Title

An automated workflow for label-free and multiplexed single cell proteomics sample preparation at unprecedented sensitivity.

Authors

David Hartlmayr[1,*], Claudia Ctortecka[1,*,§], Anjali Seth[4], Sasha Mendjan[2], Guilhem Tourniaire[4] and Karl Mechtler[1,2,3,§]

* these authors contributed equally to this work.

Affiliation

1 Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, 1030 Vienna, Austria.

2 Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria.

3 The Gregor Mendel Institute of Molecular Plant Biology of the Austrian Academy of Sciences (GMI), Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria.

4 Cellenion SASU, 60F avenue Rockefeller, 69008 Lyon, France.

§Correspondence

Claudia Ctortecka

Claudia.ctortecka@imp.ac.at

Karl Mechtler

Karl.mechtler@imp.ac.at

Research Institute of Molecular Pathology (IMP)

Vienna BioCenter (VBC)

Campus-Vienna-Biocenter 1

1030 Vienna, Austria

Abstract

The analysis of single cell proteomes has recently become a viable complement to transcript and genomics studies. Proteins are the main driver of cellular functionality and mRNA levels are often an unreliable proxy of such. Therefore, the global analysis of the proteome is essential to study cellular identities. Both multiplexed and label-free mass spectrometry-based approaches with single cell resolution have lately attributed surprising heterogeneity to believed homogenous cell populations. Even though specialized experimental designs and instrumentation have demonstrated remarkable advances, the efficient sample preparation of single cells still lacks behind. Here, we introduce the proteoCHIP, a universal option for single cell proteomics sample preparation at surprising sensitivity and throughput. The automated processing using a commercial system combining single cell isolation and picoliter dispensing, the cellenONE®, allows to reduce final sample volumes to low nanoliters submerged in a hexadecane layer simultaneously eliminating error prone manual sample handling and overcoming evaporation. With this specialized workflow we achieved around 1,000 protein groups per analytical run at remarkable reporter ion signal to noise while reducing or eliminating the carrier proteome. We identified close to 2,000 protein groups across 158 multiplexed single cells from two highly similar human cell types and clustered them based on their proteome. In-depth investigation of regulated proteins readily identified one of the main drivers for tumorigenicity in this cell type. Our workflow is compatible with all labeling reagents, can be easily adapted to custom workflows and is a viable option for label-free sample preparation. The specialized proteoCHIP design allows for the direct injection of label-free single cells via a standard autosampler resulting in the recovery of 30% more protein groups compared to samples transferred to PEG coated vials. We therefore are confident that our versatile, sensitive, and automated sample preparation workflow will be easily adoptable by non-specialized groups and will drive biological applications of single cell proteomics.

Introduction

Within recent years single cell analysis has demonstrated valuable insights into heterogeneous populations. However, proteins and especially their post translational modifications are the main driver of cellular identity and their function. The complementation of single cell transcriptomics and genomics approaches with a global protein analysis is therefore regarded essential. Most protein profiling techniques with single cell resolution, however, still rely on the availability of antibodies. Continuous technological advances drive sensitivity and accuracy of mass spectrometry (MS)-based single cell proteomics for hypothesis-free measurements. Despite that, three main aspects, throughput, measurement variability and most importantly sample preparation efficiency still lack behind comparable sequencing techniques.

The combination of dedicated instrumentation with sensitive acquisition strategies for label-free single cell analysis has been demonstrated highly accurate but very limited in throughput.[1,2] In label-free MS experiments single cells are processed and subjected to analysis individually, which is not only prone to peptide losses during the workflow and chromatographic separation but also requires about 50 minutes of measurement time per single cell (100 cells in 83 hours). This was addressed through isobaric labeling (i.e. tandem mass tags – TMT), allowing to uniquely barcode individual cells for simultaneous analysis and relative quantification upon MS-based analysis.[3] TMT reagents are available with several multiplexing capacities enabling the analysis of up to sixteen samples in one experiment.[4] This combined analysis not only reduces the measurement time per single cell to merely 5 minutes (i.e. 100 cells in 8 hours) but also increases the input material per sample. For multiplexed experiments, single cells are processed individually, TMT-labeled after tryptic digestion and combined into one sample for MS analysis. The identical mass of the TMT-labels allows for the simultaneous elution of a peptide from all samples, therefore multiplying the precursor signal and contributing ions for peptide identification. The differently equipped heavy isotopes of the TMT-reagents allow for relative quantification upon fragmentation in the MS. This has been adopted to perform first single cell proteomics experiments, through the combination of single cells with an abundant carrier spike comprised of up to 200 cells (SCoPE-MS).[5] The abundant carrier sample improves precursor selection, serves fragment ions for identification and overcomes peptide losses throughout the workflow.

Such detrimental sample losses during sample preparation were most successfully addressed using the nanoPOTS system (nanodroplet processing in one-pot for trace samples).[1] Their photolytically etched reaction vessels allow sample volume reduction to less than 200 nL for both label-free and multiplexed samples.[6,7] Minimizing sample volumes not only reduces peptide losses to vessel surfaces but improves enzymatic kinetics and reduces required chemicals at constant concentration.

Previously, most nanoPOTS approaches relied on home built robotic dispensing, which was overcome with the nested nanoPOTS (N2).[8] Here the cellenONE® a commercial liquid dispensing instrument is used for both cell isolation and single cell preparation on the N2 substrate to further reduce reaction volume down to ~20 nL. Additionally, the N2 overcomes the need for manual combination of TMT-multiplexed sets using sparsely distributed arrays of single cell reaction sites, which can be unified with a microliter droplet on top of each sample set. The authors demonstrated the reproducible analysis of TMT-multiplexed single cells in conjunction with an abundant 10 ng carrier sample transferring the combined sample to a vial or their custom built nanoPOTS autosampler.[8,9]

We here describe a highly versatile and automated workflow for both label-free and isobaric multiplexed single cell proteomics sample preparation at unprecedented sensitivity. We introduce the proteoCHIP encompassing sixteen nanowells for cross-contamination-free simultaneous processing of up to 192 single cells per chip. The proteoCHIP overcomes all manual sample handling steps including the combination of multiplexed sample sets and can be directly interfaced with standard autosamplers for MS-based analysis. The nanowells allow to reduce final sample volumes to ~200 nL for efficient processing while overcoming evaporation entirely. Further, the chip can be used with all currently available multiplexing reagents or for the efficient processing of label-free single cells. Finally, our specialized sample processing protocol allows to reduce the carrier spike to a minimum for accurate quantification while yielding comparable protein identifications to published techniques.[8,10]

Main

We here introduce the proteoCHIP as a viable option for automated single cell proteomics sample preparation within a platform combining single cell isolation and picoliter dispensing, the cellenONE®, for both label-free and multiplexed samples. The proteoChip is a complete system in the size of standard microscopy slides which is constituted of two parts. First, single cell isolation and sample preparation is performed in the nanowell part, which entails twelve fields to process twelve label-free single cells or twelve multiplexed sample sets with up to sixteen cells per set (total 192 cells, Fig. 1a). Second, the funnel part is designed for pooling of the samples (Fig. 1b) and directly interfacing with the HPLC autosampler for injection (Fig. 1c). The proteoCHIP has four main advantages over other published sample preparation workflows.[8–10,14] First, to overcome known peptide losses to plastic or glass ware,[15,16] the proteoCHIP is fabricated out of PTFE. We observe similar GRAVY indices of bulk samples from glass autosampler vials, single cell samples directly injected from the PTFE proteoCHIP, after proteoCHIP preparation and transfer to a PP vessel or preparation in PP plates (Supplemental Fig. 1). Despite that, we regularly observe beneficial peptide recovery from samples processed in PTFE. Second, the nanowells within each field hold up to 600 nL allowing to readily adapt the sample

preparation protocol without cross-contamination of the samples. Third, in contrast to successfully miniaturized sample preparation strategies[1] we overcome sample evaporation with a hexadecane layer. Hexadecanes freezing point is close to 18 ˚C, the oil covering the final sample in the 10 ˚C autosampler therefore freezes and does not interfere with the subsequent analysis. This results in constant reagent and enzyme concentrations in relation to the cells for reproducible processing efficiencies. Additionally, all fields are surrounded by elevated walls, physically separating each sample set during the workflow and storage in the autosampler. Fourth, the proteoCHIP funnel part allows convenient pooling of multiplexed single cell samples using a standard benchtop plate centrifuge (Fig. 1b). In contrast to the N2 workflow, where nested single cells are elegantly combined via the addition of a drop of sample buffer,[8] the funnel-like proteoCHIP lid, can be directly interfaced with a standard autosampler. This allows for direct injection of the samples without drying or transferring them to another vessel. Taken together, the reduction in processing volumes, manual sample handling and exposed surface areas combined with the direct interface to a standard autosampler provide single cell proteome measurements at remarkable sensitivity.

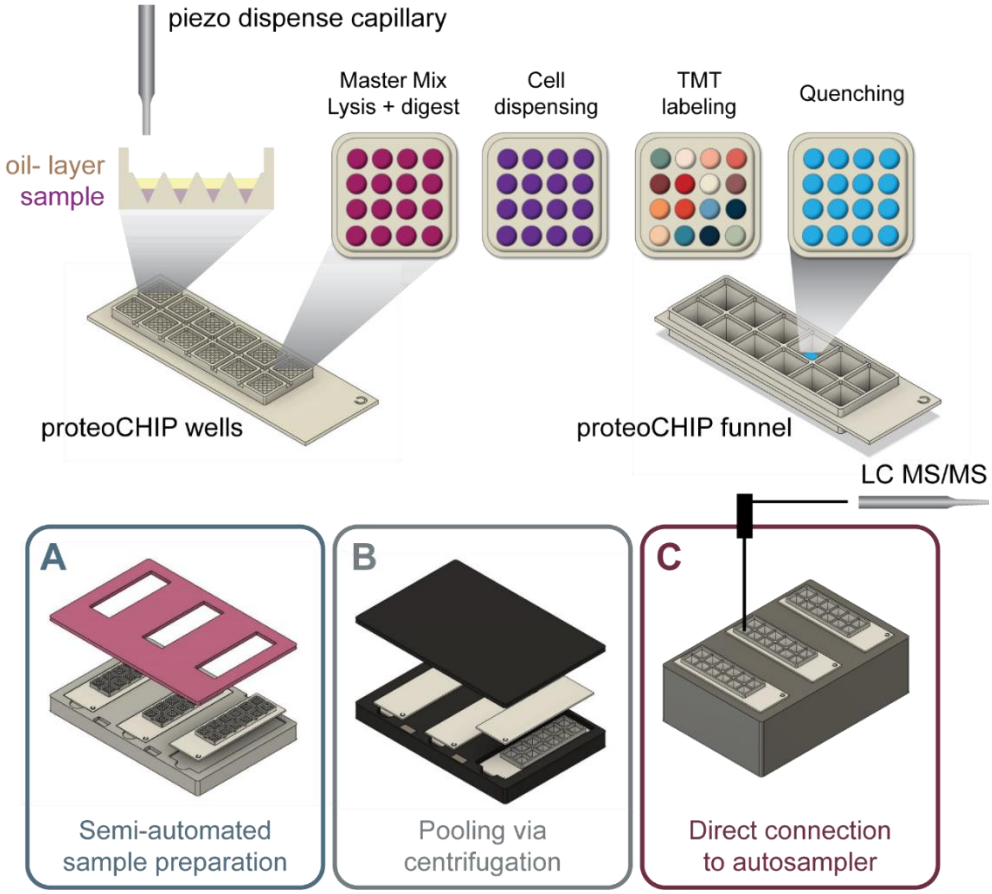

Figure 1: Illustration of the proteoCHIP based TMT-labeling workflow. (a) Up to sixteen nanowells/single cells per TMT set are prepared inside the cellenONE®, (b) are automatically combined via centrifugation and (c) directly interfaced with a standard autosampler for loss-less acquisition.

Single cell proteomics sample preparation workflow with the proteoCHIP

We perform the entire sample preparation workflow inside the cellenONE® starting with dispensing of a master mix for lysis and enzymatic digestion followed by image-based single-cell isolation, directly into the master mix. We use a combination of a MS compatible detergent to ensure efficient lysis with simultaneous tryptic digestion at a 10:1 enzyme:substrate ratio. Lysis and digestion incubation steps at 50 and 37 ˚C are performed at high humidity (i.e. 85%) while the sample is submerged under a hexadecane layer to overcome evaporation (detailed in the method section). Subsequent steps are performed at dew point to further reduce sample evaporation and residual enzymatic activity during the labeling. Afterwards, excess TMT is quenched with hydroxylamine and hydrochloric acid to avoid drastic changes in pH. Of note, this protocol allows for a final sample volume after lysis, digestion, TMT-labeling and quenching of sub-microliter without drying the sample to completeness. Subsequently, the chip is covered with the proteoCHIP funnel part, pooled in a centrifuge within only a minute, covered with adhesive aluminum foil, which can be easily pierced by the HPLC puncturer and finally injected for LC-MS/MS analysis (Fig. 1b-c).

Multiplexed single cell proteome measurements.

First, we evaluated the required abundance of the carrier spike for comparable protein identifications to state of the art techniques.[8,10] Our optimized workflow using the proteoCHIP with reduced sample volume, manipulation and surface area exposure allows to reduce the carrier to merely 20x or lower yielding around 1,000 protein groups per analytical run (Fig. 2a). We hypothesize that this ratio reduction of the carrier to single cells close to the reported ratio limit of TMT10-plex reagents will improve quantification accuracy.[17,18] In detail, we readily identify on average 1,175 and 897 protein groups based on 4,832 and 3,444 peptides per multiplexed TMT10-plex set using a 20x carrier or no carrier, respectively (Fig. 2a). All TMT10-plex single cell runs combined (i.e. 306 single cells) yield 1,789 protein groups based on 11,467 peptides. Similarly, the 20x and no carrier TMTpro samples (i.e. 276 single cells) result in an average of 1,017 and 924 protein groups from 3,873 and 3,833 peptides per analytical run, respectively (Fig. 2a). Across all TMTpro single cell sets we identify over 1,974 protein groups from 11,013 peptides. This indicates that we find multiple protein groups uniquely in some analytical runs and not across all replicates for both TMT reagents. Nevertheless, our specialized workflow resulted in highly comparable identifications for both TMT reagents and with the reduced or omitted carrier.

Recently, Cheung and co-workers proposed a signal to noise (S/N) filtering for more accurate quantification of multiplexed single cell proteomics experiments.[18] We therefore extracted the S/N value of all single cell channels using our in-house software Hyperplex (details in method section) and evaluated the S/N distribution for our experimental setup. The average single cell S/N in all conditions from cells prepared with the proteoCHIP on our instrument setup is comparable or outperforms previous reports. In detail, across multiple replicates we observe median single cell reporter ion S/N values of 40 and 100 for TMT10-plex samples or 133 and 255 for TMTpro samples, with and without the 20x carrier, respectively (Fig. 2b). Despite being acquired on different instruments, our setup and the carrier reduction vastly improves S/N reporter values compared to 7-15 S/N of the nanoPOTS or N2.[6,8,19] Interestingly, TMTpro experiments with and without the carrier resulted in higher S/N of the single cell channels compared to the TMT10-plex (Fig. 2b). We have regularly observed this phenomenon in trace samples and speculate, that this is due to the reduced NCE needed to fragment the TMTpro over the TMT10-plex reagent. The TMTpro NCE of 32 efficiently fragments the tag and is close to the energy required for fragmentation of the peptide backbone. This contrasts with the slightly higher NCE of 34 required to suitably fragment the TMT10-plex tag, possibly increasing the noise level in each MS/MS scan. Furthermore, we observed a reduction by 50% in single cell S/N in the 20x carrier compared to the no-carrier samples, for both TMT10-plex and TMTpro experiments (Fig. 2b). Despite the low ratio of the carrier to the single cells, we speculate that this is due to the increased proportion of ions sampled from the carrier[18] or compression of the single cell reporter ion signals into the noise. Additionally, a pairwise correlation of two single cell reporter ion channels demonstrates increased variance between the 20x carrier compared to the no carrier samples for both TMT10-plex and TMTpro (Fig. 2c-f). Despite the obvious beneficial aspects of a carrier spike[20], based on our findings, we agree with literature to reduce the carrier to a minimum or if possible, remove it entirely from the TMT set.[6,18]

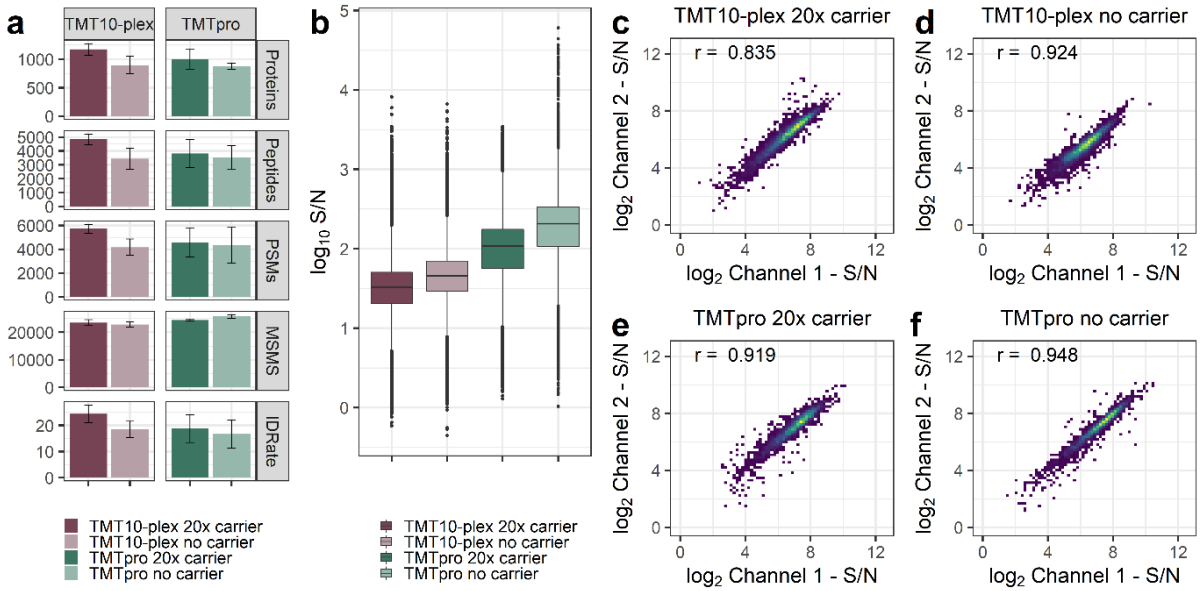

Figure 2: Application of the proteoCHIP for single cell proteomics sample preparation with TMT10-plex and TMTpro reagents at different carrier compositions. (a) Number of identified proteins, peptides, PSMs, all MS/MS scans and the ID rate for TMT10-plex (red) and TMTpro (green). Error bars represent median absolute deviation. (b) Log_{10} S/N of all single cell reporter ions at indicated condition over five replicates. Log_2 S/N correlation between two single cell samples for (c) TMT10-plex 20x carrier, (d) TMT10-plex no carrier, (e) TMTpro 20x carrier and (f) TMTpro no carrier. r = Pearson correlation estimate.

Next, based on the presumed low identification overlap between analytical runs (i.e. biological replicates), we evaluated the unique peptide sequence intersections and percentage of missing data within our single cell runs. Interestingly, we observed less overlap in unique peptide sequences for the TMTpro compared to TMT10-plex samples for 20x and no carrier setups, ranging from 50 to 85% (Fig. 3a-d). We hypothesize that both, the stochastic precursor selection of the employed data dependent acquisition (DDA) strategy and the direct injection of the sample after TMT labeling without a cleanup compromise reproducibility. Furthermore, we speculate that nearly double the TMT reagent in the final TMTpro compared to the TMT10-plex sample, contributes background signal, interferes with precursor selection and MS/MS identification, additionally decreasing peptide sequence overlap (Fig. 3a-d). Accordingly, we evaluated both the missingness of reporter ion signal per PSM and the cumulative missing data across multiple analytical runs. The high reporter ion S/N already suggested that the signal of our single cells is well above the noise therefore resulting in almost no missing values per PSM for all experimental setups (Fig. 3e-h). Even with the low missingness per analytical run, the high variance between analytical runs leads to cumulative missing quantitative data (Fig. 3i-l). The data aggregation of five analytical runs (i.e. ~50-80 single cells) reduces the number of confidently quantified proteins by 50% without imputation (Fig. 3i-l), as reported by others.[8,10] This demonstrates that despite the high quality quantitative data per run, the untargeted DDA results in accumulation of

missing data in large sample cohorts.[21,22] As a result the acquired dataset is either drastically reduced or a large proportion of quantitative data is computationally generated.

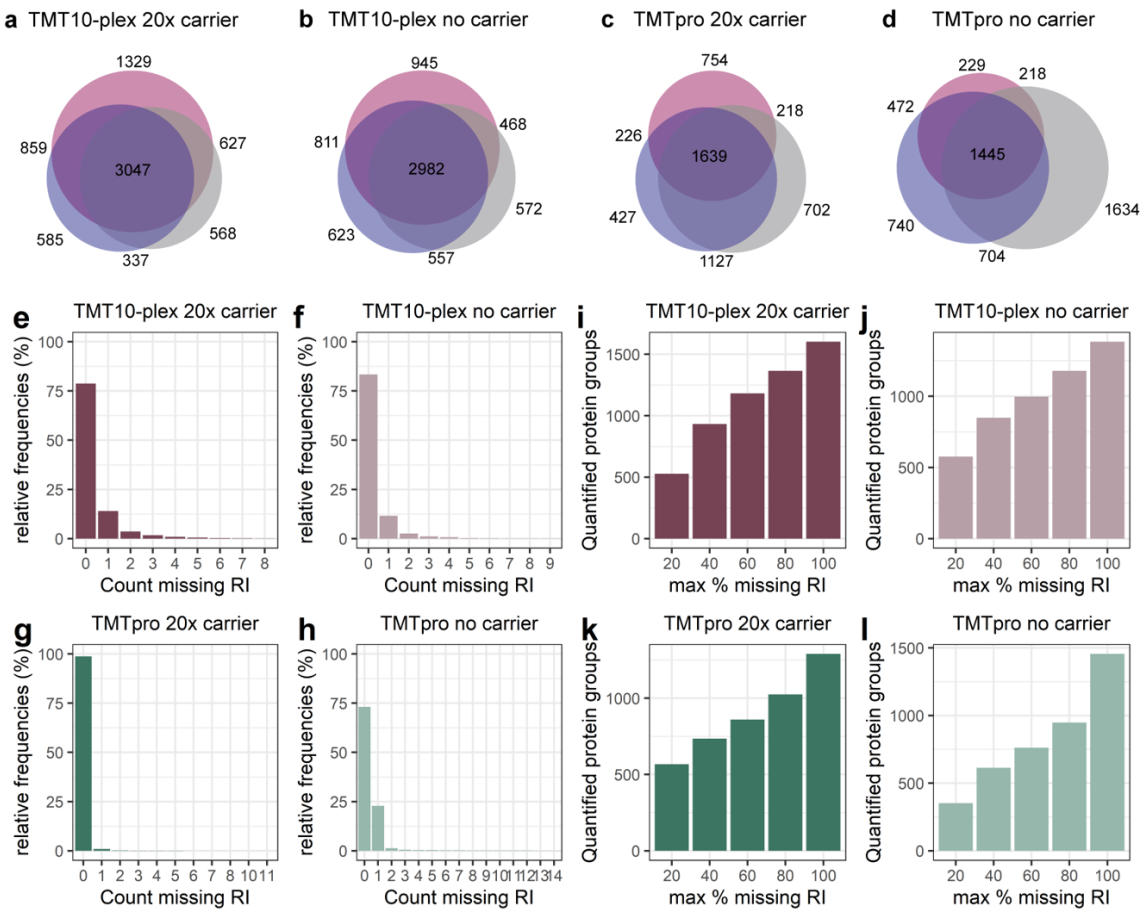

Figure 3: Data completeness and reproducibility evaluation of multiplexed single cell proteomes. Unique peptide sequence overlaps for (a) TMT10-plex 20x carrier, (b) TMT10-plex no carrier, (c) TMTpro 20x carrier and (d) TMTpro no carrier samples. Percentage of relative missing reporter ions (RI) across five analytical runs per PSM for (e) TMT10-plex 20x carrier, (f) TMT10-plex no carrier, (g) TMTpro 20x carrier and (h) TMTpro no carrier samples. Cumulative missing reporter ions (RI) per quantitfied protein across five analytical runs for (i) TMT10-plex 20x carrier, (j) TMT10-plex no carrier, (k) TMTpro 20x carrier and (l) TMTpro no carrier samples.

Differentiating two similarly sized human cell types based on their single cell proteome.

Following the surprising data quality stemming from our optimized sample preparation workflow, we tested if two similarly sized human cell types can be differentiated based on their proteome (Supplementary Fig. 2). We generated 158 HeLa and HEK single cell samples using our proteoCHIP workflow and distributed them equally across eleven TMT10-plex sets. Both, the 20x carrier and no carrier samples yield on average around 1,300 protein groups based on 5,000 peptides per analytical run (Fig. 4a). All 110 TMT10-plex labeled single cells combined yield 1,894 protein groups based on 10,665 peptides. Further, we confirmed that the similarly sized cells contain comparable protein

amounts, resulting in equally distributed reporter ion intensities in all TMT10-plex channels across multiple analytical runs (Fig. 4b). This not only indicates highly reproducible sample preparation but also strengthens our confidence that the differences we observe between the cell types originate from changes in the proteome and not different sample input. We therefore performed a principal component analysis of 110 no carrier single cells and observed a cell type specific separation via the first two components (Fig. 4c). Of note, even though we filtered for at least 70% quantitative data completeness, the cell type cluster density decreases the more analytical runs were accumulated. We speculate that this is in part due to the reduced sample overlap introduced by stochastic precursor picking and elevated background signals as described earlier (Fig. 3a-d). Consequently, we strongly believe that the optimization of in-line, loss-less sample clean-up in conjunction with an efficient data independent acquisition (DIA) strategy will further improve our results.

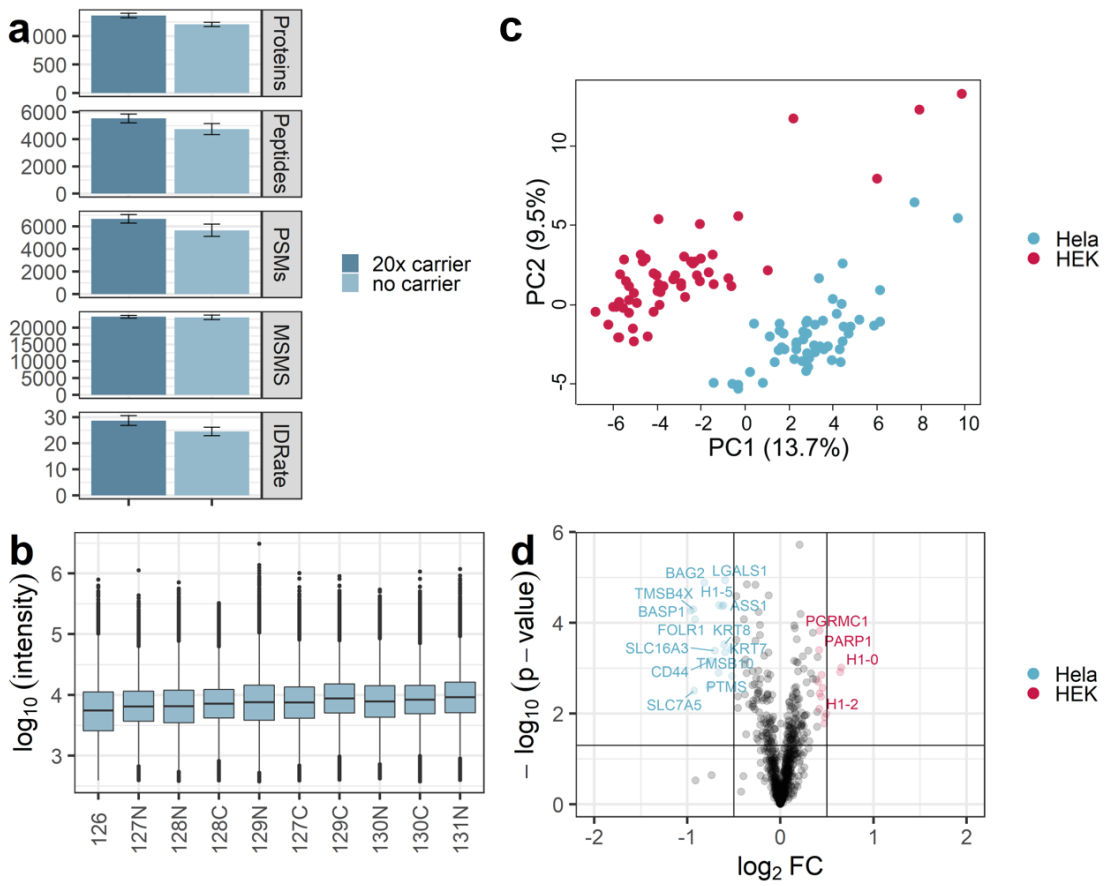

Figure 4: Comparison of HeLa and HEK single cell proteomes. (a) Protein groups, peptides, PSMs, MS/MS scans and ID-rate of TMT10-plex HeLa/HEK samples. (b) Intensity distribution of all reporter ions for both HeLa and HEK single cells across several analytical runs in $\log_{10}$ (n=11). (c) PCA clustering of single HeLa (blue) and HEK (red) TMT10-plex labeled, no carrier 110 single cells across 976 protein groups. (d) Vulcano plot of differential expressed proteins between HeLa and HEK single cells from TMT10-plex no carrier samples. Log_2 fold change and $-\log_{10}$ p-value is shown. Colors indicate protein regulation and top up- or down regulated proteins are labeled with their gene names.

Aiming at examining the cluster loadings in more detail, we investigated top differentially expressed proteins between the two cell types (Fig. 4c-d). Interestingly, one of the top hits in HeLa cells compared to HEK cells is the brain acid soluble protein 1 (BASP1), which is downregulated in most tumor cell lines except some cervical cancer lines (Fig. 4d). In contrast to other cancer cell lines, the elevated levels of the tumor suppressor BASP1 in HeLa cells even promotes tumor growth.[24] Alongside BASP1, cross referencing of our top regulated proteins to normalized expression data obtained from the Human Protein Atlas[23] (http://www.proteinatlas.org) revealed strong agreement (i.e. CD44, FOLR1, KRT7, KRT8, LGALS1, PARP1, PGRMC1, SLC7A5, TMSB4X, TMSB10). Following this, we are confident to accurately represent quantitative differences between the two cell types and that our analysis distinguishes the two solely based on their proteome. Of note, using our experimental setup, we can directly correlate changes in the proteome to the acquired image during cell sorting by the cellenONE®. This allows to estimate if an expected or unexpected clustering behavior is a result of the respective proteome or can be traced back to the preparation and cellular morphology.

Label-free single cell proteome acquisition with the proteoCHIP

We next asked how our multiplexed sample preparation workflow compares in the generation of label-free single cells. Label-free proteome analysis has several advantages over multiplexed sample workflows, like the direct MS1 based quantification, the possibility of highly confident feature matching between analytical runs and the reduced chemical noise introduced by the labeling.[25,26] We therefore evaluated the proteoCHIP protocol in the analysis of label-free single cell samples, using shorter gradients based on the vastly reduced sample input (i.e. 30 minutes compared to 60 minutes for TMT-labeled samples). This still drastically reduces the throughput of the acquired samples, however, the gradient length and overhead times between the samples is still subject to further improvement. First, we processed increasing numbers of HeLa cells starting from only one up to 6 cells, either transferring the sample to a standard PP vial for injection or measuring directly via the proteoCHIP funnel part (Fig. 1c). As expected, the sample transfer results in slight peptide losses more prominent in the lower cell input samples compared to five cells and more (Fig. 5a). Even though the samples are processed identically and transferred to a PEG pre-treated PP vial, the vessel exchange results on average in 30% decreased protein identifications for single cells. We speculate that these differences are especially striking at such low input, as this readily declines to only 10% for two cells and merely 5% in the analysis of three cells (Fig. 5a). This indicates that the direct connection and reduced sample manipulation enabled by the proteoCHIP is critical in the analysis of single cell proteomes.

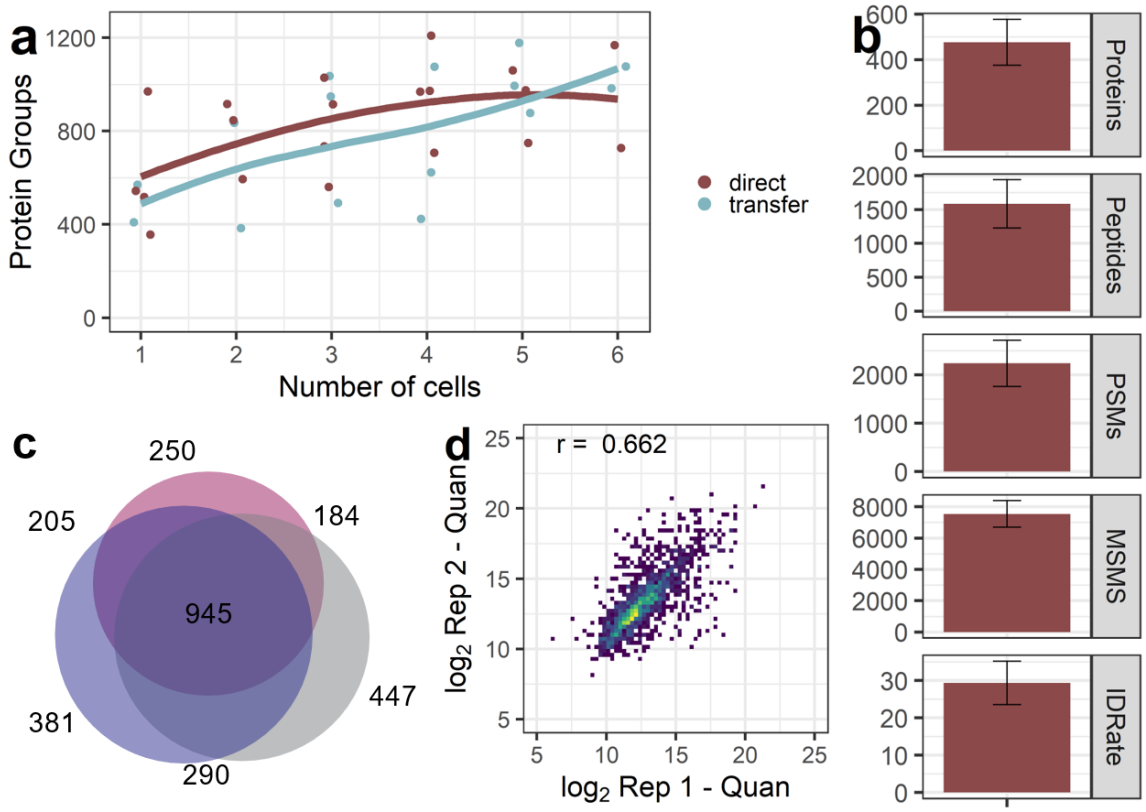

Figure 5: Label-free proteome analysis of single HeLa cells. (a) Protein groups of indicated cell numbers via direct injection from the funnel (red) or transferred to a standard PCR vial (blue). (b) Protein groups, peptides, PSMs, all MS/MS scans and the ID-rate of label-free single cells (n=32). Error bars represent median absolute deviation. (c) Unique peptide sequence overlap of three label-free single HeLa measurements. (d) Label-free protein quantification correlation of two analytical runs in log_2. r = Pearson correlation estimation.

Our optimized label-free proteoCHIP workflow reproducibly yields around 500 protein groups per single HeLa cell and 1,422 protein groups across all 30 single cell measurements. Interestingly, similarly to the TMT-labeled samples (Fig. 3a-d), the unique peptide sequence overlap between three replicates is only around 50% (Fig. 5c). Despite that, peptides that were identified across replicates positively correlated with a pearson correlation estimate of 0.662 (i.e. peptide area, Fig. 5d). We speculate that this drastic reduction in replicate overlap again is mostly caused by the stochastic selection of precursors in DDA strategies. Even though label-free measurements now allow for FDR-controlled match between runs based on MS1 features[25], we are confident that the transition to DIA measurements will improve replicate overlap and quantification correlation. Further, we speculated that the fast duty cycles and the increased usage of the ion beam of PASEF on the timsTOF Pro will benefit our label-free single cell analysis.[27] Taken together, we have successfully extended the sample processing capabilities of the proteoCHIP to label-free single cell samples at surprising sensitivity. However, we hypothesize that the data reproducibility can be further advanced via specialized acquisition strategies.

Discussion

We here demonstrate the automation of single cell sample preparation using the proteoCHIP in conjunction with a commercial single cell isolation and picoliter dispenser, the cellenONE®. Our proposed sample preparation workflow of single cells for MS-based analysis is highly adaptable and allows for the preparation of label-free or multiplexed single cells. The optimized protocol drastically reduces the digest volumes compared to previously published well-based techniques and is comparable to those successfully applied in nanoPOTS.[1,10,14] This not only limits chemical noise but as a result of the hexadecane layer covering the sample we achieve constant enzyme and chemical concentrations increasing efficiency of the sample preparation. Further, the specialized design of the proteoCHIP allows automatic pooling of multiplexed samples using a standard benchtop centrifuge, final sample collection in the proteoCHIP funnel part and direct interfacing with a standard autosampler for LC-MS/MS analysis. This semi-automated processing, pooling, and injecting eliminates error prone manual sample handling often resulting in peptide losses and additional variance.

Our efficient single cell sample preparation retains comparable protein identifications and enhanced S/N of single cell reporter ions even at reduced or eliminated carrier (Fig. 2a-b).[8,10] This not only allows increased throughput by labelling single cells with all available TMT reagents but also improves the confidence of identifying peptides stemming from the single cells and not the carrier.[18] We further show that two highly similar human cell types can be differentiated based on their proteome using our platform (Fig. 3c-d). We are therefore confident that biologically similar cell types (e.g. originating from the same organ) can be classified and profiled using our workflow. We, however, acknowledge that despite the good correlation of individual samples (Fig. 2c-f) the correlation and replicate overlap between analytical runs is still subject to improvements (Fig. 3a-d). Despite the suboptimal replicate correlation, the over 75% data completeness within one analytical run (Fig. 3e-h) leaves us confident that DIA workflows will further advance present results. We hypothesize that specialized DIA methods for the Orbitrap Exploris or diaPASEF on the timsTOF Pro will drive reproducibility at similar quantification accuracy. Further, we are confident that our optimized sample processing strategy in conjunction with the more sensitive, second generation timsTOF Pro will further increase identifications especially of label-free single cell samples.[2]

In conclusion, our miniaturized single cell proteomics sample preparation workflow with the novel proteoCHIP utilizes standard chemicals for MS-based sample preparation. Employing a versatile picoliter dispensing robot, the cellenONE®, we have achieved efficient single cell proteomics sample

preparation which can be readily adapted, addressing multiple shortcomings of previously published label-free and multiplexed methods.

Material and Methods

Sample preparation

HeLa and HEK293T cells were cultured at 37 ˚C and 5% CO_2 in Dulbecco's Modified Eagle's Medium supplemented with 10% FBS and 1x penicillin-streptomycin (P0781-100ML, Sigma-Aldrich, Israel) and L-Glut (25030-024, Thermo Scientific, Germany). After trypsinization (0.05% Trypsin-EDTA 1x, 25300-054, Sigma-Aldrich, USA/Germany), cells were pelleted, washed 3x with phosphate-buffered saline (PBS) and directly used for single cell experiments.

40-200 nL lysis buffer (0.2% DDM (D4641-500MG, Sigma-Aldrich, USA/Germany), 100 mM TEAB (17902-500ML, Fluka Analytical, Switzerland), 20 ng/uL trypsin (Promega Gold, V5280, Promega, USA) was dispensed into each well using the cellenONE® (cellenion, France) at high humidity. After single cell deposition (gated for cell diameter min 22 µm and diameter max 33 µm, circularity 1.1, elongation 1.84) a layer of Hexadecane (H6703-100ML, Sigma-Aldrich, USA/Germany) was added to the chips. The chip was then incubated at 50 ˚C for 30 minutes followed by 4 hrs at 37 ˚C, directly on the heating deck inside the cellenONE®. For TMT multiplexed experiments 100-200 nL of 22 mM TMT10-plex or TMTpro in anhydrous ACN was added to the respective wells and incubated for 1 hour at room-temperature. TMT was subsequently quenched with 50 nL 0.5 % hydroxylamine (90115, Thermo Scientific, Germany) and 3 % HCl followed by sample pooling via centrifugation using the proteoCHIP funnel part. After tryptic digest label-free samples were quenched with 0.1% TFA and both label-free or multiplexed samples were either transferred to 0.2 mL PCR-tubes coated with 1e-3 % Poly(ethylene glycol) (95172-250G-F, Sigma-Aldrich, Germany), directly injected from the proteoCHIP funnel part or kept at -20 ˚C until usage.

LC-MS/MS analysis

Samples were measured on a Orbitrap Exploris™ 480 Mass Spectrometer (Thermo Fisher Scientific) with a reversed phase Thermo Fisher Scientific Ultimate 3000 RLSC-nano high-performance liquid chromatography (HPLC) system coupled via a Nanospray Flex ion source equipped with FAIMS (operated at -50 CV). Labeled peptides were first trapped on an Acclaim™ PepMap™ 100 C18 precolumn (5 µM, 0.3 mm X 5 mm, Thermo Fisher Scientific) and eluted to the analytical column nanoEase M/Z Peptide BEH C18 Column (130Å, 1.7 µm, 75 µm X 150 mm, Waters, Germany) developing a two-step solvent gradient ranging from 2 to 20 % over 45 min and 20 to 32 % ACN in 0.08 % formic acid within 15 min, at a flow rate of 250 nL/min. Label-free samples were measured on the

same setup as described above but separated using a twostep gradient from 2 to 20 % over 15 min, 20 to 32 % ACN in 0.08 % formic acid within 5 minutes, at 250 nL/min.

Full MS data of multiplexed experiments were acquired in a range of 375-1,200 m/z with a maximum AGC target of 3e6 and automatic inject time at 120,000 resolution. Top 10 multiply charged precursors (2-5) over a minimum intensity of 5e3 were isolated using a 2 Th isolation window. MS/MS scans were acquired at a resolution of 60,000 at a fixed first mass of 110 m/z with a maximum AGC target of 1e5 or injection time of 118 ms. Previously isolated precursors were subsequently excluded from fragmentation with a dynamic exclusion of 120 seconds. TMT10-plex precursors were fragmented at a normalized collision energy (NCE) of 34 and TMTpro at a NCE of 32.

Data analysis

Peptide identification was performed using the standard parameters in Spectromine™ 2.0 against the human reference proteome sequence database (UniProt; version: 2020-10-12). N-terminal protein acetylation and oxidation at methionine were set as variable modifications and the respective TMT reagents were selected as fixed modification. Peptide spectrum match (PSM), peptide and protein groups were filtered with a false discovery rate (FDR) of 1%. S/N levels of reporter ions were extracted using the in-house developed Hyperplex (freely available: pd-nodes.org) at 10 ppm and intersected with the Spectromine™ results. Post-processing was performed in the R environment if not indicated otherwise. For quantification PSMs were summed to peptides and protein groups. Single cell reporter ion intensities are normalized to their sample loading within each analytical run. For HeLa versus HEK clustering, the raw reporter ion intensities were $\log_2$ transformed, protein groups with less than 70% missing data across the entire dataset were imputed with random values from a normal distribution shifted into the noise. The reporter ion intensities were then quantile normalized, batch corrected using ComBat for the analytical run and the TMT channel using the Perseus interface.[11] Venn Diagrams are based on unique peptide sequences and are calculated using BioVenn.[12] GRAVY scores were calculated for every unique peptide sequence identified from the respective condition, based on the Amino Acid Hydropathy Scores.[13]

References

1. Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
2. Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv* 2020.12.22.423933 (2020) doi:10.1101/2020.12.22.423933.
3. McAlister, G. C. *et al.* Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Anal. Chem.* **84**, 7469–7478 (2012).

4. Thompson, A. *et al.* TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Anal. Chem.* **91**, 15941–15950 (2019).
5. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, (2018).
6. Dou, M. *et al.* High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Anal. Chem.* **91**, 13119–13127 (2019).
7. Cong, Y. *et al.* Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chem. Sci.* (2021) doi:10.1039/D0SC03636F.
8. Woo, J. *et al.* High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *bioRxiv* 2021.02.17.431689 (2021) doi:10.1101/2021.02.17.431689.
9. Williams, S. M. *et al.* Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatography-Mass Spectrometry for High-Throughput Single-Cell Proteomics. *Anal. Chem.* (2020) doi:10.1021/acs.analchem.0c01551.
10. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
11. Tyanova, S. & Cox, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. in *Cancer Systems Biology* 133–148 (Humana Press, New York, NY, 2018). doi:10.1007/978-1-4939-7493-1_7.
12. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).
13. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
14. Schoof, E. M. *et al.* A Quantitative Single-Cell Proteomics Approach to Characterize an Acute Myeloid Leukemia Hierarchy. *bioRxiv* 745679 (2019) doi:10.1101/745679.
15. Goebel-Stengel, M., Stengel, A., Taché, Y. & Reeve, J. R. The importance of using the optimal plastic and glassware in studies involving peptides. *Anal. Biochem.* **414**, 38–46 (2011).
16. Stejskal, K., Potěšil, D. & Zdráhal, Z. Suppression of Peptide Sample Losses in Autosampler Vials. *J. Proteome Res.* **12**, 3057–3062 (2013).
17. Kelstrup, C. D. *et al.* Limits for Resolving Isobaric Tandem Mass Tag Reporter Ions Using Phase-Constrained Spectrum Deconvolution. *J. Proteome Res.* (2018) doi:10.1021/acs.jproteome.8b00381.
18. Cheung, T. K. *et al.* Defining the carrier proteome limit for single-cell proteomics. *Nat. Methods* 1–8 (2020) doi:10.1038/s41592-020-01002-5.
19. Tsai, C.-F. *et al.* An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy for Precise Quantitative Single-cell Proteomics. *Mol. Cell. Proteomics* **19**, 828–838 (2020).
20. Specht, H. & Slavov, N. Transformative Opportunities for Single-Cell Proteomics. *J. Proteome Res.* (2018) doi:10.1021/acs.jproteome.8b00257.
21. Brenes, A., Hukelmann, J., Bensaddek, D. & Lamond, A. I. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Mol. Cell. Proteomics* **18**, 1967–1980 (2019).
22. Ctortecka, C. *et al.* Comparative proteome signatures of trace samples by multiplexed Data-Independent Acquisition. *bioRxiv* 2021.02.11.430601 (2021) doi:10.1101/2021.02.11.430601.
23. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, (2015).
24. Hartl, M. & Schneider, R. A Unique Family of Neuronal Signaling Proteins Implicated in Oncogenesis and Tumor Suppression. *Front. Oncol.* **9**, (2019).

25. Yu, F., Haynes, S. E. & Nesvizhskii, A. I. Label-free quantification with FDR-controlled match-between-runs. *bioRxiv* 2020.11.02.365437 (2020) doi:10.1101/2020.11.02.365437.
26. Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics MCP* **13**, 2513–2526 (2014).
27. Meier, F. *et al.* Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **14**, 5378–5387 (2015).

Data availability

All mass spectrometry-based proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD025387].

Acknowledgements

We thank all members of our laboratories for helpful discussions. We specifically thank Elisabeth Roitinger for critical input on the manuscript. This work has been supported by EPIC-XS, project number 823839, funded by the horizon 2020 program of the European Union and the Austrian Science Fund by ERA-CAPS I 3686-B25-MEIOREC international project.

Author contributions

D.H. and C.C. prepared and acquired samples. A.S. conceptualized and designed the proteoCHIP. C.C. performed data analysis and wrote the manuscript. G.T., S.M. and K.M. supervised the research.

Conflict of interest

A.S. and G.T. are employees of Cellenion.