

METHOD

Open Access



SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation

Bogdan Budnik^{1*}, Ezra Levy², Guillaume Harmange² and Nikolai Slavov^{2,3*} 

Abstract

Some exciting biological questions require quantifying thousands of proteins in single cells. To achieve this goal, we develop Single Cell ProtEomics by Mass Spectrometry (SCoPE-MS) and validate its ability to identify distinct human cancer cell types based on their proteomes. We use SCoPE-MS to quantify over a thousand proteins in differentiating mouse embryonic stem cells. The single-cell proteomes enable us to deconstruct cell populations and infer protein abundance relationships. Comparison between single-cell proteomes and transcriptomes indicates coordinated mRNA and protein covariation, yet many genes exhibit functionally concerted and distinct regulatory patterns at the mRNA and the protein level.

Background

Cellular systems, such as tissues, cancers, and cell cultures, consist of a variety of cells with distinct molecular and functional properties. Characterizing such cellular differences is key to understanding normal physiology, combating cancer recurrence, and enhancing targeted stem cell differentiation for regenerative therapies [1–5]; it demands quantifying the proteomes of single cells.

However, quantifying proteins in single mammalian cells has remained confined to fluorescent imaging and antibodies. Fluorescent proteins have proved tremendously useful but are limited to quantifying only a few proteins per cell and sometimes introduce artifacts [5, 6]. Multiple antibody-based methods for quantifying proteins in single cells have been recently developed, including CyTOF [7, 8], single-cell Western blots [9], and Proseek Multiplex, an immunoassay readout by PCR [10]. These methods can quantify up to a few dozen endogenous proteins recognized by highly specific cognate antibodies and have enabled exciting research avenues [5]. Still, the throughput and accuracy of antibody-based methods are limited by cellular permeability, molecular crowding, epitope accessibility, and the availability of

highly specific antibodies that bind their cognate proteins stoichiometrically [5, 11].

On the other hand, the application of liquid chromatography (LC) and tandem mass spectrometry (MS/MS) to bulk samples comprised of many cells allows for the confident identification and quantification of thousands of proteins [12–18]. To develop approaches that may bring at least some of this power of LC-MS/MS to single mammalian cells, we considered all steps of well-established bulk protocols and how they may be adapted to much more limited samples. We were motivated by the realization that most proteins are present at over 50,000 copies per cell [19, 20] while modern MS instruments have sensitivity to identify and quantify ions present at hundreds of copies [21, 22]. Thus, if we manage to deliver even 1% of the protein copies from a single cell as ions for MS analysis, we may quantify them accurately [22].

Most protocols for bulk LC-MS/MS begin by lysing the cells with detergents or urea [23]. Since these chemicals are incompatible with MS, they have to be removed by cleanup procedures. These cleanup procedures can result in substantial losses of protein, and colleagues have developed advanced methods, such as SP3 [24] and iST [25], that minimize cleanup losses and allow for quantifying thousands of proteins from samples having just a few micrograms of total protein [23, 26]. Indeed, the SP3 method has been successfully used for purifying

* Correspondence: bbudnik@mcb.harvard.edu; nslavov@alum.mit.edu

¹MSPRL, FAS Division of Science, Harvard University, Cambridge, MA 02138, USA

²Department of Biology, Northeastern University, Boston, MA 02115, USA

Full list of author information is available at the end of the article



and quantifying proteins from single human oocytes (~100 μm diameter) [27]. Still, most mammalian cells are smaller (10–15 μm diameter) [19], and we were not confident that we could clean up their cell lysates (having about 500 pg of total protein) without incurring large protein losses. Thus, we sought to obviate cleanup (and therefore eliminate cleanup-related losses) by replacing chemical lysis with mechanical lysis by focused acoustic sonication [23, 28].

Before being ionized and sent for MS analysis, peptides have to be separated [12, 15, 16]. The separation for bulk samples is usually accomplished by nanoliquid chromatography (nLC). To reduce losses due to proteins adhering to the large surface area of nLC columns, low-input samples can also be separated by capillary electrophoresis [29]. We sought to minimize nLC losses by mixing labeled peptides from single cells with labeled carrier peptides so that many of the peptides lost due to nLC adhesion will be carrier peptides rather than single-cell peptides. This strategy deviates from standard protocols for bulk LC-MS/MS.

Once injected into an MS instrument, peptide ions need at least two rounds of MS analysis for confident sequence identification [14, 30, 31]. The first MS scan (MS1) determines the mass over charge ratio (M/z) for ions that entered the instrument. Then, selected ions are accumulated and fragmented, and their fragments are analyzed by an MS2 scan [12, 31]. The most commonly used fragmentation methods break peptides at the peptide bonds with efficiency that varies much from bond to bond [31]. Since some fragments are produced with low efficiency, they will not be detected if the peptide ions have low abundance; if not enough fragments are detected, the peptide cannot be sequenced. We sought to alleviate this limitation by sending for MS2 analysis-labeled peptide ions having the same M/z (and thus the same sequence labeled with sample-specific barcodes) from multiple single cells and from carrier cells so that a larger number of peptide ions are fragmented and used for sequence identification. This strategy is built upon the foundational ideas of isobaric tandem mass tags (TMT) [31–33]. TMT labels are used with conventional bulk LC-MS/MS to label samples of equal total protein amount [15, 31, 34] and offer many advantages, albeit quantification can be affected by ion co-isolation [35]; our implementation of TMT, as described below, uses a carrier channel with much higher total protein abundance than the single cells and deviates from the standard protocols.

MS instruments have expanding but limited capacity for parallel ion processing and analysis [12, 36, 37]. Thus increase in throughput has been driven in part by decreasing the time for each step, reaching low millisecond ranges for MS scans and for ion accumulation for bulk

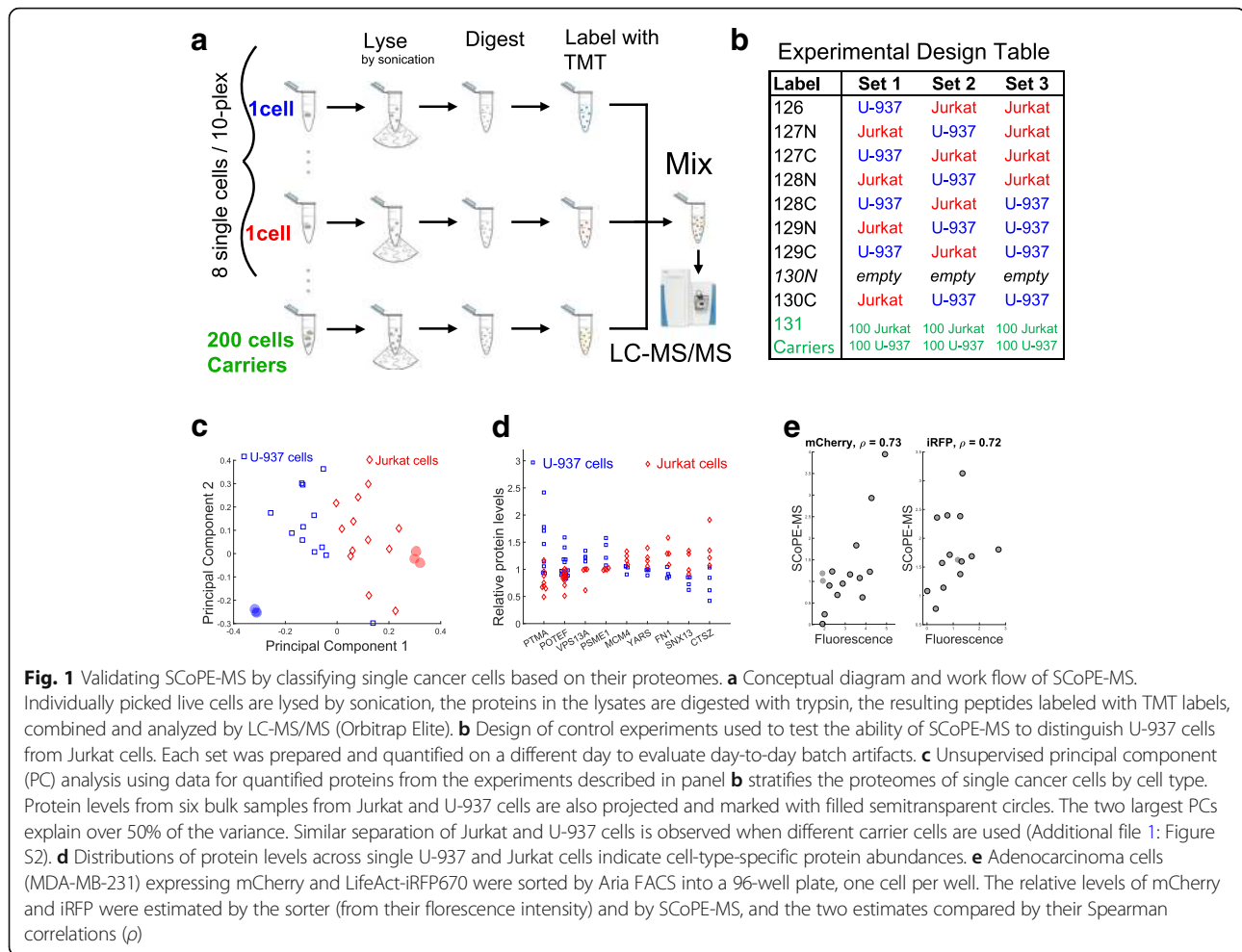
LC-MS/MS analysis [15, 36]. On the other hand, nLC elution peaks have widths on the order of seconds [22, 28]. Thus, if a peptide elutes from the nLC for 8 s and is accumulated (sampled) for only 50 ms by an MS instrument, the instrument will measure only a small fraction of the peptide molecules in the sample [22]. This inefficient sampling is compensated for in standard bulk methods by the large input amount but becomes problematic for low-input samples; counting noise alone can undermine quantification [22]. In this work, we sought to alleviate the sampling limitation by increasing the ion accumulation (sampling) time at the expense of quantifying fewer peptides per unit time. We have discussed additional strategies for increasing sampling and mitigating its trade-offs in a recent perspective [22].

Results

Thus, to develop a high-throughput method for Single Cell Proteomics by Mass Spectrometry (SCoPE-MS), we had to alter substantially the LC-MS/MS methods for bulk samples. In particular, we had to resolve two major challenges: (i) delivering the proteome of a mammalian cell to a MS instrument with minimal protein losses and (ii) simultaneously identifying and quantifying peptides from single-cell samples. To overcome the first challenge, we manually picked live single cells under a microscope and lysed them mechanically (by Covaris sonication in glass microtubes) (Fig. 1a). This method was chosen to obviate chemicals that may undermine peptide separation and ionization or sample cleanup that may incur significant losses. The proteins from each cell lysate were quickly denatured at 90 °C and digested with trypsin at 45 °C overnight (Fig. 1a). Special care was taken to ensure that each tube contained only one cell. See “Methods” for full experimental details.

To overcome the second challenge, we made novel use of tandem mass tags (TMT). This technology was developed for multiplexing [32, 33], which is usually employed for cost-effective increase in throughput. Even more crucial to our application, TMT allows quantifying the level of each TMT-labeled peptide in each sample while identifying its sequence from the total peptide amount pooled across all samples [32, 33]. SCoPE-MS capitalizes on this capability by augmenting each single-cell set with a sample comprised of about 200 carrier cells that provide enough ions for peptide sequence identification (Fig. 1a). The carrier cells also help with the first challenge by reducing losses from single cells, since most of the peptides lost due to surface adhesion will likely originate from the carrier cells. Thus, the introduction of labeled carrier cells into single-cell TMT sets helps overcome the two major challenges.

Quantification of TMT-labeled peptides relies on reporter ions (RI) whose levels reflect both peptide



abundances and noise contributions, such as coisolation interference and background noise [31, 33, 35]. The low protein abundance poses extreme challenges to the signal-to-noise ratio (SNR) and requires careful evaluation even of aspects that are well established and validated in bulk MS measurements. To evaluate the contribution of background noise to single-cell RI quantification, we estimated the signal-to-noise ratio (SNR) (Additional file 1: Figure S1). The estimates indicated that RI intensities are proportional to the amount of labeled single-cell proteomes, and very low for channels left empty. These data suggest that the signal measured in single cells exceeds the background noise by 10-fold or more. As an added SNR control for every TMT set, SCoPE-MS leaves the 130N channel empty, so that 130N RI reflect both isotopic cross-contamination from channel 131 and the background noise. We further verified that RI intensities in a channel are proportional to the protein amount labeled in that channel for both lowly and highly abundant RIs (Additional file 1: Figure S1b, c, d).

To evaluate the ability of SCoPE-MS to distinguish different cell types, we prepared three label-swapped and

interlaced TMT sets with alternating single Jurkat and U-937 cells, two blood cancer cell lines with average cell diameter of only 11 μm (Fig. 1b). The levels of all 767 proteins quantified in single cells were projected onto their principal components (PC) [38, 39]. The two-dimensional projections of single-cell proteomes are clustered by cell type and in proximity to the projection of bulk samples from the same cell type (Fig. 1c), suggesting that SCoPE-MS can identify cell types based on their proteomes. This cell-type stratification is not driven just by highly abundant proteins since the mean levels of each protein across the single cells was set to one; thus, highly and lowly abundant proteins contributed equally to cell clustering. To further test the quantification of cell-type specific protein expression, we identified proteins whose levels vary less within a cell type than between cell types. Based on a two-sample *t*-test, we found 107 proteins showing such trends at FDR < 2%; see representative distributions for such proteins in Fig. 1d.

In Fig. 1, the cell types of the carrier cells and the single cells are matched. If the proteomes of the carrier cells are significantly different from the proteomes of the

single cells, the set of analyzed proteins will change. This is because in shotgun proteomics, peptide ions sent for MS/MS are chosen based on their abundance in the MS1 survey scan. Thus, only peptides with significant abundance in the carrier channel are likely to be sent for MS2 analysis and quantified in the single cells. Therefore, the composition of the carrier channel can affect the sets of peptides quantified across the single cells, i.e., SCoPE-MS samples analyzed by a shotgun method will preferentially provide relative quantification for proteins that are abundant in the carrier cells. However, the relative quantification of a peptide in the single cells, i.e., its RI intensities in the single-cell channels, should not be affected by its abundance in the carrier cells. We tested this expectation with SCoPE-MS sets whose carrier channels contained only U-937 cells, only Jurkat cells, or only HEK-293 cells (Additional file 1: Figure S2). These changes of the carrier cells changed the probability of quantifying some proteins; those with lower abundance in the carrier cells, but hundreds of abundant proteins, were quantified across all cells and carrier channels. Since most proteins have comparable (within an order of magnitude) abundances across different cell and tissue types [16, 40], many cell types can provide useful material for the carrier channel. This carrier dependence can be partially mitigated if SCoPE-MS samples are analyzed by targeted, as opposed to shotgun, LC-MS/MS [22].

Next, we sought to compare SCoPE-MS quantification against an orthogonal and reliable method for quantifying proteins in single cells, the fluorescence of mCherry and iRFP. To this end, the relative levels of the two proteins were quantified in each single cell by a fluorescence-activated cell sorting (FACS) sorter and by SCoPE-MS (Fig. 1e). For both proteins, the Spearman correlations between the SCoPE-MS and FACS measurements exceed 0.7, suggesting that estimates of relative protein levels by SCoPE-MS are comparable to those derived by FACS.

Given the difficulty of measuring extremely low protein levels, we further evaluated SCoPE-MS data by comparing the mean estimates across single cells from Fig. 1b to the corresponding estimates from bulk samples for both integrated precursor-ion-areas (Additional file 1: Figure S3a) and relative (fold-change; Additional file 1: Figure S3b) protein levels. The correlations between bulk and single-cell estimates indicate good agreement despite the noise inherent in single-cell measurements. The relative quantification by SCoPE-MS was further evaluated by correlating protein fold-changes estimated from different pairs of Jurkat and U-937 cells labeled with different TMT tags, demonstrating good consistency of relative quantification for all cells and TMT tags (mean correlation $\rho > 0.5$; Additional file 1: Figure S3c). To eliminate the contribution of biological

variability and estimate the reproducibility of the MS measurement alone, we split a SCoPE-MS set in two and quantified each half separately. Comparisons of corresponding protein ratios estimated from each half indicated reliability between 60 and 85% depending on the magnitude of the fold changes, Additional file 1: Figure S3d. This reliability is achieved with 1/3 of a single-cell proteome (about 100–150 ng of total protein) [19] and compares favorably to reliability for bulk datasets [40]. Taken together, these estimates of quantification accuracy and reproducibility demonstrate that while SCoPE-MS measurements are noisier than bulk MS measurements, they are accurate and reproducible, especially for larger fold-changes.

Protein covariation across differentiating ES cells

Using SCoPE-MS, we quantified single-cell proteome heterogeneity and dynamics during ES cell differentiation. To initiate differentiation, we withdrew leukemia inhibitor factor (LIF) from ES cell cultures and transitioned to suspension culture; LIF withdrawal results in complex and highly heterogeneous differentiation of epiblast lineages in embryoid bodies (EB). We used SCoPE-MS to quantify over a thousand proteins at FDR = 1%, and their pair-wise correlations (averaging across single cells) in days 3, 5, and 8 after LIF withdrawal (Fig. 2a); data are available at MassIVE [38] and at ProteomeXchange [39]. Cells from different days were processed together to minimize batch biases [41]. To explore the protein covariation across the differentiating single cells, we computed and clustered all pairwise protein-protein correlations, Fig. 2a. The clustered correlation matrices exhibit clusters of correlation vectors, and we sought to evaluate their similarity across days. To do so, we computed the correlations between corresponding correlation vectors (i.e., the vector of pair-wise correlations of the i th protein from 1 day was correlated to the i th vector of pair-wise correlations from another day); see ref. [42] for more details. The results shown in Fig. 2b indicate that most correlation vectors from day 3 are positively correlated to the corresponding correlation vectors from days 5 and 8. The corresponding correlation vectors from days 5 and 8 are substantially more similar to each other (Fig. 2b), perhaps reflecting the more advanced differentiation changes on those days.

As cells differentiated and became more distinct from each other, so did the clusters of correlation vectors (Fig. 2a). Gene set enrichment analysis of the clusters indicated that functionally related proteins are over-represented. As expected, proteins forming protein complexes are strongly correlated to each other. For example, most ribosomal proteins (RPs) correlate positively to each other (Fig. 2c). A small subset of RPs covaries as a distinct cluster in the bottom right corner of Fig. 2c,

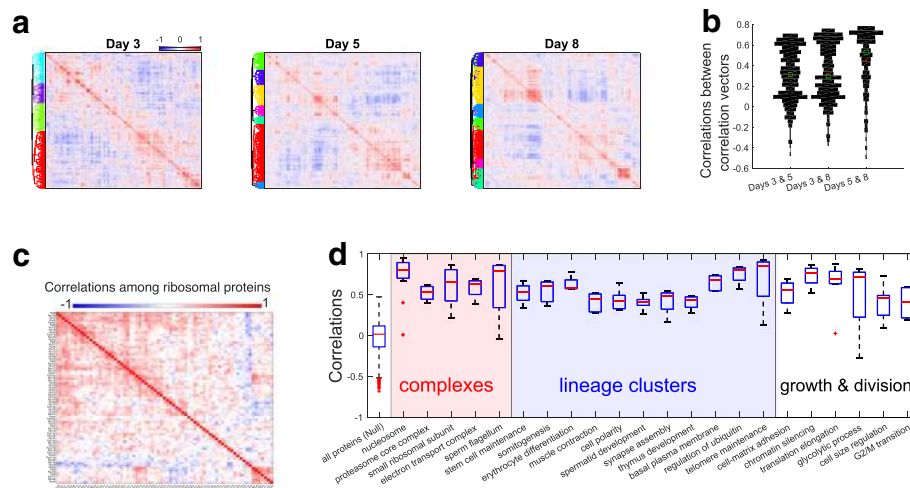


Fig. 2 Identifying protein covariation across differentiating ES cells. **a** Clustergrams of pairwise protein-protein correlations in cells differentiating for 3, 5, and 8 days after LIF withdrawal. The correlation vectors were hierarchically clustered based on the cosine of the angles between them. All single-cell sets used the same carrier channel which was comprised of cells mixed from different time points. **b** The similarity between the correlation matrices shown in panel **a** is quantified by the distribution of correlations between corresponding correlation vectors, as we previously described [42]. Medians are marked with green squares and means with red pluses. **c** All pairwise Pearson correlations between ribosomal proteins (RPs) were computed by averaging across cells. The correlation matrix was clustered, using the cosine between the correlation vectors as a similar measure. **d** To evaluate the similarity in the relative levels of functionally related proteins, we computed the Pearson correlations within sets of functionally related proteins as defined by the gene ontology (GO). These sets included protein complexes, lineage-specific proteins, and proteins functioning in cell growth and division. The distribution of correlations for all quantified proteins is also displayed and used as a null distribution. To remove a positive bias from the null distribution, we subtracted the contribution of the first pair of singular vectors from the matrix of protein levels since this pair often concentrates global effects, which include batch effects and other system-wide trends [42, 54]. The difference between the distributions of correlations for the protein clusters and the null distribution is present in the raw data before this normalization

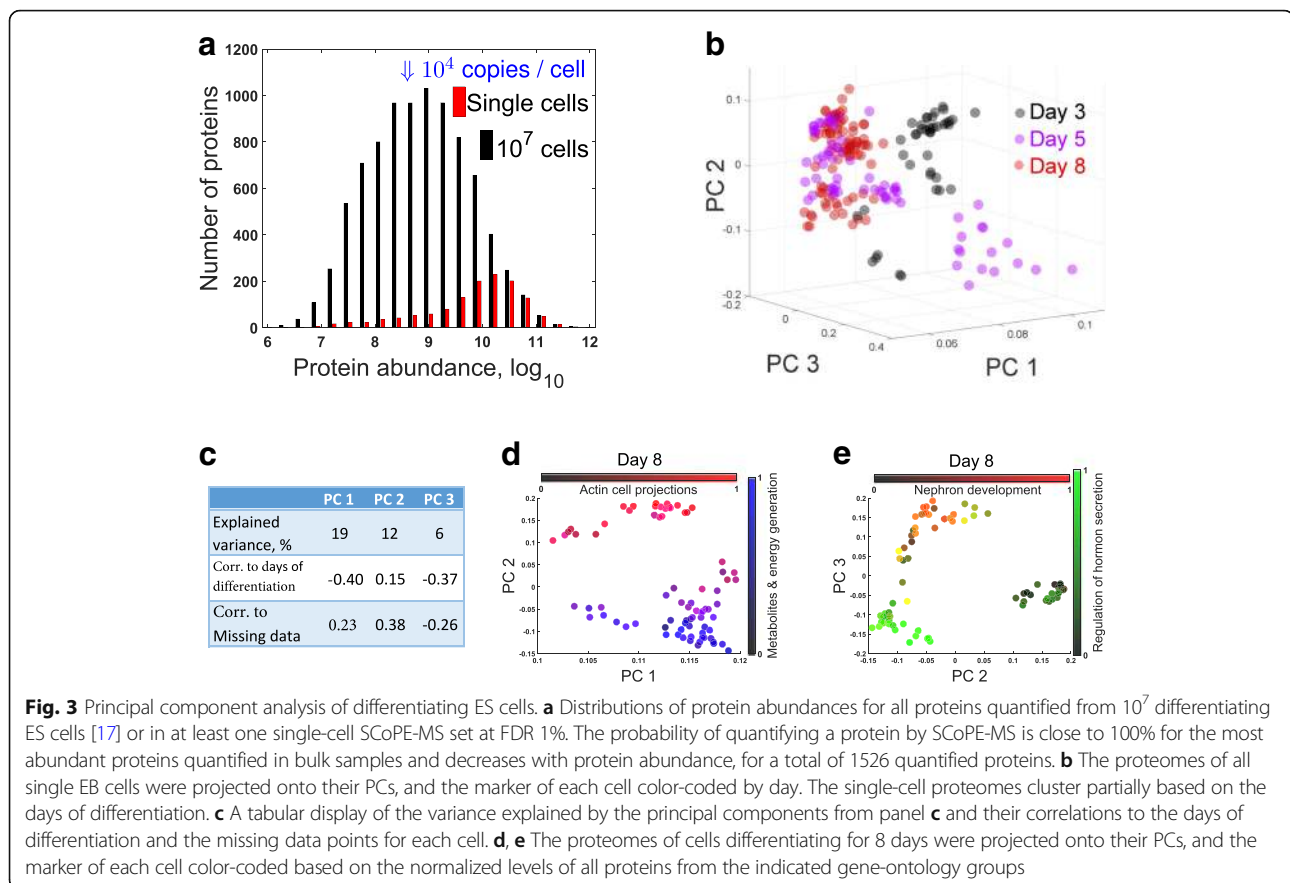
and this might reflect ribosome specialization, i.e., variation among the RP stoichiometry across the cell lineages that contributes to specialized translation functions [43–45]. Alternatively, the cluster might reflect extra-ribosomal functions [46], and these possibilities need to be evaluated more directly with isolated ribosomes [34, 45]. The subunits from other complexes, e.g., the proteasome and the electron transport complex, also covary as indicated by the positive correlations within these complexes (Fig. 2d). A similar pattern of covariation is observed for sets of lineage-specific proteins, including proteins with functions specific to neuronal, blood, and muscle cells (Fig. 2d). Proteins functioning in mRNA translation, metabolism, and cell division also covary, most likely reflecting differences in cell growth and division among the single cells as they differentiate and slow their growth rate.

Principal component analysis of differentiating ES cells

To estimate the abundance of proteins quantified in single cells, we compared the distributions of abundances for over 10,000 proteins quantified in a bulk sample [17] and for the subset of these proteins quantified in SCoPE-MS sets (Fig. 3a). Most of the proteins quantified in the single cells tend to be abundant, mostly above the median of the bulk sets, which corresponds to about

50,000 copies per cell [19]. This is expected given that we used shotgun MS, but combining improvements in SCoPE-MS and targeted MS approaches will enable quantifying substantially less-abundant proteins [22].

Next, we sought to classify single cells using all proteins identified and quantified by SCoPE-MS in single ES and EB cells. We projected the proteomes of single cells from all days (190 cells) onto their PCs (Fig. 3b). The cells partially cluster by time of differentiation; indeed, the loadings of the first three PCs correlate to the days post LIF withdrawal (Fig. 3c). However, the clustering by time of differentiation is incomplete, at least in part because of asynchrony in the differentiation [47]. Similar to single-cell RNA-seq, SCoPE-MS did not quantify each gene in each cell. The number of genes with missing quantification varies from cell to cell for single-cell RNAseq methods and this variation is one of the primary sources of variance in the estimated RNA levels [41]. To test if this is the case for SCoPE-MS, we computed the fraction of proteins with missing data for each cell and correlated that fraction to the PCs. The correlations shown in Fig. 3c suggest that the degree of missing data contributes to the variance but less than what has been described for some RNA datasets [41]. The degree of missing data can be substantially reduced by using targeted MS [22] or its influence mitigated by



simply filtering out the proteins with the most missing data or perhaps by more sophisticated normalization approaches. Since the mechanisms generating missing data differ between RNAseq and SCoPE-MS, we expect that the effects of missing data and their management will be different as well.

The clusters of lineage-specific proteins in Fig. 2 suggest that we have quantified proteomes of distinct cell lineages; thus, we attempted to identify cell clusters by projecting the proteomes of cells from day 8 onto their PCs and identifying sets of proteins that are concertedly regulated in each cluster (Fig. 3d, e). The projection resulted in clusters of cells, whose identity is suggested by the dominant proteins in the singular vectors. We identified biological functions over-represented [40] within the distribution of PC loadings and color-coded each cell based on the average levels of proteins annotated to these functions. These results suggest that SCoPE-MS data can meaningfully classify cell identity for cells from complex and highly heterogeneous populations.

Coordinated mRNA and protein covariation in single cells

Klein et al. [47] recently quantified mRNA heterogeneity during ES differentiation, and we used their inDrop data to simultaneously analyze mRNA and protein covariation

and to directly test whether genes coexpressed at the mRNA level are also coexpressed at the protein level. To this end, we computed all pairwise correlations between RNAs (Fig. 4a) and proteins (Fig. 4b) for all genes quantified at both levels in cells undergoing differentiation for 7 and 8 days. Clustering hierarchically the correlation matrices results in three clusters of genes. To compare these clusters, we computed the pairwise Jaccard coefficients, defined as the number of genes present in both classes divided by the number of genes present in either class, i.e., intersection/union. The results (Fig. 4c) indicate that the largest (green) cluster is 55% identical and the medium (blue) cluster is 33% identical. This cluster stability is also reflected in a positive correlation between corresponding mRNA and protein correlations (Fig. 4d). The magnitude of this correlation is comparable to protein-mRNA correlations from bulk datasets [16, 40] and testifies to the quantitative accuracy of both inDrop and SCoPE-MS.

Having established a good overall concordance between mRNA and protein covariation, we next explored whether and how much this concordance varies between genes with different biological functions. The covariation concordance of a gene was estimated as the similarity of its mRNA and protein correlations, using as a similarity metric the correlation between the corresponding

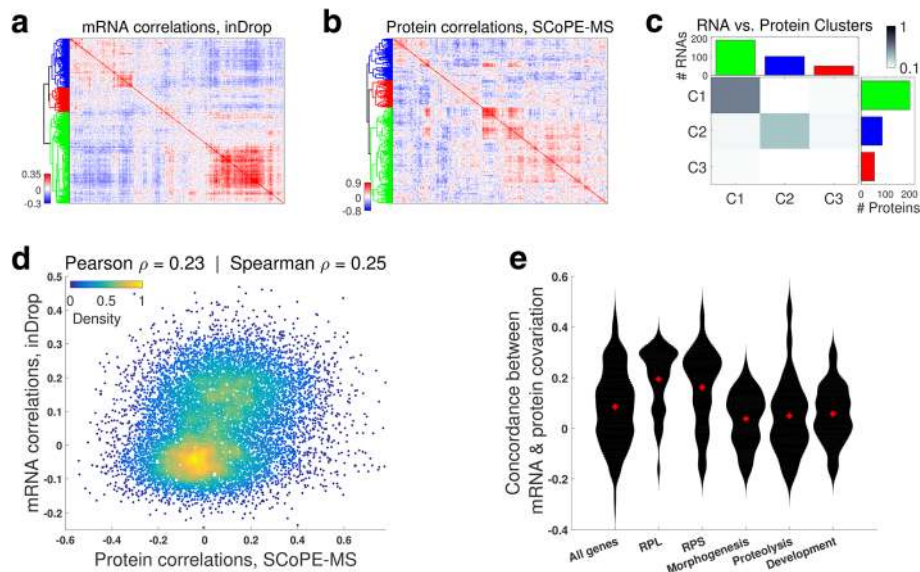


Fig. 4 Coordinated mRNA and protein covariation in differentiating ES cells. **a** Clustergram of pairwise correlations between mRNAs with 2.5 or more reads per cell as quantified by inDrop in single EB cells [47]. **b** Clustergram of pairwise correlations between proteins quantified by SCoPE-MS in 12 or more single EB cells. **c** The overlap between corresponding RNA from **a** and protein clusters from **b** indicates similar clustering patterns. **d** Protein-protein correlations correlate to their corresponding mRNA-mRNA correlations. Only genes with significant mRNA-mRNA correlations were used for this analysis. **e** The concordance between corresponding mRNA and protein correlations (computed as the correlation between corresponding correlations [42]) is high for ribosomal proteins (RPL and RPS) and lower for developmental genes; distribution medians are marked with red pluses. Only the subset of genes quantified at both RNA and protein levels were used for all panels

correlation vectors as we have done previously [42, 48]. The median concordance of ribosomal proteins of both the 60S (RPL) and 40S (RPS) is significantly higher than for all genes (Fig. 4e). This result indicates that RPL and RPS genes have significantly ($p < 10^{-20}$) more similar gene-gene correlations at the mRNA and the protein levels than the other quantified genes. In contrast to RPs, genes functioning in tissue morphogenesis, proteolysis, and development have significantly ($p < 10^{-3}$) lower concordance at the mRNA and protein levels than all genes (Fig. 4e). This difference may reflect both differences in the degree of post-transcriptional regulation or measurement noise for the different sets of genes [40].

Discussion

Until now, the power of LC-MS/MS proteomics has been circumscribed to samples comprised of many cells. Indeed, the TMT manufacturer recommends 100 μ g of protein per channel, almost 10^6 more than the protein content of a typical mammalian cell [19]. SCoPE-MS bridged this gap by clean sample preparation and by introducing TMT-labeled carrier cells. These innovations open the gates to many further improvements (e.g., increased multiplexing) that will make single-cell MS proteomics increasingly powerful [22].

Answering exciting biological questions demands quantifying proteins in many thousands of single cells, and we believe that the ideas described and demonstrated here

will make such throughput practical and affordable [22]. At the moment, the cost per cell is about \$15–30, but it can be reduced to \$1–2 per cell if Covaris tubes are washed and reused and the MS analysis is done on an in-house MS instrument. We expect that automation and improvements in sample preparation as well as increased number of tandem mass tags can reduce the cost well below \$1 per cell. Also, the fraction of missing data can be substantially reduced by using targeted MS approaches [22] and by using retention time (RT) evidence to increase the confidence in correct peptide-spectrum matches [49].

The floor of protein detectability and quantification with SCoPE-MS (as well as any other bottom-up MS method) depends not only on the abundance of a protein but also on its sequence, i.e., the number of peptides produced upon digestion and their propensities to be well separated by the chromatography and efficiently ionized by the electrospray. The implementation of SCoPE-MS in this work allowed us to quantify mostly abundant proteins present at $\geq 10^5$ copies/cell and only a few proteins present at $\geq 10^4$ copies/cell (those producing the most flyable peptides); see the distribution of abundances of the quantified proteins shown in Fig. 3a. However, we are confident that the core ideas underpinning SCoPE-MS can extend the sensitivity to most proteins in a mammalian cell, down to proteins present at ~ 1000 copies/cell. Such extension requires more efficient delivery of proteins to the MS instruments, and we

described specific approaches that can increase the efficiency by orders of magnitude [22]. These approaches include reduced lysis volume and thus protein loss [50], and increased sampling of the elution peaks. Such increased sampling is very practical in the context of SCoPE-MS samples analyzed by MS targeting proteins of interest, e.g., transcription factors. Since proteins are substantially more abundant than mRNAs, estimates of their abundance are less likely to be undermined by sampling (counting) noise. Thus, we believe that building upon this work, future developments in single-cell MS have the potential to accurately quantify most proteins in single mammalian cells, including lowly abundant ones [22].

Conclusion

SCoPE-MS enabled us to classify cells and explore the relationship between mRNA and protein levels in single mammalian cells. This first foray into single mammalian proteomes demonstrates that mRNA covariation is predictive of protein covariation even in single cells. It further establishes the promise of SCoPE-MS to quantitatively characterize single-cell gene regulation and classify cell types based on their proteomes.

Methods

Cell culture

Mouse embryonic stem cells (E14 10th passage) were grown as adherent cultures in 10-cm plates with 10 ml Knockout DMEM media supplemented with 10% ES-certified FBS, nonessential amino acids (NEAA supplement), 2 mM L-glutamine, 110 μ M β -mercapto-ethanol, 1% penicillin and streptomycin, and leukemia inhibitory factor (mLIF; 1000 U LIF/ml). ES cells were passaged every 2 days using StemPro Accutase on gelatin-coated tissue culture plates. ES differentiation was triggered by passaging the ES cells into media lacking mLIF in low adherence plates and growing the cells as suspension cultures. Jurkat and U937 cells were grown as suspension cultures in RPMI medium (HyClone 16777-145) supplemented with 10% FBS and 1% pen/strep. Cells were passaged when a density of 10^6 cells/ml was reached, approximately every 2 days.

Harvesting cells for SCoPE-MS

To harvest cells, embryoid bodies were dissociated by treatment with StemPro Accutase (Thermo Fisher #A1110501) and gentle pipetting. Cell suspensions of differentiating ES cells, Jurkat cells, or U-937 cells were pelleted and washed quickly with cold phosphate buffered saline (PBS). The washed pellets were diluted in PBS at 4 °C. The cell density of each sample was estimated by counting at least 150 cells on a hemocytometer, and an aliquot corresponding to 200 cells was placed in a Covaris microTUBE-15, to be used for

the carrier channel. For picking single cells, two 200- μ l pools of PBS were placed on a cooled glass slide. Into one of the pools, 2 μ l of the cell dilution was placed and mixed, to further dilute the solution. A single cell was then picked under a microscope into a micropipette from this solution. Then, to verify that only one cell was picked, the contents of the micropipette were ejected into the other pool of PBS, inspected, then taken back into the pipette and placed in a chilled Covaris microTUBE-15. Cell samples in Covaris microtubes were frozen as needed before cell lysis and labeling.

Sorting cells by FACS

Adenocarcinoma cells (MDA-MB-231) expressing mCherry and LifeAct-iRFP670 were sorted by Aria FACS into PCR strip-tubes, one cell per tube. Each tube contained 2 μ l of water and had a max volume of 200 μ l. The fluorescence of each protein was measured and the protein abundance estimated after compensation for the spectral overlap between mCherry and iRFP.

Cell lysis and digestion

Each sample—containing a single cell or carrier cells—was lysed by sonication in a Covaris S220 instrument (Woburn, MA) [28]. Samples were sonicated for 180 s at 125 W power with 10% peak duty cycle, in a degassed water bath at 6 °C. During the sonication, samples were shaken to coalesce droplets and bring them down to the bottom of the tube. After lysis, the samples were heated for 15 min at 90 °C to denature proteins. Then, the samples were spun at 3000 rpm for 1 min, and (50 ng/ μ l) trypsin was added: 0.5 μ l to single cells and 1 μ l to carrier cells. The samples were digested overnight, shaking at 45 °C. Once the digest was completed, each sample was labeled with 1 μ l of 85 mM TMT label (TMT10 kit, Thermo Fisher, Germany). The samples were shaken for 1 h in a tray at room temperature. The unreacted TMT label in each sample was quenched with 0.5 μ l of 5% hydroxylamine for 15 min according to the manufacturer's protocol. The samples corresponding to one TMT10 plex were then mixed in a single-glass HPLC vial and dried down to 10 μ l in a speed vacuum (Eppendorf, Germany) at 35 °C.

Bulk set

The six bulk samples of Jurkat and U-937 cells contained 2500 cells per sample. The cells were harvested, lysed, and processed using the same procedure as for the single cells but with increased amount of trypsin and TMT labels. The samples were labeled, mixed, and run as a 6-plex TMT set.

Mass spectrometry analysis

Each TMT labeled set of samples was submitted for single LC-MS/MS experiment that was performed on a LTQ Orbitrap Elite (Thermo Fisher) equipped with a Waters (Milford, MA) NanoAcquity HPLC pump. Peptides were first trapped and washed onto a 5 cm × 150 μm inner diameter microcapillary trapping column packed with C18 Reprosil resin (5 μm, 10 nm, Dr. Maisch GmbH, Germany). The peptides were separated on an analytical column 20 cm × 75 μm of C18 TPP beads (1.8 μm, 20 nm, Waters, Milford, MA) that was heated to 60 °C. Separation was achieved through applying an active gradient from 7 to 27% ACN in 0.1% formic acid over 170 min at 200 nl/min. The active gradient was followed by a 10-min 27–97% ACN wash step. Electrospray ionization was enabled through applying a voltage of 1.8 kV using a homemade electrode junction at the end of the microcapillary column and sprayed from fused silica pico-tips (20 μm ID, 15 μm tip end New Objective, MA). The LTQ Orbitrap Elite was operated in data-dependent mode for the mass spectrometry methods. The mass spectrometry survey scan (MS1) was performed in the Orbitrap in the range of 395–1,800 *m/z* at a resolution of 6×10^4 , followed by the selection of up to 20 most intense ions (TOP20) for HCD-MS2 fragmentation in the Orbitrap using the following parameters: precursor isolation width window of 1 or 2 Th, AGC setting of 100,000, a maximum ion accumulation time of 150 ms or 250 ms, and 6×10^4 resolving power. Singly charged and 4+ charge ion species were excluded from HCD fragmentation. Normalized collision energy was set to 37 V and an activation time of 1 ms. Ions in a 7.5-ppm *m/z* window around ions selected for MS2 were excluded from further selection for fragmentation for 20 s.

Analysis of raw MS data

Raw data were searched by MaxQuant [14, 51] 1.5.7.0 against a protein sequence database including all entries from a SwissProt database and known contaminants such as human keratins and common lab contaminants. The SwissProt databases were the human SwissProt database for the U-937 and the Jurkat cells and the mouse SwissProt database for the differentiating ES cells. MaxQuant searches were performed using the standard work flow [52]. We specified trypsin specificity and allowed for up to two missed cleavages for peptides having from 5 to 26 amino acids. Methionine oxidation (+ 15.99492 Da) was set as a variable modification. All peptide-spectrum matches (PSMs) and peptides found by MaxQuant were exported in the msms.txt and the evidence.txt files.

In addition to a standard search with the full SwissProt databases, we also searched the MS data with custom

sequence databases since such searches have advantages when the sequences can be better tailored to the peptides analyzed by MS [18, 53]. In the case of SCoPE-MS, we can remove sequences for lowly abundant proteins since their peptides are very unlikely to be sent for MS2. Indeed, searches with the full databases did not identify peptides from the least abundant proteins Fig. 2a. Excluding such proteins from the search can narrow down the search space and increase the statistical power for identifying the correct peptide-spectrum matches [18, 53]. To take advantage of this approach, we searched the MS data with custom databases comprised from all proteins for which MaxQuant had identified at least one peptide across many single-cell and small-bulk sets in searches against the full SwissProt databases. These reduced fasta databases contained 5267 proteins for mouse and 4961 proteins for human. Searches with them slightly increased the number of identified peptides from SCoPE-MS sets but such customized databases are not essential for SCoPE-MS.

The shotgun approach results in identifying different peptides in different SCoPE-MS sets at different levels of confidence. Because of the lower protein levels in SCoPE-MS sets compared to bulk sets, fewer fragment ions are detected in the MS2 spectra and thus peptide identification is more challenging than with bulk datasets. As a result, the 1% FDR threshold that is optimal for bulk MS data may not be optimal for SCoPE-MS datasets. To determine the FDR threshold that is optimal for single-cell data, we plotted the number of identified peptides at all levels of posterior error probability (PEP) (Additional file 1: Figure S4a). This analysis suggests that a slight increase in the arbitrary FDR threshold or 1% results in a significant increase in the peptides that can be usefully analyzed across single cells while still keeping false positives low. Thus, peptides from SCoPE-MS sets were filtered to 3% FDR computed as the mean of the PEP of all peptides below the PEP cutoff threshold [40]. To validate the protein ratios derived from peptides having $PEP \in (0.01, 0.03]$, we correlated them to the corresponding protein ratios derived from peptides having $PEP < 0.01$ (Additional file 1: Figure S4b). The positive correlation in Additional file 1: Figure S4b indicates that peptides identified with lower confidence carry quantitative information. Still, this correlation is lower than the correlations between ratios derived from two subsets of peptides having $PEP < 0.01$ (Additional file 1: Figure S4c). This may be due at least in part to the fact that factors reducing the confidence of identification, such as lower abundance or higher co-isolation, are also likely to undermine quantification. All razor peptides were used for quantifying the proteins to which they were assigned by MaxQuant. The average number of identified peptides per TMT set is shown for a few SCoPE-MS sets in

Additional file 1: Figure S4a as a technical benchmark but it has much less practical significance than the number of proteins that are quantified across enough single cells to be useful for analysis [22]. This number of genes quantified across multiple sets is the standard measure for single-cell RNA sequencing methods [47], and we have adopted it for SCoPE-MS as the more meaningful measure of the proteins whose levels can be analyzed across multiple single cells.

Data analysis

We estimated relative peptide/protein levels from the TMT reporter ions (RI), and protein abundances from the precursor areas distributed according to the RI levels. While such estimates are well validated with bulk samples, extremely low input amounts pose unique challenges that may result in artifacts, e.g., RI intensities may reflect only background noise or the isotopic impurities of TMT tags may cross contaminate TMT channels. We evaluated the degree of background noise and found it significantly below the signal coming from the labeled peptides (see Additional file 1: Figure S1). To compensate for different amounts of total protein per channel or other channel-specific variability, the median RI intensities in each channel was set to one by dividing all RI intensities by their median. In the FACS experiment, the normalization for mCherry was performed using iRFP as a control, analogous to loading controls in western blots. After this column normalization, the vector of RI intensities for each peptide was divided by its mean or median, to remove the large differences in abundances between different peptides. The relative level of each quantified razor protein was estimated as the median of the relative levels of its peptides. All analysis relied on relative levels, i.e., the level of protein in a cell relative to its mean or median level across all cells in which the protein is quantified. Missing peptide and protein levels were imputed using the k -nearest neighbors algorithm, with k being set to 1 and the similarity measure for distance being the cosine of the angle between the proteome vectors.

Relative quantification across SCoPE-MS sets

SCoPE-MS allows quantifying only eight cells per set (Fig. 1), but combining multiple sets can quantify the proteomes of hundreds and thousands of cells. We were able to successfully combine relative protein levels across SCoPE-MS sets in two different ways: (i) When the carrier material used across sets is the same (Fig. 1b), we used the carrier channel as a reference as established with bulk TMT samples [15]. (ii) When the carrier material differed across carrier channels (Additional file 1: Figure S2), we excluded the carrier channel from the analysis and normalized the relative levels of each peptide to a mean 1 across the eight single cells in each set,

four Jurkat and four U-937 cells. Approach (ii) worked well in this case because the single-cell composition of the different SCoPE-MS sets was balanced. Combining SCoPE-MS sets based on a reference channel that is kept the same across all sets is a more versatile strategy that generalizes to any experimental design and single-cell distribution across sets.

Additional file

Additional file 1: Figure S1. Contribution of background noise to quantification of peptides in single cells. **Figure S2.** Relative quantification is independent from the carrier channel. **Figure S3.** Accuracy of SCoPE-MS quantification. **Figure S4.** Confidence of peptide identification and its effect on quantification. (PDF 805 kb)

Acknowledgements

We thank R. G. Huffman, S. Semrau, R. Zubarev, J. Neveu, H. Specht, A. Phillips, and A. Chen for assistance, discussions, and constructive comments, as well as the Harvard University FAS Science Operations for supporting this research project. We thank A. Raj and B. Emert for the gift of MDA-MB-231 cell line expressing mCherry and LifeAct-iRFP670.

Funding

This work was funded by startup funds from Northeastern University and a New Innovator Award from the NIGMS from the National Institutes of Health to N.S. under Award Number DP2GM123497. Funding bodies had no role in data collection, analysis, and interpretation.

Availability of data and materials

The raw MS data and the search results were deposited in MassIVE (ID: MSV000082077) [38] and in ProteomeXchange (ID: PXD008985) [39]. Supplemental website can be found at: northeastern.edu/slavovlab/2016_SCoPE-MS/.

Authors' contributions

BB, EL, GH, and NS performed experiments and collected data; NS supervised the research, analyzed the data, and wrote and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Ethics approvals were not needed for the study.

Consent for publication

Not applicable

Competing interests

The authors have filed a provisional patent on the entire protocol, application number: 62/618,301. The method is freely available for research and all other non-commercial purposes.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹MSPRL, FAS Division of Science, Harvard University, Cambridge, MA 02138, USA. ²Department of Biology, Northeastern University, Boston, MA 02115, USA. ³Department of Bioengineering, Northeastern University, Boston, MA 02115, USA.

Received: 20 February 2018 Accepted: 19 September 2018

Published online: 22 October 2018

References

- Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. *Nat Rev Cancer*. 2005;5(4):275–84.

2. Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, et al. Dynamic proteomics of individual cancer cells in response to a drug. *Science*. 2008;322(5907):1511–6.
3. Semrau S, van Oudenaarden A. Studying lineage decision-making in vitro: emerging concepts and novel tools. *Annu Rev Cell Dev Biol*. 2015;31:317–45.
4. Symmons O, Raj A. Whats luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Mol Cell*. 2016;62(5):788–802.
5. Levy E, Slavov N. Single cell protein analysis for systems biology. *Essays Biochem*. 2018;62. <https://doi.org/10.1042/EBC20180014>.
6. Landgraf D, Okumus B, Chien P, Baker TA, Paulsson J. Segregation of molecules at cell division reveals native protein localization. *Nat Methods*. 2012;9(5):480–2.
7. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem*. 2009;81(16):6813–22.
8. Bendall SC, Simonds EF, Qiu P, El-ad DA, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. 2011;332(6030):687–96.
9. Hughes AJ, Spelke DP, Xu Z, Kang CC, Schaffer DV, Herr AE. Single-cell western blotting. *Nat Methods*. 2014;11(7):749–55.
10. Darmanis S, Gallant CJ, Marinescu VD, Niklasson M, Segerman A, Flavourakis G, et al. Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep*. 2016;14(2):380–9.
11. Marcon E, Jain H, Bhattacharya A, Guo H, Phanse S, Pu S, et al. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat Methods*. 2015;12(8):725.
12. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198–207. <https://doi.org/10.1038/nature01511>.
13. De Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Fröhlich F, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*. 2008;455(7217):1251.
14. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367–72.
15. Slavov N, Budnik B, Schwab D, Airolidi E, van Oudenaarden A. Constant growth rate can be supported by decreasing energy flux and increasing aerobic glycolysis. *Cell Rep*. 2014;7:705–14.
16. Wilhelm M, Schlegl J, Hahne H, Gholami A, Lieberenz M, et al. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014;509:582–7.
17. van den Berg PR, Budnik B, Slavov N, Semrau S. Dynamic post-transcriptional regulation during embryonic stem cell differentiation. *bioRxiv*. 2017;1. <https://doi.org/10.1101/123497>.
18. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods*. 2014;11(11):1114.
19. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res*. 2010;38:D750–3.
20. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Corrigendum: global quantification of mammalian gene expression control. *Nature*. 2013;495:126127.
21. Zubarev RA, Makarov A. Orbitrap mass spectrometry. *Anal Chem*. 2013;85(11):5288–96.
22. Specht H, Slavov N. Transformative opportunities for single-cell proteomics. *J Proteome Res*. 2018;17:2563–916. <https://doi.org/10.1021/acs.jproteome.8b00257>.
23. Dhabaria A, Cifani P, Reed C, Steen H, Kentsis A. A high-efficiency cellular extraction system for biological proteomics. *J Proteome Res*. 2015;14(8):3403–8.
24. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsvelde J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol*. 2014;10(10):757.
25. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods*. 2014;11(3):319.
26. Sielaff M, Kuharev J, Bohn T, Hahlbrock J, Bopp T, Tenzer S, et al. Evaluation of FASP, SP3, and iST protocols for proteomic sample preparation in the low microgram range. *J Proteome Res*. 2017;16(11):4060–72.
27. Virant-Klun I, Leicht S, Hughes C, Krijgsvelde J. Identification of maturation-specific proteins by single-cell proteomics of human oocytes. *Mol Cell Proteomics*. 2016;15(8):2616–27.
28. Li S, Plouffe BD, Belov AM, Ray S, Wang X, Murthy SK, et al. An integrated platform for isolation, processing, and mass spectrometry-based proteomic profiling of rare cells in whole blood. *Mol Cell Proteomics*. 2015;14(6):1672–83.
29. Lombard-Banek C, Moody SA, Nemes P. Single-cell mass spectrometry for discovery proteomics: quantifying translational cell heterogeneity in the 16-cell frog (*Xenopus*) embryo. *Angew Chem Int Ed*. 2016;55(7):2454–8.
30. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*. 1994;5(11):976–89.
31. Sinitcyn P, Rudolph JD, Cox J. Computational methods for understanding mass spectrometry-based shotgun proteomics data. *Annu Rev Biomed Data Sci*. 2018;1:207–34.
32. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*. 2003;75(8):1895–904.
33. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004;3(12):1154–69.
34. Slavov N, Semrau S, Airolidi E, Budnik B, van Oudenaarden A. Differential stoichiometry among core ribosomal proteins. *Cell Rep* 2015;13:865 – 873.
35. Savitski MM, Mathieson T, Zinn N, Sweetman G, Doce C, Becher I, et al. Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *J Proteome Res*. 2013;12(8):3586–98.
36. Michalski A, Damoc E, Hauschild JP, Lange O, Wiegand A, Makarov A, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics*. 2011;10(9):M111–011015.
37. Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, et al. Parallel accumulation–serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J Proteome Res*. 2015;14(12):5378–87.
38. Budnik B, Levy E, Harmange G, Slavov N. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *MassIVE*, MSV000082077. <https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=4f30cbe81fc440f79bd73f6c27f1816b> (2018).
39. Budnik B, Levy E, Harmange G, Slavov N. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *ProteomeXchange*, PXD008985. <http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX008985> (2018).
40. Franks A, Airolidi E, Slavov N. Post-transcriptional regulation across human tissues. *PLoS Comput Biol*. 2017;13(5):e1005535.
41. Hicks SC, Teng M, Irizarry RA. On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-Seq data. *bioRxiv*. 2015;1:025528.
42. Slavov N, Dawson KA. Correlation signature of the macroscopic states of the gene regulatory network in cancer. *Proc Natl Acad Sci*. 2009;106(11):4079–84. <https://doi.org/10.1073/pnas.0810803106>.
43. Mauro VP, Edelman GM. The ribosome filter hypothesis. *Proc Natl Acad Sci*. 2002;99(19):12031–6.
44. Preiss T. All ribosomes are created equal. Really? *Trends Biochem Sci*. 2015;41(2). <https://doi.org/10.1016/j.tibs.2015.11.009>.
45. Emmott EP, Jovanovic M, Slavov N. Ribosome stoichiometry: from form to function. *Peer J Preprints*. 2018;6:e26991v1. <https://doi.org/10.7287/peerj.preprints.26991v1>.
46. Wool IG. Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci*. 1996;21(5):164–5.
47. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015;161(5):1187–201.
48. Silverman SJ, Petti AA, Slavov N, Parsons L, Briehof R, Thiberge SY, et al. Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. *Proc Natl Acad Sci*. 2010;107(15):6946–51.
49. Chen A, Franks A, Slavov N. DART-ID increases single-cell proteome coverage. *bioRxiv*. 2018. <https://doi.org/10.1101/399121>.
50. Specht H, Harmange G, Perlman DH, Emmott E, Niziolek Z, Budnik B, Slavov N. Automated sample preparation for high-throughput single-cell proteomics. *bioRxiv*. 2018. <https://doi.org/10.1101/399774>.
51. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794–805.

52. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc.* 2016;11(12):2301.
53. Woo S, Cha SW, Merrihew G, He Y, Castellana N, Guest C, et al. Proteogenomic database construction driven from large scale RNA-seq data. *J Proteome Res.* 2013;13(1):21–8.
54. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. Random matrix approach to cross correlations in financial data. *Phys Rev E.* 2002;65(6):066126.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

