Mass spectrometry-based proteomic analysis of NSCLC tumor and biopsy samples

Janne Lehtiö (✉ janne.lehtio@ki.se)

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Taner Arslan

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Ioannis Siavelis

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Yanbo Pan

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Fabio Socciarelli

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Olena Berkovska

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden <https://orcid.org/0000-0002-8811-0591>

Husen M. Umer

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Georgios Mermelekas

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Mohammad Pirmoradian

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Mats Jönsson

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Hans Brunnström

Department of Pathology, Laboratory Medicine Region Skåne, Lund, Sweden, Division of Pathology, Department of Clinical Sciences, Lund, Lund University, Lund, Sweden

Odd Terje Brustugun

Section of Oncology, Drammen Hospital, Vestre Viken Health Trust, Drammen, Norway, Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

Krishna Pinganksha Purohit

University of Edinburgh Centre for Inflammation Research, Institute for Regeneration and Repair, Queen's Medical Research Institute, Edinburgh bioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK, MRC Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh bioQuarter, 5 Little France Drive, Edinburgh EH16 4UU, UK

Richard Cunningham

University of Edinburgh Centre for Inflammation Research, Institute for Regeneration and Repair, Queen's Medical Research Institute, Edinburgh bioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK, MRC Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh bioQuarter, 5 Little France Drive, Edinburgh EH16 4UU, UK

Hassan Foroughi Asl

Genomic Medicine Center, Karolinska University Hospital, Stockholm, Sweden. Clinical Genomics Facility, Department of Microbiology, Tumour and Cell Biology, Karolinska Institutet, Stockholm, Sweden.

Sofi Isaksson

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Elsa Arbajian

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Mattias Aine

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Anna Karlsson

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Marija Kotevska

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden, Department of Respiratory Medicine and Allergology, Skåne University Hospital, Lund, Sweden

Carsten Gram Hanson

University of Edinburgh Centre for Inflammation Research, Institute for Regeneration and Repair, Queen's Medical Research Institute, Edinburgh bioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK, MRC Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh bioQuarter, 5 Little France Drive, Edinburgh EH16 4UU, UK

Vilde Drageset Haakensen

Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, Department of Oncology, Oslo University Hospital, Oslo, Norway

Åslaug Helland

Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, Department of Oncology, Oslo University Hospital, Oslo, Norway, Faculty of Medicine, University of Oslo, Norway

David Tamborero

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Henrik J. Johansson

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Rui M. Branca

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Maria Planck

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden, Department of Respiratory Medicine and Allergology, Skåne University Hospital, Lund, Sweden

Johan Staaf

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Lukas M. Orre

Department of Oncology and Pathology, Karolinska Institutet, Science for Life Laboratory, Solna, SE-17165, Sweden

Method Article

Keywords: cancer, mass spectrometry, proteomics

Posted Date: November 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.pex-1560/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The associated publication reports proteogenomic analysis of non-small cell lung cancer, where we identified molecular subtypes with distinct immune evasion mechanisms and therapeutic targets, and validated our classification method in separate clinical cohorts. This protocol describes the sample preparation and mass spectrometry (MS)-based in-depth and rapid proteomic analyses of tumor and biopsy samples. We deployed single-pot solid-phase-enhanced sample preparation (SP3). For the in-depth analysis, we used TMT labeling, followed by high-resolution isoelectric focusing (HiRIEF) prefractionation and LC-MS with data-dependent acquisition (DDA). The reported protocol achieved analytical depth of close to 14,000 quantified proteins and almost 10,000 across the entire cohort of 141 samples. The rapid analysis was label-free, based on LC-MS with data-independent acquisition (DIA). The median number of identified proteins was 3,967 and 3,552 in two independent cohorts of tumor samples (n = 141 and 208, respectively), and 2,494 in another cohort of biopsy material (n = 84).

Introduction

Reagents

Equipment

Procedure

HiRIEF-LC-MS TMT-DDA based analysis of early-stage NSCLC cohort

Sample selection and preparation.

DNA, RNA and protein from 192 fresh frozen tissue pieces were extracted using the AllPrep Kit (QIAGEN, cat. No. 80204), as described previously¹. For the current proteomics analysis, 35 samples were excluded due to insufficient protein amount or deviating protein-RNA or protein-DNA concentration correlation resulting in 157 samples remaining for protein digestion and further MS analysis. Four volumes (relative to the sample volume) of ice-cold (-20 °C) acetone were added to each protein fraction from the Allprep kit to precipitate the proteins. The tubes were inverted three times and incubated at -20 °C for 60 min, followed by centrifugation for 10 min at 12,000 × g in a pre-cooled centrifuge at 4 °C. The supernatant was discarded, and the pellet was washed once with 100 µl of ice-cold ethanol. The pellet was then dispersed in 100 µl ice-cold ethanol by ultrasonication (Program: Am 50%, time 10 s, pulse 1.0 s on the Bandelin Sonoplus probe sonicator, from Heco, Norway), centrifuged, and the resulting pellet was air-dried for 10 min. The pellet was subsequently dissolved in 200 µl of reconstitution buffer (4% (w/v) SDS, 25 mM HEPES pH 7.6), and protein concentration was determined using Bio-Rad DC protein assay kit (cat. No. 500-0116). For each sample, 300 µg (2 µg/µl) of reconstituted protein were reduced by addition of dithiothreitol (DTT) for a final concentration of 1 mM. Free thiols were subsequently alkylated with excess chloroacetamide at a final concentration of 4–10 mM.

Protein clean-up and digestion were then performed using a modified SP3 (single-pot, solid-phase-enhanced sample-preparation)² protocol. Namely, proteins were captured on SP3 beads (GE Healthcare Sera-Mag SpeedBeads™ Carboxyl Magnetic Beads, hydrophobic 65152105050250, hydrophilic 45152105050250) by adding stock beads suspension (10 µg/µl, 1:10 bead to sample volume) and acetonitrile (ACN) for a final concentration of 70%. The mixture was incubated under rotation for 30 min at room temperature (RT). To remove the lysis buffer, the tubes were placed on a magnetic rack and incubated for 2 min at RT. The supernatant was discarded, the tubes were removed from the magnetic rack and the bead-attached-proteins were washed twice with 500 µl of 70% ethanol (incubated for 30 s on the magnetic rack, followed by supernatant removal). Thereafter, 200 µl of ACN were added and the samples were incubated for 15 s on the magnetic rack. The supernatant was then discarded, and the beads were air-dried for 30 s. The proteins were digested by sequential addition of LysC and trypsin enzymes for a total incubation time of minimum 20 h at 37 °C. The first digestion solution contained LysC (1:50 enzyme to protein ratio) in 1 M Urea/50 mM HEPES. Thereafter, trypsin (1:50 enzyme to protein ratio) in 50 mM HEPES was added. Digested peptides were collected as the supernatant after placing the tube on a magnetic rack. Finally, 50 µl of water was added twice to collect the remaining peptides and the peptide concentration was measured using Bio-Rad DC protein assay. Four out of 157 samples had insufficient peptide amount (< 100 µg) for TMT labeling and were excluded. To identify outlier samples, the remaining 153 samples were pre-screened by LC-MS/MS on a Q Exactive HF using short-gradient (60 min) DDA runs. Based on analysis of the short-gradient data, 10 samples with extensive blood contamination were excluded, resulting in 143 samples remaining for tandem mass tag (TMT) labeling. Subsequent re-analysis of clinical data resulted in the exclusion of two additional samples after MS data generation due to uncertain primary tumor origin. This resulted in a final cohort size of 141 lung cancer samples for subsequent analysis.

Tandem Mass Tag (TMT) labeling and HiRIEF pre-fractionation of peptides.

A total of 143 samples were TMT-labeled. Before labeling, a reference pool was prepared to function as denominator in each TMT set. The pool comprised peptides from 77 AC samples pooled together to form 1-mg AC sub-pool; the same amount of peptides from 32 SqCC samples that were pooled together to form a 1-mg SqCC sub-pool; and peptides from 22 LCC and 10 LCNEC samples that were pooled together to form a 1-mg LCC+LCNEC sub-pool; these sub-pools were then pooled together to form the final 3-mg reference pool. 100 µg of peptides from each tumor sample and reference pool was labeled with TMT 10-plex reagent according to the manufacturer's protocol (Thermo Scientific). The 143 tumor samples were distributed across 16 TMT 10-plex sets, with 9 tumor samples and one reference pool, except in set 16, which had two reference pools. An additional TMT set, No. 17, was designed to include 4 reference pool samples and 6 tumor sample replicates also present in the primary 16 TMT sets. Labeled samples in each TMT set were pooled, cleaned by strata-X-C-cartridges (Phenomenex) and dried in a vacuum centrifuge (Electron Savant SpeedVac Concentrator, Thermo Fisher Scientific).

The TMT-labeled peptides, were separated by High-Resolution Isoelectric Focusing (HiRIEF) on pH 3.7–4.9 and 3–10 strips (300 µg per strip) as described previously^{3,4}. Peptides were extracted from the strips

by a liquid handling robot (Etan digester from GE Healthcare Bio-Sciences AB, which is a modified Gilson liquid handler 215). A polypropylene well-former with 72 wells was put onto each strip and 50 µl of MilliQ water was added to each well. After a 30-min incubation, the liquid was transferred to a 96-well plate (V-bottom, polypropylene, Greiner 651201), and the extraction was repeated 2 more times with 35% ACN and 35% ACN/0.1% formic acid (FA) in MilliQ water, respectively. The extracted peptides were dried in the 96-well plate in a SpeedVac.

MS-based quantitative proteomics.

For each LC-MS run of a HiRIEF fraction, the autosampler (Ultimate 3000 RSLC system, Thermo Scientific Dionex) dispensed 20 µl of 3% ACN/0.1% FA solvent into the corresponding well of the microtiter plate, mixed by aspirating/dispensing 10 µl ten times, and finally injected 10 µl into a C18 trap desalting column (Acclaim pepmap, C18, 3 µm bead size, 100 Å, 75 µm x 20 mm, nanoViper, Thermo Scientific). Peptides were separated using a gradient of mobile phase A (5% DMSO, 0.1% FA) and B (90% ACN, 5% DMSO, 0.1% FA), ranging from 6% to 37% B in 30–90 min (depending on immobilized pH gradient- isoelectric focusing, IPG-IEF, fraction complexity) with a flow of 250 nl/min. The Q Exactive HF was operated in data-dependent acquisition (DDA) mode, selecting top 5 precursors for fragmentation by high-energy collision dissociation (HCD). The survey scan was performed at 60,000 resolution from 300–1500 m/z, with a maximum injection time of 100 ms and a target of 1×10^6 ions. For generation of HCD fragmentation spectra, a maximum ion injection time of 100 ms and AGC of 1×10^5 were used before fragmentation at 30% normalized collision energy and 30,000 resolution. Precursors were isolated with a width of 2 m/z and put on the exclusion list for 60 s. Single and unassigned charge states were rejected from precursor selection.

Peptide and protein identification.

Peptide and protein identification were performed as described previously⁴. Briefly, Orbitrap raw MS/MS files were converted to mzML format using msConvert from the ProteoWizard tool suite (v.3.0.19127). Spectra were then searched using MSGF+ (v2017.07.21) and Percolator (v3.1), where search results from all HiRIEF fractions of each TMT set were grouped for Percolator target/decoy analysis. All searches were done against the human protein database of Ensembl 92 in a Nextflow pipeline (<https://github.com/lehtiolab/nf-workflows>, commit: 898bb20). MSGF+ settings included precursor mass tolerance of 10 ppm, fully tryptic peptides, maximum peptide length of 50 amino acids and a maximum charge of 6. Fixed modifications were TMT-10plex on lysines and peptide N-termini, and carbamidomethylation on cysteine residues. A variable modification was used for oxidation on methionine residues. Quantification of TMT-10plex reporter ions was done using OpenMS project's IsobaricAnalyzer (v2.0). Peptide spectrum matches (PSM) found at 1% FDR (false discovery rate) were used to infer gene identities.

Protein quantification by TMT 10-plex reporter ions was calculated using TMT PSM ratios to the reference TMT channels and normalized to the sample median. The median PSM TMT reporter ratio from

peptides unique to a gene symbol was used for quantification. Protein FDRs were calculated using the picked-FDR method using gene symbols as protein groups and limited to 1% FDR.

DIA-based analyses of NSCLC cohorts

Sample preparation

For the early-stage cohort, each of the peptide samples prepared for the DDA-based analysis described above was aliquoted prior to TMT-labeling. The peptides underwent an additional SP3 peptide clean-up step described below.

For each of the late-stage cohort samples, 450 μ l of protein extract were obtained using the AllPrep Kit (QIAGEN, cat no 80204), 225 μ l of which was used for further processing. Whereas, for the validation cohort protein extracts were prepared by cutting each of the tumor pieces to obtain a 2 $\times$ 2 mm slice which was washed in PBS (1 ml, thrice), homogenized, and lysed. The tissue pieces in 200 μ l of lysis buffer (4% w/v SDS, 25 mM HEPES pH 7.6, 1 mM DTT) were placed in Precellys lysing kit "Tissue homogenizing CKMix" tubes (Bertin Technologies) and shaken at 30 s⁻¹ for 20 min (TissueLyser, Qiagen). The samples were then heated on a shaker (95 °C, 500 rpm, 5 min, Thermomixer comfort, Eppendorf) and sonicated (50% amplitude, 1 s pulse, 1 min). The protein extracts were transferred to Eppendorf tubes and centrifuged at 14,000 $\times g$ for 15 min. The centrifugation and tube transfer steps were repeated until complete removal of debris. The total protein concentration was measured using the Bio-Rad DC protein assay kit. For the validation cohort, 200 μ g of protein were aliquoted for further processing, whereas the entire sample was used for the late-stage cohort.

SP3 protein clean-up and digestion was performed for the late-stage and validation cohort samples as described above for the early-stage cohort. The protocol was scaled for the late-stage cohort samples to account for the variable amounts of material. Thereafter, a peptide clean-up was performed for all samples using the SP3 method. Briefly, fresh SP3 beads suspension (10 μ g/ μ l, 1:10 bead to sample volume) and ACN (final concentration of 95%) were added to 50–200 μ g of peptides and incubated under rotation at RT for 30 min. The tubes were then placed on a magnetic rack, the supernatant was discarded, and the beads were washed twice with 200 μ l of ACN. The beads were briefly air-dried, after which the peptides were eluted with 100 μ l of 3% ACN/0.1% FA and transferred to a new tube. The peptide concentration was measured using the Bio-Rad DC protein assay. The required quantities for further LC-MS analysis were aliquoted and dried in a SpeedVac.

Spectral library preparation

A pooled sample containing peptides from 129 different tumor samples from the early-stage cohort was combined for spectral library generation. A total of 2 mg of pooled peptides were aliquoted into two parts, each one was subjected to the fractionation of peptides, one by HiRIEF and one by high-pH peptide

fractionation. For HiRIEF pre-fractionation, peptides were separated by IPG-IEF on pH 3–10 strips as described above in “HiRIEF pre-fractionation of peptides”. The extracted peptides were dried in SpeedVac, dissolved in 3% ACN/0.1% FA and consolidated to a final of 40 fractions (as described in the HiRIEF fraction scheme file in the PXD dataset, PXD020191). For high-pH pre-fractionation, peptides were fractionated with basic-pH reverse-phase (BPRP) high-performance liquid chromatography (HPLC). Peptides were loaded and separated on a 25 cm C18 packed column (XBridge Peptide BEH C18, 300 Å, 3.5 µm, 2.1 mm x 250 mm). 96 fractions were collected from the column and consolidated to a final of 40 fractions.

MS data acquisition.

Peptides were separated using an Ultimate 3000 RSLCnano system coupled to a Q Exactive HF (Thermo Fischer Scientific, San Jose, CA, USA). Samples were trapped on an Acclaim PepMap nanotrap column (C18, 3 mm, 100 Å, 75 µm x 20 mm, Thermo Scientific), and separated on an Acclaim PepMap RSLC column (C18, 2 µm bead size, 100 Å, 75 µm x 50 cm, Thermo Scientific). Peptides were separated using a gradient of mobile phase A (5% DMSO, 0.1% FA) and B (90% ACN, 5% DMSO, 0.1% FA), ranging from 6% to 30% B in 180 min with a flow of 250 nl/min.

To create the spectral library, each of the 80 fractions was analyzed in a data-dependent acquisition manner (DDA). The method was set for selecting top 10 precursors for fragmentation by HCD. The survey scan was performed at 120,000 resolution from 400–1200 m/z, with a max injection time of 100 ms and a target of 1×10^6 ions. For generation of HCD fragmentation spectra, a max ion injection time of 100 ms and AGC of 2×10^5 were used before fragmentation at 25% normalized collision energy, 30,000 resolution. Precursors were isolated with a width of 2 m/z and put on the exclusion list for 15 s. Single and unassigned charge states were rejected from precursor selection.

For the DIA-based analysis of the individual tumor samples, the samples were dissolved in phase A (5% DMSO, 0.1% FA) and 5 µg of peptides were injected into the LC-MS system. The data was acquired using a variable window strategy. The survey scan was performed at 120,000 resolution from 400–1200 m/z, with a max injection time of 200 ms and target of 1×10^6 ions. For generation of HCD fragmentation spectra, maximum ion injection time was set as auto and AGC of 2×10^5 were used before fragmentation at 25% normalized collision energy, 30,000 resolution. The sizes of the precursor ion selection windows were optimized to have similar density of precursors m/z based on identified peptides from the spectral library. The median size of windows was 18.3 m/z with a range of 15–88 m/z covering the scan range of 400–1200 m/z. Neighbor windows had 2 m/z overlap.

DIA-based peptide and protein identification and quantification.

Spectral library generation as well as peptide and protein identification and quantification were performed on the Spectronaut software package (version 13.10) from Biognosys. For spectral library generation, all 80 MS raw files (40 HiRIEF + 40 high pH RP fractions) were searched by the integrated search engine Pulsar. Files were searched against ENSEMBL protein database (GRCh38.92.pep.all.fasta). All

parameters were set as default and for each peptide, the best 3 to 6 fragments were used. Results were filtered at all the precursor, peptide, and protein levels with 1% FDR. Out of 213,392 precursors, the peptide library consisted of 160,185 peptides representing 11,915 protein groups.

For protein identification and quantification, all DIA raw files were analyzed by Spectronaut using the above generated spectral library. All parameters were kept as default for protein identification. Briefly, runs were recalibrated using iRT standard peptides in a local and non-linear regression. Precursors, peptides and proteins were filtered with FDR 1%. The decoy database was created by mutation method. For quantification, only peptides unique to a protein group were used. Protein groups were defined base on gene symbols to obtain a gene symbol-centric quantification. Stripped peptide quantification was defined as the top precursor quantity. Protein group quantification was calculated by the median value of up to 3 most abundant peptides. Normalization was performed at the MS2 level and quantification at the MS1 level based on the peak area. The data filtering was set as Q value for each sample. Some identifications did not have true quantifications at the MS1 level and the instrument's software automatically imputed these with 1, thus, these values of 1 were treated as NAs for further quantitative analysis.

Troubleshooting

Time Taken

Anticipated Results

References

1. Karlsson, A. *et al.* Gene Expression Profiling of Large Cell Lung Cancer Links Transcriptional Phenotypes to the New Histological WHO 2015 Classification. *J Thorac Oncol* **12**, 1257-1267, doi:10.1016/j.jtho.2017.05.008 (2017)
2. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **14**, 68-85, doi:10.1038/s41596-018-0082-x (2019).
3. Branca, R. M. *et al.* HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods* **11**, 59-62, doi:10.1038/nmeth.2732 (2014)
4. Zhu, Y. *et al.* Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow. *Nat Commun* **9**, 903, doi:10.1038/s41467-018-03311-y (2018)