Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS

Yangyang Bian1,2, Runsheng Zheng1, Florian P. Bayer1, Cassandra Wong3, Yun-Chien Chang1, Chen Meng4, Daniel P. Zolg1, Maria Reinecke1,5,6, Jana Zecha1, Svenja Wiechmann1,5,6, Stephanie Heinzlmeir1, Johannes Scherr7, Bernhard Hemmer8,9, Mike Baynham10, Anne-Claude Gingras3, Oleksandr Boychenko11 & Bernhard Kuster1,4,5*

Nano-flow liquid chromatography tandem mass spectrometry (nano-flow LC–MS/MS) is the mainstay in proteome research because of its excellent sensitivity but often comes at the expense of robustness. Here we show that micro-flow LC–MS/MS using a 1 × 150 mm column shows excellent reproducibility of chromatographic retention time (<0.3% coefficient of variation, CV) and protein quantification (<7.5% CV) using data from >2000 samples of human cell lines, tissues and body fluids. Deep proteome analysis identifies >9000 proteins and >120,000 peptides in 16 h and sample multiplexing using tandem mass tags increases throughput to 11 proteomes in 16 h. The system identifies >30,000 phosphopeptides in 12 h and protein-protein or protein-drug interaction experiments can be analyzed in 20 min per sample. We show that the same column can be used to analyze >7500 samples without apparent loss of performance. This study demonstrates that micro-flow LC–MS/MS is suitable for a broad range of proteomic applications.

1 Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany. 2 Medical Research Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China. 3 Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada. 4 Bavarian Biomolecular Mass Spectrometry Center (BayBioMS), Technical University of Munich, Freising, Germany. 5 German Cancer Consortium (DKTK), partner site Munich, Munich, Germany. 6 German Cancer Research Center (DKFZ), Heidelberg, Germany. 7 Centre for Preventive and Sports Medicine, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany. 8 Department of Neurology, Klinikum Rechts der Isar, Medical Faculty, Technical University of Munich, Munich, Germany. 9 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. 10 Thermo Fisher Scientific, Runcorn, UK. 11 Thermo Fisher Scientific, Germering, Germany. *email: kuster@tum.de
nano-flow liquid chromatography (nano-flow LC) has been the mainstay in proteome research for >20 years, primarily because low flow rates improve peptide ionization by electrospray (ESI) for mass spectrometry (MS) and, hence, sensitivity. However, this comes at the cost of the challenge of manufacturing reproducible and long-lasting columns, maintaining stable ESI over extended periods of time, rapid chromatographic overloading, mass spectrometric saturation and often long, unproductive overhead times for sample transfer at low flow rates. These factors can limit the reproducibility of peptide identification and quantification as well as the comprehensive- ness, robustness and throughput of proteome analysis, particularly when analyzing samples of high complexity or wide dynamic range of protein concentrations as represented by tissues and body fluids.

Particularly for targeted quantitative MS assays such as selected reaction monitoring (SRM) or, more recently, parallel reaction monitoring (PRM), where the mass spectrometer is focused on a small number of analytes to maximize sensitivity and quantitative accuracy/precision, standard analytical HPLC columns (2.1 mm inner diameter, ID) are frequently used to address the aforementioned challenges. Also for so-called data-independent acquisition (DIA) methods that aim to catalog the peptides present in a sample systematically, the field is increasingly adopting 300 μm ID columns as a compromise between sensitivity and robustness.

As the sensitivity of mass spectrometers has greatly improved over the years as a result of, e.g., more efficient ionization, ion transfer and detection, further advances in untargeted (also referred to as discovery-type) and data dependent (DDA) proteome analysis may be sought by improving peptide separations. For example, Gonzalez et al. employed a standard 2.1 mm ID analytical column to identify about 800 proteins and 4,000 peptides from 40 μg E. coli protein digest using a 120 min LC gradient. However, sample quantities of that order may often not be available from biological sources. More recently, Lenço et al. reported the identification of about 2,800 human proteins in 60 min from 2 μg HeLa protein digests using an online LC–MS/MS method employing a 1 mm ID column. In a series of elegant experiments, that report demonstrated that discovery proteomics is feasible in principle using such a micro-flow LC–MS/MS system. Another recent interesting approach was presented by Bache et al. who introduced specialized new chromatographic hardware that aims to combine the advantages of micro-flow and nano-flow LC. Here, complex digests were separated at flow rates of 10–20 μL/min at very low pressure using stage tips, embedded in a pre-formed LC-gradient and subsequently analyzed by online nano-flow LC–MS/MS. The authors showed that the system identified nearly 10,000 human proteins and 130,000 peptides from fractionated HeLa protein digests within 18 h and that the system is stable across over 2,000 injections.

Here, we report on the systematic evaluation of the merits of online micro-flow LC–MS/MS for quantitative discovery proteome analysis using standard HPLC equipment available in any analytical laboratory. At the heart of the method is a commercial 1 × 150 mm reversed phase HPLC column operating at a flow rate of 50 μL/min coupled online to a sensitive and rapid mass spectrometer. Data collected from >2,000 samples show that most of the limitations of nano-flow LC can be overcome at a very moderate loss of practical sensitivity. The approach markedly improves robustness, throughput and reproducibility of quantification without the need for specialized equipment. The results suggest that this approach has the potential to transform the field because of the ease of its technical implementation, the wide range of feasible applications and the very high data quality which makes the system suitable for the analysis of clinical specimen.

Results and discussion
Basic performance characteristics of micro-flow LC–MS/MS.
The cross-sectional area of a 1 mm ID micro-flow LC column is 178 times larger than that of a 75 μm ID nano-flow LC column typically used in proteome research and the optimal flow-rate scales in the same way (Fig. 1a). While a wider column diameter improves separation efficiency by eliminating column overloading, the higher flow rate needed for a 1 mm ID column compared to a nano-flow LC column massively dilutes analyte concentration which should lead to a strong loss of electrospray ionization (ESI) efficiency and, as a result, sensitivity. We found that this can be partially offset by the very narrow LC peaks afforded by the higher flow rate which increases peptide concentration (Fig. 1b, c and Source Data File) and by adding traces of DMSO that we have shown to enhance peptide ionization (Supplementary Fig. 1 and Source Data File). As a result, only 5× more sample was required on the micro-flow compared to the nano-flow LC system when using a 28 Hz MS data acquisition method to obtain similar numbers and quality of peptide and protein identifications in single-shot analysis of complex HeLa protein digests while maintaining superior chromatographic performance throughout (Fig. 1d–g and Source Data File). The faster 41 Hz MS data acquisition method available on the Orbitrap HF-X may also be used but requires 10× more material in the micro-flow setup vs nano-flow LC–MS/MS (Supplementary Fig. 2 and Source Data File), which is why all of the data presented below (except for the TMT analysis, see methods) was collected using the 28 Hz method. A serial dilution analysis of the same HeLa protein digest showed that >1000 proteins could be identified from 200 ng of protein digest (quantified on protein level) when optimizing LC gradient times (30 min) and MS data acquisition parameters (28 Hz, Supplementary Fig. 3 and Source Data File) which is sufficient for a wide range of proteomic applications (see also below). We next tested the micro-flow system for the deep characterization of proteomes by off-line fractionation of digests using high pH reversed phase HPLC. From 200–400 μg of a HeLa or human placenta protein digest respectively (quantified on protein level), we identified nearly 10,000 proteins and between 120,000 and 140,000 peptides within 16 h of total analysis time. This is very comparable in quantity and quality to results from the recent nano-flow LC literature (Fig. 2a, b and Source Data File). We note that digestes in these reports were quantified on the peptide level. Hence, the actual differences between the quantities used between the different laboratories are likely smaller than the stated values. Because the micro-flow setup allowed for direct sample injection onto the column at 100 μL/min, the overhead times needed for sample loading, column equilibration etc. could be drastically reduced compared to popular nano-flow LC systems operating at low flow rates. This, in turn, increases the effective use of the mass spectrometer particularly for short gradient times, which supports the analysis of up to 96 samples per day (Fig. 2c, d).

Deep-scale analysis of TMT multiplexed proteomes.
Next, we set up two independent micro-flow LC systems on an Orbitrap Q Exact HF-X and an Orbitrap Fusion Lumos and found very similar performance and overlap for both label-free single shot and deep-scale proteomics analysis (Supplementary Fig. 4). A multiplexed and deep-scale analysis of 11 human cancer cell lines using tandem mass tags (TMT) resulted in identification of ~7800 and ~6400 proteins using MS2 (HF-X) and MS3 (Lumos) methods, respectively, from 250 μg of peptides within 16 h of total
Fig. 1 Qualitative performance characteristics of the micro-flow LC-MS/MS system. 

a Comparison of the cross-sectional areas of LC columns of different inner diameters (ID). 
b Boxplots summarizing the chromatographic peak width distributions (full width at half-maximum, FWHM) of all identified peptides for different LC gradient times (2 µg HeLa protein digest injected). Boxes and whiskers cover 50% and 1.5 times the interquartile range of the data respectively. Numbers above boxes denote the median FWHM values (in seconds), numbers below boxes represent the number of peptides contained in the analysis. 
c Example base peak chromatogram of 2 µg HeLa protein digest separated by a 60 min LC gradient. Selected chromatographic peaks are labeled with the m/z and FWHM values of the underlying peptide. 
d Bar charts comparing peptide identification results obtained for different sample loadings and LC gradient times using either nano-flow29, 31 (red) or micro-flow LC-MS/MS (blue). White numbers inside bars denote the number of peptides. 
e Same as panel d but for proteins. 
f Box plots comparing Andromeda peptide identification scores obtained for different sample loadings and LC gradient times using either nano-flow (red, n = 24,113 for 30 min, 36,442 for 60 min) or micro-flow (blue, n = 19,514 (1 µg), 23,047 (5 µg), and 23,630 (10 µg) for 30 min, and n = 14,992 (1 µg), 34,356 (5 µg), and 37,130 (10 µg) for 60 min) LC-MS/MS. 
g Same as panel f but for peptide chromatographic peak widths. The number of peaks used for each box are 22,254 (nano), 19,104 (micro, 1 µg), 21,820 (micro, 5 µg), and 22,065 (micro, 10 µg) for 30 min, and 33,772 (nano), 12,831 (micro, 1 µg), 32,835 (micro, 5 µg), and 34,884 (micro, 10 µg) for 60 min. Boxes and whiskers are defined as in b. Source data are provided as a Source Data file for b, f, and g.
an improved practical dynamic range of protein expression and separation performance of the micro-

LC–MS/MS system. a Summary of key experimental parameters and results comparing deep-scale proteome analysis of HeLa and placenta protein digests using published nano-flow LC–MS/MS data and data obtained by micro-flow LC–MS/MS in this study. b Box plots comparing Andromeda peptide identification scores of the data shown in a. Boxes and whiskers are defined as in Fig. 1b. The number of peaks used for each box are 303,597 (Kelstrup et al.23), 282,215 (Bache et al.29), 287,834 (This study, 400 µg HeLa), and 276,306 (This study, Placenta). Boxes and whiskers are defined as in Fig. 1b. Source data are provided as a Source Data file for b. c Summary of the actual sample throughput that can be achieved by the micro-flow LC–MS/MS system presented in this study compared to a typical nano-flow LC–MS/MS setup. d Overlay of base peak chromatograms of the same high pH reversed phase chromatography fraction of a TMT11 labeled peptides from a HeLa. e Summary of key experimental parameters and results comparing deep-scale proteome analysis of the micro-flow LC–MS/MS system. a Summary of key experimental parameters and results comparing deep-scale proteome analysis of the micro-flow LC–MS/MS system. The performance gap between the MS2 and MS3 measurements on the Lumos could be closed by extending the LC gradient time per fraction from 15 to 25 min leading to an increase of total analysis time from 16 to 24 h for the 11 proteomes. Again, the peptide quantities required to achieve this performance level were only 2–5 times higher than those for nano-flow LC (Fig. 2e). The improved chromatographic separation performance of the micro-flow system also led to an improved practical dynamic range of protein expression quantification between the 11 human cell lines (Fig. 2f). Specifically, dynamic range (defined to cover 95% of all protein expression ratios between cell lines, see methods) increased from 1:50 (nano-flow) to 1:75 (micro-flow) for MS2 measurements, to 1:100 (micro-flow) for MS3 measurements all using 15 min gradients, and to 1:150 (micro-flow) when extending the gradient time from 15 to 25 min (Fig. 2f). These above results demonstrate that deep-scale proteome analysis of higher organisms can be envisaged at a throughput of 11–16 proteomes per day.
Robustness and reproducibility of micro-flow LC-MS/MS. To demonstrate robustness and to explore the quantitative performance of the micro-flow LC system, we set up an experiment consisting of 1550 consecutive injections organized into 10 identical cycles (or batches) of 155 injections each (Fig. 3a) that were analyzed over the course of ~40 days. In each cycle, we analyzed 20 replicates of 2 µg of HeLa, 5 µg of urine, 5 µg of cerebrospinal fluid (CSF) and 5 µg of plasma (4 replicates from 5 individuals) protein digests as well as one deep-scale human placenta protein digest using 200 µg starting material. In total 500 fmol of synthetic peptide digests as well as one deep-scale human placenta protein digest (CSF) and 5 µg of plasma (4 replicates from 5 individuals) protein replicates of 2 µg of HeLa, 5 µg of urine, 5 µg of cerebrospinal fluid (CSF), and 5 µg of plasma (4 replicates from 5 individuals) protein digests were analyzed using micro-flow LC-MS/MS using the stated gradient times. Between each sample type, 3 replicates of 500 fmol PROCAL runs were added between each sample type to assess carry-over. We observed very stable peptide and protein identification numbers for the first 8 cycles, after which mass spectrometric performance dropped as a result of the accumulation of contaminants (Supplementary Figs. 5, 6 and Source Data File). In contrast, the performance of the micro-flow LC system was stable throughout (Fig. 3b and Source Data File, Supplementary Fig. 7 and Source Data File). The retention times (RT) of the spiked PROCAL peptides showed an average CV of 0.26% (n > 40,000 data points) demonstrating very high chromatographic reproducibility. Remarkably, the RTs of PROCAL peptides measured alone or spiked into body fluid samples were essentially the same (Pearson correlation coefficient of 0.997, Supplementary Figs. 5 and 6). To further investigate the robustness and reproducibility of the micro-flow LC system, we developed a long-term performance test consisting of 10 cycles of 155 injections in ~4 days. In each cycle, we analyzed 20 replicates of 2 µg of HeLa, 5 µg of urine, 5 µg of cerebrospinal fluid (CSF), and 5 µg of plasma (4 replicates from 5 individuals) protein digests from five individuals, as well as one deep-scale placenta protein digest (200 µg protein digest separated into 48 fractions) were analyzed using micro-flow LC-MS/MS. Between each sample type, 3 replicates of 500 fmol PROCAL synthetic retention time standards were injected. a Design of a long-term performance test consisting of 10 cycles of 155 injections each. In each cycle, 20 replicates of 2 µg HeLa, 5 µg urine, 5 µg cerebrospinal fluid (CSF), and 5 µg each of 4 replicates of plasma digests from five individuals, as well as one deep-scale placenta digest (200 µg protein digest separated into 48 fractions) were analyzed using micro-flow LC-MS/MS using the stated gradient times. Between each sample type, 3 replicates of 500 fmol PROCAL synthetic retention time standards were injected. b Retention time stability of PROCAL peptides measured alone or spiked into body fluids (urine, CSF, plasma samples; main plot). The equation represents the linear model that was fitted to the data ($R^2$, squared Pearson correlation coefficient). The bottom right inset shows an expanded view of the main plot showing the retention time distribution of three closely eluting PROCAL peptides (in different colors) across all experiments. The top left inset summarizes the carry-over analysis across all 10 cycles (columns denote average carry-over and error bars denote the standard deviation). Source data are provided as a Source Data file. c Cumulative density plot showing the inter-cycle reproducibility of protein quantification for 200 injections of HeLa, urine, and CSF, and 40 injections of five individual plasma samples (common proteins only). Dotted lines denotes the percentage of proteins that show <20% coefficient of variation (CV) in the analysis. The inset shows a t-SNE analysis of the 40 plasma injections measured from each of five individuals across the 10 cycles. d Same as in c but at the peptide level.
$R^2$ of 1.00; slope of 1.00, intercept of <0.01 min), demonstrating the near absence of chromatographic matrix effects$^{33}$ (Fig. 3b and Source Data File). In addition, sample carry-over was extremely low (average of 0.16% for CSF, 0.11% for plasma, 0.05% for placenta and 0.01% for urine and HeLa; Fig. 3b and Source Data File) removing a common issue of nano-flow LC particularly for the analysis of tissues and body fluids$^{34}$. The low carry-over is likely owing to the very low amount of sample loaded on the column relative to its capacity and the high volume of solvents passing over the column.

The exquisite chromatographic reproducibility also led to a very high reproducibility of protein and peptide quantification (Fig. 3c, d). Across all single shot samples, the median CV of common quantified peptides was between 4.6% (HeLa) and 7.2% (urine). Variation within cycles was even smaller (Supplementary Fig. 8) and practically all proteins had CVs below 20%. At the peptide level, the median CV of common quantified peptides was between 11.6% (HeLa) and 14.2% (CSF) and variation within cycles was again smaller (Supplementary Fig. 9). Between 75% and 85% of all quantified peptides showed CVs of <20% (Fig. 3d). Reproducibility of quantification was further assessed by a t-SNE analysis$^{35}$ at both protein (Fig. 3c) and peptide (Fig. 3d) level that clustered the 40 replicates of the 5 individuals from whom plasma was repetitively analyzed across the 10 cycles. Batch effects between the 10 cycles were observed for each sample type, but on a much smaller scale compared to the differences between human subjects (Supplementary Fig. 10).

The above figures of merit are very encouraging, as most quantified proteins would pass guidelines on quality specifications for clinical assays$^{36}$. Such performance characteristics would be difficult if not impossible to achieve by nano-flow LC separations because of the very high dynamic range of protein concentrations present particularly in body fluids. Such high dynamic range compromises the quality of chromatographic separations at sample loadings that also yield high numbers of peptide and protein identifications. Both issues can be overcome using the micro-flow setup presented here. Using 30 min LC gradients, about 250 plasma proteins (2300 peptides), 600 CSF proteins (4500 peptides) and 1100 urine proteins (5000 peptides) were identified from 5 µg of protein digest (Supplementary Figs. 5, 6 and Source Data File, and Supplementary Fig. 11 and Source Data File). Such quantities are easily obtained from biological sources and the number of identifications is comparable to figures reported in the recent nano-flow LC literature$^{37-39}$. In contrast to nano-flow LC, the micro-flow LC system showed no obvious sign of overloading even when injecting as much as 20 µg protein digest of non-depleted protein digest, (Supplementary Fig. 12a). However, there is evidence that the mass spectrometric signal may saturate at high loading of plasma samples for some high abundant peaks (Supplementary Fig. 12b).

An further important consideration for high-throughput applications or experiments requiring reproducible results over extended periods of time is column lifetime. During the 40 days of the long-term performance test described above, the column showed very high separation reproducibility throughout as demonstrated by the overlay of the base peak chromatograms of 10 urine samples (one from each cycle; Supplementary Fig. 13a). At the time of writing, the column that was used in this study had separated >7500 samples over the course of about 1 year with no apparent loss of performance (Supplementary Fig. 13b and Source Data File) further attesting to the high potential of this set-up for implementation in clinical proteomics research.

**Sub-proteome analysis by micro-flow LC–MS/MS.** We extended the range of applications of micro-flow LC–MS/MS to the analysis of proteomes of lesser complexity, notably to protein–protein interactions using affinity purification (AP) of tagged proteins$^{40}$ or employing proximity labeling (BioID)$^{31}$. Single-shot analysis of AP-MS and BioID-MS experiments (2 biological × 3 technical replicates for each bait protein) on the micro-flow LC system (15 min gradients each) recovered between 86% and 96% of the high-confident interactors identified in a previous publication$^{40}$ and in the Human Cell Map project$^{32}$ (Fig. 4a, Supplementary Data 1, 2). Both previous analysis used nano-flow LC systems and much longer gradient times. We also downloaded the ten most highly confident interaction partners from each bait from the String database$^{33}$ (Supplementary Data 3) and found that most of these interaction partners were also identified as high confidence interactors in the micro-flow LC–MS/MS data and subsequent analysis by the software package SAINTP$^{44,45}$ (Fig. 4b, c, Supplementary Fig. 14).

We next applied the micro-flow approach to the analysis of drug–protein interactions using the kinobeads approach$^{46}$. Illustrated by the kinase inhibitor AT-9283 as an example (Fig. 4d), the micro-flow LC system covered >90% of the measured protein kinases but in less than a third of the gradient time compared to the same experiment performed by nano-flow LC (Supplementary Fig. 15a, b). More importantly, the main targets of the drug were all identified and the effective concentration of drug needed to compete 50% of the bound kinase (EC$_{50}$) obtained from the dose-response curves characterizing the interactions were very similar between experiments measured by micro- or nano-flow LC–MS/MS (Supplementary Fig. 15c–f).

As shown above (Fig. 1b), micro-flow LC separations using short gradient times generated very sharp LC peaks (here, median of 3.0 s, FWHM) and 8–10 data points across an LC peak are typically required for accurate determination of the LC peak area and thus quantification. Therefore, it is important to match the cycle time of the mass spectrometer (i.e. the time between two MS1 scans) to the chromatographic resolution (here ~0.6 s; see also Supplementary Fig. 2). Figure 4e shows extracted ion chromatograms for the AURKA peptide QWALEDFEIGRPPLLK as an example and the data shows that sufficient data points cover the LC peak at all drug doses allowing the determination of the peak area with good confidence. This is even more important for the analysis of post-translational modifications such as phosphopeptides, as their quantification cannot be stabilized by aggregating several peptide quantification measurements into one protein quantification value. The dose response curves shown in Fig. 4f representing two peptides containing the phosphorylation site S212 (ARTSpS-FAEPGGGGGGGGGSASGPPTGGGK) and two peptides containing the site Y729 (QLVRGEPNVSPPYICSR and RGEVPNVSPPYICSR) of the kinase GSK3A show that the EC$_{50}$ values obtained from the individual phospho-peptides are very similar to that of the aggregated protein indicating very good quantification also at the peptide level. Given such qualitative and quantitative performance levels, it becomes feasible to screen the targets of hundreds to possibly thousands of kinase inhibitors in this way.

**Analysis of phosphoproteomes by micro-flow LC–MS/MS.** Finally, we investigated the merits of the micro-flow LC system for phosphoproteome analysis. To this end, 2 mg of HeLa protein digest were fractionated by off-line high pH reverse phase chromatography into 96 fractions that were pooled into 12 fractions followed by IMAC phosphopeptide enrichment of each of the 12 fractions. Two workflow replicates were prepared and the phosphopeptides were analyzed by either micro-flow or nano-flow LC–MS/MS coupled to an Orbitrap QE HF-X mass spectrometer and using the same (60 min) LC gradient and MS data acquisition
methods. Somewhat surprisingly given the overall low abundance of phosphopeptides (typically ~1% relative to the total), micro LC–MS/MS identified 32,493 unique phosphopeptides and 27,639 phosphorylation sites corresponding to 4,886 phosphoproteins within 12 h of gradient time. The same sample analyzed by nano-flow LC resulted in the identification of 28%, 14%, and 18% more phosphopeptides, phosphorylation sites and phosphoproteins respectively, confirming that nano-flow LC plays out its advantages when sample quantities are low (Fig. 5a). This is also reflected by a lower identification score (Fig. 5d and Source Data File) as a result of the lower absolute signal intensity in micro-flow LC–MS/MS compared to nano-flow LC–MS/MS. As expected, LC peaks were sharper for micro-flow LC compared to nano-flow LC separations (Fig. 5e and Source Data File) which may improve the separation of phosphorylation site isomers. Using the peptide SGAQASSTPLSPTR of Lamin A/C as an example, Fig. 5f shows that separation of multiple such singly phosphorylated peptide isomers is indeed possible (for further examples, see Supplementary Fig. 16). Also, the micro-flow LC system is more efficient in separating the S18 and S22 phosphorylation isomers as deduced from the ~40% higher ratio of the difference in retention time (deltaRT) divided by the chromatographic peak width. From our data, it is, however, not clear if this higher separation efficiency generally translates into a clear advantage for the micro-flow LC setup for the separation of phosphorylation isomers as we did not find enough cases of closely eluting phosphopeptide isomers on which this hypothesis could be tested. Nevertheless, it is noteworthy that the micro-flow LC setup performed quite well for the analysis of phosphoproteomes, which makes this approach worth considering for high-throughput applications or for projects in which sample availability is not a concern.

In conclusion, this study showed that micro-flow LC–MS/MS is a very versatile alternative to the conventional nano-flow LC approach for a broad range of proteomic applications. Because of its robust qualitative and quantitative performance characteristics, the simplicity of implementation and the very broad range of available high-quality micro-flow columns, the authors expect that the approach will be broadly enabling for many experts as well as non-specialized laboratories. The approach also paves the way for more routinely translating proteomics into clinical applications, particularly for quantitative and high-throughput body fluid analysis. In addition, the very large amount of data collected in this project, should be a useful resource for the scientific community to further investigate aspects of the methodology that are not covered in this report.

Methods

Sample selection and preparation. Human specimen used in this study were obtained following informed consent and observing the appropriate ethics approval process of the Technical University of Munich. The study was approved by the ethics committee of the faculty of medicine of the Technical University of Munich.

Fig. 4 Application of the micro-flow LC–MS/MS system to the analysis sub-proteomes. a Recovery analysis of high-confidence interactors obtained by replicate analysis (n = 6; two biological replicates and three technical replicates) of affinity purifications performed using FLAG-tagged human MEPCE and EIF4A2, and BioID proximity labeled human LMNA, NIFK, and CTNN1 and analyzed by on-one shot micro-flow LC–MS/MS using 15 min gradients (AP-MS and BioID-MS), compared to results of the same experiments published previously or part of the Human Cell Map project. Numbers inside the bars represent the number(274,61),(463,84) of interactors identified by micro-flow LC–MS/MS vs those annotated in the aforementioned resources. b Interaction networks based on the ten most confident interaction partners in STRING for the bait proteins MEPCE (FLAG affinity purification). FC denotes the fold change values of an interaction partner (over control pulldowns) assessed by SAINTEXpress analysis. Proteins without FC annotations indicate that they were not identified as high confident interactors in this micro-flow LC–MS/MS study. c Same as b but for NIFK (BioID proximity labeled). d Example dose-response curves of protein kinases that are the targets of the kinase inhibitor AT-9823 obtained by kinobeads competition pulldown experiments and analyzed by micro-flow LC–MS/MS system using a 15 min gradient for each drug dose. EC50 values (effective concentration 50) denote the drug concentration necessary to compete 50% of the binding of a protein to kinobeads. e Extracted ion chromatograms of the AURKA peptide QWALEDFEIGRPLGK from the kinobead experiment in d illustrating the quantification of this peptide as a function of the applied drug dose. f Dose–response curves akin to d but for four phosphopeptides containing the phosphorylation sites pS21 and pY279 of the protein kinase GSK3A, as well as the aggregated data for the entire protein.
In total 1 mL plasma samples collected from five healthy donors were centrifuged for 10 min at 4000 g. In total 50 μL supernatant were taken out and diluted by 5 volumes of 8 M urea buffer containing 80 mM Tris–HCl, pH 7.6, and then stored at −80 °C until further use. Second-morning mid-stream urine was collected from healthy donors and centrifuged at 4000 × g for 30 min to remove cell debris. The supernatant was vacuum concentrated 5-fold using a SpeedVac and proteins were precipitated using ethanol, incubated at −80 °C for 30 min. The resulting supernatant was stored at −80 °C until further use.

The 10 pancreatic cell lines (BxPC-3 (ATCC, CRL-1687), Dan-G (DSMZ, ACC 249), HPAC (ATCC, CRL-2119), HuPta-4 (DSMZ, ACC 223), IMIM-PC-1 (PRID, CVCL_4061), MiaPaCa2 (ATCC, CRM-CRL-1420), Panc-10.05 (ATCC, CRL-2547), PaTu-8998-S (DSMZ, ACC 204), PaTu-8998-T (DSMZ, ACC 162) and PSN-1 (ATCC, CRL-3211)) were provided by Günter Schneider. The cell lines were cultured according to the cell line provider's recommendations to 80–100% confluence. The cells were lysed with 8 M urea, 40 mM Tris/HCl (pH 7.6), 1 × EDTA-free protease inhibitor mixture (Complete Mini, Roche), and 1 × Phosphatase inhibitors (Sigma Aldrich). The cell lysate was clarified by centrifugation at 20,000 × g for 20 min. The supernatants were used for in solution trypsin digestion.

### Protein digestion and peptide desalting

Protein concentration was measured by the Bradford assay. Proteins were reduced by 10 mM DTT at 37 °C for 1 h, and alkylated by 55 mM chloroacetamide (CAA) at room temperature for 30 min in the dark. For the CSF sample, the protein solution was mixed with one volume of 40 mM Tris–HCl (pH 7.6). For all the other samples, the protein solution were mixed with five volumes of 40 mM Tris/HCl (pH 7.6). Proteins were digested with sequencing grade trypsin (Roche) at a protease-to-protein ratio of 1:100 (w/w) for 4 h, followed by adding further trypsin (1:100) and incubating overnight at 37 °C. Digestion was quenched by addition of formic acid (FA) to a final concentration of 1%, and the resulting peptide mixture was centrifuged at 5000 × g for 15 min. Peptides in the supernatant were loaded on Sep-Pak C18 Cartridges (Waters) and eluted by 50% ACN, 0.1% FA in water and dried in a SpeedVac. Samples were stored at −80 °C until further use.

### TMT labeling

For TMT11-plex labeling, desalted peptides from HeLa, and ten pancreatic cell line protein digests were reconstituted in 61% FA and peptide concentration was determined by NanoDropTM 2000 (Thermo Scientific). The peptides from the cell lines of MiaPaCa2, HuPta-4, BxPC-3, HPAC, Panc-10.05, BxPC-3. Therefore, the TMT labeling procedure was performed as following:

- **TMT labeling**: 100 picomoles of each sample were labeled with TMT 11-plex reagents. The labeled samples were mixed in equimolar amounts and incubated at room temperature for 1 h. Following the labeling reaction, the samples were desalted by using a Source: Data File for a. **Extracted ion chromatograms of phosphorylation site isomers of the peptide SGAQASSTPLSPTR from human lamin A/C**

Fig. 5 Micro vs nano-flow LC-MS/MS for analysis of phosphoproteomes. a Venn diagram showing the number and overlap of phosphopeptides identified by nano-flow LC-MS/MS (red) and micro-flow LC-MS/MS (blue) using 2 mg of HeLa protein digest, separated into 12 high-pH reversed phase chromatography fractions, enriched for phosphopeptides using IMAC, and analyzed by 60 min LC gradients. b Same as in panel a but for phosphorylation sites. c Same as in a but for phosphoproteins. d Boxplots showing the Andromeda score distribution for peptides identified by nano-flow LC-MS/MS (red, n = 109,417) and micro-flow LC-MS/MS (blue, n = 77,763). e Box and whiskers are defined as in Fig. 1b. Numbers above box plots denote the median Andromeda score. f Same as in d but for the chromatographic peak width of the peptides (red, n = 102,066, blue, n = 83,908). Source data are provided as a Source Data file for e and f. 

**Key Figure 5**: This figure illustrates the comparison of phosphoproteomes analyzed using micro-flow LC-MS/MS and nano-flow LC-MS/MS. The Venn diagram (a) shows the overlap and unique phosphopeptides identified from HeLa protein digest using 12 high-pH reversed phase chromatography fractions, enriched for phosphopeptides using IMAC, and analyzed by 60 min LC gradients. The boxplots (b) display the Andromeda score distribution for peptides identified by micro-flow LC-MS/MS and nano-flow LC-MS/MS. The extracted ion chromatograms of phosphorylation site isomers of the peptide SGAQASSTPLSPTR from human lamin A/C are also shown (c). The key figure also includes boxplots (d) for the chromatographic peak width of the peptides, with numbers above the box plots denoting the median Andromeda score. Extracted ion chromatograms of phosphorylation site isomers of the peptide SGAQASSTPLSPTR from human lamin A/C are also shown (e). Finally, boxplots (f) illustrate the distribution of the chromatographic peak width of the peptides identified by micro-flow LC-MS/MS and nano-flow LC-MS/MS.
PSN-1, Dan-G, Pa-Tu-8998-S, Pa-Tu-8998-T, IMIM-PC-1 and HeLa were labeled. BioID samples were prepared according to the previous protocol. Briefly, 200 μg peptides of each of the eleven cell lines were dried in a SpeedVac and reconstituted in 40 μl of 50 mM HEPES buffer (pH 8.5). Then, 0.2 mg TMT reagent in 10 μl dry ACN was added to each sample and mixed with a pipette. The mixture was incubated at 25 °C and 400 rpm on a thermostorer for 15 min. After 5 μl of 5% Hydroxyamine solution was added to each sample to stop the reaction, the reaction was performed at 25 °C and 400 rpm on a thermostorer for 15 min. Finally, the labeled peptides were pooled together and 40 μl of 10% FA in 100 ACN were added. To avoid one SpeedVac step before peptide desalting, peptide solutions were diluted by 20 volumes of 0.1% FA, and purified by the Sep-Pak C18 Cartridge and the eluate was dried in a SpeedVac and stored at −80 °C until further use.

Off-line high pH reversed phase peptide fractionation. A Dionex Ultra 3000 HPLC system operating a Waters XBridge BEH310 C18 3.5 μm × 250 mm column was used to fractionate peptides at a flow rate of 200 μl/min. Buffer A was 25 mM ammonium bicarbonate (pH = 8.0), buffer C was 100% ultrapure water (ELGA), buffer D was 100% ACN, buffer B was not used in this system. The proportion of buffer A was kept at 99% during the separation. Fraction collection was every minute and fractions were collected into a 96 well plate. For non-labeled peptides, the 200 or 400 μg protein digests were separated by a linear gradient from 5% D to 50% D in 87 min, and followed by a linear gradient from 50% D to 80% D in 5 min.

To be able to compare results obtained in this study to data from the literature, 92 fractions were collected (2 min to 94 min) and subsequently pooled into 46 fractions by adding fraction 47 to fraction 1, fraction 48 to fraction 2 and so forth. The HeLa protein digest (200 μg and 400 μg) and placenta protein digest (200 μg) were fractionated into 46 fractions to compare this with the published data. For the placenta digest used for the 10 cycles’ long-term test, we collected 96 one minute fractions (between 1 min and 97 min) and pooled these into 48 fractions (as above). For TMT labeled peptides, 500 μg pooled peptides were separated by a linear gradient from 9% D to 42% D in 86 min, followed by a linear gradient from 42% D to 80% D in 12 min. 96 fractions were collected and pooled into 48 fractions (as above). Peptide fractions were frozen at −80 °C freezer for at least 1 h and dried in a SpeedVac without prior desalting.

Phosphopeptide separation and Fe-IMAC enrichment. In total 2 mg HeLa protein digest was separated on a 2.1 × 150 mm Waters XBridge BEH310 C18 3.5 μm column with a linear gradient from 4% D to 32% D in 45 min, ramped to 80% D in 6 min, and kept there for 3 min before ramped back to 5% D in 2 min and 96 fractions were collected at 0.5 min intervals. Peptides were pooled in a step-wise fashion from 96 to 24 to 12 fractions during the separation. Fractions were dried in a SpeedVac and stored at −80 °C until performing phosphopeptide enrichment. Phosphopeptides were enriched from each of the 12 fractions using Fe(III)-IMAC-NTA (Agilent Technologies) on the AssayMAP Bravo Platform (Agilent Technologies). IMAC cartridges were primed with 100 μl 1% (v/v) ACN/buffer with 50 mM 0.1% TFA, samples were dissolved in 200 μl of loading buffer and loaded onto cartridges. The cartridges were washed with 50 μl loading buffer, and phosphopeptides were eluted with 40 μl of 1% ammonium bicarbonate. Phosphopeptides were dried down and stored at −80 °C until subjected to LC-MS/MS analysis.

Kinobeads, FLAG based AIPs and BioID pull-downs. Kinobeads selectivity profiling of AIP-7283 was performed with the standard protocol. The K-562 (ATCC, CCL-243), COLO-205 (ATCC, CCL-222) and MV-4-11 (ATCC, CRL-9591) cells were cultured in RPMI 1640 medium (Biologchrom GmbH) supplemented with 10% (v/v) FBS (Biologchrom GmbH) and 1% (v/v) antibiotics. SK-N-BE(2), K-562 and IMIM-PC-1 cells were cultured in IMDM medium (Biochrom GmbH) supplemented with 10% (v/v) FBS (Biochrom GmbH) and 1% (v/v) antibiotics. 2.5 mg of a protein mixture from the sample loop was incubated with 500 μl of the IMIM-PC-1 and IMIM-PC-1 cell extracts and incubated at 4 °C for 3 h with rotating. Cell pellets for FLAG AIPs were resuspended in a 1:4 cell buffer ratio. FLAG affinity purification (on magentic M2-antibody conjugated anti-Flag beads, Sigma-Aldrich, M8828) was performed for 2 h at 4 °C on a nutator. After that, with washes of 1 ml lysis buffer, the samples were washed and followed by an additional wash with 1 ml of 20 mM TrisHCl pH 8, and 2 mM CaCl2. Finally, samples were trypsinized on-beads overnight at 37 °C, rotating and without alkylolation/reduction, and dried in a SpeedVac.

Setup of the online micro-flow LC-MS/MS system. A Dionex Ultimate3000 RSL.Cnano System was coupled online to a Q Exactive HF-X or an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) in this study. For the final LC setup, we used nanoViper capillaries for all the connections. The pump outlet of the nanoViper capillary (sample loop) was connected to sample loop (sample loop) to the column inlet. A 20 μl sample loop was used with the micro-flow LC system in direct injection mode. The sample loop was 100 μl in volume and 0.1 mm ID × 300 mm OD. The ion spray voltage was used to connect the column outlet to the ground metal union of the Ion Max API source. Another 50 μm ID × 150 mm nanoViper capillary was used to connect the other end of the ground metal union to the sample inlet of HESI-II probe (50 μm ID) of the Ion Max API Source. The probe depth was set to 0.1 m of the Ion Max API source.

The results reported in this study were all obtained on a commercially available Thermo Fisher Scientific Acclaim PepMap 100 C18 LC column (2 μm particle size, 1 mm ID × 150 mm; catalog number 164711). Column temperature was maintained at 55 °C using the integrated column oven. As a side note, we evaluated five different PepMap columns which were produced by Thermo Scientific (Scientist, Nov. 2017, and 2018, respectively and found that the 2012 and 2013 batch showed better separation efficiency, which is why the 2013 batch column was used throughout this study.

Three LC pumps available on Dionex Ultimate3000 RSL.Cnano System (the loading pump, the NC gradient pump and a modified Vanquish pump) were used to deliver the gradient. The micro-flow LC-MS/MS system was initially developed by delivering gradients using the loading pump. The DMSO titration experiment was performed using the loading pump at a flow rate of 68 μl/min, and using linear gradients of 5–28% B, 4–27% B, 3–26% B, 2–25% B, 1–24% B, 0.1–23.1% B for solvent A with 0%, 1%, 2%, 4% and 10% solvent B, respectively. Although the loading pump can be used, we observed a rather long gradient delay, resulting from gradient mixing before the loading pump head and resulting in about 220 μl dead volume. Such gradient delays are unacceptably long for 15 min and 30 min gradient volumes. Therefore, we installed a micro flowmeter (catalog number 6041.7903 A, maximum flow rate 50 μl/min) on the NC pump to deliver the gradient, improving particularly for short gradients. Therefore, a modified Vanquish pump was used for the method development parts of the manuscript, including gradient testing, serial HeLa dilution tests, deep-scale fractionated HeLa and placenta protein digests tests, etc. The presented data demonstrates that the NC pump/micro flowmeter setup was very robust.

However, as the highest flow rate for the NC pump/micro flowmeter combination is 50 μl/min, LC overhead time was still not optimal and should be improved particularly for short gradients. Therefore, a modified Vanquish pump capable of delivering gradients of 1 μl/min was used.

This modified Vanquish pump is a binary gradient pump, and has technical characteristics similar to standard high-pressure binary gradient pumps in the NCS-3500RS module (https://assets.thermofisher.com/TFS-Assets/CMD/Specification-Sheets/PS-71899-LC-UltiMate-3000-RSL.Cnano-PS71899-EN.pdf) with a pump 1 µl flow rate of <25 nL min−1 and maximum column volume of 1000 nL. There was no additional mixer installed between the pump outlet and fluidics. All observed delays in elution are associated only with the volume of the column, the injection loop and the capillary connections between pump, column, auto-sampler, injection loop, and HESI probe.

This makes it possible to flush the column at a flow rate of 100 μl/min and decrease the total overhead time to 5 min, including 3 min sample injection and 2 min column equilibration time. A flow rate of 50 μl/min was used to deliver the linear gradients. This setup enabled a throughput of 96 samples per day using 10 min gradient time. The modified Vanquish pump was used for the long-term performance test encompassing 1550 injections of different samples, the deep-scale
**Data processing and analysis.** Except otherwise noted, the data were searched by MaxQuant v1.6.2.32 against the UniProtKB Human Reference Proteome database (v22.07.13, 88,381 entries). Default MaxQuant parameters were used. Trypsin was specified as the enzyme, cleaving after all lysine and arginine residues and allowing for one missed cleavages. Cysteine carbamidomethylation was specified as fixed modification and protein N-terminal acetylation and oxidation of methionine were considered as variable modifications. The false discovery rate (FDR) was set to 1% on the peptide-spectrum match (PSM) and protein levels. For the 1,550 injection experiments. For the comparison with previously reported high confidence interactions, we obtained the dataset of AP pull-down samples (EIF4A2 and MECPE) from the supplemental information of published paper49, and the dataset of the BioID pull-down samples (LMNA, NIFK, and CTNNA1) was downloaded from the Human Cell Map website (https://human.cellmap.org). Online STRING analysis of the five baits41, the line thickness between each protein indicates the strength of the data support, all the active interaction sources including textmining, experiment, databases, co-expression, gene fusion, neighborhood and co-occurrence were used. The minimum required interaction score was set at 0.4 with the minimum number of interactors for the 1st shell, and none interactors for the 2nd shell. The structure previews inside the network bubbles were enabled. The fold change (FC)
values of the interactors (over control pull-downs) assessed by SIAINT analysis were annotated in the STRING network (Fig. 4b, c and Supplementary Fig. 14, proteins without FC annotation were not identified as high confident interactors in SIAINT analysis).

Data analysis downstream of MaxQuant output results was performed in R44. The evidence result file was used for the carry-over analysis. After removing all the reverse peptides, the peptide intensities of each raw file were summarized. As albumin is one of the most abundant proteins in the body fluid samples, we kept the peptides matched to potential contaminating proteins for carry-over analysis. The evidence result files were used for the PROCAL retention time (RT) analysis, as we searched the raw files with both human and PROCAL sequences, we firstly removed all the peptides identified from the human database. For technical reasons, three raw files HeLa_P035214_BA1_S00_A00_R11, Plasma_P035250_BE1_S00_A00_R2 and CSF_P035234_PCI_S00_A00_R15 had to be excluded from the retention time analysis. As the MBR search was enabled, there was sometimes more than one retention time value for one peptide in the same raw file. In such cases, only the retention time with the highest intensity value was kept. In addition, peptides whose intensity values were below SE7 were excluded from the analysis. Regarding the dynamic range analysis of the TMT sample: dynamic range was defined as the ratio of the maximum and minimum intensity values of a peptide in the TMT11 channels representing different human cancer cell lines. If there were zero value channels, the intensity of the channel with smallest value above zero was used as the minimum intensity value. To ensure a fair comparison, only peptides identified in all the four LC–MS/MS experiments were considered. If peptides were represented by more than one MS/MS spectrum as in the MS/MS spectra as internal standards for biomarker discovery and/or verification studies. Methods 81, 24–33 (2015).

Gillet, L. C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell Proteom. 11, 016717 (2012).

Vowinkel, J. et al. Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microLC and data-independent acquisition. Sci. Rep. 8, 4346 (2018).

Bruderer, R. et al. Analysis of 1508 plasma samples by capillary-flow data-independent acquisition profiling proteomes of weight loss and maintenance. Mol. Cell Proteom. 18, 1242–1254 (2019).

Hahne, H. et al. DMSO enhances electrospy response, boosting sensitivity of proteomic experiments. Nat. Methods 10, 989 (2013).

Hebert, A. S. et al. Improved precursor characterization for data-dependent mass spectrometry. Anal. Chem. 90, 2333–2340 (2018).

Ridgway, M. E., Bleiderer, C., Mann, M. & Park, M. A. Trends in trapped ion mobility–mass spectrometry instrumentation. Trends Anal. Chem. 116, 324–331 (2019).

Meier, F. et al. Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. Mol. Cell Proteom. 17, 2534–2545 (2018).

Hebert, A. S. et al. Comprehensive single-shot proteomics with FAIMS on a hybrid orbitrap mass spectrometer. Anal. Chem. 90, 9529–9537 (2018).

Shishkova, E., Hebert, A. S. & Coon, J. J. Now, more than ever, proteomics needs better chromatography. Cell Syst. 3, 321–326 (2019).

Shishkova, E., Hebert, A. S., Westphall, M. S. & Coon, J. J. Ultra-high pressure (<30,000 psi) packing of capillary columns enhancing depth of shotgun proteome analyses. Anal. Chem. 90, 11503–11508 (2018).

Töth, G., Paník-Jankóvari, T. & Mitulovic, G. Pillar array columns for peptide separations in nanoscale reversed-phase chromatography. J. Chromatogr. A 1603, 426–432 (2017).

Gonzalez Fernandez-Nino, S. M. et al. Standard flow liquid chromatography for shotgun proteomics in bioenergy research. Front. Bioeng. Biotechnol. 3, 44 (2015).

Lenco, J. et al. Conventional-flow liquid chromatography-mass spectrometry for exploratory bottom-up proteome analyses. Anal. Chem. 90, 5381–5389 (2018).

Bache, N. et al. A Novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. Mol. Cell Proteom. 17, 2284–2296 (2018).

Wiszniowski, J. R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J. Proteome Res. 8, 5674–5678 (2009).

Kelstrup, C. D. et al. Performance evaluation of the Q exactive HF-X for shotgun proteomics. J. Proteome Res. 17, 727–738 (2017).

Zolg, D. P. et al. PROCAL: a set of 40 peptide standards for retention time calibration of 97 proteins in dried blood spots. Circ. Cardiovasc Genet. 10, 1475–1481 (2017).

van der Ende, E. L. et al. Novel CSF biomarkers in genetic frontotemporal dementia identified by proteomics. Ann. Clin. Transl. Neuro. 6, 987–707 (2019).

Saito, S. et al. The Optimized Workflow for Sample Preparation in LC/MS- MS-Based Urine Proteomic. Methods Protoc. 2, 46 (2019).

Niu, S. et al. Plasma profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Mol. Syst. Biol. 15, 6879 (2019).

Tsou, C.-C. et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat. Methods 12, 258 (2015).

Goyaud, E. et al. BioID-based identification of Skp Cullin F-box (SCF) complex substrates. J. Proteome Res. 9, 5960 (2010).

van der Ende, E. L. et al. Novel CSF biomarkers in genetic frontotemporal dementia identified by proteomics. Ann. Clin. Transl. Neuro. 6, 987–707 (2019).

Saito, S. et al. The Optimized Workflow for Sample Preparation in LC/MS- MS-Based Urine Proteomic. Methods Protoc. 2, 46 (2019).

Niu, S. et al. Plasma profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Mol. Syst. Biol. 15, 6879 (2019).

Tsou, C.-C. et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat. Methods 12, 258 (2015).

Goyaud, E. et al. BioID-based identification of Skp Cullin F-box (SCF) complex substrates. J. Proteome Res. 9, 5960 (2010).
43. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607–D613 (2018).
44. Teo, G. et al. SAINTEXpress: improvements and additional features in Significance Analysis of INTeractome software. *J. Proteom.* **100**, 37–43 (2014).
45. Choi, H. et al. SAINT: probabilistic scoring of affinity purification–mass spectrometry data. *Nat. Methods* **8**, 70 (2011).
46. Kleyer, S. et al. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol. Cell Proteom.* **18**, 1468–1478 (2019).
47. Yu, P. et al. Trimodal Mixed Mode Chromatography That Enables Efficient Offline Two-Dimensional Peptide Fractionation for Proteome Analysis. *Anal. Chem.* **89**, 8884–8891 (2017).
48. Zecha, J. et al. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol. Cell Proteom.* **18**, 1468–1478 (2019).
49. Reinecke, M. et al. Chemoproteomic selectivity profiling of PIKK and PI3K kinase inhibitors. *ACS Chem. Biol.* **14**, 655–664 (2019).
50. Lambert, J.-P. et al. Mapping differential interacomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat. Methods* **10**, 1239 (2013).
51. Hesketh, G. G., Youn, J. Y., Samavarchi-Tehrani, P., Raught, B. & Gingras, A. C. Parallel Exploration of Interaction Space by BioID and affinity purification coupled to mass spectrometry. *Methods Mol. Biol.* **1550**, 115–136 (2017).
52. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
53. Youn, J. Y. et al. High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell* **69**, 517–532 (2018). e511.
54. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comput. Graph Stat.* 5, 299–314 (1996).
55. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

**Acknowledgements**

Y.B. is grateful for postdoctoral fellowships by the Alexander von Humboldt Foundation, the Carl Friedrich von Siemens Foundation and the National Natural Science Foundation of China (81600046). R.Z. is grateful for a PhD scholarship by the Chinese Scholarship Council (CSC). Parts of this work were funded by grants from the German Science foundation (DFG-SFB1309), the ProteomeTools project (BMBF; grant no. 031L0008A) and the German Excellence Initiative. Work in ACG’s laboratory is supported by a Canadian Institutes of Health Research Foundation Grant (CIHR FDN 143301) and the Government of Ontario and Genome Canada and Ontario Genomics (OGI-139). The authors wish to thank Christopher D. Go for sharing the BioID cell lines. We wish to thank all members of the Kuster group for technical assistance and fruitful discussions.

**Author contributions**

B.K. conceived the study. O.B., Y.B., and R.Z. set up and optimized the micro-flow LC system. Y.B., R.Z., F.P.B., A.-C.G., and B.K. designed experiments. Y.B., R.Z., F.P.B., C.W., Y.-C.C., M.R., S.W., and D.P.Z. performed experiments. Y.B., C.M., C.W., F.P.B., M.R., J.Z., S.H., J.S., B.H., M.B., and A.-C.G. analyzed data. Y.B. and B.K. wrote the manuscript.

**Competing interests**

B.K. is a founder and shareholder of OmicScouts and msAId. D.P.Z. is a founder and shareholder of msAId. They have no operational role in either company. O.B. and M.B. are employees of Thermo Fisher Scientific. R.Z. is currently an employee of Thermo Fisher Scientific. The other authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13973-x.

**Correspondence**

and requests for materials should be addressed to B.K.

**Peer review information** *Nature Communications* thanks Oleg Krokhin, and the other anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

© The Author(s) 2020