bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation

Andreas-David Brunner[1], Marvin Thielert[1], Catherine Vasilopoulou[1], Constantin Ammar[1], Fabian Coscia[1, 2], Andreas Mund[2], Ole B. Horning[3], Nicolai Bache[3], Amalia Apalategui[4], Markus Lubeck[4], Oliver Raether[4], Melvin A. Park[5], Sabrina Richter[6, 7], David S. Fischer[6, 7], Fabian J. Theis[6, 7], Florian Meier[1, 8], Matthias Mann[1, 2, *]

[1] Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

[2] NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark

[3] EvoSep Biosystems, Thriges Pl. 6, 5000 Odense, Denmark

[4] Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany

[5] Bruker Daltonics Inc., 40 Manning Road, Billerica, MA 01821, United States of America

[6] Helmholtz Zentrum München – German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany

[7] TUM School of Life Sciences Weihenstephan, Technical University of Munich, 85353 Freising, Germany

[8] Functional Proteomics, Jena University Hospital, Am Klinikum 1, 07747 Jena, Germany

* Corresponding author. Tel: +49 89 8578 2557; E-mail: mmann@biochem.mpg.de

Keywords

Quantitative accuracy, proteomics at single-cell resolution; low flow LC MS; single-cell proteomics, cell cycle, drug perturbation, systems biology, single-cell heterogeneity

Abstract

Single-cell technologies are revolutionizing biology but are today mainly limited to imaging and deep sequencing. However, proteins are the main drivers of cellular function and in-depth characterization of individual cells by mass spectrometry (MS)-based proteomics would thus be highly valuable and complementary. Chemical labeling-based single-cell approaches introduce hundreds of cells into the MS, but direct analysis of single cells has not yet reached the necessary sensitivity, robustness and quantitative accuracy to answer biological questions. Here, we develop a robust workflow combining miniaturized sample preparation, very-low flow-rate chromatography and a novel trapped ion mobility mass spectrometer, resulting in a more than ten-fold improved sensitivity. We accurately and robustly quantify proteomes and their changes in single, FACS-isolated cells. Arresting cells at defined stages of the cell cycle by drug treatment retrieves expected key regulators such as CDK2NA, E2 ubiquitin ligases such as UBE2S and highlights potential novel ones. Comparing the variability in more than 420 single-cell proteomes to transcriptome data revealed a stable core proteome despite perturbation. Our technology can readily be applied to ultra-high sensitivity analysis of tissue material, including post-translational modifications and to small molecule studies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

In single-cell analysis, biological variability can directly be attributed to individual cells instead of being averaged over a population or a tissue [1–3]. While microscopy has always been single-cell based, specialized deep sequencing technologies have achieved this for systems biological approaches [4–7]. At the level of proteins, the functional actors of cells, single cells are currently studied by antibody-based technologies, which are by necessity directed against previously chosen targets [8–10]. In contrast, mass spectrometry (MS)-based proteomics is unbiased in the sense that it measures all proteins within its range of detection [11,12]. Thus it would be highly desirable to apply this technology to single cells if the required sensitivity and robustness could be achieved. Previous approaches that employed chemical multiplexing of peptides have labeled a small number of single cells but combined them with a dominant booster channel for MS-analysis [13,14], which can hamper signal deconvolution [15,16]. Alternatively, proof of principle has been demonstrated for unlabeled approaches using sophisticated sample preparation methods in pico-liter devices [17,18]. However, a single-cell proteomics technology that answers biological questions is still outstanding.

Sensitivity evaluation of our current MS setup

We recently introduced parallel accumulation – serial fragmentation (PASEF), a mass spectrometric acquisition scheme in which peptide ions are released from a trapped ion mobility (TIMS) device in the vacuum system in concentrated packages [19,20]. Chemical noise is widely distributed as a result of its heterogeneous nature and the ten-fold increased peak capacity due to TIMS (**Fig. 1a, b**). These precursors can be fragmented in a highly sensitive manner, either in data dependent (ddaPASEF) or data independent (diaPASEF) mode, resulting in very high ion utilization and data completeness [21]. To explore sensitivity limits, we measured a dilution series of HeLa cell lysate from 25 ng down to the equivalent of a few single cells on a quadrupole time-of-flight instrument (TIMS-qTOF). This identified more than 550 proteins from 0.8 ng HeLa lysate with the dda acquisition mode and a conservative MaxQuant analysis (**Methods, Fig. 1B**) [22–24]. Proteins were quantified with the linear signal response expected from the dilution factors (**Fig. 1D**). Furthermore, quantitative reproducibility in replicates at the lowest level was still excellent (R = 0.96, **Fig. 1F**). Given that the protein amount of a single HeLa cell is as low as 150 pg [25], and accounting for inevitable losses in sample preparation including protein digestion, we estimated that we would need to increase sensitivity by at least an order of magnitude to enable true single cell proteomics.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

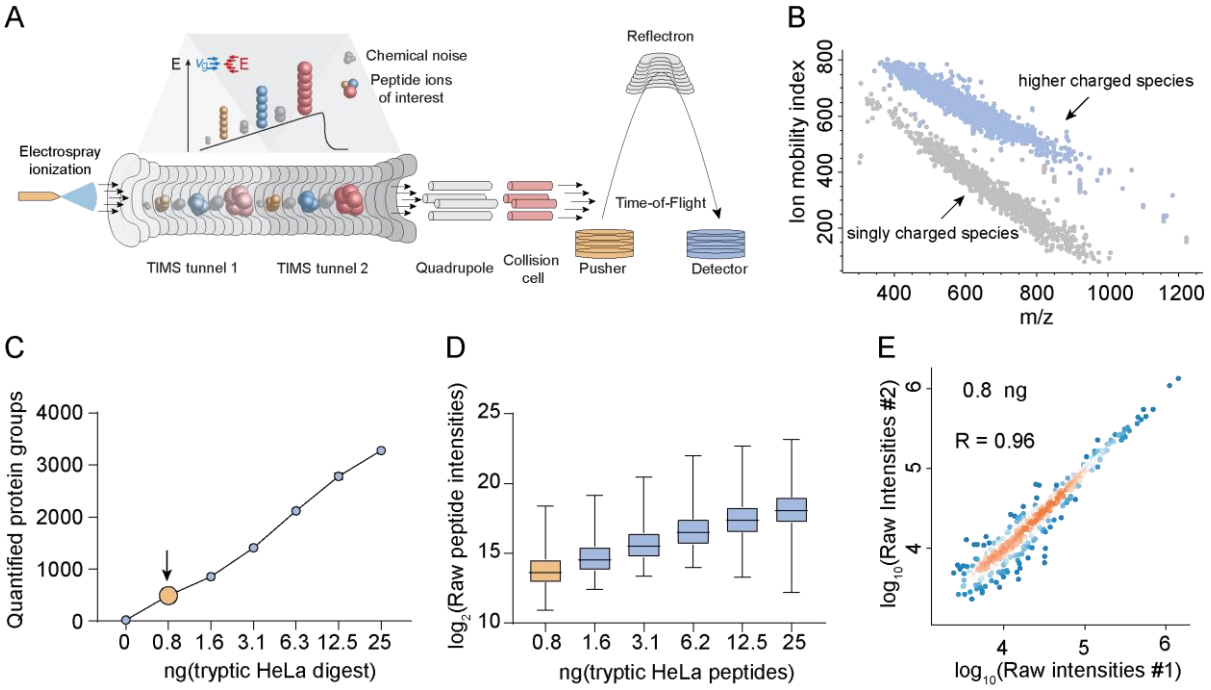


Figure 1: A, B: The TIMS-qTOF principle separating singly charged background peaks from multiply charged peptide precursor ions, making precursor ions visible at extremely low signal levels (0.8 ng HeLa digest). **C**: Quantified proteins from a HeLa digest dilution series from 25 ng peptide material down to 0.8 ng, roughly corresponding to the protein amount contained in three HeLa cells [25]. **D:** Linear quantitative response curve of the HeLa digest experiment in B. **E:** Quantitative reproducibility of two successive HeLa digest experiments at the lowest dilution (technical LC-MS/MS replicates).

A mass spectrometer with much increased sensitivity

Three main factors govern MS sensitivity: ionization efficiency, transfer efficiency into the vacuum system and ion utilization by the instrument [26]. We first constructed an instrument with a brighter ion source, introduced different ion optic elements and optimized parameters such as detector voltage. Together, this led to a 4.7-fold higher ion current (**Fig. 2A**). Next, we FACS sorted 0, 1 and up to 6 HeLa cells in triplicate into individual 384-wells, processed them separately and analyzed them on this modified mass spectrometer. This resulted in 824, 1364 and 1946 identified proteins for one, two and six cells, respectively. Note that this analysis benefited from transferring peptide identifications on the MS1 level, as expected from extremely low sample amounts [27] (**Fig. 2B**). Quantification accuracy was high when comparing single cells, not much reduced from comparing six cells (**Fig. 2 C, D**). A rank order abundance plot revealed that the measured single-cell proteome preferentially mapped to the higher abundant part of the six-cell proteome, indicating that proteome coverage depended deterministically on overall sensitivity (**Fig. 2E**). Inspecting the lowest abundant peptide that was shared between all samples showed that clearly interpretable fragment ion series were still present at high signal to noise levels (**Fig. 2F**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

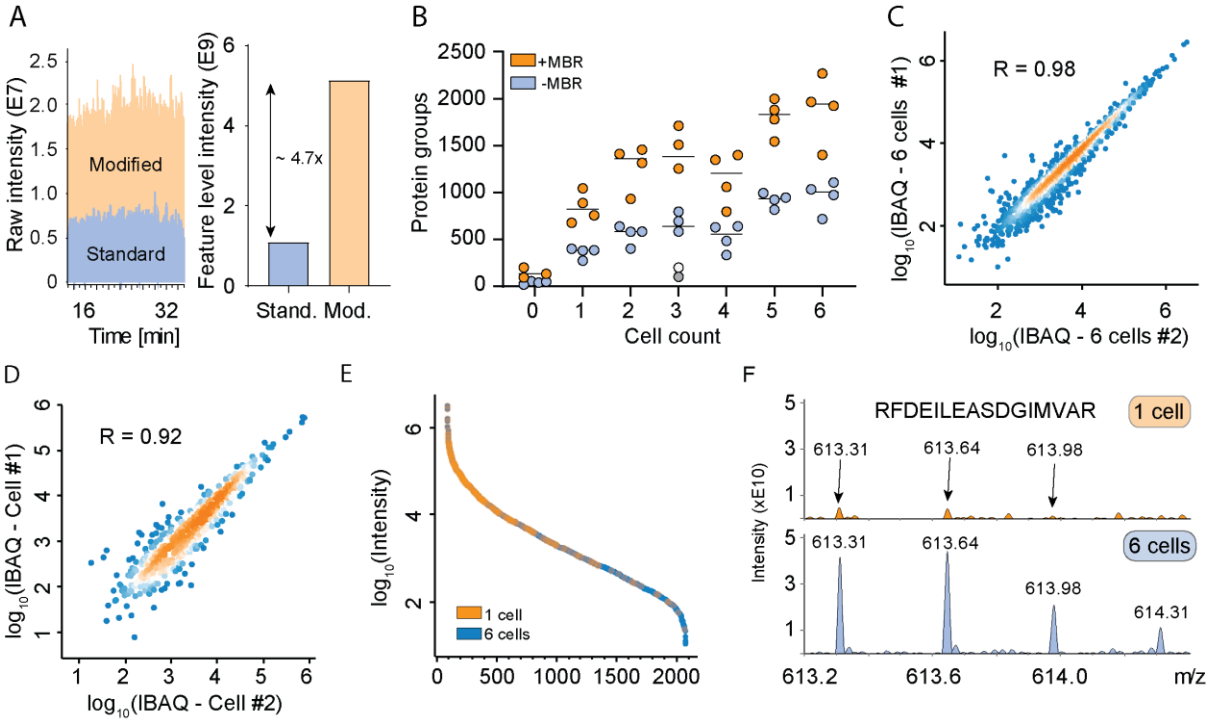


Figure 2: A: Raw signal increase from standard versus modified TIMS-qTOF instrument (left) and at the isotope pattern level (quantified features in MaxQuant) (right). **B:** Proteins quantified from one to six single HeLa cells, either with 'matching between runs' in MaxQuant (orange) or without matching between runs (blue). The outlier in the three cell measurement in grey is likely due to failure of FACS sorting as it identified a similar number as blank runs. **C:** Quantitative reproducibility in a rank order plot of a six-cell replicate experiment. **D:** Same as C for one cell against another. **E:** Rank order of protein signals in the six-cell experiment (blue) with proteins quantified in a single-cell colored in orange. **F:** Raw spectrum of one precursor isotope pattern of the indicated sequence and shared between the one-cell and six-cell experiments.

Single-cell protein extraction coupled to low flow chromatography

As electrospray (ES) is concentration dependent, sensitivity increases with decreasing flow-rate, however, very low flow systems are challenging to operate robustly and are consequently not widely available [28–31]. We recently described a chromatography system that decouples sample loading and gradient formation from the LC-MS run and operates at a standardized flow-rate of 1 ul/min for high reproducibility [32]. This flow is fully controlled by a single pump instead of the binary gradients produced by other systems. We found that it worked robustly when scaled down to 25 nl/min but we standardized on 100 nl/min, which enabled stable operation for the entire project with the same column (**Suppl. Fig. 1**). ES sprayer diameter and gradient length (30 min with 35 min between injections, 40-samples per day) were optimized for turnover, minimizing carry-over and stability.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

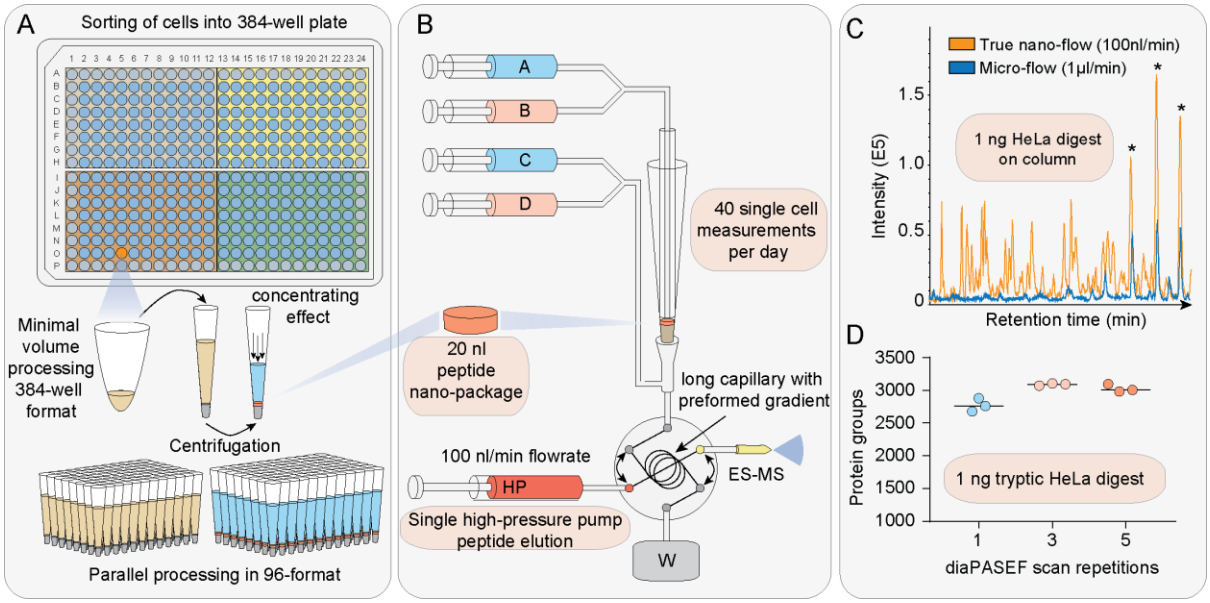


Figure 3: A: Single cells are sorted in a 384-well format into 1 µl lysis buffer by FACS, with outer wells serving as qualitative and quantitative controls. Single cells are lysed, proteins are solubilized and digested at 72 °C in 20 % acetonitile. Petides are concentrated into 20 nl volumes (nanopackages) in StageTips in a 96-well format. **B:** These tips are automatically picked and peptide nanopackages are eluted in a sub-100 nl volume. After valve switching, the peptide nanopackage is pushed on the analytical column and separated by the single high-pressure pump at 100 nl/min. **C:** Basepeak chromatogram of the standardized nano-flow (100 nl/min, orange) and micro-flow (1 µl/min, blue) gradients with 1 ng of HeLa digest on the stage-tip. Asterices indicate polyethylene glycole contaminants in both runs. **D:** Nano-flow (100 nl/min) and short gradient diaPASEF method combined. Summation of 1, 3 or 5 diaPASEF scan repetitions was used to find the optimum for high-sensitivity measurements at 1 ng of HeLa digest.

Single-cell proteomics requires nearly loss-less sample preparation from cells through protein digestion and to purified peptides ready for MS-analysis [17,18,33]. We found that the small volumes of 384-well plates intended for polymerase chain reactions provided a versatile and automatable environment for cell lysis and protein digestion in minimal volumes [34] (Methods, **Fig. 3A**). Briefly, single cells were sorted into wells containing 1 uL lysis buffer, followed by a heating step and further addition of buffer containing digestion enzymes to a total of 2 uL, all in an enclosed space. Peptides were concentrated in StageTip devices (standard EvoTips) into 20 nL nanopackages, from which they were eluted in minimal volumes (**Fig. 3B**). To benchmark the effect of reduced flow-rate and the concentrated elution, we directly compared signal traces of the normal 1 uL/min to the 100 nL/min set up. For 1 ng peptide material this resulted in a ten-fold increase in signal (**Fig. 3C**). To achieve high data completeness between single-cell measurements, we next replaced dda by diaPASEF, in which fragment level matching is further supported by ion mobility data [35]. We found that combining diaPASEF scan repetitions further improved protein identification numbers (**Fig. 3D**). Together, the very low flow chromatography and this diaPASEF acquisition mode resulted in the highly reproducible identification and quantification of more than 3,000 HeLa proteins from only 1 ng, a drastic increase from the 550 identified in our initial

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

set-up from a similar amount. Data completeness was more than 75% and coefficient of variation (CV) less than 11%. This demonstrated that diaPASEF provides its advantages also at extremely low sample amounts, prompting us to adopt this acquisition mode for the single-cell workflow in the remainder of this work.

Single-cell proteomics dissects drug-arrested cell-cycle states

The cell cycle is an important and well-studied biological process that has frequently been used as a test case in single-cell studies [36,37]. To investigate if our proteomics workflow could detect biological responses to drug perturbation at the single-cell level, we treated HeLa cells with thymidine and nocodazol to produce four cell populations enriched in specific cell cycle stages (221 cells; Methods, **Fig. 4A**). We quantified up to 1,209 proteins per single-cell and 1,305 overall, using a HeLa dia spectral library with about 29,000 precursors (Methods). This number ranged from a median of 568 in G1 to 1,081 in G1/S, 842 in G2, and 946 in G2/M (cutoff minimum for analysis is 400 proteins, **Fig. 4B**). To estimate total protein amount, we summed all protein signals based on their identifying peptides. G1 and G1/S cells showed least variation whereas by protein amounts G2 and G2/M cells varied four-fold (**Fig. 4C**). Judged by protein amount, G2 cells were approximately twice as large as G1 and G1/S cells; thus single-cell proteomics correctly reflected the proliferation state, while highlighting a substantial heterogeneity. The proteomes of the different cell cycle states grouped together in a Principal Component Analysis (PCA) plot (**Fig. 4D**). In addition to these drug-perturbed cells, we measured more than 200 untreated ones. The proteomes of these asynchronous cells were distributed over the cell cycle states (**Fig. 4E**).

Next, we asked whether single-cell proteome measurements can be used to assign cellular states, similar to how single-cell RNA-sequencing (scRNA-seq) measurements have frequently been applied to cell type and state discovery, highlighted by cellular atlas projects [38]. In previous proteomics, cell populations had been enriched for cell cycle states and sets of regulated proteins had been extracted [36,37]. We here selected cell cycle stage marker proteins as the top 20 most differentially expressed in either the G2/M, G1 and S protein set from ref [36,37], as it used similar drug treatment on bulk populations and compared how likely our single-cell G2/M phase cells can be discriminated based on those (Methods). We used these marker proteins to set up a cell cycle stage classifier that we had previously applied to scRNA-seq cell cycle state prediction, which clearly distinguished cells from G2/M and G1/S based on the G2/M phase score (**Fig. 4E**) [39].

To directly compare single-cell proteomes between G2/M and G1/S cells, we median normalized for protein amounts and stringently filtered our data set for completeness (Methods). Among the significantly regulated proteins was a large number of known cell cycle regulators, some of which are highlighted (Volcano plot; **Figure 4F)**. Quantitative MS data at the fragment ion level was highly significant for these as illustrated by the cell cycle regulator CDKN2A and further examples (FDR < $8.3\ 10^{-5}$, **Fig. 4G**; **Suppl. Fig. 3A, 4**). Our single-cell data set also highlighted proteins not previously associated with the cell cycle and the G2/M transition. For instance, the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

putative pseudogene NACAP1 was clearly identified and regulated (FDR < 10^{-14}, **Fig. 4H, Suppl. Fig. 3B**). It might have escaped previous detection because of its small size (213 amino acids) as we have noticed previously [40]

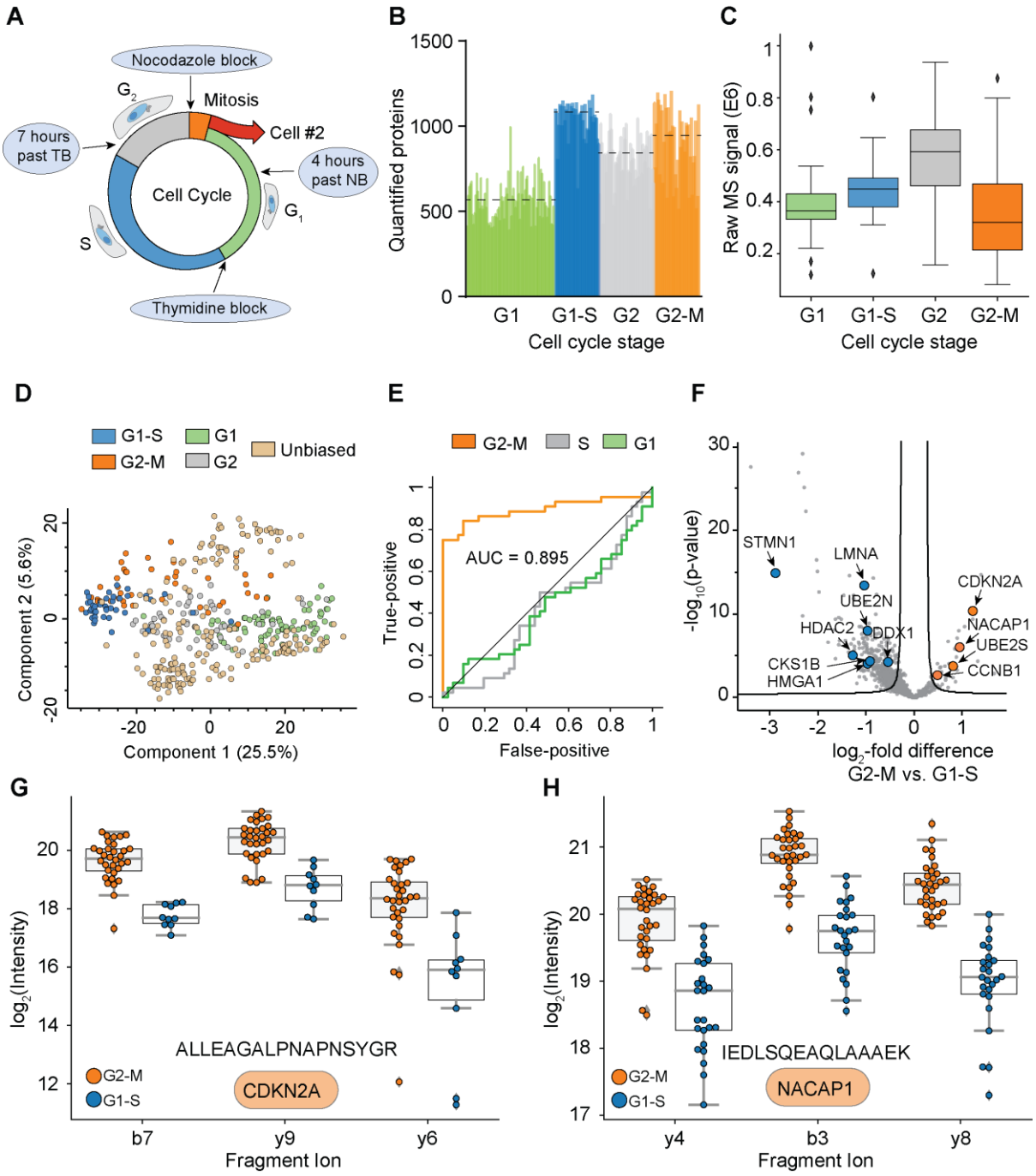


Fig. 4: A: Arresting single cells by drug perturbation. **B:** Numbers of protein identifications across 221 cells in the indicated cell cycle stages as enriched by the drug treatments in A. **C:** Violin plot of total protein signals of the single cells in B. **D.** PCA of single-cell proteomes of B with non-drug treated single cells overlaid in yellow. **E.** Receiver Operator Curve (ROC) for the prediction of G2/M cells against G1/S based on G2/M marker proteins with the indicated area under the curve (AUC) score. The other two curves, based on S and G1 marker proteins, respectively, indicate the inverse predictive power of these scores. **F.** Volcano plot of quantitative protein differences vs. reproducibility in the two drug

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

arrested states. Arrows points point towards colored significantly regulated key proteins of interest with at least 0.5-$\log_2$ fold-change. **G.** Quantitative fragment ion level data for CDKN2A and its peptide ALLEAGALPNAPNSYGR. **H.** Quantitative fragment ion level data for the pseudogene NACAP1 and its peptide IEDLSQEAQLAAAEK.

Differences between single-cell proteomes and transcriptomes

Given our set of more than 420 single-cell proteomes, we compared the single-cell proteomics measurements with similar single-cell RNA sequencing data (scRNA-seq). To minimize bias, we selected assays from two widespread scRNA-seq technologies, Drop-seq [41] and the lower throughput SMART-seq2 [42], on the same cellular system. The Drop-seq assay is based on unique molecular identifiers (UMIs) [43] to control for amplification biases in library preparation, whereas the SMART-seq2 assay is not UMI-controlled [44]. Note that MS-based proteomics inherently does not involve any amplification and is not subject to associated artifacts.

The HeLa cell culture should be relatively homogenous in gene expression states, except cell cycle heterogeneity. This biological homogeneity assumption allows us to assess self-consistency of the measurement technologies. First we computed the distribution over all pairwise correlation coefficients on mutual non-zero observation of cells within a technology (Methods). We found that in the protein measurement, cells have higher correlation on average than in the droplet-based method and similar correlation to the SMARTseq2 method (**Fig 5A**). The apparent cell-cell correlation in SMARTseq2 is an upper bound to the actual transcript based correlation values because of lacking amplification bias control.

When we considered the fraction of non-zero observations across technologies we found that on average in 60% of the more than 1,200 observed proteins by MS-based proteomics cells had non-zero expression values (**Fig 5B**). This profile was similar to that of SMART-seq2 and both had much less zero observations than the droplet-based protocol. Note that this analysis depends on the sequencing depth of the scRNA-seq library, whereas there is no such limiting sequencing step in the MS-based proteomics experiment. Next, we investigated whether there were detection limiting effects in the protein measurements. Such effects are discussed for scRNA-seq measurements as “drop-out events” or “zero-inflation”, but are now much reduced in UMI-based protocols [45]. We did find a bimodality in protein measurements (**Suppl. Fig 4**), suggesting that single-cell protein measurements could benefit from imputation and that likelihood-based parameter estimation methods could be based on bi-modal likelihood functions [46].

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

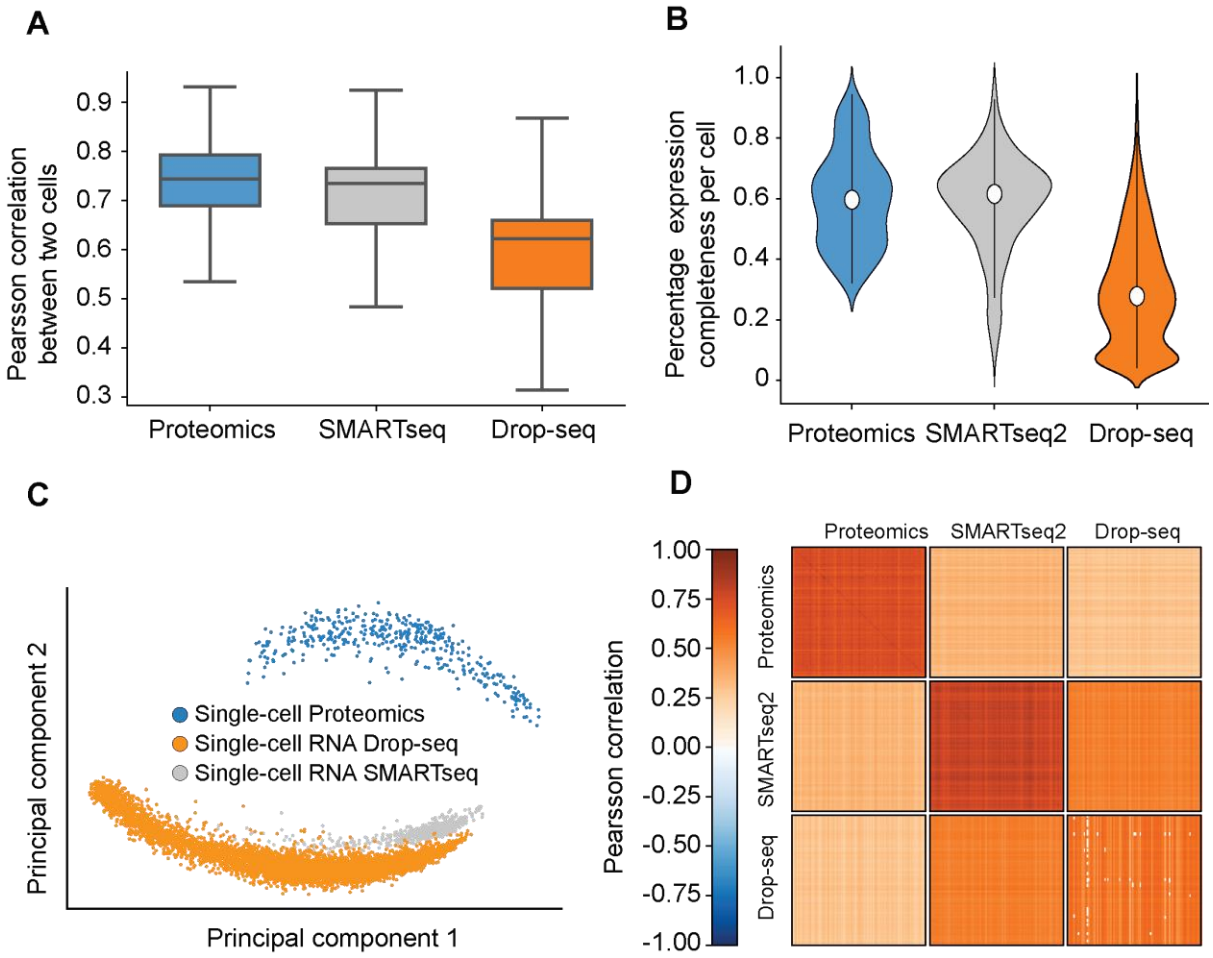


Fig. 5: A: Overall Pearson correlations for MS-based proteomics (1,281 proteins), SMART-seq2 (11,289 genes) and Drop-seq (8,146 genes). **B:** Violin plot of expression completeness in measurements for the three single-cell technologies **C.** Principal component analysis of single-cell gene and protein expression measurements. **D:** Heat map of cell-cell correlations across individual cells measured by proteomics and by both transcriptome technologies.

For bulk measurements, transcript levels generally correlate moderately with the corresponding protein levels, however, this correlation strongly depends on the biological situation [47]. Especially at the single-cell level, this issue is further convoluted with technical differences in the measurement technologies. We therefore asked to what degree scRNA-seq measurements could be used as proxy for protein measurements in our data and found that protein measurements separate strongly from RNA in a principal component analysis (**Fig 5 C).** Single-cell transcript expression levels correlate well across scRNA-seq technologies, but not with single-cell protein measurements (**Fig 5 D**). This suggests that single-cell protein and RNA levels are very different, re-emphasizing that protein measurements yield complementary information to RNA measurements and do not simply re-iterate similar gene expression states. This implies distinct RNA and protein abundance regulation mechanisms on both modalities, dissection of which would not be possible with RNA measurements alone. Further analyses with matched RNA and protein samples could help shed light onto this.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Single-cells have a stable core proteome

Prompted by the divergent correlation values between the proteome and transcript levels, we next investigated the variability of gene expression as a function of abundance. For low expressed proteins, the coefficients of variation (CVs) followed the Poisson distribution, whereas this variability measure became much smaller at high mean expression levels (**Fig. 6A**). This suggests that measured variability for the proteins below this level is currently dominated by sensitivity limitations of the technology. In contrast, the same analysis for UMI-controlled scRNAseq data revealed much higher overall transcriptome variability (**Supp. Fig. 4**). Remarkably, this difference is already very apparent with the current sensitivity of MS-based proteomics. Comparing single-cell proteome measurements with six-cell proteomes (**Fig. 2C**), suggests that a moderate increase in MS sensitivity would reveal a large part of the proteome to be quantitatively stably expressed. Based on these observations, we defined a 'core-proteome' subset in the MS-based proteomics data by selecting the top 150 proteins with the lowest CVs of the proteins shared between at least 70% of the more than 420 single-cell measurements, including the drug perturbations. As expected, these proteins where in the non-Poisson dominated part of the distribution of the proteome (**Fig. 6A**). Interestingly, this did not appear to be the case for the corresponding transcripts, many of which still followed Poisson statistics (**Fig. 6B**). This difference is also apparent when separately visualizing the overall variability of the core proteome and the corresponding transcriptome (**Fig. 6C**). This highlighted proteins frequently used for normalization such as GAPDH, ACTB and PKM, providing a positive control. The rank plot of the core proteome by MS-signal also reveals a diverse set of proteins, including representatives of the translation and folding machineries, as well as transporters and solute carriers.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

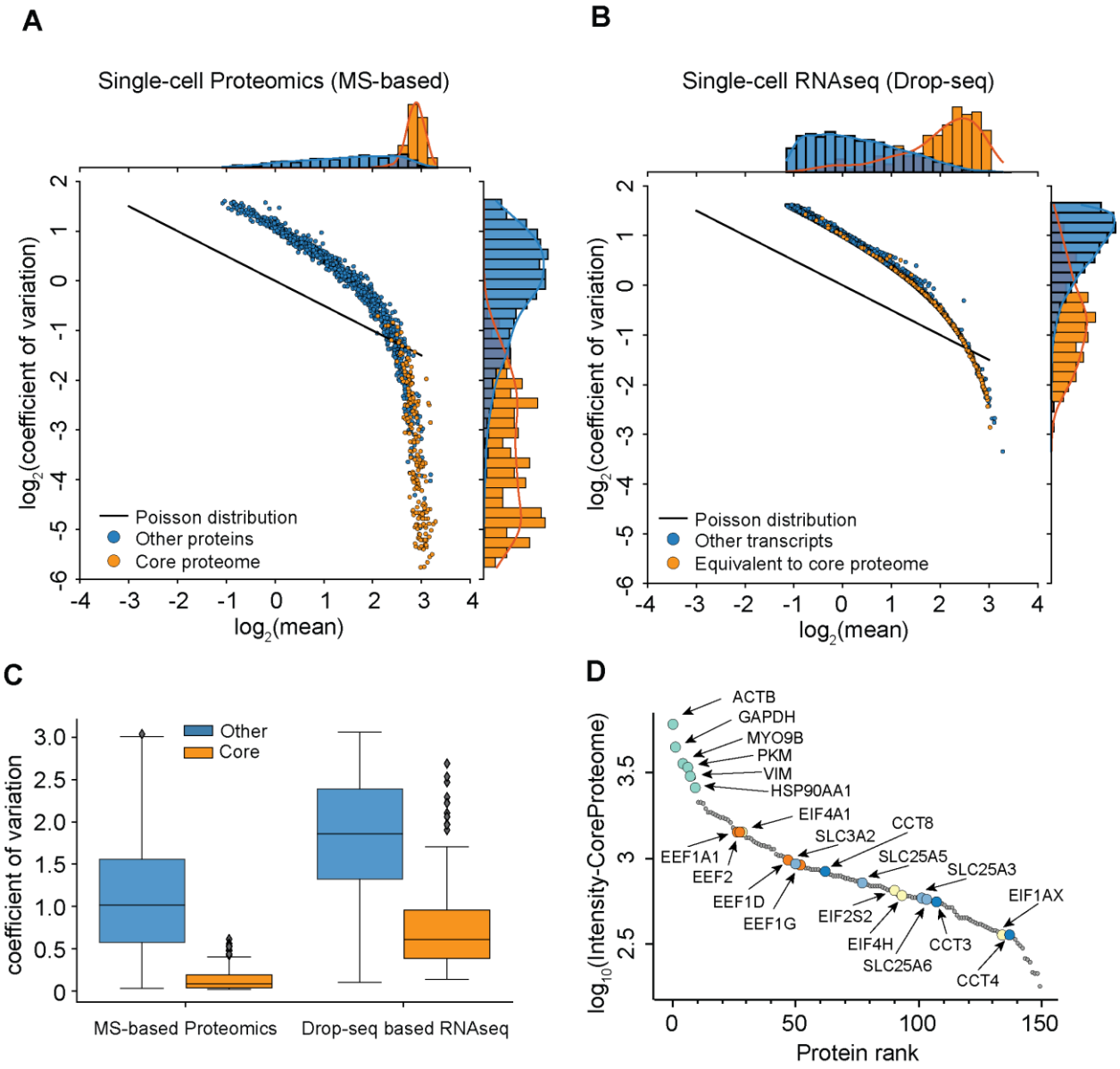


Fig. 6: A: Coefficient of variation of protein levels as a function of mean expression levels with the 'core proteome' colored in orange. **B.** CV of transcript expression values as a function of mean expression levels with transcripts corresponding to the core proteome in orange. **C.** Box plot of variability of protein signals and scRNA-seq across the cell populations for the core proteome and transcriptome and for the remainder of the protein measurements and corresponding transcript measurements. **D.** Rank order abundance plot for the core proteome with protein classes color coded (Green: Top six highest expressed core proteome members; Orange: Translation initiation factors; Yellow: Translation elongation factors; Light blue: Solute carriers (SLCs); Dark blue: Chaperonin family members).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Outlook

Although mainly demonstrated here for single-cell total proteome measurements, the sensitivity gain achieved in our workflow will be advantageous in any situation that is sample limited. This includes investigation of post-translational modifications from small numbers of cells or from *in vivo* material, measurements directly from paraffin embedded formalin fixed (FFPE) pathology specimens and the analysis of other compound classes such as metabolites or drugs. Our workflow is also compatible with chemical multiplexing with the advantage that the booster channel causing reporter ion distortions could be omitted or reduced. Furthermore, there are many opportunities for increasing overall sensitivity, including even brighter ion sources, improved chromatography and better data analysis and modeling tools, similar to the rapid recent advances in the scRNAseq field.

Acknowledgements

A.D.B. acknowledges support from the International Max Planck Research School for Life Sciences – IMPRS-LS. S. R. is supported by the Helmholtz Association under the joint research school "Munich School for Data Science" - MUDS. D.S.F. acknowledges support from a German Research Foundation (DFG) fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). F.C. acknowledges the European Union's Horizon 2020 research and innovation program (Marie Skłodowska-Curie individual fellowship under grant agreement 846795). We thank our colleagues in the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, at the Center for Protein Research at Copenhagen University, and at Bruker Daltonik for discussions and help. In particular, we thank I. Paron, A. Piras and C. Deiml for technical support and J.B. Müller for column production. We also thank Mark Ridgeway, Craig Whitehouse, Andreas Brekenfeld and Christoph Gebhardt of Bruker for their contribution to the development, installation and trouble-shooting of the modified mass spectrometer.

Author contributions. A.-D.B. and M.M. conceptualized and designed the study. A.-D.B., M.C.T., F.C., and A.M. performed experiments. M.A.P. and O.R. designed the new mass spectrometer. A.D-B., O.B.H. and N.B. conceived the new EvoSep gradient. A.D.-B., M.C.T., O.B.H., and N.B. evaluated the new EvoSep gradient and optimized it for proteomics performance. D.F. and F.J.T. conceptualized the single cell analyses and models. A.D.-B., S. R., D. F., M.C.T., F.M., O.B.H., N.B., C.V. and M.M. analyzed the data; A.-D.B. and M.M. wrote the manuscript.

Competing interests. M. L., O.R., A.A. and M.A.P. are employees of Bruker Daltonik. O. B. H. and N.B. are employees of EvoSep Biosystems. M. M. is an indirect shareholder in EvoSep Biosystems. F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH and Cellarity Inc., and ownership interest in Cellarity, Inc. and Dermagnostix. All other authors have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science (80-.).* **343**, 776–779 (2014).
2. Plass, M. *et al.* Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. doi:10.1126/science.aaq1723
3. Regev, A. *et al.* The human cell atlas. *Elife* **6**, (2017).
4. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).
5. Izar, B. *et al.* A single-cell landscape of high-grade serous ovarian cancer. doi:10.1038/s41591-020-0926-0
6. Tirosh, I. *et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq*.
7. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* (2011). doi:10.1038/nbt.2038
8. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).
9. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer Spatially resolved single-cell phenotypes. *Nature* **578**, 615 (2020).
10. Uhlén, M. *et al.* Tissue-based map of the human proteome. doi:10.1126/science.1260419
11. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
12. Larance, M. & Lamond, A. I. Proteins form the structural fabric of cells and under-pin all metabolic processes and regulatory mechanisms. Protein properties, including abundance levels, protein-protein interactions, post-translational modifications. (2015). doi:10.1038/nrm3970
13. Specht, H. *et al.* Single-cell mass-spectrometry quantifies the emergence of macrophage heterogeneity. doi:10.1101/665307
14. 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).
15. Cheung, T. K. *et al.* Defining the carrier proteome limit for single-cell proteomics. *Nat. Methods* (2020). doi:10.1038/s41592-020-01002-5
16. Brenes, A., Hukelmann, J., Bensaddek, D. & Lamond, A. I. *Multi-batch TMT reveals false positives, batch effects and missing values*. **16**, (2019).
17. Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat. Commun.* **9**, 1–10 (2018).
18. Li, Z.-Y. *et al.* Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. (2018). doi:10.1021/acs.analchem.8b00661
19. 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).
20. Meier, F. *et al. Online parallel accumulation – serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer*. **28**,
21. Meier, F. *et al.* Parallel accumulation-serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage. doi:10.1101/656207

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. 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**, (2008).
23. Prianichnikov, N. *et al. MaxQuant software for ion mobility enhanced shotgun proteomics*. **17**,
24. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry–based shotgun proteomics. *Nat. Protoc.* **11**, (2016).
25. Volpe, P. & Eremenko-Volpe, T. Quantitative studies on cell proteins in suspension cultures. *Eur. J. Biochem.* 195–200 (1970).
26. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1–8 (1996).
27. Prianichnikov, N. *et al.* MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics. *Mol. Cell. Proteomics* **19**, 1058–1069 (2020).
28. Yamashita, M. *et al. Analytical Properties of the Nanoelectrospray Ion Source A C R e s e a r c h*. *Int. J. Mass Spectrom. Ion Processes* **68**, (UTC, 1996).
29. Emmett, M. R. & Caprioli, R. M. Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* **5**, 605–613 (1994).
30. Greguš, M., Kostas, J. C., Ray, S., Abbatiello, S. E. & Ivanov, A. R. Improved Sensitivity of Ultralow Flow LC−MS-Based Proteomic Profiling of Limited Samples Using Monolithic Capillary Columns and FAIMS Technology. (2020). doi:10.1021/acs.analchem.0c03262
31. Xiang, P. *et al.* Picoflow Liquid Chromatography−Mass Spectrometry for Ultrasensitive Bottom-Up Proteomics Using 2-μm-i.d. Open Tubular Columns. *Cite This Anal. Chem* **92**, 4711–4715 (2020).
32. Bache, N. *et al.* A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-Robust Proteomics. **15**, 2020 (2018).
33. 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**, 161 (2018).
34. Specht, H., Emmott, E., Koller, T. & Slavov, N. High-throughput single-cell proteomics quantifies the emergence of macrophage heterogeneity. doi:10.1101/665307
35. Meier, F. *et al.* Deep learning the collisional cross sections of the peptide universe from a million training samples. (2020). doi:10.1101/2020.05.19.102285
36. Ly, T. *et al.* Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *Elife* **6**, (2017).
37. Aviner, R., Shenoy, A., Elroy-Stein, O., Geiger, T. & Snyder, M. Regulation through Integrative Multi-omic Analysis. *PLoS Genet* **11**, 1005554 (2015).
38. Regev, A. *et al.* The human cell atlas. *Elife* **6**, (2017).
39. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
40. Chen, J. *et al. Pervasive functional translation of noncanonical human open reading frames*.
41. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
42. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* (2013). doi:10.1038/nprot.2014.006
43. Schwabe, D., Formichetti, S., Junker, J. P., Falcke, M. & Rajewsky, N. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transcriptome dynamics of single cells during the cell cycle. *Mol. Syst. Biol.* **16**, (2020).
44. Hu, W. *et al.* HeLa-CCL2 cell heterogeneity studied by single-cell DNA and RNA sequencing. *PLoS One* **14**, e0225466 (2019).
45. Svensson, V. Droplet scRNA-seq is not zero-inflated. *Nat. Biotechnol.* 147–150 (2020).
46. Risso, D., Perraudeau, F., Gribkova, S. & Dudoit, S. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.* (2018). doi:10.1038/s41467-017-02554-5
47. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* doi:10.1038/s41576-020-0258-4

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Sample preparation for bulk dilution experiments. For all benchmark experiments purified peptides from bulk HeLa cells were used. HeLa was cultured in Dulbecco's modified Eagle's medium at 10% fetal bovine serum, 20 mM glutamine and 1% penicillin–streptomycin. Cells were collected by centrifugation, washed with phosphate-buffered saline (PBS), flash-frozen in liquid nitrogen and stored at −80 °C. Cells were resuspended in SDC lysis buffer buffer[1] and boiled for 20 min at 95°C, 1500 rpm to denature and reduce and alkylate cysteins, followed by sonication in a Branson, cooled down to room temperature and diluted 1:1 with 100 mM TrisHCl pH 8.5. Protein concentration was estimated by Nanodrop measurement and 500 µg were further processed for overnight digestion by adding LysC and trypsin in a 1:50 ratio (µg of enzyme to µg of protein) at 37° C and 1500 rpm. Peptides were acidified by adding 1% TFA 99% isopropanol in a 1:1 ratio, vortexed, and subjected to StageTip [2] clean-up via SDB-RPS. 20 µg of peptides were loaded on two 14-gauge StageTip plugs. Peptides were washed two times with 200 µL 1% TFA 99% isopropanol followed 200 µL 1% TFA 99% Isopropanol in an in-house-made StageTip centrifuge at 2000 xg, followed by elution with 100 µL of 1% Ammonia, 80% ACN, 19% ddH2O into PCR tubes and dried at 60°C in a SpeedVac centrifuge (Eppendorf, Concentrator plus). Peptides were resuspended in 0.1% TFA, 2% ACN, 97.9% ddH2O.

Sample preparation for single-cell experiments. HeLa cells were cultured as described above. Supernatant was removed, cells were detached with trypsin-treatment, followed by strong pipetting for cell aggregate dissociation. Cells were washed three times with ice-cold Phosphat-buffered Saline (PBS), pelleted by centrifugation, and the supernatant was removed. For fluorescent-activated cell-sorting (FACS), DAPI was added and sorting performed on the DAPI-negative cell population. Single cells were sorted into 384-well plates containing 1µl of 20% acetonitrile (ACN), 100mM TrisHCl pH 8.5, centrifuged briefly, sealed with aluminum foil and frozen at -80°C until further use. When needed, single-cell containing 384-well plates were incubated for 30 min at 72°C in a PCR cycler, followed by sonication. Protein digestion was performed overnight at 37°C in a PCR cycler after adding 1 µl of 20% ACN, 100mM TrisHCl pH 8.5, 1ng trypsin/lysC mix. For the peptide bulk and cell count dilution experiments, peptides were resuspended in 4 µl of 2% ACN, 0.1% TFA, 97.9% ddH_2O and injected directly via NanoLC.

For all other single cell experiments, samples were dried in a SpeedVac, resuspended in 20 µl 0.1% formic acid (FA), 99.9% ddH2O (Buffer A) before transfer into activated EvoTips. These were activated following the standard EvoSep protocol [3]. 50 µl buffer A was added to each EvoTip followed by centrifugation at 200 xg for 1 min. The sample was transferred into the EvoTip, followed by centrifugation at 600xg for 1min, and two centrifugation steps after adding 50 µl buffer A. Last, 150µl buffer A was added to each EvoTip and spun for 30 sec at 300 xg.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cell cycle experiments. The drug-perturbed cell-cycle arrest experiment was designed to trap cells in four cell-cycle stages – G1, the G1/S-transition, G2 and the G2/M-transition. HeLa cells were grown to approximately 30 % confluence as described above, washed and treated for 24 h with 5 mM thymidine, released for 4.5 h and treated again with 5 mM thymidine, or 0.1µg/ml nocodazol for 13 h. Cells for the G1/S phase (thymidine block) or G2/M phase (nocodazol block) were washed in PBS, trypsinated, subjected to strong pipetting to dissociate cell aggregates and ice-cold PBS washes before DAPI-negative single-cell FACS sorting. A second batch of G1/S phase and G2/M phase blocked cells was washed and cultured for 7 h or 2.5 h to yield early G2 and G1-phase HeLa cells. These were washed with PBS, trypsinated and subjected to DAPI-negative single-cell FACS sorting into 384-well plates pre-loaded with 1 µl 20% acetonitrile, 100 mM TrisHCl pH 8.5 lysis buffer.

High-pH reversed-phase fractionation. To generate a deep library of HeLa precursors for all data-dependent benchmark experiments, peptides were fractionated at pH 10 with the spider-fractionator[4]. 50 µg of purified peptides were separated on a 30 cm C_{18} column in 96 min and concatenated into 24 fractions with 2 min exit valve switches. Peptide fractions were dried in a SpeedVac and reconstituted in 2% ACN, 0.1% TFA, 97.9% ddH2O for LC-MS analysis.

Liquid-chromatography. For the initial benchmark experiments with HeLa bulk dilution and the cell count dilution, liquid chromatography analysis was performed with an EASY nanoLC 1200 (Thermo Fisher Scientific). Peptides were loaded on a 50 cm in-house packed HPLC-column (75µm inner diameter packed with 1.9µm ReproSil-Pur C18-AQ silica beads, Dr. Maisch GmbH, Germany). Sample analytes were separated using a linear 60 min gradient from 5-30% B in 47.5 min followed by an increase to 60% for 2.5 min, by a 5 min wash at 95% buffer B at 300nl/min and re-equilibration for 5 min at 5% buffer B (Buffer A: 0.1% Formic Acid, 99.9% ddH2O; Buffer B: 0.1% Formic Acid, 80% CAN, 19.9% ddH2O). The column temperature was kept at 60°C by an in-house manufactured oven.
For all other proteome analyses, we used an Evosep One liquid chromatography system[5] and analyzed the single-cell proteomes with a novel 35 min stepped pre-formed gradient eluting the peptides at 100 nl/min flow-rate. We used a 15 cm × 75 µm ID column with 1.9 µm C18 beads (EvoSep) and a 10 µm ID electrospray emitter (Bruker Daltonik). Mobile phases A and B were 0.1 vol% formic acid in water and 0.1 vol% formic acid in ACN, respectively.
Both LC systems were coupled online to a modified trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Pro, Bruker Daltonik GmbH, Germany) via a nano-electrospray ion source (Captive spray, Bruker Daltonik GmbH).

Construction of a novel mass spectrometer with higher sensitivity. We modified our ion source to draw more ions into the vacuum system of the instrument by modifying the glass capillary that conducts gas and ions between the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ionization region at atmospheric pressure and the first pumping region of the mass spectrometer. Additional gas is eliminated via an extra pumping stage. Novel prototype ion optics, a high pressure ion funnel and a radio frequency (RF) multipole confine the ions and transport them to the next vacuum region where the analysis by trapped ion mobility mass spectrometry (TIMS) occurs. The glass capillary is oriented orthogonal to the high pressure funnel so that neutral contaminants and droplets are first directed away from the funnel by the gas flow. Furthermore, the high pressure funnel and RF multipole are oriented orthogonal to the TIMS, maintaining the gas dynamics of our original design. Remaining neutral contaminants are guided away from the TIMS entrance. To accommodate the increased ion current, the TIMS analyzer was updated to a new stacked ring (SRIG) design. We use a higher order RF field in the ion accumulation region to create a larger effective ion storage volume than the low order fields of previous designs. A low order quadrupolar field is maintained in the analyzer region to compress the ions towards the analyzer axis during elution to maintain high mobility resolution. The transition between the high order and low order parts of the device was optimized compared to prior designs to further improve peak shape and ion mobility resolution. This results in about a factor of three gain in ion capacity and therefore about a factor of three in the instrument's dynamic range.

Mass spectrometry. Mass spectrometric analysis was performed either in a data-dependent (dda) or data-independent (dia) PASEF mode. For ddaPASEF, 1 MS1 survey TIMS-MS and 10 PASEF MS/MS scans were acquired per acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 50 ms each and we analyzed the ion mobility range from $1/K_0$ = 1.6 Vs cm^{-2} to 0.6 Vs cm^{-2}. Precursor ions for MS/MS analysis were isolated with a 2 Th window for m/z < 700 and 3 Th for m/z >700 in a total m/z range of 100-1.700 by synchronizing quadrupole switching events with the precursor elution profile from the TIMS device. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at $1/K_0$ = 1.6 VS cm^{-2} to 20 eV at $1/K_0$ = 0.6 Vs cm^{-2}. Singly charged precursor ions were excluded with a polygon filter (otof control, Bruker Daltonik GmbH). Precursors for MS/MS were picked at an intensity threshold of 1.500 arbitrary units (a.u.) and resequenced until reaching a 'target value' of 20.000 a.u taking into account a dynamic exclusion of 40 s elution. For DIA analysis, we made use of the correlation of Ion Mobility (IM) with m/z and synchronized the elution of precursors from each IM scan with the quadrupole isolation window. We used the Py3-method for library acquisition of the single-cell experiments and the short gradient diaPASEF method as described in Meier et al. [6], but performed up to 5 consecutive diaPASEF cycles before the next MS1-scan (see main text). The collision energy was ramped linearly as a function of the IM from 59 eV at 1/K0 = 1.6 Vs cm^{-2} to 20 eV at 1/K0 = 0.6 Vs cm^{-2}.

Raw data analysis. ddaPASEF data for tryptic HeLa digest dilution series and the cell count experiment were analyzed in the MaxQuant environment (version 1.6.7) and searched against the human Uniprot databases (UP000005640_9606.fa,

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

UP000005640_9606_additional.fa), which extracts features from four-dimensional isotope patterns and associated MS/MS spectra [7,8]. False-discovery rates were controlled at 1% both on peptide spectral match (PSM) and protein levels. Peptides with a minimum length of seven amino acids were considered for the search including N-terminal acetylation and methionine oxidation as variable modifications and cysteine carbamidomethylation as fixed modification, while limiting the maximum peptide mass to 4,600 Da. In case of single-cell count experiment no cysteine carbamidomethylation was searched since no alkylation was performed. Enzyme specificity was set to trypsin cleaving c-terminal to arginine and lysine. A maximum of two missed cleavages were allowed. Maximum precursor and fragment ion mass tolerance were searched as default for TIMS-DDA data. Peptide identifications by MS/MS were transferred by matching four-dimensional isotope patterns between the runs (MBR) with a 0.7-min retention-time match window and a 0.05 $1/K_0$ ion mobility window in case of the single cell-count dilution experiment into a deep ddaPASEF library consisting of 24 fractionations of tryptic HeLa digest. These data were also searched without matching between runs to access the MBR-mediated identification increase. Either intensity-based absolute quantification (IBAQ) or label-free quantification was performed with the MaxLFQ algorithm and a minimum ratio count of one [9].

For all other single-cell experiments, we used a small precursor library consisting of 29,087 precursors mapped to 23,198 peptides and 4,251 protein groups, which was acquired with the Py3 diaPASEF method[6] and generated with the Spectronaut software (version 14.9.201124.47784; Biognosys AG, Schlieren, Switzerland). A minimum of three fragments per peptide, and a maximum of six fragments were included. All single-cell measurements were searched against the human UniProt reference proteome of canonical and isoform sequences. Searches used protein N-terminal acetylation, oxidation of the N-terminus and oxidation of cysteins as variable modifications. We generated one decoy precursor per precursor in the spectral library and used a conservative normal distribution estimator approach for p-value estimation. Protein intensities were normalized using the “Local Normalization” (Q-value complete) algorithm in Spectronaut based on a local regression model [10]. A protein and precursor FDR of 1% was used. Default settings were used for other parameters. In brief, a trypsin/P proteolytic cleavage rule was used, permitting a maximum of two missed cleavages and a peptide length of 7–52 amino acids.

Visualization and FDR estimates of fragment ion intensities.

Quantitative fragment ion profiles were generated from the Spectronaut output table via the “F.PeakArea” column. Only fragment ions used for quantification in Spectronaut were included (EG.UsedForProteinGroupQuantity = True, EG.UsedForPeptideQuantity = True, F.ExcludedFromQuantification = False). To cancel out cell-size dependent abundance changes, one normalisation factor was estimated per cell, using fold-change based normalization of the whole dataset, as described in the MS-EmpiRe method, which we also used for FDR control [11]. The intensities were log2 transformed and subsequently visualized.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Proteomics downstream data analysis. Proteomics data analysis was performed in the Perseus environment (version 1.6.7) and GraphpadPrism (version 8.2.1)[12]. MaxQuant output tables were filtered for ”Reverse“, ”Only identified by site modification“, and ”Potential contaminants“ before further processing. For single-cell analysis on the Spectronaut output, data were filtered first for at least 400 protein observations per cell and at least 15% quantification events across rows and log_2-tranformed. Principal component analysis (PCA) was performed on column-wise median-normalized proteins with coefficient of variation of less or equal than 200%. Differential expression analysis by two-sided unpaired t-test was performed on two groups filtered for at least 50% row-wise quantification events within one group followed by column-wise median normalization. False-discovery rate control due to multiple hypothesis testing was performed by a permutation-based model and SAM-statistic with an S0-parameter of 0.3. For cell-size estimation based on raw MS-signal, intensity outputs within cell cycle resolved single-cell proteomics results were summed up and visualized as boxplots. The core proteome was calculated by filtering the whole single-cell proteomics data set for at least 70% quantification events for each protein followed by selection of the top 150 proteins with the smallest coefficient of variation across the dataset.

Single-cell protein and RNA comparison. The SMART-Seq2 [13,14] data set measured 720 HeLa cells in 3 different batches, with a total of 24,990 expressed genes. The Drop-seq [15,16] data set contained 3 batches with a total of 5665 cells and 41,161 expressed genes. We performed the single cell analysis with scanpy v1.6.0 [17]. We used standardized filtering across all datasets, removed cells with less than 400 genes expressed and removed genes detected in less than 10% of the remaining cells, resulting in 11,289 transcripts in 720 cells in the SMART-Seq2 dataset and 8146 transcripts and 5625 cells measured with Drop-seq technology. All abundance entries were linearly scaled to sum to 1E6 per cell and then *log(x+1)* transformed. Correlation values between the expressions of two cells were computed as the Pearson correlation on all entries that were non-zero in both vectors based on the preprocessed data.

Cell Cycle State Prediction. Cell cycle predictions were performed using the scanpy method score_genes[18,19] based on three sets of proteins that are specifically expressed in the G1 (EIF4A2, GDA, HIST1H1E, KRT18, HNRNPA1, DBNL), S (NOLC1, ATP2A2, CANX, BCAP31, CPT1A, CYC1, CKB) or G2/M (TOP2A, HMGB1, CCNB1, EIF5B, TMSB10) phase, respectively. The cell phase specific protein sets were selected based on the z-scored fold-change ratios provided in Geiger et al.[20]. The top20 highest differentially expressed genes were selected, but only the aforementioned ones were also identified in our data. This scoring method yields the average expression on the provided set of genes minus the average expression on a

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reference set of genes, for each cell. The reference set is chosen to mirror the average expression of the target gene set.

References

1. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
2. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. (2007). doi:10.1038/nprot.2007.261
3. Sample loading protocol for Evotips. Available at: https://www.evosep.com/wp-content/uploads/2020/08/Sample-loading-protocol_A6_v4_28.02_WEB.pdf.
4. Kulak, N. A., Geyer, P. E. & Mann, M. Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol. Cell. Proteomics* **16**, 694–705 (2017).
5. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2**, 185–195 (2016).
6. Meier, F. *et al.* Parallel accumulation-serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage. doi:10.1101/656207
7. 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**, (2008).
8. Prianichnikov, N. *et al.* MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics. *Mol. Cell. Proteomics* **19**, 1058–1069 (2020).
9. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
10. Callister, S. J. *et al.* Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics. (2006). doi:10.1021/pr050300l
11. Ammar, C., Gruber, M., Csaba, G. & Zimmer, R. MS-EmpiRe utilizes peptide-level noise distributions for ultra-sensitive detection of differentially expressed proteins. *Mol. Cell. Proteomics* **18**, 1880–1892 (2019).
12. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731–740 (2016).
13. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* (2013). doi:10.1038/nprot.2014.006
14. Hu, W. *et al.* HeLa-CCL2 cell heterogeneity studied by single-cell DNA and RNA sequencing. *PLoS One* **14**, e0225466 (2019).
15. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
16. Schwabe, D., Formichetti, S., Junker, J. P., Falcke, M. & Rajewsky, N. The transcriptome dynamics of single cells during the cell cycle. *Mol. Syst. Biol.* **16**, (2020).
17. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
18. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

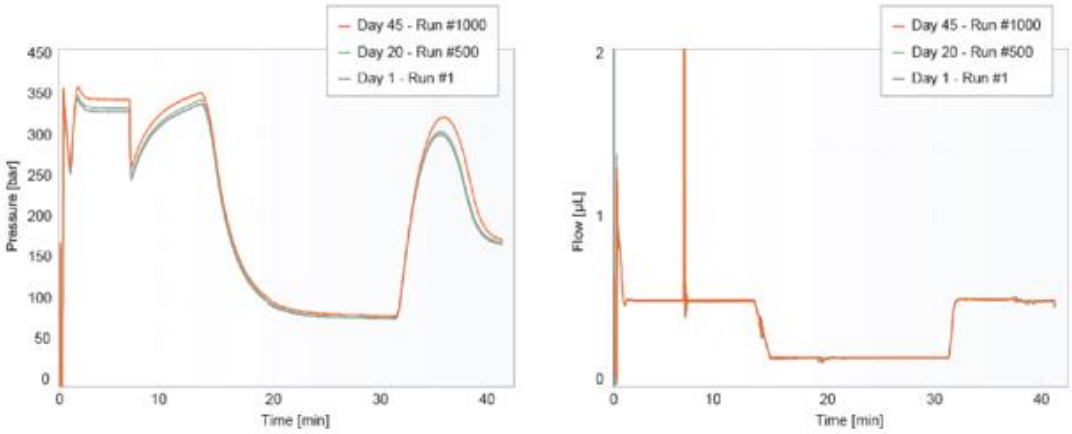
19. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol. Vol.* **33**, (2015).
20. Aviner, R., Shenoy, A., Elroy-Stein, O., Geiger, T. & Snyder, M. Regulation through Integrative Multi-omic Analysis. *PLoS Genet* **11**, 1005554 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

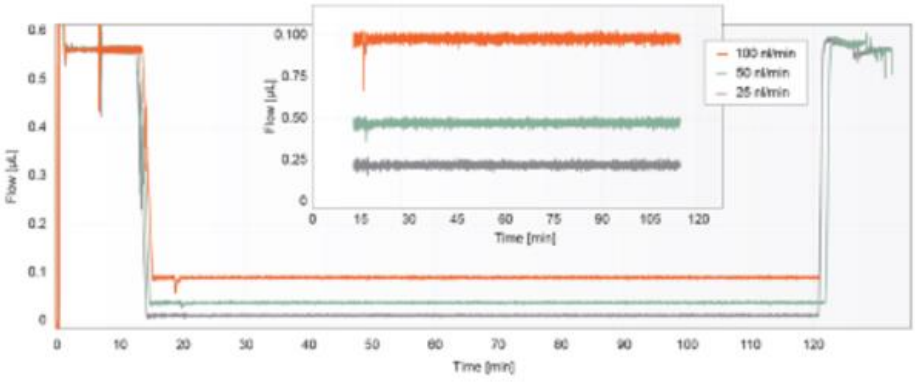
Supplementary Figures

A

B



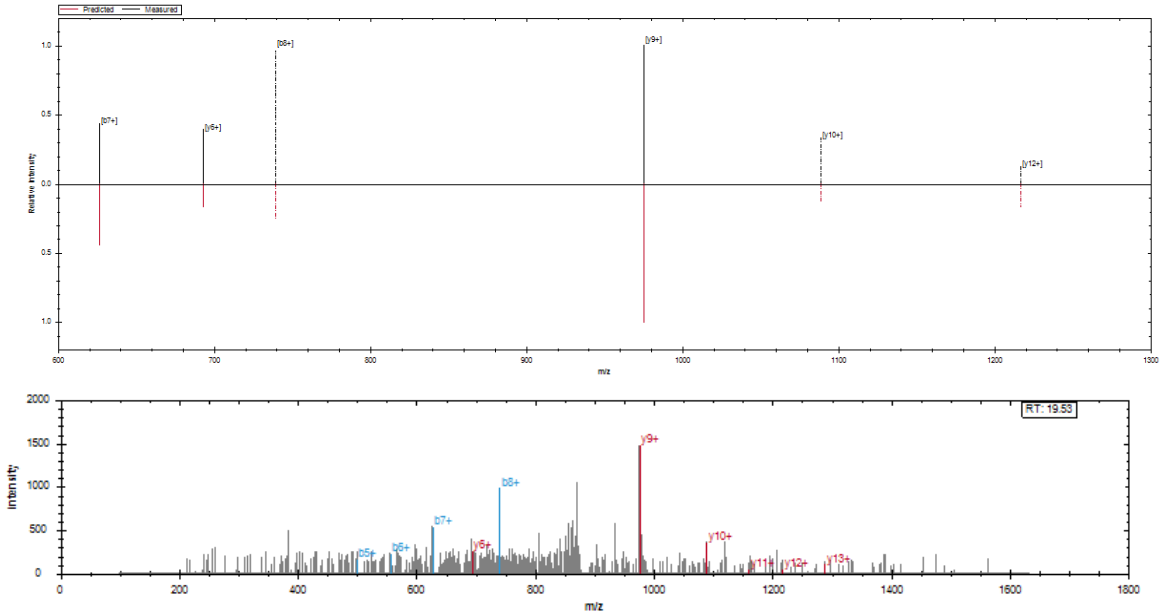
C



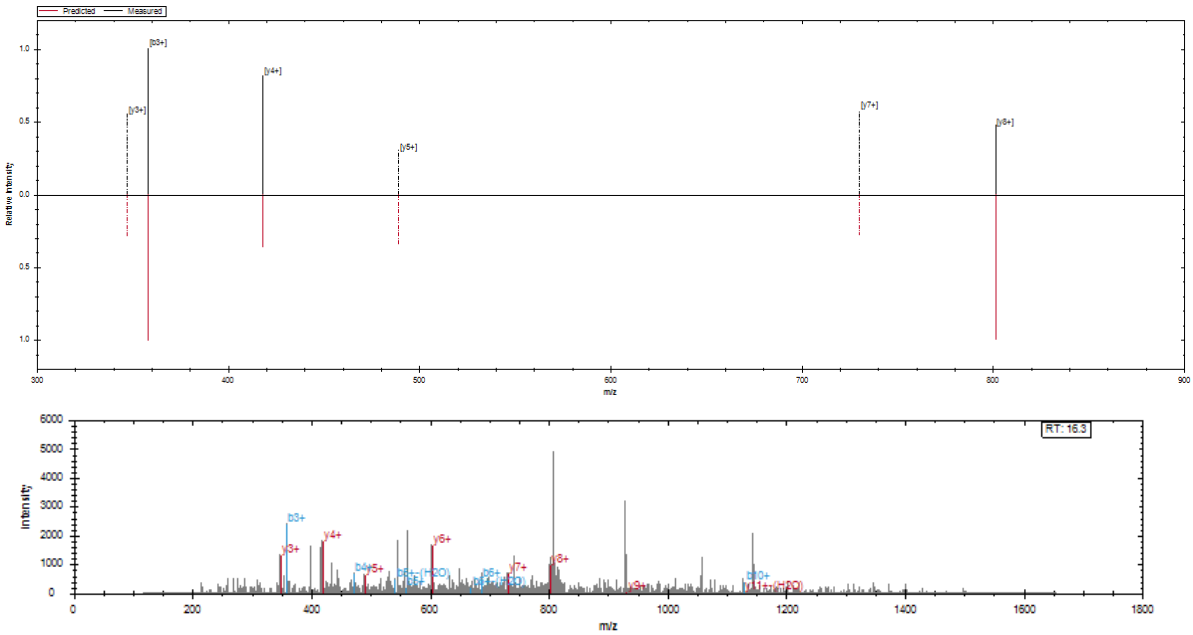
Supplementary Figure 1: 100 nl/min true nanoflow gradient. **A:** Column and emitter setup. **B:** Pressure and flow profile of the gradient of more than 1000 consecutive runs. **C:** Flow profiles of the 100/50/25 nl/min prototype test gradients.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A

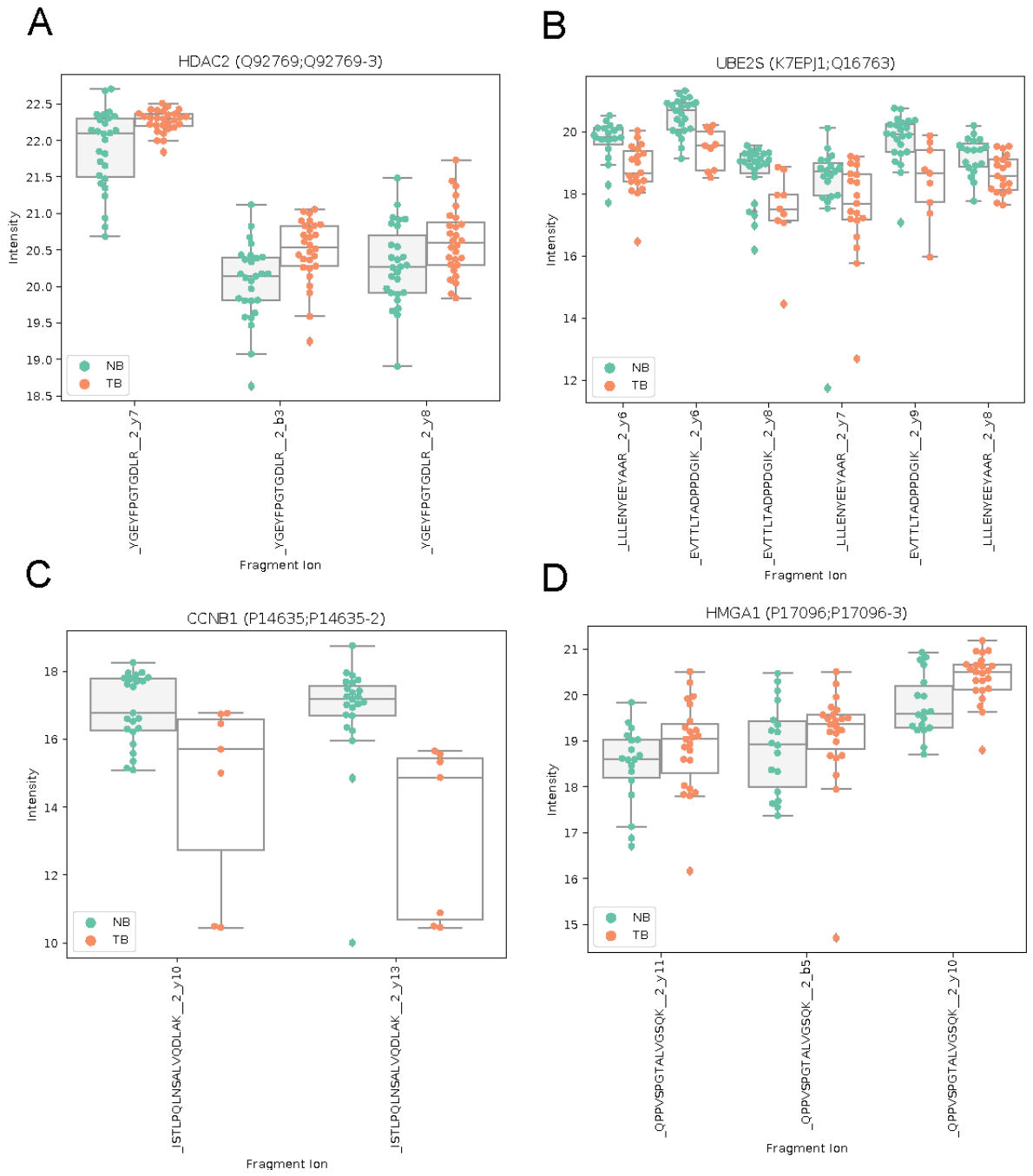


B



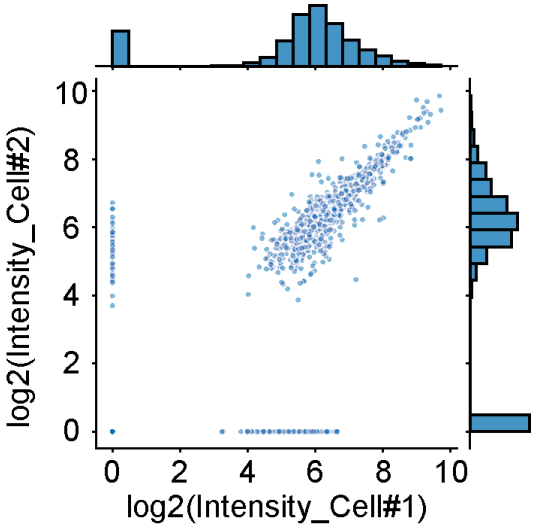
Supplementary Figure 2: A: Comparison of the theoretical and experimental MS2 spectra of a CDKN2A peptide. **B:** Comparison of the theoretical MS2 spectrum with the experimental one for a NACAP1 peptide.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

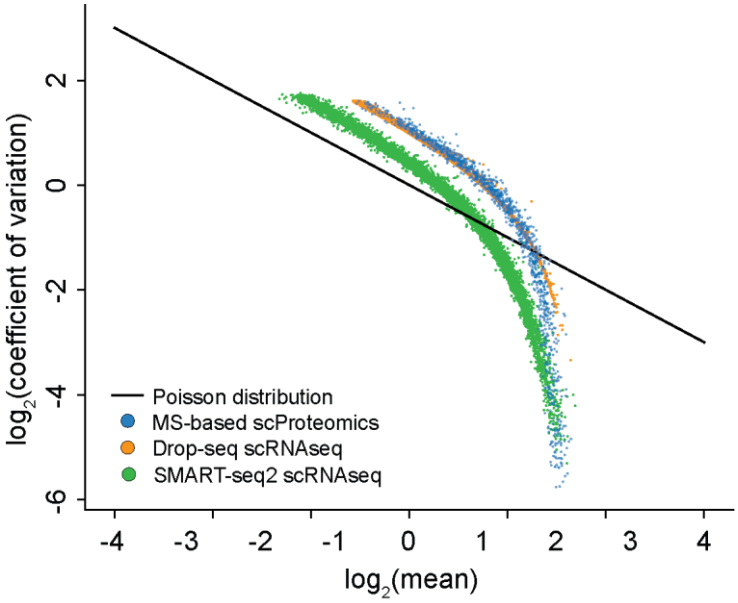


Supplementary Figure 3: A: Fragment ion intensities for a peptide of HDAC2. **B:** Fragment ion intensities for a peptide of UBE2S. **C:** Fragment ion intensities for a peptide of CCNB1.**D:** Fragment ion intensities for a peptide of HMGA1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.423933; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 4: Correlation of two independently measured single-cell proteomes to each other.



Supplementary Figure 5: Distribution of MS-based single-cell proteomics (blue) and transcriptomics data of two technologies (Drop-seq (orange), SMART-seq2 (green)).