Single-cell mass-spectrometry quantifies the emergence of macrophage heterogeneity

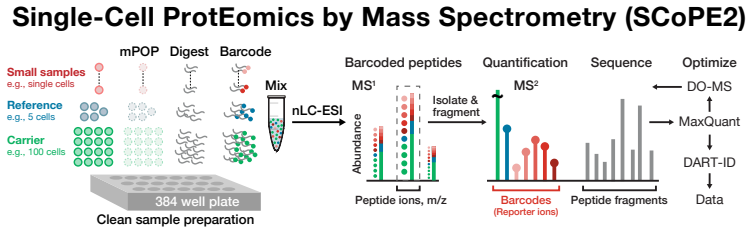
Harrison Specht,[1] Edward Emmott,[1,2] Aleksandra A. Petelski,[1] R. Gray Huffman,[1] David H. Perlman,[1,3] Marco Serra,[4] Peter Kharchenko,[4] Antonius Koller,[1] Nikolai Slavov[1,✉]

[1]Department of Bioengineering and Barnett Institute, Northeastern University, Boston, MA 02115, USA
[2]Current address: Centre for Proteome Research, Department of Biochemistry, University of Liverpool, Liverpool, L69 7ZB, UK
[3]Current address: Merck Exploratory Sciences Center, Merck Sharp & Dohme Corp., 320 Bent St. Cambridge, MA 02141
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

✉ Correspondence: nslavov@alum.mit.edu
∈ Data, code & protocols: scope2.slavovlab.net

Abstract

The fate and physiology of individual cells are controlled by protein interactions. Yet, our ability to quantitatively analyze proteins in single cells has remained limited. To overcome this barrier, we developed SCoPE2. It lowers cost and hands-on time by introducing automated and miniaturized sample preparation while substantially increasing quantitative accuracy. These advances enabled us to analyze the emergence of cellular heterogeneity as homogeneous monocytes differentiated into macrophage-like cells in the absence of polarizing cytokines. SCoPE2 quantified over 2,700 proteins in 1,018 single monocytes and macrophages in ten days of instrument time, and the quantified proteins allowed us to discern single cells by cell type. Furthermore, the data uncovered a continuous gradient of proteome states for the macrophage-like cells, suggesting that macrophage heterogeneity may emerge even in the absence of polarizing cytokines. Parallel measurements of transcripts by 10x Genomics scRNA-seq suggest that SCoPE2 samples 20-fold more copies per gene, thus supporting quantification with improved count statistics. Joint analysis of the data indicated that most genes had similar responses at the protein and RNA levels, though the responses of hundreds of genes differed. Our methodology lays the foundation for automated and quantitative single-cell analysis of proteins by mass-spectrometry.



Introduction

Tissues and organs are composed of functionally specialized cells. This specialization of single cells often arises from the protein networks mediating physiological functions. Yet, our ability to comprehensively quantify the proteins comprising these networks in single-cells has remained relatively limited.[1] As a result, the protein levels in single cells are often inferred from indirect surrogates – sequence reads from their corresponding mRNAs.[2–4]

Single-cell RNA sequencing methods have illuminated cellular types and states comprising complex biological tissues, aided the discovery of new cell types, and empowered the analysis of spatial organization.[2,5] These methods depend on the ability to capture and detect a representative set of cellular transcripts. Many transcripts are present at low copy numbers, and with the existing scRNA-seq protocols capturing around 10-20% of molecules in a cell, the resulting sampling is very sparse for many transcripts. Because of this, the estimates of mRNA abundances are notably affected by sampling (counting) errors.[2,3] THus, even with an ideal scRNA-seq protocol, the estimates of abundance for many transcripts would be limited by stochastic effects.[6]

Sampling many protein copies per gene may be feasible since most proteins are present at over 1,000-fold more copies per cell than their corresponding transcripts.[1,6,7] This high abundance also obviates the need for amplification. Since amplification may introduce noise, obviating amplification is a desirable aspect. Thus, the high copy number of proteins may allow their quantification without amplification.

However most technologies for quantifying proteins in single cells rely on antibodies, which afford only limited specificity.[1] Although tandem mass spectrometry(MS/MS) combined with liquid chromatography (LC) and electrospray ionization (ESI) has enabled accurate, high-specificity and high-throughput quantification of proteins from bulk samples,[8–11] its application to single cells is in its infancy.[1,6] To apply these powerful technologies to the analysis of single cells, we developed Single-Cell ProtEomics by Mass Spectrometry (SCoPE-MS).[6,12] SCoPE-MS introduced the concept of using carrier proteins barcoded with tandem-mass-tags (TMT), which serves three important roles: (i) reducing sample loss, (ii) enhancing the detectability of ions during MS1 survey scans and (iii) providing fragment ions for peptide sequence identification. By combining this concept with MS-compatible cell lysis, we established the feasibility of applying multiplexed LC-ESI-MS/MS to quantify proteins from single cells.

While SCoPE-MS and its ideas have been reproduced and adopted by others[13–18], the cost, throughput, and reliability of the data fall short of our vision of single-cell proteomics.[6] Our vision requires quantifying thousands of proteins and proteoforms across thousands of single cells at

an affordable cost and within a reasonable time frame. Such data could support clinical applications such as biomarker discovery. Moreover, these data could permit inferring direct causal mechanisms underlying the functions of protein networks.[6] The more cells and proteoforms are quantified, the fewer assumptions are needed for this analysis. Thus, our goal in developing SCoPE2 was to increase the number of cells and proteins analyzed at affordable cost while sampling a sufficient number of ion copies per protein to make quantitative measurements.

To achieve this goal, we followed previously outlined opportunities.[6] In particular, we overhauled multiple experimental steps, including cell isolation, lysis, and sample preparation.[4,19] Furthermore, we developed methods for optimizing the acquisition of MS data (DO–MS; Data-driven Optimization of MS)[20] and for interpreting these data once acquired, e.g., for enhancing peptide identification (DART-ID; Data-driven Alignment of Retention Times for IDentification).[21] These advances combine synergistically into a next-generation SCoPE-MS version, SCoPE2, that affords substantially improved quantification and throughput.

SCoPE2 enabled us to ask fundamental questions: do homogeneous monocytes produce homogeneous macrophages in the absence of polarizing cytokines? Are macrophages inherently prone to be heterogeneous, or is their heterogeneity simply reflecting different progenitors and polarizations induced by different cytokines? These questions are cornerstones to our understanding of macrophage heterogeneity that plays important roles in human pathophysiology: depending on their polarization, macrophages can play pro-inflammatory (usually ascribed to M1 polarization) or anti-inflammatory roles (usually ascribed to M2 polarization), and be involved in tissue development and maintenance.[22] Some studies suggest that rather than separating into discrete functional classes, the M1 and M2 states represent the extremes of a wider spectrum of end-states.[22,23] We found that the individual macrophage-like cells were highly heterogeneous even though they originated from homogeneous monocytes exposed to identical environmental conditions.

Results

The overall work-flow of SCoPE2 is illustrated in Fig. 1a. Single cells are isolated in individual wells, lysed, and their proteins digested to peptides. The peptides from each single cell are covalently labeled (barcoded) with isobaric tandem-mass-tags (TMT), and therefore labeled peptides with the same sequence (and thus mass) appear as a single mass/charge cluster in the MS1 scans. The MS instrument isolates such clusters and fragments them, see Box 1. In addition to generating peptide fragments which facilitate peptide identification, fragmentation generates reporter ions (RI),

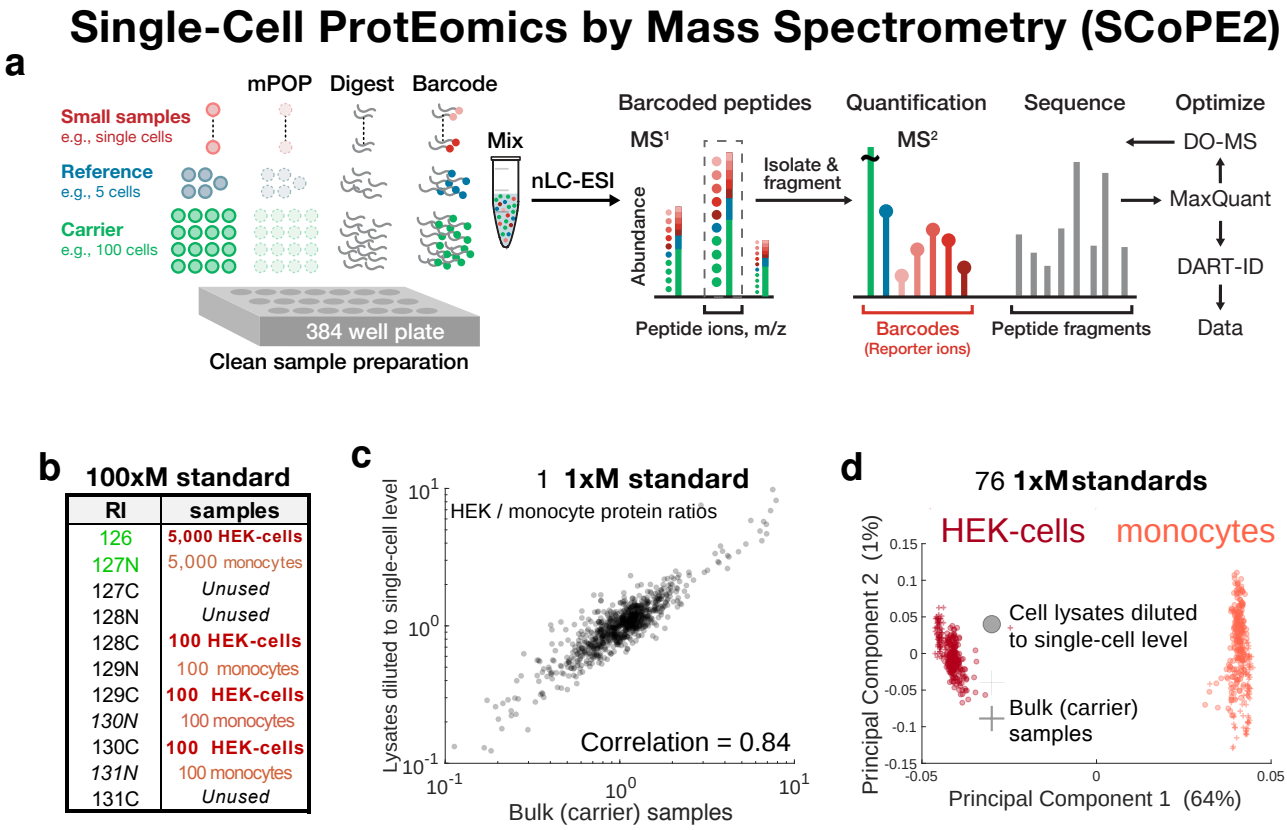


Figure 1 | Developing and benchmarking SCoPE2 with standards
(**a**) Conceptual diagram and work flow of SCoPE2. Cells are sorted into multiwell plates and lysed by mPOP[19]. The proteins in the lysates are digested with trypsin, the resulting peptides labeled with TMT, combined, and analyzed by LC-MS/MS. SCoPE2 sets contain reference channels that allow merging single-cells from different SCoPE2 sets into a single dataset. The LC-MS/MS analysis is optimized by DO-MS,[20] and peptide identification enhanced by DART-ID.[21] (**b**) Schematic for the design of a 100xM standard set. Monocytes (U937 cells) and embryonic kidney cells (HEK-293) were serially diluted to the indicated numbers, lysed, digested, and labeled with tandem-mass-tags having the indicated reporter ions (RI). (**c**) Comparison of protein fold-change between the embryonic kidney cells and monocytes estimated from the small-samples and from the carrier samples of a 1xM standard, i.e., 1 % sample from the 100xM standard described in panel a. The relative protein levels measured from samples diluted to single-cell levels are very similar to the corresponding estimates from the carrier (bulk) samples. (**d**) Principal component analysis separates samples corresponding to embryonic kidney cells (HEK-293) or to monocytes (U-937 cells). The small samples (which correspond to 100 cells diluted 100-fold to single-cell level) cluster with the carrier samples, indicating that relative protein quantification from all samples is consistent and based on cell type. All quantified proteins were used for this analysis and each protein was normalized separately for the carrier channels and the small sample channels.

whose abundances reflect protein abundances in the corresponding samples (single cells), Fig. 1a. Key advances of SCoPE2 over SCoPE-MS include:

Automated and miniaturized cell lysis. Instead of lysing cells by focused acoustic sonication, SCoPE2 lyses cells by Minimal ProteOmic sample Preparation (mPOP).[19] mPOP uses a freeze-heat cycle that extracts proteins efficiently in pure water, thus obviating cleanup before MS analysis. mPOP allows sample preparation in multiwell plates, which enables simultaneous processing of many samples in parallel with inexpensive PCR thermocyclers and liquid dispensers. This advance over SCoPE-MS allows SCoPE2 to decrease lysis volumes 10-fold, from $10\mu l$ to $1\mu l$, to reduce the cost of consumables and equipment over 100-fold, and to increase throughput of sample preparation over 100-fold by parallel processing.

Improved data integration across single cells. SCoPE2 also introduces a reference channel composed of a reference sample used in all sets. This reference is about 5-fold more abundant than a small sample channel (i.e., equivalent to a single cell proteome) so that the higher abundance results in improved ion-counting statistics while remaining comparable to that of single cells, and thus likely to be within the linear range of quantification

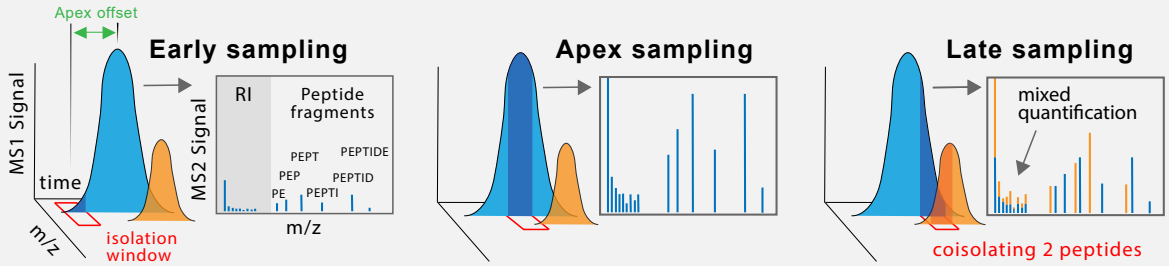
Increased throughput and improved quantification. SCoPE2 introduces shorter nLC gradient, which allows analyzing more cells per unit time. Furthermore, improved peptide separation and a narrower isolation window (0.7 Th) result in improved ion isolation and thus improved quantification, see Box 1.

Systematic parameter optimization. The sensitivity and quantification of LC-MS/MS experiments depends on numerous instrument parameters whose tuning often depends on trial-and-error approaches.[20] For example, tuned parameters may increase the number of chromatographic peaks sampled at their apexes, thus increasing the copies of peptide ions used for quantification, see Box 1. To make such optimizations more systematic, SCoPE2 employs DO-MS to interactively display data required for the rational optimization of instrument parameters and performance.

Enhanced peptide sequence identification. Once the MS data are acquired, SCoPE2 can use additional features of the data to enhance their interpretation. For example, SCoPE2 uses a principled Bayesian framework (DART-ID) to incorporate retention time information for increasing the confidence of assigning peptide sequences to MS spectra while rigorously estimating the false discovery rate.[21]

These advances work synergistically to enhance the ability of SCoPE2 to quantify proteins in single cells. Below, we exemplify the improvements, starting with the application of DO-MS.

Box 1: Optimizing isobaric-mass-tags quantification by SCoPE2



During LC-MS/MS, each labeled peptide elutes from the chromatographic column as an elution peak over a time period typically ranging from 10 to 40 seconds (5 - 20 seconds at mid-height) while its ions are isolated (sampled) for MS2 analysis over much shorter intervals, typically ranging from 5 to 80 milliseconds. If the elution peak is sampled too early (left panel) or too late (right panel), the fraction of the peptide ions used for quantification and sequence identification is smaller compared to sampling the apex (middle panel). To increase the fraction of sampled ions per peptide, we used DO-MS to increase the probability of sampling the apex (Fig. 2c), decreased the elution peak width (see methods), and increased the MS2 fill time to 300 milliseconds.

SCoPE2 quantifies peptides sequentially, one peptide at a time. For each analyzed peptide, the MS instrument aims to isolate only ions from the peptide by applying a narrow mass filter (m/z isolation window) denoted by a red rectangle in the sketch above. Yet, ions from other peptides might also fall within that window and thus become coisolated, as shown with the blue and orange peptides in the third panel. Since coisolated peptides contribute to the measured reporter ions (RI), coisolation reduces the accuracy of quantification. To minimize coisolation, we reduced the isolation windows to 0.7 Th and improved apex targeting as described above. The success of these optimizations was evaluated by the Precursor Ion Fraction (PIF), a benchmark computed by MaxQuant as an estimate for the purity of the ions isolated for fragmentation and MS2 analysis, Fig. 2c,

Optimizing SCoPE2 with standards

The quality of LC-MS/MS data strongly depends on numerous interdependent parameters (e.g., chromatographic packing, LC gradient steepness, and ion accumulation time), see Box 1. To optimize such parameters, we applied DO-MS on replicate samples, termed "master standards". Each 1xM standard is a 1% injection of a bulk sample (100xM) composed of eight barcoded samples: two 5,000-cell carrier samples (one for each cell type) and six 100-cell samples (three for each cell type), as shown in Fig. 1b. Each $1\mu l$ injection of this bulk sample constitutes a 1xM standard which contains peptide input equivalent to 50-cells in each carrier channel, with the remaining six channels each containing peptide input equivalent to a single cell. Thus, the 1xM standards approximate idealized SCoPE2 sets that enabled us to focus on optimizing LC-MS/MS parameters using identical samples, i.e., independently of the biological variability between single cells.

First, we optimized our analytical column configuration and LC gradient settings. Each 1xM sample was analyzed for only 60 minutes since our goal was to optimize the number of proteins quantified across many cells, rather than merely the number of proteins quantified per sample.[6] By varying chromatographic parameters and benchmarking their effects with DO-MS, we minimized elution peak widths. Sharper elution peaks increase the sampled copies of each peptide per unit time and reduce the probability that multiple peptides are simultaneously isolated for MS2 analysis, see Box 1. Concurrent with optimizing peptide elution profiles, we optimized the data-dependent acquisition MS settings, such as minimum MS2 intensity threshold, MS2 injection time, and number of ions sent for MS2 analysis per duty cycle (i.e., TopN), to increase the probability of sampling the apex of the elution peak of each peptide. This optimization increased the number of ion copies sampled from each peptide.[20]

Benchmarking SCoPE2 with standards

The 1xM standards also permitted estimating the instrument measurement noise – independently of biological and sample preparation noise – in the context of SCoPE2 sets. This noise estimate was motivated by our concern that factors unique to ultra-low abundance samples, such as counting noise,[1,3,6] may undermine measurement accuracy. To estimate the measurement reproducibility, we correlated the RI intensities of lysates corresponding to the same cell type: an average correlation of 0.98 suggested good reproducibility. However, this reproducibility benchmark cannot evaluate the accuracy of relative protein quantification, i.e., the ability of SCoPE2 to quantify changes of proteins across cell-types.[24] To benchmark accuracy, we compared the fold-changes of proteins

between human embryonic kidney cells and monocytes (HEK-293 / U-937 protein ratios) estimated from two cell lysates diluted to single-cell levels against the corresponding ratios estimated from the bulk samples used as carriers, Fig. 1b. The high concordance of these estimates (Spearman $\rho = 0.84$) strongly indicates that the instrument noise in quantifying proteins by SCoPE2 is small, consistent with our arguments that the abundance of proteins in single mammalian cells is high enough to minimize the sampling (counting) noise.[6]

To further evaluate relative quantification, beyond the results for two samples diluted to single-cell level (Fig. 1b), we consolidated the data from 76 1xM standards and computed all pairwise correlations. This 592-dimensional matrix was projected onto its first two principal components (PC). The largest PC accounts for 64% of the total variance in the data and perfectly separates all samples corresponding to HEK-293 cells or monocytes. Crucially, the cell lysates diluted to single-cell levels cluster the same way as the carrier samples, indicating that the clustering is driven by cell-type specific protein differences rather than by reproducible artifacts. Thus, the SCoPE2 design can reliably quantify protein abundances at the single-cell level.

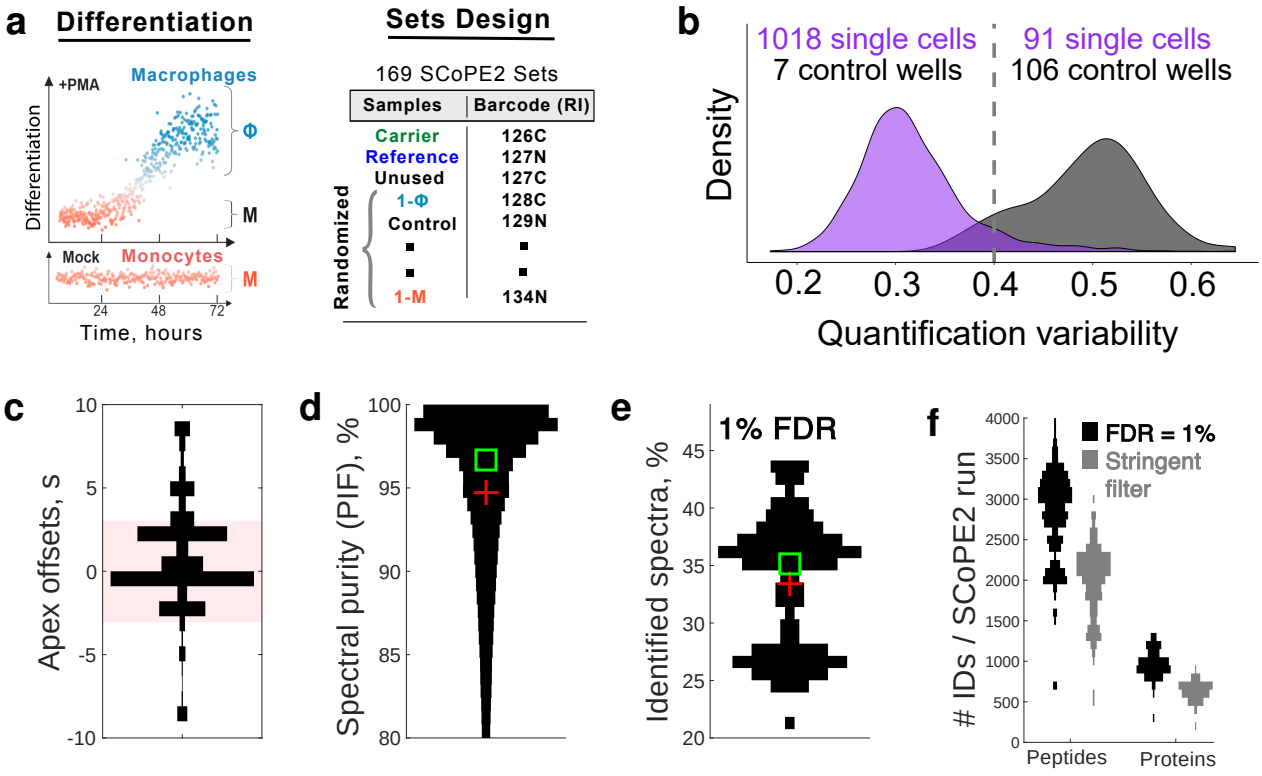


Figure 2 | Model system and technical benchmarks for analyzed single cells and proteins (**a**) Monocytes were differentiated into macrophages by PMA treatment, and FACS-sorted cells prepared into 179 SCoPE2 sets, 62 labeled with TMT 11-plex and 117 labeled with TMT 16-plex. (**b**) Distributions of coefficients of variation (CVs) for the fold-changes of peptides originating from the same protein. The CVs for single cells are significantly lower than for the control wells. (**c**) A distribution of time differences between the apex of chromatographic peaks and the time when they were sampled for MS2 analysis, see Box 1. Over 80% of ions (shaded box) were sampled within 3 seconds of their apexes. (**d**) A distribution of precursor ion fractions (PIF) for all peptides across all SCoPE2 sets. For most MS2 scans, over 97 % of the ions isolated for fragmentation and MS2 analysis belonged to a single precursor (peptide sequence), see Box 1. The square and the cross mark the median and the mean respectively. (**e**) About 35% of MS2 spectra are assigned to peptide sequences at 1 % false discovery rate (FDR). (**f**) Number of identified and quantified peptides and proteins in single cells from SCoPE2 sets analyzed on 60min nLC gradients. All peptide and protein identifications are at 1% FDR and are supported by DART-ID[21]. The criteria for stringent filtering are described in Methods. See Fig. S2b for number of peptides and proteins identified from MS spectra alone.

Model system and technical benchmarks for single cells and proteins

Having demonstrated that proteins from 1xM standards can be quantified with low noise, we next applied SCoPE2 to the analysis of single cells. As a model system, we chose monocytes differentiated into macrophage-like cells in the presence of phorbol-12-myristate-13-acetate (PMA), Fig. 2a. We chose this system since it provides a clear benchmark – the ability to identify two closely related but distinct cell types. This system also presents an open research question: are macrophage-like cells produced from this differentiation as homogeneous as the monocytes from which they originate or more heterogeneous? To answer this question independently from the heterogeneity inherent to primary monocytes, we used a homogeneous monocytic human cell line, U-937.[25]

The SCoPE2 work-flow (Fig. 1a) can be used with manually picked cells, FACS-sorted cells or cells isolated by microfluidic technologies that minimize the volume of the droplets containing cells.[6] Here, we used a BD FACSAria I cell sorter to sort single cells into 384-well plates, one cell per well, see Fig. 2a and Fig. S1. The single cells from two biological replicate preparations of the differentiation protocol were sorted and prepared into 179 SCoPE2 sets, as depicted in Fig. 1a and Fig. 2a. The sorting followed a randomized layout to minimize biases, and sample preparation was automated as described in Methods. Each set has a carrier corresponding to 200 cells and a reference corresponding to 5 cells. The carriers of some sets were sorted individually, with 200 cells per well, while the carriers and reference channels of other sets were diluted from a larger bulk sample.

First, we sought to identify successfully analyzed single cells based on low variability for the relative protein quantification derived from different peptides, Fig. 2b. Specifically, different peptides originating from the same protein should provide similar estimates for the protein fold changes across single cells. The variability between estimates from different peptides was quantified by the coefficient of variation (CV) for all peptides originating from the same protein, i.e., standard deviation / mean for the RI ratios. Then the median CV (across all proteins) of a single cell provides a measure for the consistency of relative protein quantification in that cell. The distribution of CVs for all single cells and control wells indicates that 1,018 single cells have lower quantification variability than the control wells, Fig. 2b. These single cells were analyzed further. See **Supplementary note 1**.

As explained in Box 1, the quantitative accuracy of LC-MS/MS measurements depends on sampling the apex of chromatographic peaks. Thus, we used DO-MS to optimize the instrument parameters so that chromatographic peaks are sampled close to their apexes.[20] As a result, most ions were sampled for MS2 analysis within 3 seconds of their apexes (Fig. 2c). The combination of sampling

close to the apex, sharp chromatographic peaks, and narrow isolation widows (0.7 Th) minimized the simultaneous coisolation of different peptide ions for MS2 analysis. Indeed, MaxQuant estimated over 97 % median purity of the ions sent for MS2 analysis, Fig. 2d. The high spectral purity and the use of retention times to bolster peptide sequence identification[21] allowed assigning a peptide sequence to about 35% of the MS2 spectra from each SCoPE2 run at 1% FDR, Fig. 2e. The high spectral purity supports accurate quantification, consistent with a linear dependence between the measured signal and the input amount shown in Fig. S2a.

On average, SCoPE2 quantifies over 2,500 peptides corresponding to about 1,000 proteins per SCoPE2 set analyzed on one hour chromatographic gradient, Fig. 2f. While longer gradients can increase this coverage, they will reduce the number of cells analyzed per unit time. Since most single-cell analysis requires analyzing large numbers of single cells, we focused on shorter nLC gradients that increase the number of proteins quantified across many single cells rather than merely the number of proteins per cell.[6] We applied further filtering of peptides and proteins to ensure FDR below 1% within each peptide, and within each protein, not just across all peptides and datapoints; see Methods.

Types of missing values in SCoPE2 data. The first type of missing values is due to peptides sent for MS2 analysis but resulting in too low signal to noise ratio (SNR) for quantification. This is a minor contribution to missingness. Indeed, only 10% of the data for peptides identified in a SCoPE2 set were missing because the RI signal in some single cells is below the SNR threshold. This missingness can be reduced further by increasing ion sampling (accumulation) times or by narrowing chromatograpic peaks, see Box 1.[6] The second and major type of missing values is due to peptides not sent for MS2 analysis. This is because the MS instrument does not have time to send for MS2 scans all ions detected at MS1 survey scans, and thus it isolates for MS2 analysis different peptide subsets during different SCoPE2 runs (see Fig. 1a and Box 1), a well-described phenomenon for data-dependent acquisition (DDA).[10,11] As a consequence, we obtained quantification for a diverse set of 2,791 proteins. Thus in the matrix of 2,791 proteins $\times$ 1018 cells, approximately 70% of the values were missing. This type of missing data, unlike the first, is a product of experimental design and not necessarily a lack of sensitivity. To increase reproducibility of protein measurement between SCoPE2 sets, a specific list of peptides can be targeted for analysis in each set.[6]

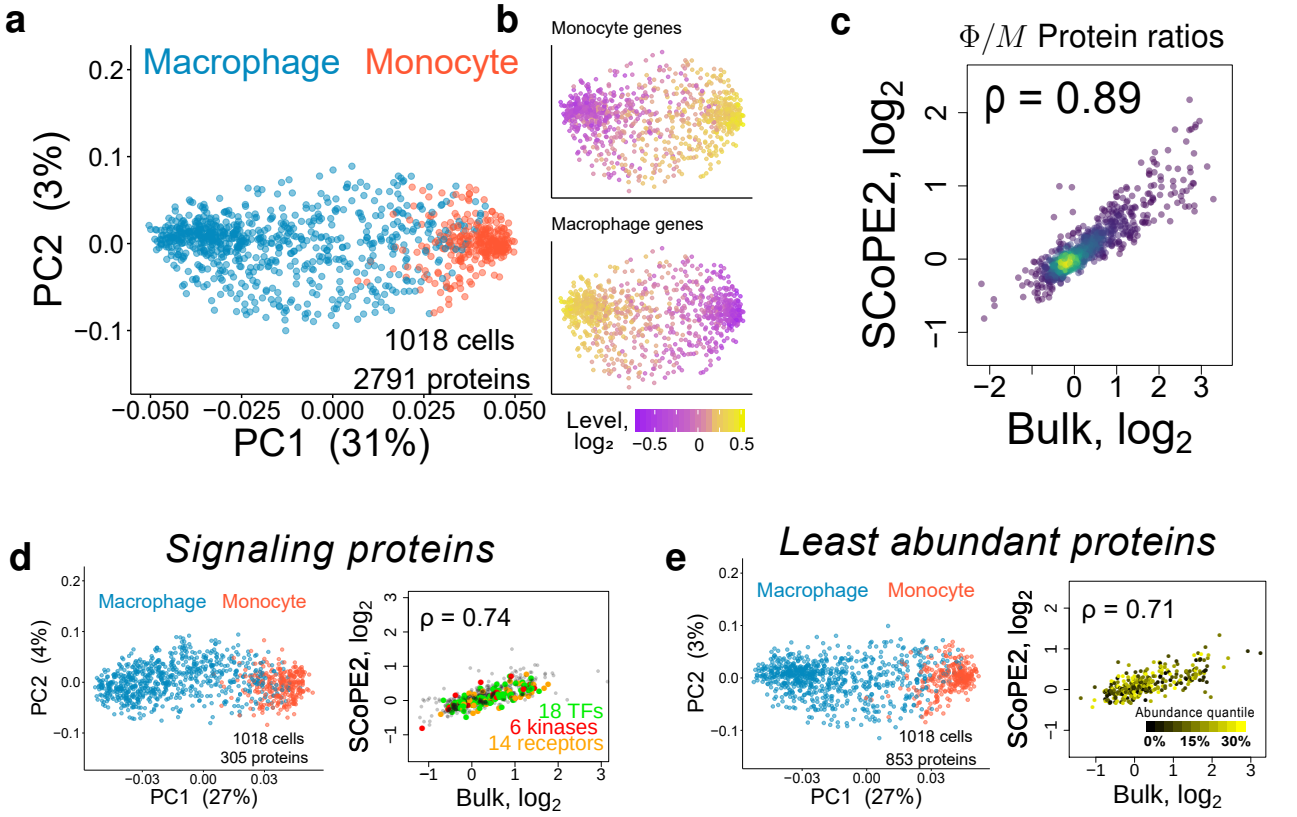


Figure 3 | **Identifying cell clusters by principal component analysis and benchmarking SCoPE2 quantification.** (**a**) A weighted principal component analysis (PCA) of 1,018 single cells using all 2,791 proteins quantified across multiple single cells. Missing values were imputed using k-nearest neighbor (k=3). Cells are colored by cell type. (**b**) The cells from the PCA in panel a are colorcoded based on the abundance of monocyte and macrophage genes, defined as the 30 most differential proteins between bulk samples of monocytes and macrophages. (**c**) The relative protein levels (macrophage / monocyte protein ratios) estimated from the single cells correlate to the corresponding measurements from bulk samples. Proteins functioning in signaling (**d**) as well as the least abundant proteins quantified by SCoPE2 (**e**) allow clustering cells by cell type. The protein fold-changes between monocytes and macrophages for these protein sets are consistent between single cells and bulk samples, similar to panel c.

Cell-type classification

High-throughout single-cell measurements are commonly used to identify and classify cell types. Thus, we sought to test the ability of SCoPE2 data to perform such classification. As a first approach, we performed principal component analysis (PCA), Fig. 3a. Using all 2,791 quantified proteins, PCA separates the cells into two mostly discrete clusters along PC1, which accounts for 31 % of the total variance of the data. Color-coding the cells by their labels indicates that the clusters correspond to the monocytes and the macrophages, Fig. 3a. The macrophage cluster appears more spread-out, suggesting that the differentiation increased the cellular heterogeneity. To evaluate the abundance

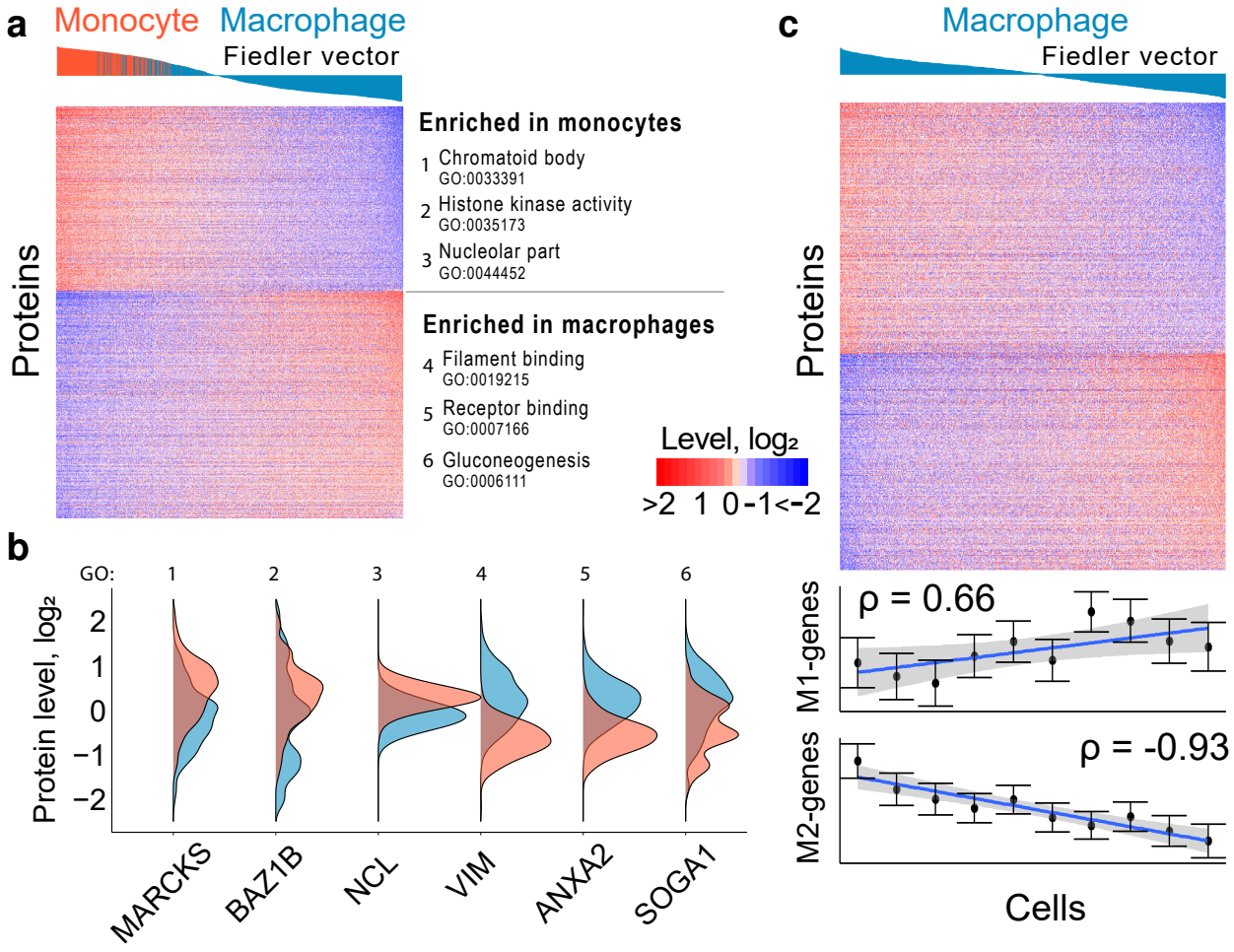


Figure 4 | **Single-cell proteomes define a continuum of macrophage-like states** (**a**) Heatmap of the top 20 % most variable proteins (555) between two clusters of cells identified by unsupervised spectral clustering of all quantified proteins and cells. The cells are ordered based on their rank in the corresponding Fiedler vector from the spectral clustering, see eq. 1. The color bars above the heatmap indicate the loadings of each cell in the Fiedler vector. (**b**) Gene set enrichment[26] identified overrepresented functions for the proteins enriched within each cell type. These functions are displayed alongside representative protein distributions from each gene set. (**c**) The unsupervised spectral analysis from panel a was applied only to the macrophage-like cells, revealing a gradient of macrophage heterogeneity. Cells were ordered based on the corresponding elements of the Fiedler eigenvector, eq. 1. The color bars above the heatmap indicate the loadings of each cell in the Fiedler vector. The top 25% of proteins with the largest fold change between the first 40 cells and last 40 cells are displayed (693 proteins). The single-cell levels of genes in the protein data set previously reported to be enriched in M1 or M2 polarized primary human macrophages[27] are displayed at the bottom; each datapoint represents the median value over bins of 102 cells and error bars denote standard error of data points in each bin.

of monocyte and macrophage associated proteins within the single cells, we color-coded each cell by the median abundance of proteins identified to be differentially abundant from analyzing bulk samples of monocytes and macrophages, Fig. 3b.

While the PCA clustering is consistent with the known cell types, it is inadequate to benchmark protein quantification. As a direct benchmark, we averaged *in silico* the single-cell data to compute monocytes / macrophages protein fold-changes and compared these estimates to the the corresponding fold-changes measured from bulk samples, i.e., averaging across single cells by physically mixing their lysates, Fig. 3c. The correlation ($\rho = 0.89$) indicates that SCoPE2 can accurately measure protein fold-changes in single cells.

In principle, the cell-type classification in Fig. 3a,b may be driven by relatively few abundant structural proteins while less abundant regulatory proteins, such as kinases, receptors and transcription factors, might be poorly quantified. To evaluate this possibility, we applied the PCA analysis using only proteins functioning in signaling (Fig. 3d) or only the least abundant proteins quantified by SCoPE2 (Fig. 3e). The results indicate that these proteins groups also correctly classify cell types, albeit the fraction of variance captured by PC1 is lower, (27 %) for both the signaling proteins, Fig. 3d, and the least abundant proteins, Fig. 3e. The relative quantification of proteins from both sets correlates positively ($\rho = 0.74$, $\rho = 0.71$) to the corresponding bulk protein ratios.

Macrophages exhibit a continuum of proteome states

Next we turn to the question of whether the homogeneous monocytes differentiated to similarly homogeneous macrophages, Fig. 2a. To this end, we performed unsupervised spectral analysis of the cell population, which allowed us to characterize cellular heterogeneity without assuming that cells fall into discrete clusters. To do so, the cells were connected into a graph based on the correlation of their proteome profiles. We then examined the eigenvector of the graph Laplacian matrix with the smallest non-trivial eigenvalue (Fiedler vector, eq. 1), which captures the most prominent structural axis of the cell similarity graph; see Methods. We then sorted cells according to their Fiedler vector values, thus capturing the most prominent aspect of cellular heterogeneity Fig. 4a. This cell clustering is based on hundreds of proteins with differential abundance between monocytes and macrophage-like cells, shown in Fig. 4a: Proteins with higher abundance in monocytes are enriched for proliferative functions, including histone kinase activity.[26] Proteins with higher abundance in macrophages are enriched for immune functions and cell adhesion proteins, including intermediate filament protein Vimentin, Fig. 4b. These enrichment results are consistent with the functional specialization of monocytes and macrophages and further validate the ability of SCoPE2 data to recapitulate known biology.

To explore the heterogeneity within the macrophages, we applied the same spectral analysis as in Fig. 4a, but this time only to the macrophage-like cells. The distribution of the elements of the

associated Fiedler eigenvector (defined by eq. 1) suggests that the cellular heterogeneity observed in this population is better described by a continuous spectrum rather than by discrete clusters, Fig. 4c. Indeed, the heatmap of protein levels for macrophage-like cells ordered based on the Fiedler eigenvector shows that most proteins change gradually across a continuous spectrum, Fig. 4c. Analyzing the proteins from this gradient, we observed a remarkable trend: Genes previously identified as differentially expressed between M1 and M2-polarized primary human macrophages[27] are also differentially expressed between single macrophage cells. For example, the cells at the left edge of Fig. 4c show high expression of genes upregulated in M2-polarized macrophages, decreasing monotonically from the left to right of Fig. 4c. Genes upregulated in M1-polarized primary human macrophages appear to be expressed in a reciprocal fashion, with lower expression at the left edge of Fig. 4c, increasing monotonically across the figure.

Joint analysis of single-cell protein and RNA data

Next, we analyzed single cells from two biological replicates of differentiating monocytes (Fig. 2a) by scRNA-seq using the 10x Chromium platform and compared the cell clustering and differential genes measured by scRNA-seq and SCoPE2. We first consider measurement noise since it may contribute to apparent differences between the protein and RNA measurements.[24] Specifically, we can expect to confidently identify differential abundance between cell types only for genes whose biological variability exceeds the measurement noise.

A major source of measurement noise for both SCoPE2 and scRNA-seq is sampling low copy number of molecules per gene. A low copy-number of sampled molecules results in significant counting noise.[6] This noise arises because scRNA-seq and SCoPE2 sample only a subset of the molecules from a single cell. This sampling process contributes to a counting error: The standard deviation for sampling n copies is $\sqrt{n}$ (from the Poisson distribution), and thus the relative sampling error, estimated as standard deviation over mean, is $\sqrt{n}/n = 1/\sqrt{n}$, Fig. 5a. Thus our optimization of SCoPE2 aimed to increase ion delivery not merely the number of identified peptides.[20] Similarly, to increase the sampling of mRNAs, we sequenced the 10x library to an average depth of 100,000 reads per cell. Furthermore, we chose for subsequent analysis only representative cells with at least 10,000 UMIs per cell, Fig. S3.

To estimate our sampling error, we sought to convert the RI abundances (i.e., the barcodes from which SCoPE2 estimates peptide abundances) into ion copy numbers. To do so, we extracted the signal to noise ratios (S/N) for RIs and multiplied these ratios by the number of ions that induces a

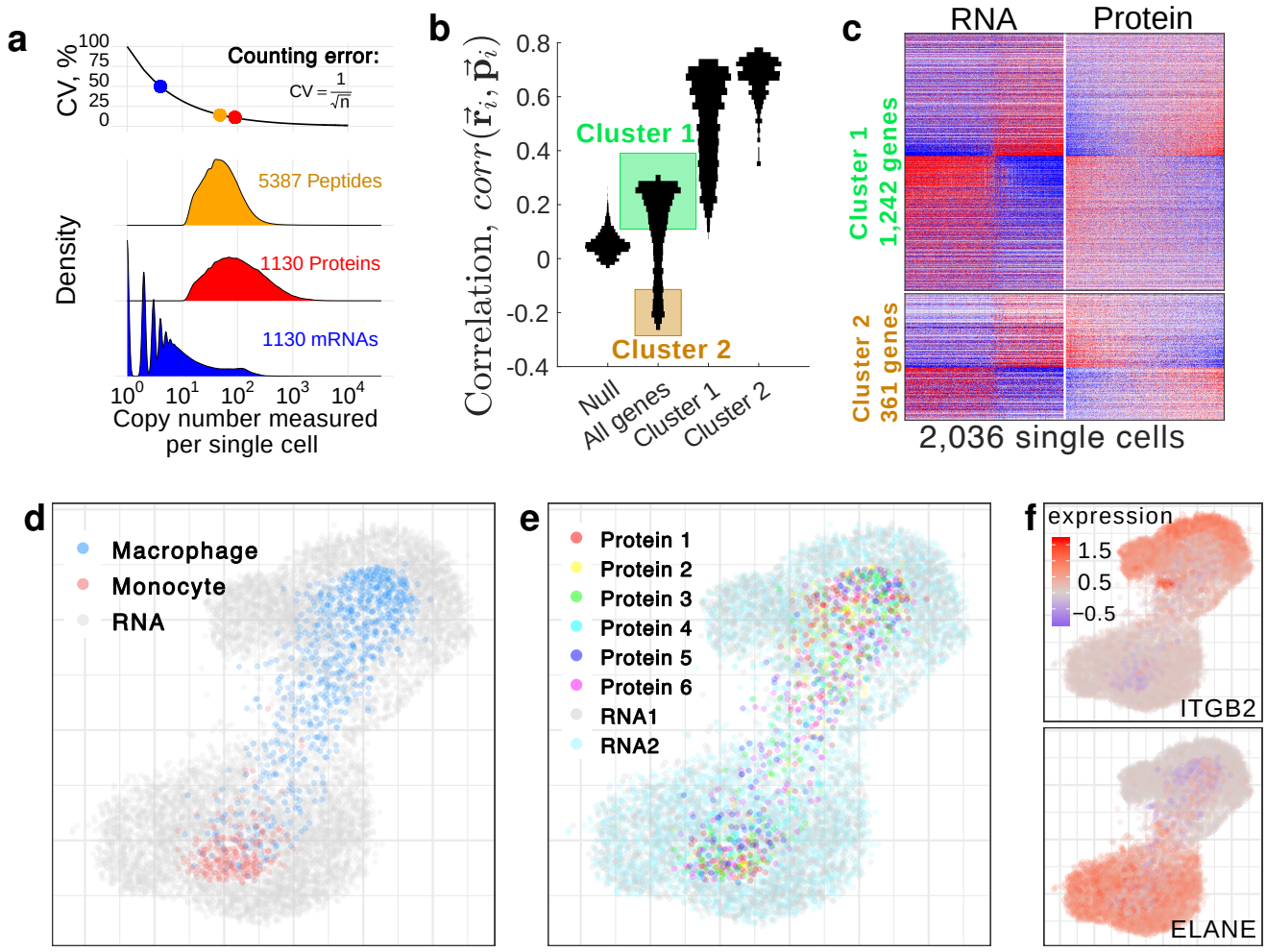


Figure 5 | **Joint analysis of single-cell RNA and protein data** (**a**) Number of unique-barcode reads per mRNA or ions per peptide/protein for a set of 1130 genes. The higher copy numbers measured for proteins support more reliable counting statistics compared to mRNAs. (**b**) Distributions of correlations between $\mathbf{r}_i$ and $\mathbf{p}_i$, where $\mathbf{r}_i$ is the vector of pairwise correlations of the i^{th} RNA to all other RNAs, and $\mathbf{p}_i$ is the vector of pairwise correlations of the i^{th} protein to all other proteins.[12,28] The null distribution correspond to permuting the order of RNAs and proteins. The two modes of the distribution of correlations for all genes were used to define gene Clusters 1 and 2. The correlations between $\mathbf{p}_i$ and $\mathbf{r}_i$ was then recomputed just within the space of genes from Clusters 1 or from Clusters 2 and displayed as separate distributions. (**c**) Genes from Cluster 1 display similar abundance profiles at both the RNA and protein levels, while genes from Cluster 2 display the opposite profiles. The columns correspond to single cells sorted by the loadings of the Fiedler vectors of each dataset, i.e., by cell type. The colormap represents log_2 fold-changes with a 2-fold dynamic range between low (blue) and high (red). (**d-f**) Joint projections of the RNA and protein data by Conos[29]. (**d**) Cells analyzed by SCoPE2 are colorcoded by cell type while cells analyzed by scRNA-seq are marked gray. (**e**) All single cells are colorcoded by biological replicate and batch. (**f**) Cells are colorcoded by the expression of marker genes for monocytes and macrophages.

unit change in S/N. Since our Q-Exactive basic Orbitrap operated at 70,000 resolving power, a S/N ratio of 1 corresponds to about 6 ions.[30,31] Thus for our system, a S/N ratio of 50 corresponds to

300 ions; see Methods and **Supplementary note 2** for details. The results in Fig. 5a indicate that SCoPE2 samples 10-100 fold more copies per gene than scRNA-seq, which corresponds to smaller sampling (counting) errors. The sampling can be increased further by increasing the ion sampling times, see Box 1.[6]

To investigate the similarities and differences at the transcriptional and post-transcriptional levels, we compared the correlation vectors of the scRNA-seq and SCoPE2 data as previously described.[12,28] Specifically, for each dataset we computed the matrices of pairwise Pearson correlations between genes, averaging across the single cells. Then to quantify the similarity of covariation of the i^{th} gene, we correlated the i^{th} vectors of RNA ($\mathbf{r}_i$) and protein ($\mathbf{p}_i$) correlations. The correlations between $\mathbf{r}_i$ and $\mathbf{p}_i$ for all genes are bimodally distributed, with a large positive mode, Fig. 5b. Yet, they become close to zero when the order of genes is permuted, see the null distribution in Fig. 5b. We used the two modes of the distribution of correlations between $\mathbf{r}_i$ and $\mathbf{p}_i$ to define genes with similar protein and RNA covariation (Cluster 1) and genes with opposite covariation (Cluster 2). The genes not included in either cluster are not differentially abundant between monocytes and macrophages at either the protein or the RNA level. The genes within a cluster correlate to each other similarly as indicated by the high correlations between $\mathbf{r}_i$ and $\mathbf{p}_i$ when these vectors are comprised only by the correlations of genes from Cluster 1 or from Cluster 2, Fig. 5b. To further confirm the coherence of these clusters, we displayed the RNA and protein levels of their genes as heatmaps in Fig. 5c. About 60 % of the genes are in Cluster 1. These genes exhibit similar RNA and protein profiles and are enriched for many biological functions, including antigen presentation, cell adhesion, cell proliferation and protein synthesis, Fig. S5. The genes from Cluster 2 have opposite RNA and protein profiles and are enriched for functions including Rab GTPase activity, and protein complex assembly, Fig. S6. Macrophage heterogeneity that we discovered in the SCoPE2 data Fig. 4c is also observed with the scRNAseq data, Fig. S4.

The similarity between the RNA and protein abundances of genes from Cluster 1 (Fig. 5b,c) suggest that the RNA and protein data might be projected jointly. To this end, we used Conos[29] to generate a joint graph integrating all analyzed cells and to project them on the same set of axes, Fig. 5d-f. The joint projection results in two distinct clusters of the cells analysed by SCoPE2, Fig. 5d. These clusters correspond to the monocytes and macrophages and are surrounded by more diffused clusters of the cells analyzed by scRNAseq. The wider spread of the cells analyzed by scRNA-seq may be due to higher biological and technical variability (noise) for transcripts. Indeed, the pairwise correlations between cells for both the monocytes and macrophages are higher for the cells analyzed by SCoPE2 compared to the cells analyzed by scRNAseq, Fig. S7. These correla-

tions between cells also corroborate that the macrophage-like cells are more heterogeneous than the monocytes.

Colorcoding the cells by biological replicate and batch reveals no significant clusters and thus no residual artifacts and batch effects, Fig. 5e. In contrast, colorcoding the cells with monocytes and macrophage markers reveals that the clusters correspond to the cell types, Fig. 5f. These results demonstrate that, at least for our model system, the low dimensional projections of single-cell protein and RNA measurements result in similar clusters albeit the clusters based on RNA measurements are more spread out, suggesting more biological and technical variability in the RNA measurements.

Discussion

SCoPE2 substantially advances the capabilities of SCoPE-MS, Table 1; it enables scalable, robust and affordable quantification of about 1,000 proteins per single cell, and over 2,700 proteins across many cells. This coverage is achieved with 90 min of analysis time per SCoPE2 set (about 6 min / cell), which allowed us to analyze over a thousand of cells on a single instrument in about 10 days. Most exciting for us, SCoPE2 succeeded in delivering and quantifying hundreds of ion copies from most detected proteins. This observation strongly supports the feasibility of single-cell LC-MS/MS protein quantification without amplification. Indeed, we good agreement between protein fold-changes estimated from the SCoPE2 data and from conventional bulk methods, Fig. 1c and Fig. 3c.

A key aim of SCoPE2 is to reduce cost and analysis time. This aim motivated many of our choices, including the use of commercial multiwell plates (as opposed to specialized tubes as we did previously[12]), the use of multiplexing, and the reduction of nLC gradients to 1 hour. These efforts allowed SCoPE2 to reduce the cost and time for sample preparation by over 10-fold, Table 1. It also reduced the LC-MS/MS time, and thus its cost. The estimated cost in Table 1 is based on MS facility fees. The cost can be lower for in-house LC-MS/MS analysis.

The reliability of data from SCoPE2 opens the potential not only to identify and classify cell subpopulations, but to go beyond such descriptive analysis: We believe that accurate protein quantification across thousands of single cells may provide sufficient data for studying post-transcriptional regulation in single cells and for inferring direct causal mechanisms in biological systems.[6]

To have such an impact, SCoPE2 analysis must be robust and accessible. A step in this direction

Benchmark	SCoPE-MS	SCoPE2	Relevant figure
Correlation to benchmark fold-changes	0.2	0.89	3c
Purity of ions isolated for quantification	79%	97%	2d
Single-cell protein measurements / hour	610	4,630	2f
Sample preparation:			
- Time, hours / cell	< 1	< 0.03	2a
- Cost, USD / cell	< 10	< 1	1a
LC-MS/MS:			
- Time, hours / cell	0.50	0.12	2a
- Cost, USD / cell†	48 - 96	10 - 20	1a

Table 1 | SCoPE2 improves quantitative accuracy, depth of proteome coverage, ease of sample preparation and cost-effectiveness 2-10 fold over SCoPE-MS. Quantitative accuracy is evaluated by comparing relative protein ratios from *in silico* averaged single cells to bulk data, shown in Fig. 3c. Rate of protein identification is calculated as the median number of unique protein data points identified per hour of LC-MS/MS time. Cost of SCoPE-MS does not include cost of sonication instrumentation. The throughput estimates for SCoPE2 are based on TMT 16-plex.
† LC-MS/MS cost assumes MS facility fee of $100 - $200 per hour.

is replacing the expensive and time-consuming lysis used by SCoPE-MS[12] with mPOP[19], Fig. 1a. Another step is the implementation of DO-MS, which makes it easier to execute and adapt SCoPE2 to different samples and LC-MS systems.[20] A further step is the analysis identifying successful cells shown in Fig. 2b. We believe that these steps bring us closer to the transformative opportunities of single-cell proteomics.[6]

We demonstrated that U-937-derived macrophages showed increased heterogeneity compared to the monocyte form, Fig. 3a. Upon exposure to identical environmental conditions, single macrophage cells exhibited coordinated protein level changes, Fig. 4b. In the absence of further treatment with polarizing cytokines or lipopolysaccharide to specifically induce macrophage polarization,[32] the differentiated macrophage population existed in a continuum, showing reciprocal loss or gain of proteins previously identified as enriched in M1 or M2 macrophages[27], Fig. 4b. This observation suggests that polarization might be a propensity inherent to macrophages.

Data Availability:

The raw MS data and the search results were deposited in MassIVE (ID: MSV000083945 and MSV000084660).

- **Facilitating LC-MS/MS evaluation:** To facilitate evaluation of our RAW LC-MS/MS data, we include detailed distribution plots generated by DO-MS.[20] These plots allow quick assessment of the nLC, ions detected at MS1 and MS2 level, apex offsets, identification rates and other important LC-MS/MS features.

- **Facilitating data reuse:** To facilitate reanalysis of our data, we also made them available in an easily reusable form, including 3 files in comma separated values (csv) format as follows:

    1. `Peptides-raw.csv` – peptides $\times$ single cells at 1 % FDR. The first 2 columns list the corresponding protein identifiers and peptide sequences and each subsequent column corresponds to a single cell. Peptide identification is based on spectra analyzed by MaxQuant[33] and is enhanced with using DART-ID[21] by incorporating retention time information.

    2. `Proteins-processed.csv` – proteins $\times$ single cells at 1 % FDR, imputed and batch corrected.

    3. `Cells.csv` – annotation $\times$ single cells. Each column corresponds to a single cell and the rows include relevant metadata, such as cell type if known, measurements from the isolation of the cell, and derivative quantities, i.e., rRI, CVs, reliability.

Supplemental website can be found at: scope2.slavovlab.net

Acknowledgments: We thank A.T. Chen, S. Semrau, A. Marneros, A. Makarov, M. Jovanovic, J. Alvarez, Z. Niziolek, Y. Katz, A. Andersen, B. Budnik, T. Hirz, and B. Karger, for assistance, discussions and constructive comments. This work was funded by a New Innovator Award from the NIGMS from the National Institutes of Health to N.S. under Award Number DP2GM123497, an iAward from Sanofi to N.S., and through a Merck Exploratory Science Center Fellowship, Merck Sharpe & Dohme Corp. to N.S.

Competing Interests: The authors declare that they have no competing financial interests.

Correspondence: Correspondence and materials requests should be addressed to nslavov@alum.mit.edu

Author Contributions

Experimental design: H.S., and N.S.

LC-MS/MS: H.S., D.H.P., R.G.H., and T.K.

Sample preparation: H.S., E.E., A.P., and M.S.

Data analysis: H.S., P.K., and N.S.

Writing & editing: H.S., E.E, and N.S.

Raising funding & supervision: N.S.

Methods

Cell culture Human embryonic kidney cells (HEK-293 cells) were grown as adherant cultures in DMEM with high glucose (Sigma-Aldrich D5796), supplemented with 10% fetal bovine serum (FBS, Millipore Sigma F4135) and 1% penicillin-streptomycin (pen/strep, ThermoFisher 15140122). U-937 (monocytes) cells were grown as suspension cultures in RPMI medium (HyClone 16777-145) supplemented with 10% fetal bovine serum (FBS, Millipore Sigma F4135) and 1% penicillin-streptomycin (pen/strep, ThermoFisher 15140122). Cells were passaged when a density of 10^6 cells/ml was reached, approximately every two days. Monocytes were differentiated to macrophage-like cells by first adding phorbol 12-myristate 13-acetate (PMA) to the culture medium at final concentration of 5nM for 24 hours. Then, these newly-adherent cells were washed with fresh PMA-free medium and allowed to recover in PMA-free medium for an additional 48 hours before harvest. Mock-treated U-937 cells were passaged with fresh PMA-free media at 24 hours and harvested along with the treated cells at 72 hours.

Harvesting cells U-937 cells that underwent PMA-induced differentiation were washed twice with ice-cold phosphate buffered saline (1x PBS) and dissociated by scraping. Cell suspensions of un-differentiated U-937 cells were pelleted and washed quickly with cold PBS at 4 ^oC. Media was removed from adherent cultures of HEK-293 cells via pipetting, and the cells were subsequently incubated with 4 ^oC 0.05% trypsin-EDTA (Gibco, ThermoFisher 25300054) at 37 ^oC for 3 minutes. Cold 1x PBS was added to the dissociated HEK-293 cells, which were then pelleted via centrifugation, washed with 1x PBS, and pelleted again. The washed pellets of HEK-293 and U-937 cells were diluted in 1x PBS at 4 ^oC. The cell density of each sample was estimated by counting at least 150 cells on a hemocytometer. When preparing cells for SCoPE2 sets, the pellets were split into two tubes: one was diluted in 1x PBS and used for cell sorting; the other was diluted in pure water (Optima LC/MS Grade, Fisher Scientific W6500) and used for carrier and reference channel preparation in bulk.

Sample randomization and sorting SCoPE2 sets were designed such that, on average, there would be 5 single macrophages, 2 single monocytes and 1 control well per set. Control wells in this context are defined as wells that experience all sample preparation steps, except that no single cell is present in the well. SCoPE2 sets were randomized over a 384-well plate such that there would be either a maximum of 32 SCoPE2 sets (when using TMTPro 16plex) or 48 SCoPE2 sets (when using TMT 11plex) produced per plate. Single U-937 monocyte and macrophage cells were isolated and distributed into $1\mu l$ of pure water with MassPREP peptide mixture ($25 fmol/\mu l$ final concentration, Waters 186002337) in 384-well PCR plates (ThermoFisher AB1384) using a BD

FACSAria I cell sorter. In 62 of the experimental sets, the carrier channels were sorted individually, such that 100 cells of macrophage-type and 100 cells of monocyte-type were sorted together into a single well, serving as the carrier channel for a single SCoPE2 set. The reference channel was prepared separately by harvesting 10,000 cells of each type into a $500\mu l$ eppendorf tube, which was then used in all sets across multiple plates.

Carrier and reference channel preparation in bulk Except for the 62 experimental sets referenced above, the carrier channel was prepared in bulk and aliquoted into carriers corresponding to 200 cells each. A single cell suspension of about 22,000 cells was transferred to a $200\mu l$ PCR tube (USA Scientific 1402-3900) and then processed via the SCoPE2 sample preparation as described below. This cell lysate was used to generate both the carrier and reference channels.

SCoPE2 sample preparation Cells were lysed by freezing at -80 ^oC for at least 5 minutes and heating to 90 ^oC for 10 minutes. Then, samples were centrifuged briefly to collect liquid; trypsin (Promega Trypsin Gold, Mass Spectrometry Grade, PRV5280) and Triethylammonium bicarbonate buffer (TEAB, Millipore Sigma T7408-100ML) were added at final concentrations of $10\ ng/\mu l$ and 100mM, respectively. The samples were digested for 3 hours in a thermal cycler at 37 ^oC (BioRad T1000). Samples were cooled to room temperature before TMT labeling (TMT11 plex kit & TMTPro 16plex, ThermoFisher, Germany). Single cells and carrier cells sorted into the plate were labeled with $1\ \mu l$ of 22mM TMT label for 1 hour at room temperature. For those samples in which the carrier and reference were both prepared in bulk, half of the cells were labeled with 85mM TMT 126C (representing the carrier channel) and half were labeled with 85mM TMT 127N (representing the reference channel). The unreacted TMT label in each sample was quenched with $0.5\mu l$ of 0.5% hydroxylamine (Millipore Sigma 467804-10ML) for 45 minutes at room temperature. Samples were centrifuged briefly following all reagent additions to collect liquid. The samples corresponding to either one TMT11plex or one TMTPro set were then mixed in a single glass HPLC insert (ThermoFisher C4010-630) and dried down to dryness in a speed-vacuum (Eppendorf, Germany) and either frozen at -80 ^oC for later analysis or immediately reconstituted in $1.2\mu l$ of 0.1% formic acid (ThermoFisher 85178) for mass spectrometry analysis.

1xM standard preparation HEK-293 and U-937 cells were harvested and counted as described above. 40,000 cells of each type were resuspended in $20\mu l$ of HPLC-grade water (Fisher Scientific W6500), frozen at -80^oC for 30 minutes, and heated to 90^oC for 10 minutes in a thermal cycler, before being brought to room temperature. $1\mu l$ of benzonase (Millepore Sigma E1014-5KU), diluted to 5 units/μl, was added to each sample, which were then placed in a sonicating bath for 10 minutes. After sonication, $4\mu l$ of a master mix containing $2.5\mu l$ of 1M triethylammonium bicarbonate

(TEAB, Millipore Sigma T7408-100ML), $1.3\mu l$ of 200ng/μl trypsin (Promega Trypsin Gold, Mass Spectrometry Grade, PRV5280), and $0.2\mu l$ of HPLC-grade water was added to each sample, and the samples were placed in a preheated 37^oC thermal cycler (BioRad T100) to digest for 3 hours. $3.13\mu l$ of each cell type was aliquoted into 4 PCR tubes and each of the eight samples were then labeled following the rubric shown in Fig. 1b with $1.6\mu l$ of 85mM TMT11plex tags for one hour at room temperature. The unreacted TMT labels in each sample were then quenched with $0.7\mu l$ of 1% hydroxylamine solution (Millipore Sigma 467804-10ML) for 30 minutes at room temperature. Both labelled carrier samples were combined in a single glass HPLC insert (ThermoFisher C4010-630), and the non-carrier samples were each brought up to $50\mu l$ total volume, of which $1\mu l$ of each was added to the HPLC insert. The combined samples in the HPLC insert were then dried down to dryness in a speed-vacuum (Eppendorf, Germany) prior to storage at -80. Before analysis by LC-MS/MS, the sample was reconsituted in $100\mu l$ of 0.1% formic acid (ThermoFisher 85178), such that a $1\mu l$ injection containing material equivalent to 50 cells in the two carrier channels and 1 cell in each of the six additional channels was analyzed by LC-MS/MS.

Ladder experiments preparation U-937 cells were counted as described above, digested as described above, and serially diluted into 2000, 10, 20, 30, 40, 50, and 60-cell equivalents which were each labeled with TMT11plex tags in two randomized designs with three replicates for each design, for a total of six replicates. These samples were diluted 10x so material corresponding to 200, 1, 2, 3, 4, 5, and 6-cell equivalents was injected and analyzed by LC-MS/MS.

SCoPE2 mass-spectrometry analysis SCoPE2 samples were separated via online nLC on a Dionex UltiMate 3000 UHPLC; $1\mu l$ out of $1.2\mu l$ of sample was loaded onto a 25cm x 75 $1\mu m$ IonOpticks Aurora Series UHPLC column (AUR2-25075C18A). Buffer A was 0.1% formic acid in water and buffer B was 0.1% formic acid in 80% acetonitrile / 20% water. A constant flow rate of 200nl/min was used throughout sample loading and separation. Samples were loaded onto the column for 20 minutes at 1% B buffer, then ramped to 5% B buffer over two minutes. The active gradient then ramped from 5% B buffer to 25% B buffer over 53 minutes. The gradient then ramped to 95% B buffer over 2 minutes and stayed at that level for 3 minutes. The gradient then dropped to 1% B buffer over 0.1 minutes and stayed at that level for 4.9 minutes. Loading and separating each sample took 95 minutes total. All samples were analyzed by a Thermo Scientific Q-Exactive mass spectrometer from minute 20 to 95 of the LC loading and separation process. Electrospray voltage was set to 2,200V, applied at the end of the analytical column. To reduce atmospheric background ions and enhance the peptide signal-to-noise ratio, an Active Background Ion Reduction Device (ABIRD, by ESI Source Solutons, LLC, Woburn MA, USA) was used at the nanospray interface.

The temperature of ion transfer tube was 250^oC and the S-lens RF level was set to 80. After a precursor scan from 450 to 1600 m/z at 70,000 resolving power, the top 7 most intense precursor ions with charges 2 to 4 and above the AGC min threshold of 20,000 were isolated for MS2 analysis via a 0.7 Th isolation window with a 0.3 Th offset. These ions were accumulated for at most 300ms before being fragmented via HCD at a normalized collision energy of 33 eV (normalized to m/z 500, z=1). The fragments were analyzed at 70,000 resolving power. Dynamic exclusion was used with a duration of 30 seconds with a mass tolerance of 10ppm.

Analysis of raw MS data Raw data were searched by MaxQuant[9,33] 1.6.2.3 against a protein sequence database including all entries from the human SwissProt database (downloaded July 30, 2018; 20,373 entries) and known contaminants such as human keratins and common lab contaminants (default MaxQuant contaminant list). MaxQuant searches were performed using the standard work flow[34]. We specified trypsin/P digestion and allowed for up to two missed cleavages for peptides having from 7 to 25 amino acids. Tandem mass tags (TMT 11plex or TMTPro 16plex) were specified as fixed modifications. Methionine oxidation (+15.99492 Da), asparagine deamidation (+0.9840155848 Da), and protein N-terminal acetylation (+42.01056 Da) were set as a variable modifications. As alkylation was not performed, carbamidomethylation was disabled as a fixed modification. Second peptide identification was disabled. Calculate peak properties was enabled. All peptide-spectrum-matches (PSMs) and peptides found by MaxQuant were exported in the evidence.txt files. False discovery rate (FDR) calculations were performed in the R programming language environment.

DART-ID search Seventy-six replicate injections of the 1xM standards, 179 SCoPE2 sets, and 69 additional runs described in the MassIVE submission metadata were analyzed together by DART-ID[21]. A configuration file for the search is included in Supplementary information.

Peptide and protein filtering The subsequent data analysis was performed in the R (v3.5.2) programming language environment, and the code used is available at github.com/SlavovLab/SCoPE2/code. The MaxQuant evidence.txt (with identification confidence updated by DART-ID) was filtered for peptide FDR $< 1\%$ across the whole data set, and reverse hits and contaminant hits were removed (contaminants denoted by MaxQuant's default contaminants list). This level of filtering was employed to produce the "FDR = 1%" distributions in Fig. 2f. In addition to the filtering done above, more stringent filters were applied to SCoPE2 runs before further analysis: peptides with precursor intensity fraction (PIF) was below 80% were removed as well as peptides for which the reporter ion intensity in the single cells exceeded 10% of the values for the carrier channel. Then, the false discovery rate (FDR) was calculated for each peptide separately and each peptides was filtered sep-

arately to FDR $< 0.01\%$. Experiments with less than 300 peptide identifications were removed (10 sets, resulting in 169 SCoPE2 sets remaining for single cell filtering steps).

Single-cell filtering Relative reporter ion intensities (rRI) were calculated for each peptide in each single cell relative to its set's reference channel. Single cells with a median rRI $<$ 1/50 were removed outright. The internal consistency of protein quantification for each single cell was evaluated by calculating the coefficient of variation (CV) for proteins (Leading razor proteins) identified with $>$ 5 peptides for that cell. Coefficient of variation is defined as the standard deviation divided by the mean. The CVs were computed for the relative reporter ion intensities, i.e., the RI reporter ion intensities of each peptide were divided by their mean resulting in a vector of fold changes relative to the mean. Control wells were used to determine a reasonable cutoff value for the median CV per cell below which we could have higher confidence that that channel truly contained cellular material and not just signal from noise or contamination. Three outlier SCoPE2 experiments were removed entirely, including single cells, due to their control wells having low CV:

`191104S-LCB7-X-APNOV16plex-Set-4`
`191107S-LCB7-X-APOCT16plex-Set-20`
`191107S-LCB7-X-APOCT16plex-Set-21`

Data transformations After filtering, the peptide-level reporter ion intensities for the remaining single cells were arranged into a matrix of peptides x single cells (rows x columns). All single cell reporter ion intensities were normalized by the reference channel intensities in their respective sets. The columns then the rows were normalized by dividing by their median and mean values, respectively (computed ignoring missing values). Peptides quantified in at least ten cells were kept, with the average peptide being quantified in 192 cells after this filtering. The values in the matrix were log_2 transformed, then the protein-level quantification was calculated by mapping each peptide to its respective (leading razor) protein and taking the median value if there was more than 1 peptide mapped to that protein. The resulting matrix has dimensions proteins x single cells (rows x columns). The data was again normalized by subtracting the column then the row medians (log_2 scale). The average protein was quantified in at least 213 cells.

Imputing missing values Missing values in the protein x single cell matrix were imputed by k-nearest neighbor imputation ($k = 3$) using Euclidean distance as a similarity measure between the cells.

Weighted principal component analysis From the protein x single cell matrix all pairwise protein correlations (Pearson) were computed. Thus for each protein there was computed a vector of correlations with a length the same as the number of rows in the matrix (number of proteins). The

dot product of this vector with itself was used to weight each protein prior to principal component analysis. Principle component analysis was performed on the correlation matrix of the weighted data.

Spectral clustering of cells Spectral clustering was performed by first computing a matrix of positive pairwise weights, $\mathbf{W}$ between all cells. We defined the weight between two cells to be their Pearson correlation plus 1 so that all weights were positive. Then the Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements contain the sum of elements in the corresponding rows of $\mathbf{W}$, i.e., $D_{i,i} = \sum_j W_{i,j}$. Then trivially, the smallest eigenvalue of $\mathbf{L}$ is 0, and its corresponding eigenvector is the constant vector, e.g., the vector of ones. The second smallest eigenvalue corresponds to the Fiedler vector, the non-constant vector $\mathbf{v}$ that minimizes eq. 1.

$$\mathbf{v}^T\mathbf{L}\mathbf{v} = \frac{1}{2}\sum_{i,j} w_{ij}(v_i - v_j)^2, \text{ so that } \mathbf{v}^T\mathbf{v} = 1 \tag{1}$$

Thus, computing this Fiedler eigenvector corresponds to global convex optimization that assigns similar $\mathbf{v}$ elements to cells connected by high weights. The Fiedler vector was used for sorting cells in Fig. 4.

Single-cell RNA-seq data A cellular mixture identical to that used for the single-cell proteomics was assessed with scRNA-seq using 10x Genomics Chromium platform and the Single Cell 3' Library & Gel Bead Kit (v2). Two biological replicates of the cell suspension (stock concentration: 1200 cells/μl) were loaded into independent lanes of the device. An average number of about 10,000 cells/lane were recovered. Following the library preparation and sample QC by Agilent BioAnalyzer High Sensitivity chip, the two libraries were pooled together, quantified by KAPA Library Quantification kit and sequenced using the Illumina Novaseq 6000 system (Nova S1 100 flow cell) with the following run parameters: Read1: 26 cycles, i7 index: 8 cycles, Read2: 93 cycles. Demultiplexing and count matrix estimation was carried out using CellRanger software. Upon QC and manual inspection, the cells containing less than 10^4 UMI barcodes were determined to be amplifying empty or background droplets and discarded. Joint alignment of proteome and scRNA-seq data was performed using Conos package[29], using CCA space. The scRNA-seq datasets were pre-processed using pagoda2 package to normalize for cell size and gene variance distributions. To ensure a balanced comparison, from each scRNA-seq replicate we sampled the number of cells equivalent to the number of cells measured in the single-cell proteome replicates. To ensure that the sampled cells were representative, they were sampled around the mode of the scRNA-seq cell-size

distribution. The proteome datasets were also pre-processed using pagoda2, however no variance normalization was performed - the RI values were used directly for downstream alignment. Conos alignment of scRNA-seq and proteome datasets was performed using a neighborhood size of k=15 using joint PCA space. The resulting alignment graph was visualized using a largeVis embedding (Fig. S3).

Converting signal-to-noise to ion counts Signal-to-noise (S/N) was extracted from the raw files by Proteome Discoverer v2.3. Ion counts were calculated by multiplying the S/N by an estimate for the number of ions that induces a unit change in S/N for our instrument. This factor was estimated as 3.5 for an orbitrap at a resolving power of $240,000$. This factor scales with the square root of the ratio of the resolving power, thus for our instrument run at $70,000$ resolving power, the number of ions that induces a unit change in S/N is $3.5 \times \sqrt{\frac{240{,}000}{70{,}000}} = 6.5$.[30,31] Since we did not search all RAW files with Proteome Discoverer, the results reported in Fig. 5 are only for experiments FP94 and FP97.

Gene set enrichment analysis Gene set enrichment analysis was performed by comparing the distributions of abundances of proteins/RNAs from a functional group to that of all proteins/RNAs and computing a probability that they are samples from the same master distribution. The effect size is summarized as the mean abundance of all proteins/RNAs from a functional group. Genes (proteins) were ordered by fold change between their means in each cluster (for comparing monocytes and macrophage-like cells) or by fold change between the mean of the first 50 and last 50 cells as ordered by spectral clustering (for comparing macrophage-like cells).

Determining genes enriched in monocytes and macrophage-like cells from bulk proteomic data For each protein profiled by bulk proteomic methods a two-sided t-test was performed comparing the relative protein level between the two cell types (20 replicates per cell type). Fold change between the two cell types was calculated by taking the difference in means and the top 60 most differential proteins (30 "up-regulated" in monocytes, 30 "up-regulated" in macrophage-like cells) with a p-value less than 0.01 from the t-test were taken. This list of proteins constitute the "monocyte genes" and the "macrophage genes" displayed in Fig. 4b. Genes up and down-regulated in M1 and M2 macrophage subtype were determined from MSigDB gene sets M13671 and M14515[35].

Supplementary note 1 | Evaluating single-cell quantification

While developing SCoPE2, we observed that many experimental failures manifest non-specifically with RI ratios between the carrier and the single-cell peptides that deviate from the expected ones, i.e., "compressed RI ratios." In particular, failure to sort the single cells or to extract and digest

their proteins results in lower than expected RI intensities in the single-cell samples. Conversely, contamination of the single-cell samples or cross labeling peptides from the carrier with TMT labels for single cells can result in unusually high RI signal for single cells (i.e., "compressed RI ratios"). Thus, large deviations from the expected relative RI intensities of single cells may indicate a variety of potential problems in sample preparation. We chose as a QC metric the consistency of relative protein quantification from different peptides as quantified by the coefficient of variation (CV). Crucially, the CV is computed from the fold-changes between single cells relative to the mean across all samples. Thus, the CVs reflect the consistency of relative quantification which is relevant to our ability to discern biologically meaningful protein variation. CV computed from the non-normalized RI intensities are less useful for this analysis since they can be low even for very noisy data; they conflate different sources of variance and are dominated by the huge difference in the abundance of different proteins that is consistent across different human tissues.[24]

Supplementary note 2 | Signal-to-noise ratio

The signal to noise ratio (SNR) for single-cell RIs is an important parameter affecting the quantitative capability of SCoPE2. The largest fold changes that can be reliably quantified are about equal to or smaller than the SNR. For the data presented here, the SNR are about 10 or higher, and thus we should be able to quantify about 10-fold changes in relative protein levels, which encompasses most protein fold-changes between different human tissues.[24] This limitation is much less severe than the corresponding limitation for scRNA-seq (where the largest fold-changes are limited by the transcript copies sampled per mRNA) and should not constrain significantly data analysis and interpretation. Still, for some analysis, it may be desirable to increase the SNR. This can be achieved by increasing the ion accumulation time, increasing the resolving power of the orbitrap, or using a newer instrument.

Supplemental Figures

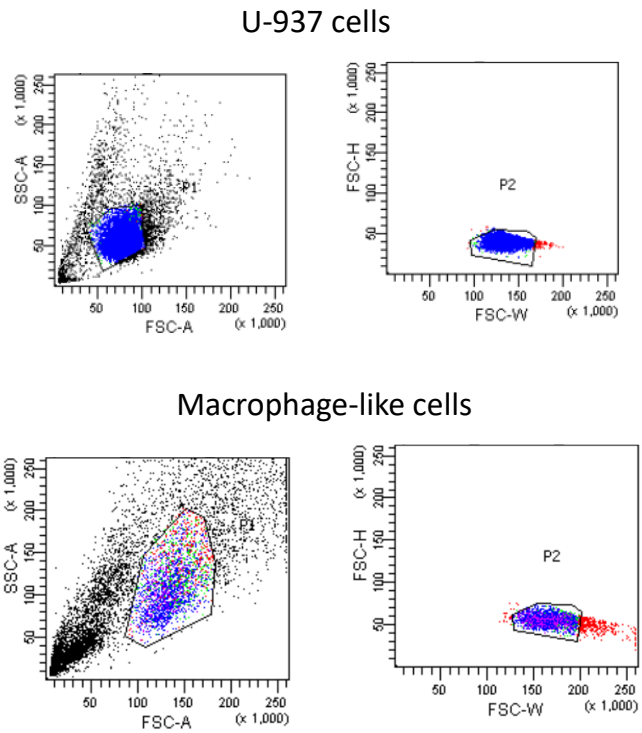


Figure S1 | FACS gating for sorting single monocyte and macrophage-like cells.
Representative FACS gatings are shown for the selection of live monocyte and macrophage-like cells, separating them from debris, dead cells, or granules. Secondary gating was used to decrease the probability of sorting doublets.

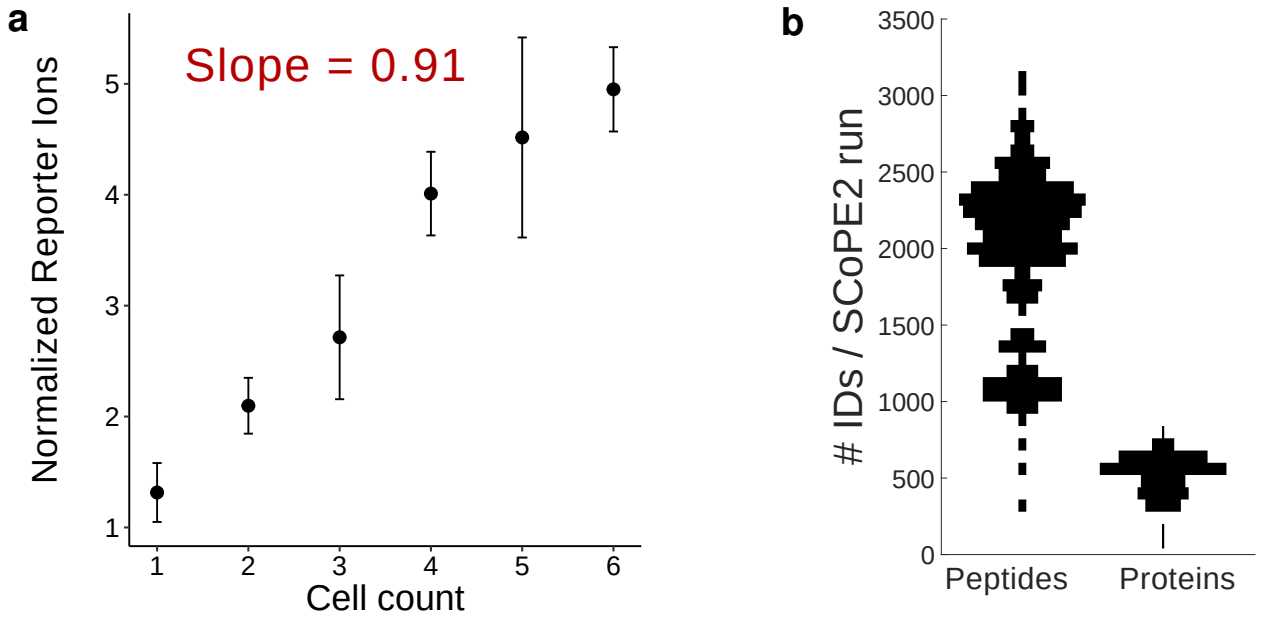


Figure S2 | Scaling of MS signal with protein input and peptide identification based on MS spectra. (**a**) A ladder of 1 to 6 single cell equivalents. The average signal measured per SCoPE2 channel scales linearly with its input. We performed 6 experiments and varied (randomised) the TMT labels used for different input amounts. The error bars denote standard deviations between replicates. (**b**) Number of peptides and proteins identified by MaxQuant per SCoPE2 set at 1% FDR. These identifications are based only on the MS spectra and do not incorporate retention time information.

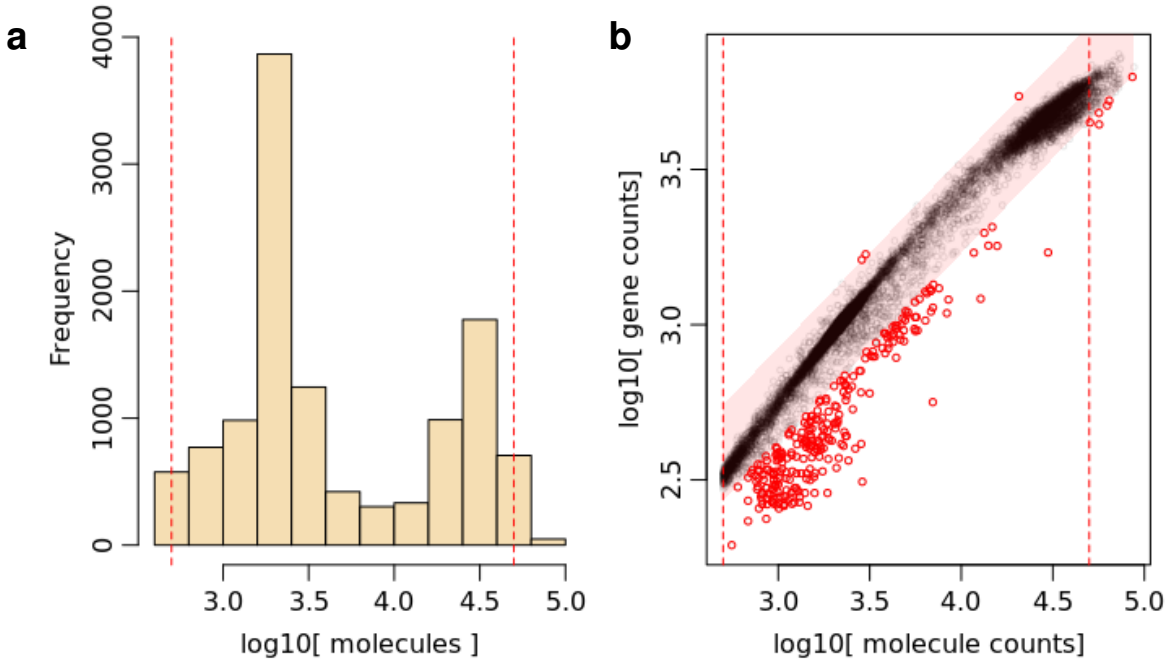


Figure S3 | Quality control plot for single-cell RNAseq data from the 10x Chromium platform.
(**a**) A distribution of number of unique molecular barcodes per cell. The left mode likely corresponds to empty droplets while the right mode corresponds to the single cells. All cells used for our analysis were sampled from the peak of the second mode. (**b**) A scatter plot of number of genes detected per cell versus the number of unique molecular barcodes per cell.

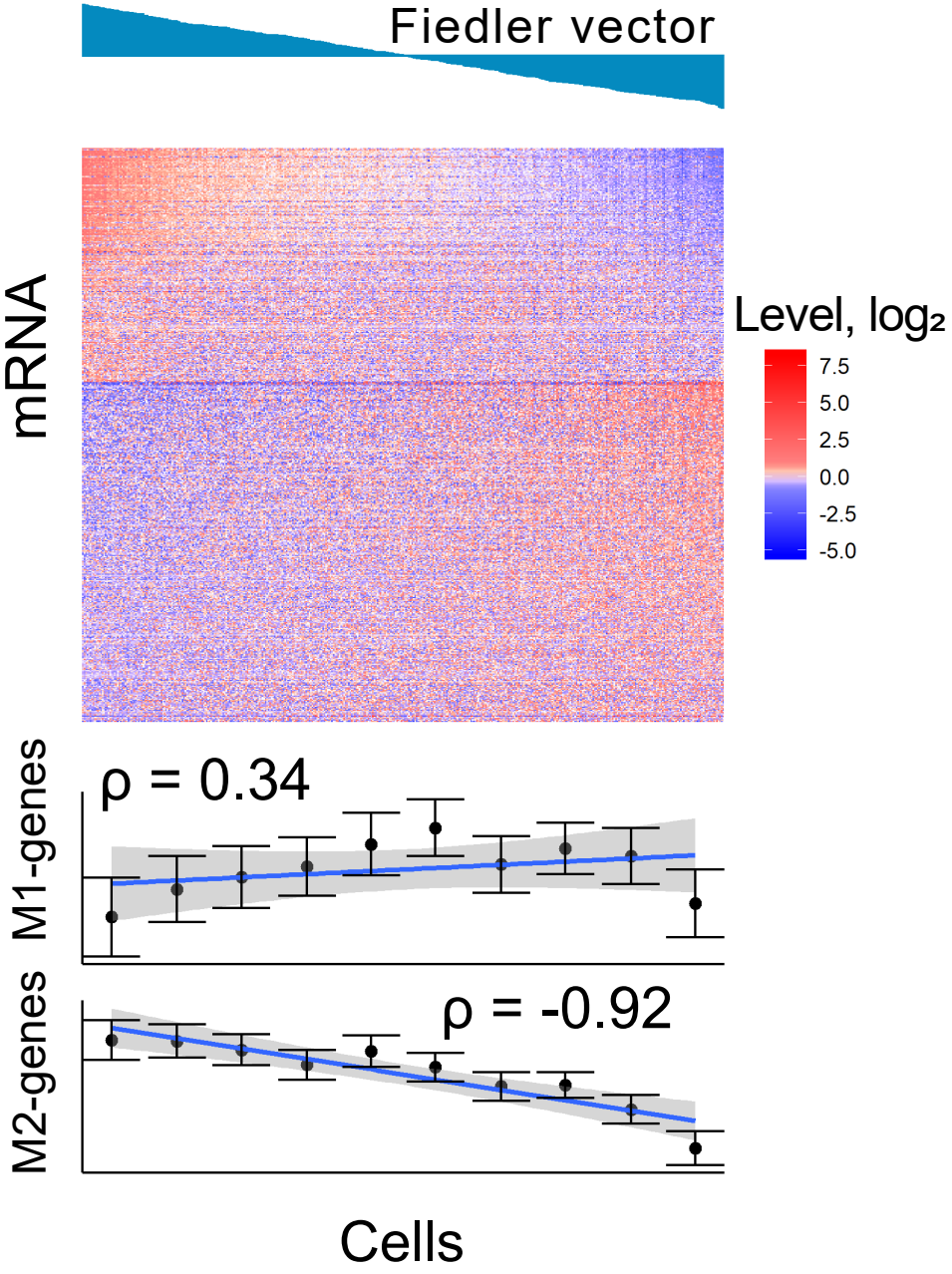


Figure S4 | Gradient of macrophage-like cells polarization observed in RNA data
Macrophage-like Single cells from the mRNA data set display a gradient of polarization. mRNA with cognate protein found in the single cell data were selected. The macrophage-like cells in the RNA data were selected using an arbitrary cutoff on the embedding from Fig. 5d. Displayed are the top 25% most variable mRNA (518) levels across 425 single cells. The level for each mRNA is relative to its mean level across the 425 single cells. The levels of genes in the selected mRNA data that are enriched in M1 or M2-polarized macrophages[27] are plotted below.

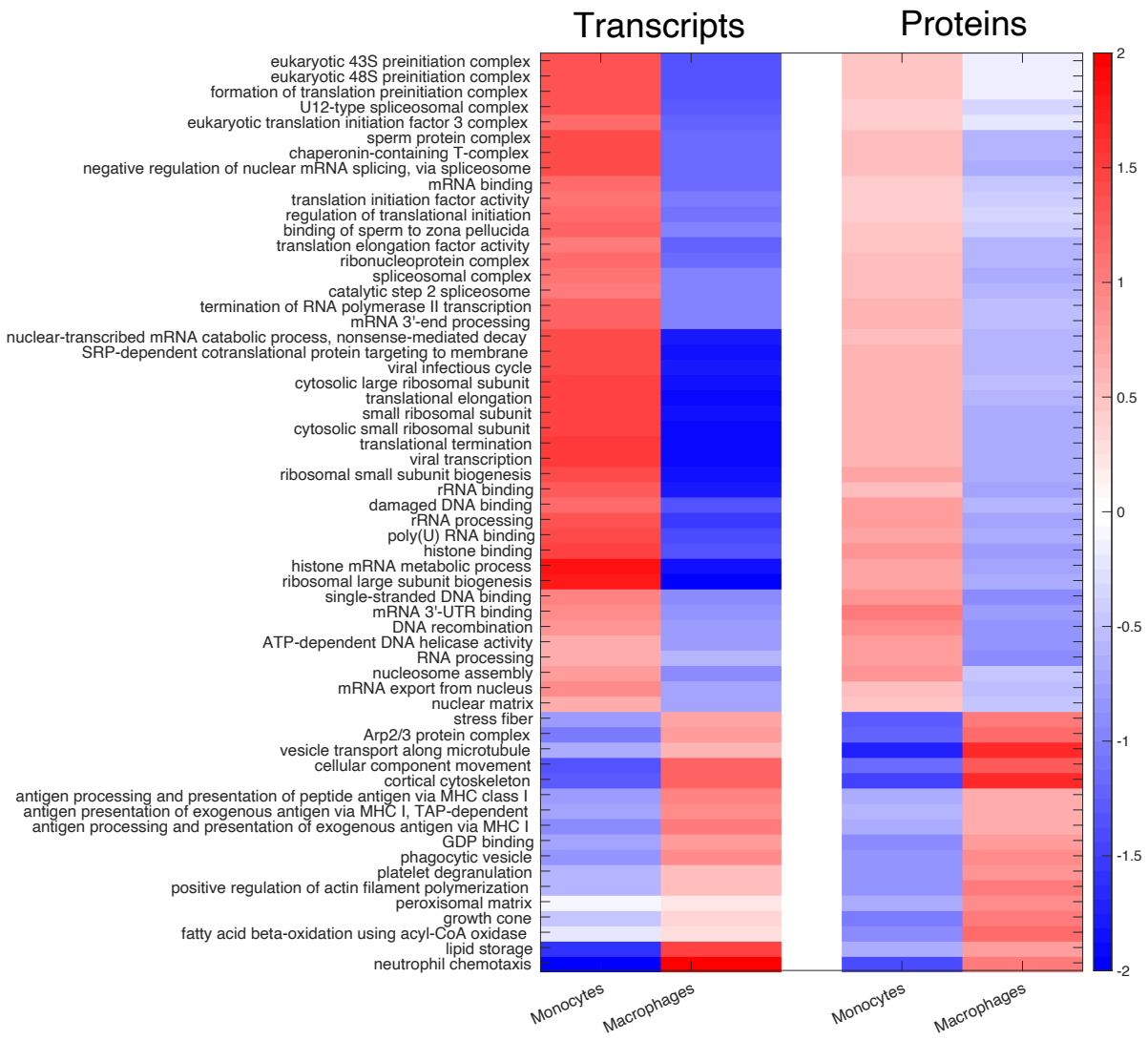


Figure S5 | Gene set enrichment analysis of cluster 1
Functional groups of genes that show differential abundance at the mRNA and protein levels in macrophages and monocytes from Cluster 1 in main Fig. 5. The colormap shows fold changes relative to the mean across all single cells on a log_2 scale.

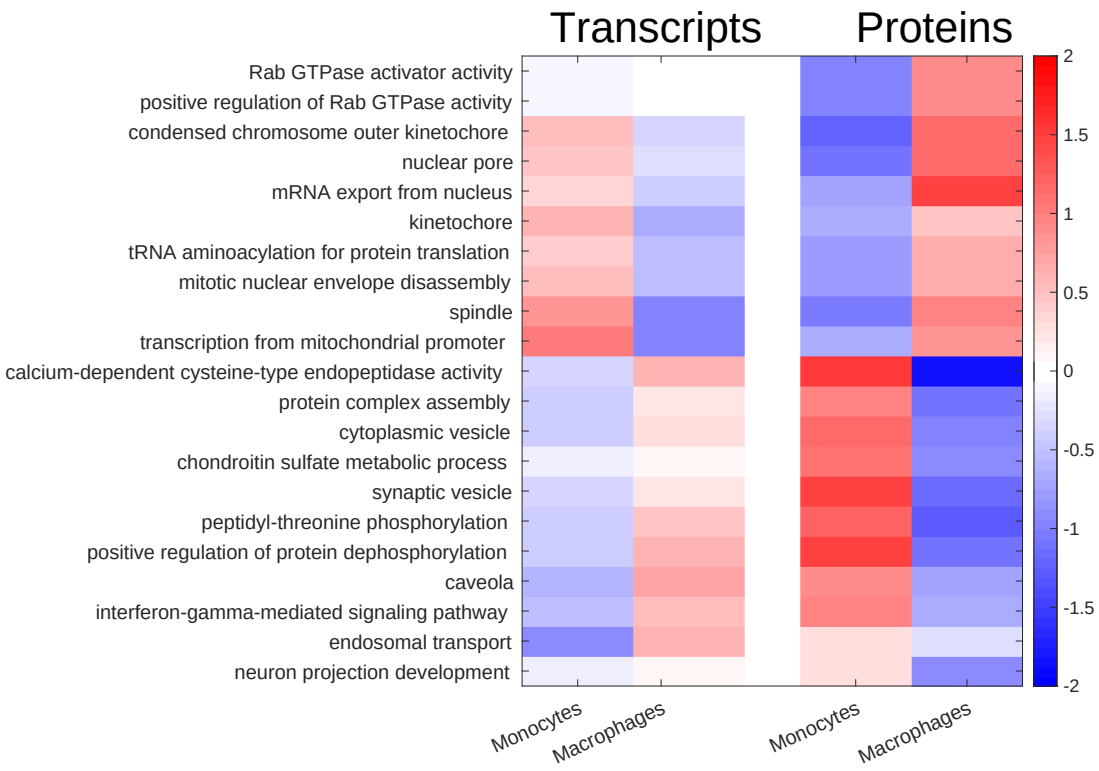


Figure S6 | Gene set enrichment analysis of cluster 2
Functional groups of genes that show differential abundance at the mRNA and protein levels in macrophages and monocytes from Cluster 2 in main Fig. 5. The colormap shows fold changes relative to the mean across all single cells on a log_2 scale.

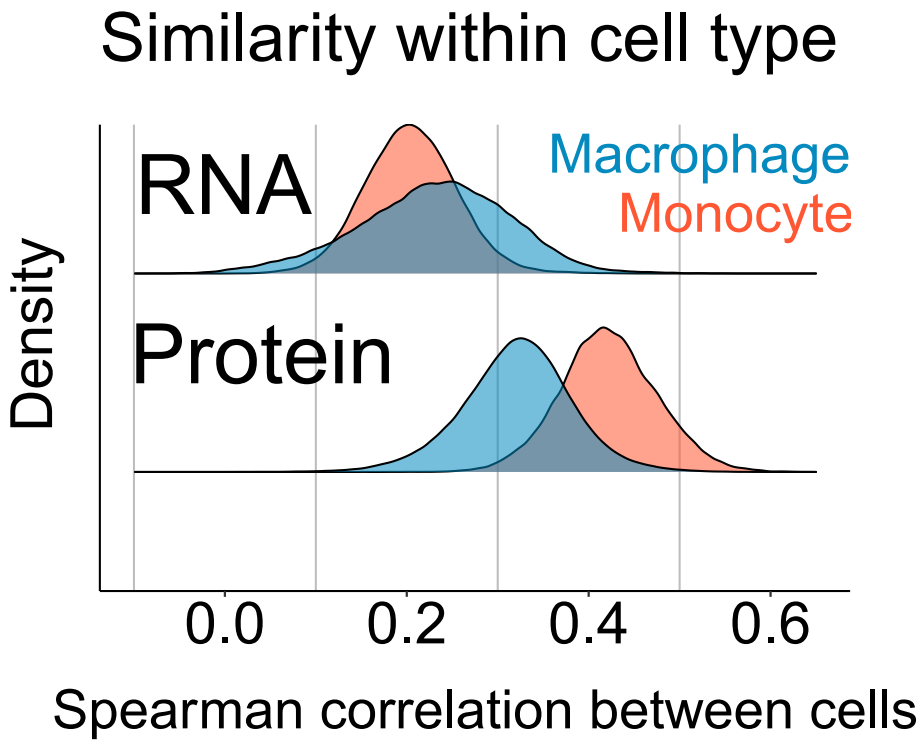


Figure S7 | Distributions of pairwise correlations within cell types.
To evaluate the homogeneity of monocytes and macrophages, we computed and plotted all pairwise Spearman correlations within each cell type, i.e., all pairwise correlations between monocytes and all pairwise correlations between macrophage-like cells. For the RNA data, which was sequenced as a mixed population of cells, cell type was assigned according to Fig. 5d. Correlations for the RNA data are computed from the fold-changes relative to the mean protein level across all cells (which is roughly 50% macrophage-like, 50% monocyte-like). For the protein data, the correlations are computed from the fold-changes relative to the reference so that they do not conflate variability between different proteins and emphasize the differences between the cells.[24] For proteins, the correlations between macrophage-like cells are significantly lower than between monocytes, consistent with the more dispersed macrophage cluster in Fig. 3a. For RNA, the correlations between macrophage-like cells display greater variability than the larger mode of the monocyte distribution. These results indicate that the macrophage-like cells are more heterogeneous than the monocytes from which they differentiate. The correlations between the proteins are slightly higher than between the RNAs which might reflect smaller biological and technical variability (noise) for the proteins.

References

1. Levy, E. & Slavov, N. Single cell protein analysis for systems biology. *Essays In Biochemistry* **62.** doi:10.1042/EBC20180014 (4 2018).
2. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics* **14,** 618 (2013).
3. Grün, D., Kester, L. & Van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature methods* **11,** 637 (2014).
4. Marx, V. A dream of single-cell proteomics. *Nature Methods* **16,** 809–812. ISSN: 1548-7105 (2019).
5. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nature Reviews Gastroenterology & Hepatology,* 1 (2019).
6. Specht, H. & Slavov, N. Transformative opportunities for single-cell proteomics. *Journal of Proteome Research* **17,** 2563–2916 (8 June 2018).
7. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic acids research* **38,** D750–D753 (2010).
8. Cravatt, B. F., Simon, G. M. & Yates Iii, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991 (2007).
9. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10,** 1794–1805 (2011).
10. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. en. *Nature* **537,** 347–355. ISSN: 1476-4687 (Sept. 2016).
11. Sinitcyn, P., Rudolph, J. D. & Cox, J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annu. Rev. Biomed. Data Sci.* **1,** 207–34 (2018).
12. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biology* **19,** 161 (2018).
13. Dou, M. *et al.* High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Analytical Chemistry* **91,** 13119–13127 (2019).
14. Schoof, E. M. *et al.* A Quantitative Single-Cell Proteomics Approach to Characterize an Acute Myeloid Leukemia Hierarchy. *bioRxiv.* doi:10.1101/745679 (2019).
15. Tan, Z., Yi, X., Carruthers, N. J., Stemmer, P. M. & Lubman, D. M. Single Amino Acid Variant Discovery in Small Numbers of Cells. *Journal of Proteome Research* **18,** 417–425 (2019).
16. Vitrinel, B., Iannitelli, D. E., Mazzoni, E. O., Christiaen, L. & Vogel, C. A simple method to quantify proteins from one thousand cells. *bioRxiv.* doi:10.1101/753582 (2019).
17. Yi, L. *et al.* Boosting to Amplify Signal with Isobaric Labeling (BASIL) Strategy for Comprehensive Quantitative Phosphoproteomic Characterization of Small Populations of Cells. *Analytical chemistry* **91,** 5794–5801 (2019).

18. Yang, L., George, J. & Wang, J. Deep Profiling of Cellular Heterogeneity by Emerging Single-Cell Proteomic Technologies. *PROTEOMICS* **n/a,** 1900226.

19. Specht, H. *et al.* Minimal sample preparation for high-throughput proteomics. *bioRxiv.* doi:10.1101/399774 (2019).

20. Huffman, G., Chen, A. T., Specht, H. & Slavov, N. DO-MS: Data-Driven Optimization of Mass Spectrometry Methods. *J. of Proteome Res.* doi:10.1021/acs.jproteome.9b00039 (2019).

21. Chen, A., Franks, A. & Slavov, N. DART-ID increases single-cell proteome coverage. *PLoS Comput Biol.* doi:10.1371/journal.pcbi.1007082 (2019).

22. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Reports* **6,** 13 (2014).

23. Ginhoux, F., Schultze, J. L., Murray, P. J., Ochando, J. & Biswas, S. K. New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nature Immunology.* **17,** 34–40 (2016).

24. Franks, A., Airoldi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS computational biology* **13,** e1005535 (2017).

25. Sundstrom, C. & Nilsson, K. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* **17,** 565–577 (1976).

26. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 (2009).

27. Martinez, F. O., Gordon, S., Locati, M. & Mantovani, A. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. eng. *Journal of Immunology (Baltimore, Md.: 1950)* **177,** 7303–7311. ISSN: 0022-1767 (Nov. 2006).

28. Slavov, N. & Dawson, K. A. Correlation signature of the macroscopic states of the gene regulatory network in cancer. *Proceedings of the National Academy of Sciences* **106,** 4079–4084 (2009).

29. Barkas, N. *et al.* Joint analysis of heterogeneous single-cell RNA-seq dataset collections. *Nature methods* **16,** 695–698 (2019).

30. Eiler, J. *et al.* Analysis of molecular isotopic structures at high precision and accuracy by Orbitrap mass spectrometry. *International Journal of Mass Spectrometry* **422,** 126–142 (2017).

31. Makarov, A. & Denisov, E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *Journal of the American Society for Mass Spectrometry* **20,** 1486–1495 (2009).

32. Abdulhadi, F. *Differentiation of U-937 Monocytes to Macrophage-Like Cells Polarized into M1 or M2 Phenotypes According to Their Specific Environment: A Study of Morphology, Cell Viability, and Cd Markers of an In Vitro Model of Human Macrophages.* MA thesis (Wright State University, 2014).

33. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26,** 1367–1372 (2008).

34. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols* **11,** 2301 (2016).

35. Coates, P. J., Rundle, J. K., Lorimore, S. A. & Wright, E. G. Indirect macrophage responses to ionizing radiation: implications for genotype-dependent bystander signaling. eng. *Cancer Research* **68,** 450–456. ISSN: 1538-7445 (Jan. 2008).