mAb 1:25 (Cell Signaling Technologies).
  - Total protein load quantification – 0.125 mg/ml sample concentration, Total Protein Assay (ProteinSimple).

RESULTS

Protein and Phosphopeptide Profiling of NSCLC

We investigated the proteomic landscape of primary tumors and matched adjacent normal samples from 772 NSCLC patients (Figure 3A). On average, 6,050 protein groups and 20,589 phosphopeptides were quantified in each lung tissue sample. The proteomic profiles obtained from all samples were analyzed by hierarchical clustering (Figure 3B) and principal component analysis of the protein intensity values where both analyses reflected a clear separation between tumor and normal tissue profiles (Figures 3C and 3D).

Figure 3: Protein and Phosphopeptide Profiling of 772 NSCLC Patients. (A) Subset of key demographic and histologic features, along with smoking status for the cohort of 772 NSCLC patients characterized in this study. One person represents 50 patients. (B) Hierarchical clustering, and (C) Principal component analysis of protein intensity values mainly revealed co-clustering according to tissue type. (D) A similar pattern was observed in principal component analysis of phosphopeptides.

Analysis of differential protein and phosphopeptide levels

We observed robust differences in the protein levels of three known dysregulated proteins of lung cancer between tumor and normal samples (Figure 4A). Next, we selected EGFR, a protein widely targeted by NSCLC therapies to gain additional insights into the different phosphorylation sites for further characterization using phosphoproteomics data (Figure 4B). We found consistent elevation of EGFR phosphorylation status on three known C-terminal sites, Ser1064, Tyr1110, and Tyr1197, reflecting the dysregulation of EGFR in NSCLC consistent with previous studies [6]. In order to experimentally validate the proteomics and phosphoproteomics data obtained through DIA LC-MS/MS we used Simple Western Size™ assay. We analyzed three patient samples from a pool of samples that showed elevated tumoral levels of total EGFR and phosphorylated EGFR-Tyr1110 in MS data and observed similar high expression levels using Simple Western Size™ (Figure 4C). Our findings show strong concordance between proteomics data obtained from high-throughput MS and support the consistency of the outcomes from Indivumed data analysis.

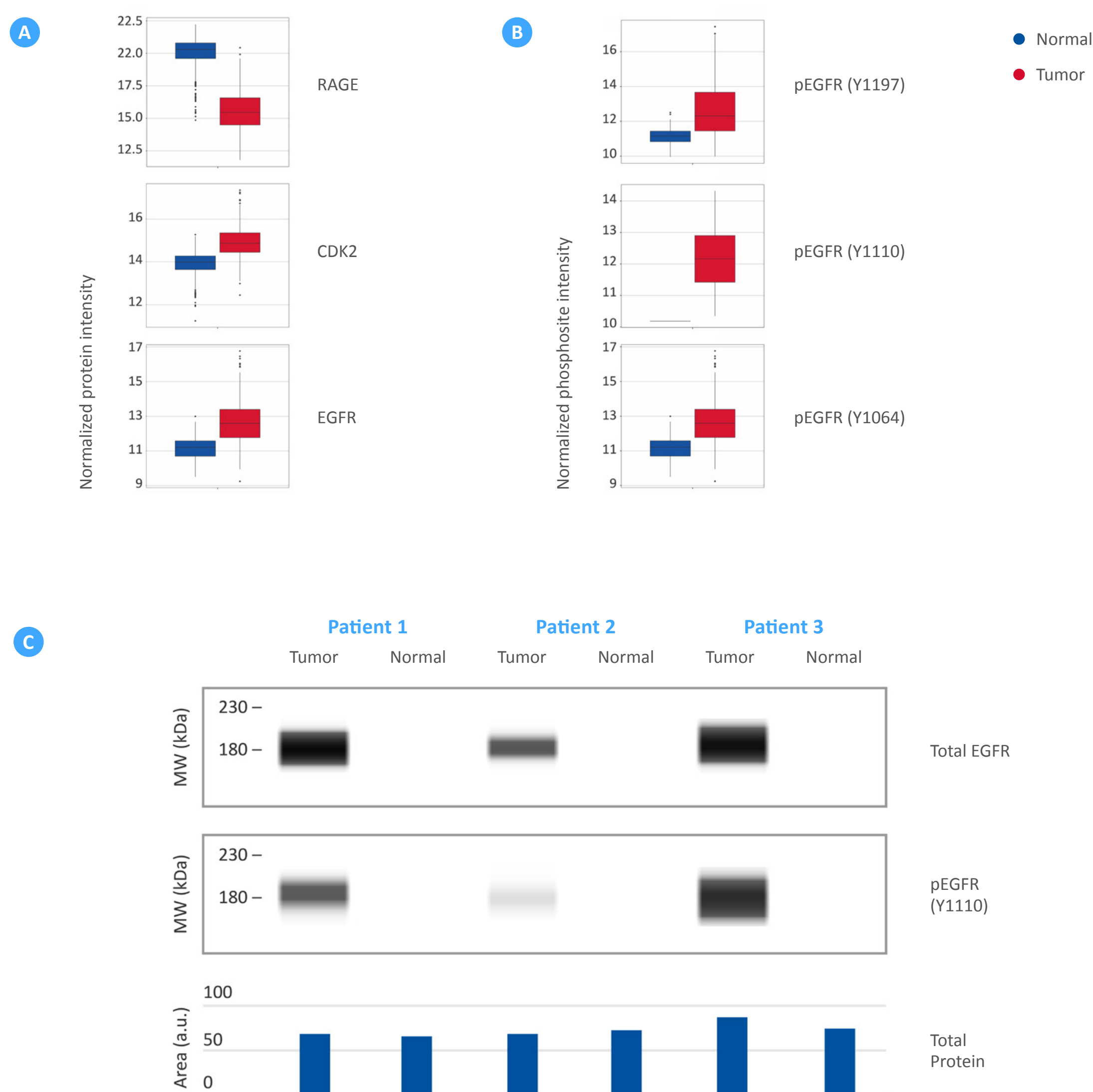

Figure 4: Differential Total Protein and Phosphoprotein Expression Levels in NSCLC. (A) Boxplots showing protein levels of RAGE, CDK2, and EGFR in tumor and normal samples. **(B)** Boxplots showing the differential phosphorylation levels of specific EGFR sites. **(C)** Detection of total EGFR (top row) and pEGFR (Y1110) (middle row) levels, together with total protein load quantification (bottom row) of tumor and normal paired samples from three different NSCLC patients using Simple Western Size™ assay.

Description of cellular pathway activation status

As an example, we describe here an overview of the EGFR pathway in the complete NSCLC cohort (Figure 5A). We observed the upregulation of proteins for which targeted therapies are either currently available or under development, namely, EGFR, KRAS, MEK, ERK, AKT, and mTOR. Other proteins, like 4-EBP1, showed a strong deregulation at both the total and phosphorylation protein level. The inhibitory effect of 4-EBP1 phosphorylation could be counteracting the high protein levels, which in a dephosphorylated form would lead to transcriptional inhibition in the cell. Overall, we were able to quantify 94 out of 120 GO annotated proteins of this pathway with protein or phosphopeptide information (Figure 5B), thereby allowing us to build a detailed picture of the alterations in the EGFR pathway of NSCLC.

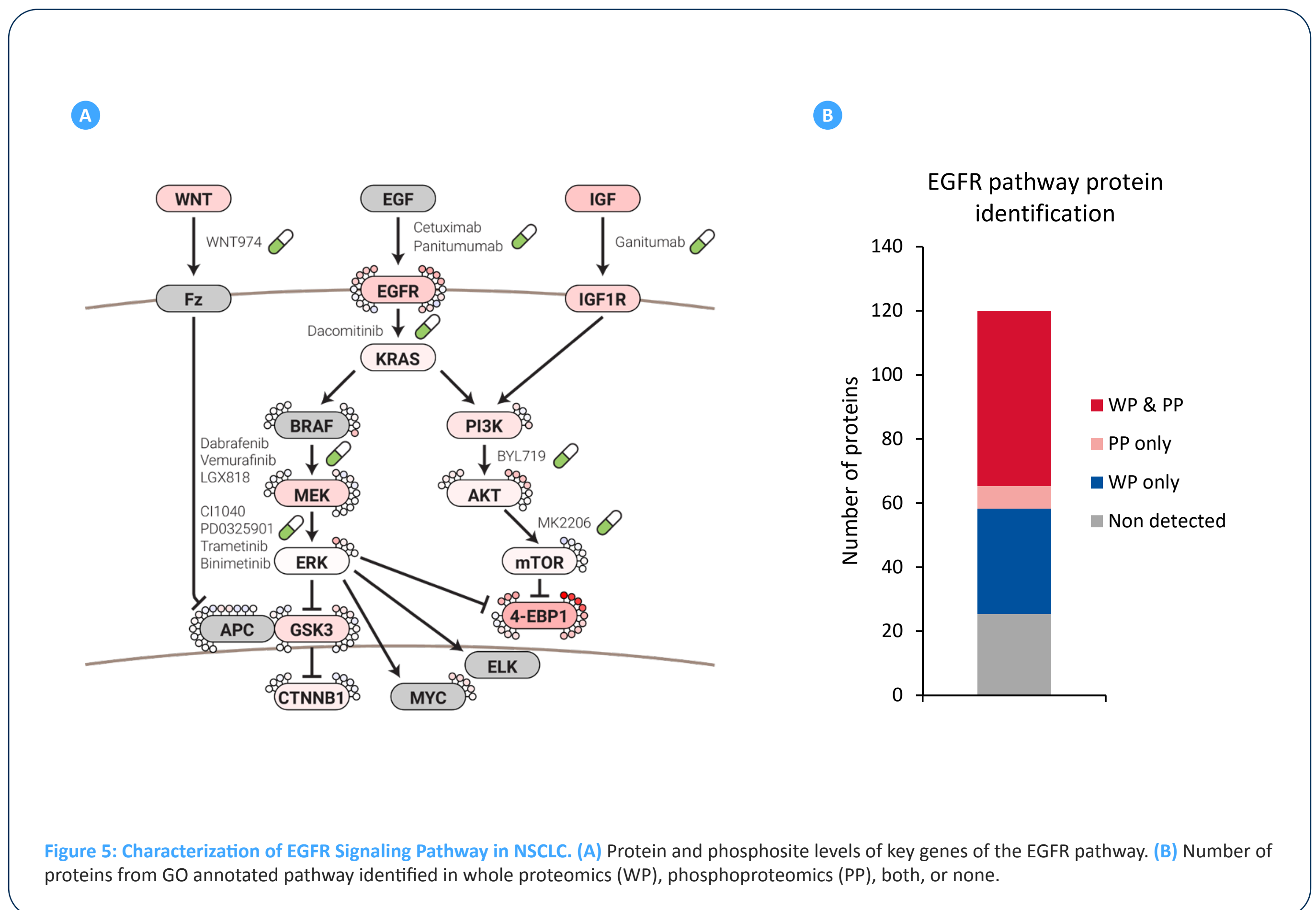

Figure 5: Characterization of EGFR Signaling Pathway in NSCLC. (A) Protein and phosphosite levels of key genes of the EGFR pathway. **(B)** Number of proteins from GO annotated pathway identified in whole proteomics (WP), phosphoproteomics (PP), both, or none.

Indivumed Therapeutics

CONCLUSIONS

Protein profiling of tumor samples by proteomics technologies has been largely underutilized in target discovery and drug development, even though most anti-cancer therapies have been developed against protein targets. A study combining transcriptome and proteome analyses of breast cancer tissue found that levels of RNA corresponded more closely with levels of proteins in tumors than in healthy tissue, suggesting that the degree to which RNA and proteins change in lockstep might provide a crucial window into cancer biology. While protein level measurements more closely resemble the biology of the cell, post-translational modifications such as protein phosphorylation provide an extra level of dynamic function control. Phosphorylation is well known to have an important role in driving cancer phenotypes and has been successfully targeted in precision medicine approaches. As such, a better understanding of both the proteomic and phosphoproteomic landscape is essential to advance targeted anti-cancer therapy development [7, 8].

Indivumed's nRavel® is the first multiomics platform to systematically include proteomics and phosphoproteomics, together with whole genome, transcriptomics, small RNA, and clinical data for each cancer case for both tumor and normal adjacent samples. The extensive depth of information provides a multi-dimensional landscape of tumor biology that can better classify a patient's cancer and predict their response to treatment.

A direct analysis of multiple protein and phosphosite levels across large cancer cohorts within the Indivumed database provides researchers with an unprecedented view of the status of complex molecular networks in tumor tissue and its normal counterpart, thus enabling the potential identification of new targets and biomarkers and designing new therapeutic approaches. Here we have identified several targets of the canonical EGFR signaling pathway in both whole proteomics and phosphoproteomics, but furthermore we can access a large proportion of the entire EGFR signaling system that can be utilized to drive cutting-edge NSCLC drug discovery programs.

REFERENCES

[1] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424

[2] Herbst, R.S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. Nature 553, 446–454

[3] Michael A. Gillette, (2020) Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma Cell 182, 200–225

[4] American Cancer Society. Lung Cancer (2020)

[5] Bruderer R., Bernhardt O., Gandhi T., et al. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues, Molecular & Cellular Proteomics 14(5), 1400-1410

[6] Mayumi Ono and Michihiko Kuwano, (2006). Molecular Mechanisms of Epidermal Growth Factor Receptor (EGFR) Activation and Response to Gefitinib and Other EGFR-Targeting Drugs, Clin Cancer Res (12) (24) 7242–7251

[7] Vasaiker, S., Huang C., Wang X., Zhang B., et al. (2019) Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities, Cell 177, 1035–1049

[8] Simon Makin, (2020). Beyond the genome, Nature (585)

Contact Information

Indivumed GmbH · Falkenried 88, Bldg. D · D-20251 Hamburg · Germany
PHONE: +49 40 4133 83 0
EMAIL: application-notes@indivumed.com

www.indivumed-therapeutics.com

© Copyright 2019-2023 by Indivumed GmbH. All Rights Reserved. No copying and/or distribution of copies allowed without prior approval by Indivumed GmbH. This material contains Confidential Information.