Published in final edited form as:

Anal Chem. 2019 April 02; 91(7): 4797–4805. doi:10.1021/acs.analchem.9b00345.

Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-cell Proteomics in Complex Tissues: Developing Cell Clones in Live *Xenopus laevis* and Zebrafish Embryos

Camille Lombard-Banek¹, Sally A. Moody², M. Chiara Manzini³, Peter Nemes^{1,2,*}

¹Department of Chemistry & Biochemistry, University of Maryland, College Park, MD 20742

²Department of Anatomy & Regenerative Biology, The George Washington University, Washington, DC 20052

³Department of Pharmacology & Physiology, The George Washington University, Washington, DC 20052

Abstract

Label-free single-cell proteomics by mass spectrometry (MS) is currently incompatible with complex tissues without requiring cell culturing, single-cell dissection, or tissue dissociation. We here report the first example of label-free single-cell MS-based proteomics directly in single cells in live vertebrate embryos. Our approach integrates optically guided *in situ* subcellular capillary microsampling, one-pot extraction/digestion of the collected proteins, peptide separation by capillary electrophoresis, ionization by an ultrasensitive electrokinetically pumped nano-electrospray, and detection by high-resolution MS (Orbitrap). With a 700-zmol (420,000 copies) lower limit of detection, this trace-sensitive technology confidently identified and quantified ~750–800 protein groups (<1% false discovery rate) by analyzing just ~5 ng of protein digest, viz. <0.05% of the total protein content from individual cells in the 16-cell *Xenopus laevis* (frog) embryo. After validating the approach by recovering animal-vegetal pole proteomic asymmetry in the frog zygote, the technology was used to uncover proteomic reorganization as the dorsal-animal (D11) cell of the 16-cell embryo gave rise to its neural-tissue fated clone in the embryo developing to the 32-, 64-, and 128-cell stage. In addition to enabling proteomics on smaller cells in *X. laevis*, we also demonstrated this technology to be scalable to single cells in live zebrafish embryos. Microsampling single-cell MS-based proteomics raises exciting opportunities to study cell and developmental processes directly in complex tissues and whole organisms at the (sub)level and of the building block of life: the cell.

Graphical Abstract

*Correspondence to: Peter Nemes, 0107 Chemistry Building, 8051 Regents Dr., College Park, MD 20742, USA; nemes@umd.edu, (Tel.) 301-405-0373.

Author contributions. P.N. and C. L.-B. designed the study. S.A.M. dissected embryonic cells and provided a portion of *Xenopus laevis* embryos. M.C.M. provided the zebrafish embryos. C. L.-B. measured the cells. P.N. and C.L.-B. analyzed the data, interpreted the results, and wrote the manuscript. All authors commented on the manuscript.

Competing Financial Interests. The authors have no competing financial interests.

Associated Content. The Supporting Information is available free of charge on the ACS Publications website at DOI: TBD. Protein identifications and quantification, results from Pearson correlation analysis, Student's *t*-test, and GProX analysis

Keywords

Single cell analysis; Proteomics; Microsampling; Mass spectrometry; Capillary electrophoresis; *Xenopus*

INTRODUCTION

Single-cell proteomics by mass spectrometry (MS) promises to revolutionize our understanding of cell biology. However, MS-based single-cell proteomics has so far been limited to cells in cultures or suspensions, typically obtained by dissection from tissues or dissociation of tissues or organisms. Additionally, these workflows present drawbacks that complicate results interpretation for biological investigations, such as the loss of cellular locational information that is important for cell identity, physiology, and/or the cellular microenvironment. Dissociation, dissection, or flow sorting can also perturb dynamic molecular processes via exposure to physical or chemical forces, as was recently evidenced at the level of the metabolome.^{1, 2} These strategies are also prone to contamination from neighboring cells and can be difficult to scale to smaller cells and complex tissues and organisms. Thus far, no single-cell MS technologies have been developed to enable the label-free detection of large numbers of proteins directly in complex tissues and whole organisms.

Single-cell mass spectrometry (MS) is a new frontier in cell and developmental biology and health studies by enabling the direct measurement of proteins, peptides, and metabolites in single cells. This field of single-cell MS was the focus of several reviews lately.^{3–14} At present, proteomics by MS pools together large numbers, usually thousands to millions, of cells to obtain deep to nearly complete coverage of proteomes.¹⁵ However, cell averaging inherently loses potentially important information specific to each cell, such as differential fates that embryonic cells assume during development of tissues, organs, and the organism. Without molecular amplification available for the whole proteome (or metabolome), technologies of single-cell MS must deliver exquisite detection sensitivity. Detection of α and β globulins in individual erythrocytes by capillary electrophoresis (CE)-MS^{16–18} pioneered MS proteomics for single cells. By using rare earth-metal tagged antibodies, the multiplexity of protein-targeted single-cell measurements were recently augmented to 23 different proteins using mass cytometry (CyTOF)¹⁹. Extension of single-cell proteomics to label-free operation would enable the detection of a large number of proteins, thus opening new potentials for cell and developmental biological studies.

We and others recently developed sufficiently sensitive label-free MS to enable proteomics on single isolated cells. Considerably sized cells in early developing embryos of the South African clawed frog (*X. laevis*) were amenable to manual dissection, followed by discovery (untargeted) proteomic measurement of each isolated cell using custom-built CE-MS instruments^{20, 21} and liquid chromatography (LC) MS^{22, 23}. For example, microanalytical CE-MS has enabled the detection of ~1,600 different protein groups between single cells that were dissected from 16-cell embryos of the frog, allowing us to discover previously unknown proteomic cell heterogeneity at such an early stage of development.^{20, 21} To process smaller cells, nanoPOTS (nanodroplet processing in one pot for trace samples platform) was recently introduced, enabling the identification of ~3,000 proteins from ~10 mammalian cells in culture,²⁴ ~670 protein groups from single HeLa cells and human lung primary cells,²⁵ and 164–607 protein groups from 1–5 LNCaP cells isolated from whole blood.²⁶

However, cell dissection, tissue dissociation, and culturing pose inherent limitations for many experiments in biology and health studies. Dissection is challenging to adapt to developing systems, such as developing embryos in *X. laevis*, zebrafish, or mouse, in which cell division leads to rapidly shrinking cell dimensions, cell adhesion is strengthened in later stages of development, and morphogenic movements can cause translocation of cells over considerable distances. Dissection/dissociation is also prone to contamination by materials from neighboring cells, calling for careful validation of the results. As already mentioned earlier, physical/chemical processing can also alter molecular states, complicating results interpretation for single cells.² Another challenge for biological studies arises from collateral damage to other cells and tissues, which in turn hinders the study of developing tissues and organs. Because the relationship between the transcriptome and proteome is complex during development and posttranslational modifications vastly expand protein heterogeneity²⁷, there is a high need for the direct characterization of the single-cell proteome to complement single-cell transcriptomics data during vertebrate embryonic development. At present, there exists no label-free single-cell proteomics MS technology capable of characterizing hundreds to thousands of proteins in single cells directly in complex tissues and whole organisms, including developing vertebrate embryos.

To fill this technological gap, we here report MS-based single-cell proteomics of identified cells in live vertebrate embryos of important model organisms. This report follows our initial disclosure of the technology in 2017²⁸ with the optimized workflow disseminated in 2018²⁹. Instead of averaging cells to boost detection sensitivity, as is classical to MS, or dissecting cells, as currently possible in single-cell MS by us and others,^{20–22} we microfabricated capillaries to aspirate a portion of identified cells directly from live embryos. Proteins from the collected aspirate were characterized label-free using a custom-built CE-ESI-MS system capable of ultrasensitive detection. As an example, we demonstrated the confident quantification of 450–800 different protein groups (<1% false discovery rate) in single cells in cleavage- and blastula-staged *Xenopus laevis* and zebrafish embryos, two popular vertebrate models in cell and developmental biology. Furthermore, for the first time in MS and developmental biology, we also quantified changes in the expression of hundreds of proteins as a single identified cell divided to give rise to its neural-tissue fated cell clone in the live frog embryo. The single-cell proteomics data from this work complement those

already accessible by single-cell transcriptomics, thus opening the door to systems biology studies of cell developmental processes.

METHODS

Reagents

Chemicals, solvents, and TCPK-modified trypsin were purchased in reagent grade or higher purity from Fisher Scientific (Pittsburg, PA). Solutions were prepared in LC-MS grade methanol, acetonitrile, water, formic or acetic acids. Bare fused silica capillaries (40/110 μm inner/outer diameter) were obtained from Polymicro technologies (Phoenix, AZ) and used without modification. Microcapillaries for microsampling and nanoESI were fabricated from borosilicate glass capillaries (0.75/1 mm inner/outer diameter) from Sutter Instruments (Novato, CA).

Solutions

For culturing *X. laevis* embryos, 100% *Steinberg's solution* (pH 7.4) was prepared following an established protocol.³⁰ For culturing zebrafish embryos, the *E2 media* was prepared following an established protocol.³¹ To extract proteins, the *lysis buffer* was prepared to contain 1% (w/v) SDS, 20 mM Tris-HCl, 5 mM EDTA, and 150 mM NaCl in reverse-osmosis water. The *CE background electrolyte* contained 25% (v/v) MeOH and 1 M formic acid. The CE-nanoESI *sheath solution* was prepared from 10% (v/v) MeOH with 0.5% (v/v) AcOH.

Animal Care and Embryo Handling

All protocols for the maintenance and handling of *Xenopus laevis* or zebrafish (*Danio rerio*) embryos were approved by the Institutional Animal Care and Use Committee of the University of Maryland (IACUC #R-DEC-17-57) or the George Washington University (IACUC #A311). Details regarding the processing of the embryos and the capillary microsampling and dissection of identified cells are provided in the SI document.

CE-nanoESI-HRMS

Digested proteins were measured in a custom-built CE platform that we recently described in detail.²⁰ In this study, ~10 nL of cell digest were electrophoretically separated in a ~90-cm fused silica capillary filled with the background electrolyte. The CE separation potential was +18 kV for 15 min and +15 kV thereafter until separation was stopped at 80 min. Most peptides were separated within a 30-min migration time window. Peptides were ionized in an electrokinetically pumped co-axial low-flow CE-nanoESI interface, which was constructed based on a recent design³². In this study, the spraying mode was operated in the cone-jet regime for efficient ion generation³³ by direct imaging of the hydrodynamic meniscus at the emitter tip using a long-working distance objective (Mitutoyo Plan Apo, Edmund Optics, Barrington, NJ) equipped with a CCD camera (EO-2018C, Edmund Optics). Further technical details are available in the SI document.

Safety

Standard chemical and biological safety protocols were followed. Electrospray emitters and capillaries, which pose puncture hazard, were handled with care. Electrically conductive parts of the CE-EK-nanoESI platform, which pose electrical shock hazard, were grounded or shielded in an interlock-enabled enclosure to prevent users from accidental exposure.

Data Analysis

To identify and quantify proteins (<1% false discovery rate), the MS¹–MS² data were analyzed using the Andromeda search engine³⁴ in MaxQuant version 1.5.5.1³⁵ against the *Xenopus* proteome (downloaded from UniProt in October 2017) or zebrafish proteome (downloaded in January 2018 from Uniprot.org). Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were performed in MetaboAnalyst version 3.0.³⁶ Cluster analysis was performed in GProX.³⁷ Gene ontology and pathway annotations were performed based in PANTHER^{38, 39}.

Data Repository

All RAW and processed MS–MS/MS data and the concatenated *Xenopus* reference proteome have been uploaded to the ProteomeExchange Consortium⁴⁰ via PRIDE under dataset identifier PXD006905.

RESULTS AND DISCUSSION

Technology Development

Our sampling strategy began with culturing embryos to a desired developmental stage. Cells were identified under a high-resolution stereomicroscope based on morphological cues in this study. Alternatively, endogenous/exogenous molecular markers may be used for cell identification. For example, Figure 1 highlights the neural-tissue fated midline animal-dorsal (D11) cell in the 16-cell *X. laevis* embryo based on cell size, cleavage pattern, and pigmentation in accord with established cell fate maps⁴¹ (see also electronic Supplementary Information, SI). To sample proteins from these cells, we adapted an optically guided capillary microsampling platform² to a remote-controlled precision three-axis translation stage, allowing for the membrane of the desired cell to be pierced with a borosilicate capillary (pulled to an ~20 µm diameter tip here). A pre-calibrated volume was aspirated from the cell using an online-hyphenated microinjector delivering negative pressure pulses to the capillary (–40 psi). While capillary microsampling has been utilized for metabolites in single cells in culture or tissues (reviewed in Refs. 3–14) and also recently in the *X. laevis* embryo by us², this work demonstrates the first example of capillary microsampling of single cells in a vertebrate embryo for MS-based proteomics. As an example, we extracted ~10 nL of the D11 cell, viz. ~10% of the total cell volume in this study. Based on protein assay data for whole embryos²⁰, this sample contains an ~1 µg of total protein. However, at this stage of development, ~90% of the cell's proteome is dominated by abundant yolk proteins (vitellogenins), which extrapolates to ~100 ng yolk-free protein in the aspirate. These protein amounts are ~10–100-times lower than those typically required for UPLC MS.^{22, 23} To enable the proteomic analysis of smaller cells that form at later stages in the

embryo, we purposefully measured only ~5 ng from the aspirated protein amount, or <0.05% of the total protein content of each identified cell in the 16-cell embryo. With scalable dimensions by capillary microsampling and ultrasensitive detection by CE-nanoESI-HRMS, our single-cell proteomic strategy should be applicable to other types of cells and models, including human cells.

To analyze the trace amounts of proteins collected, microprobe sampling was integrated to ultrasensitive CE-ESI-MS (Fig. 1a). Although this instrument recently enabled the detection of broad types of proteins in whole dissected *X. laevis* cells,²⁰ the miniscule amounts of proteins that were collected in this study required optimization of the bottom-up proteomic workflow. Microaspirated samples were pressure-ejected into individual microtubes, in which proteins were extracted and trypsin-digested in 5 μ L of buffer solution. To minimize/avoid potential peptide losses during sample handling, we here eliminated the typical bottom-up proteomic steps of reduction, alkylation, and protein purification. If desired, these steps can be implemented using limited amounts of reagents, as we recently demonstrated for dissected cells.^{20, 21} The digests were dried and reconstituted in 2 μ L of 60% acetonitrile (ACN) with 0.05% acetic acid (AcOH), chosen to preconcentrate peptides on-column via field-amplified sample stacking during CE. We deposited 500 nL of the resulting digest into a sample-loading microvial (see (I) in Fig. 1), whence ~10 nL, containing ~5 ng of total peptides, or ~0.05% of the total single-cell proteome, were loaded into the microanalytical CE platform (Fig. 1) following our established protocols.²⁰ Peptides were electrophoretically separated over an ~30-min window before ionization in an ultra-sensitive CE nano-flow ESI interface. This interface supplied a low-flow co-axial sheath flow around the CE separation capillary (Fig. 1) using an electrokinetic pump following a recent design.³² To further enhance ionization efficiency in this work, we used a long working-distance camera and employed ion current measurements to ensure that the electrospray was operated in the cone-jet spraying regime for efficient ionization³³ (see Tc in inset, Fig. 1). This approach enabled a lower limit of detection of ~700 zmol (420,000 copies) for model peptides (see Supplementary Figure 1, Fig. S1). Generated peptide ions were fragmented using a high-resolution tandem mass spectrometer equipped with a higher-energy collision induced dissociation cell (Q-Exactive Plus, Thermo Scientific) to identify proteins based on proteotypic peptides against the proteome of the species.

Benchmarking Microprobe CE-MS to Whole-cell Dissection

We streamlined and then benchmarked *microprobe CE-ESI-MS* against whole-cell dissection (Fig. 2), which is the closest neighboring single-cell proteomics technology for the *X. laevis* embryo.^{20–22} To enhance identification and quantification, different buffers were tested to extract and digest proteins: 50 mM ammonium bicarbonate (AmBic) containing no additive, 10% ACN, or 0.05% RapiGest (Waters Corp.) were compared to the reference approach, *viz.* whole-cell dissection with digestion in 1% SDS²⁰. To assess technical and biological variability, each experiment was repeated for $n = 3$ different D11 cells, each from a different embryo, with each sample analyzed in technical duplicate.

Microprobe single-cell CE-ESI-MS identified significantly more proteins than cell dissection. As shown in Figure 2A, the microprobe approach was most efficient with

digestion in AmBic without additives. The number of identified proteins is tabulated in Supplementary Table 1 (Table S1). Identified proteins are tabulated in Tables S2A–D. Analysis of ~5 ng total peptides from each cell by the microprobe approach yielded ~15,000 tandem mass spectra, including ~3,500 peptide spectral matches, which were confidently assigned to 360 different protein groups (<1% false discovery rate) per cell on average (426 proteins cumulative). Notably, in spite of consuming more than 4-times the amount of protein digest (20 ng), whole-cell dissection still identified fewer proteins (282 proteins/cell by dissection, $P = 0.008$, Fig. 1b). Additionally, microprobe CE-MS exclusively identified twice more proteins (compare Venn diagrams in Fig. 2). Gene ontology annotation of molecular processes and subcellular localization revealed no known differences between proteins that were identified by either of these approaches (see Fig. S2).

Furthermore, microprobe sampling also significantly improved protein quantification, despite an almost 10-fold reduction in starting protein amounts used in this workflow. As shown in Figure 2B (see numbers in Table S1), quantification was optimal using AmBic in the absence of additives: 341 proteins/cell (391 proteins cumulative) were quantifiable by optimized microprobe vs. 271 proteins/cell by cell dissection ($P = 0.004$) using label-free quantification (MaxLFQ⁴²). Quantified proteins are tabulated in Tables S2E–H. Based on calculated LFQ intensity values, the reproducibility of quantification was indistinguishable between microprobe CE-MS and whole-cell dissection ($P = 0.21$). As anticipated, biological variability significantly enhanced the variability of number of proteins that were quantified between cells (22.6% by microprobe vs. 16.6% RSD by CE-MS-only, $P = 5.7 \times 10^{-8}$, Fig. 2B). The observed significant sensitivity improvements may be credited to a combination of factors: Compared to cell dissection, microprobe CE-MS enhanced peptide separation by providing ~55% more theoretical plate numbers (118,000 vs. 183,000, respectively) and enhanced peptide ionization by efficiently reducing interferences from abundant salts² contained in the embryo culture media.

Technology Validation

To authenticate microprobe CE-ESI-MS, we asked whether the technology was able to find proteomic differences between the animal and vegetal poles of the *Xenopus laevis* zygote, viz. the fertilized 1-cell embryo (see Fig. 3A), which harbor known molecular differences at the level of transcripts and some proteins. An ~10 nL volume was aspirated at random positions in each hemisphere from $n = 4$ different zygotes with each sample analyzed in technical duplicate. A total of 752 proteins were identified and 485 proteins were reproducibly quantified using LFQ between each sample using CE-MS (see identification numbers in Table S3). Identified and quantified proteins are listed in Tables S4A and S4B, respectively.

These quantitative metadata were used to compare proteomic profiles between the poles. Figure 3B shows results from the unsupervised principal component analysis (PCA) the calculated LFQ intensities considering the two most important principal components (PC), which accounted for ~55% of total variance in the data. PCA loadings are listed for each quantified protein in Table S4C. Data points corresponding to the animal and vegetal samples formed clusters that were clearly distinguishable from each other, indicating

systematic proteomic differences between the two poles (see scores plot in top panel). The PCA loadings plot informed us of proteins that were comparably expressed between the poles (proteins with $PC1 = \sim 0$) as well as those that were enriched in the vegetal ($PC1 >> 0$) or animal ($PC1 << 0$) pole (see gradients for guidance in **bottom panel**). Proteins with finer differences in protein expression may be recognized along PC2.

Furthermore, the observed protein differences were statistically significant for several gene products, including some with known intracellular asymmetry in the zygote. As shown in Figure 3C, enrichment was statistically ($P < 0.05$) and biologically (fold change > 1.3) significant for 115 proteins. The results of statistical analysis are tabulated in Table S4D. For example, basic transcription factor 3l4 (Btf3l4), elongation factor 1-gamma (Eef1g), glyceraldehyde 3-phosphate dehydrogenase (Gaped), and ~40 different ribosomal proteins (e.g., Rpl4) were more abundant in the animal hemisphere (see **bottom panel**), which are known to be enriched in the animal cap of the embryo.⁴³ Conversely, perilipin 2 (Plin2) and vitellogenin chains (Vtg a1, a2, b1, and b2) were enriched in the vegetal pole (see **bottom panel**) in concordance with previous studies on transcriptomic or proteomic heterogeneity in the zygote.⁴³ These results demonstrate the ability of microprobe CE-ESI-MS to recover known or expected subcellular proteomic heterogeneity in the zygote, essentially validating the technology. In principle, thereby, our technology is scalable to smaller cells and the three-dimensionally complex structure of the embryo as its cells divide, which we also demonstrated below for a cell clone.

Single-cell Proteomics in a Developing *X. laevis* Cell Clone

Next, we asked using the microprobe CE-MS technology whether the protein signature of a single identified cell changes as a given cell divides to form its descendent cells (called a clone). Detection of such changes would be intriguing, because currently there is no information available on the proteomic state of cells in the early clone, before zygotic transcription begins with the midblastula transition. As model, we selected the clone that descends from the midline animal-dorsal cell (D11) in the 16-cell *X. laevis* embryo⁴¹ through 4 consecutive cell divisions (see Fig. 4A); this group of cells reproducibly gives rise to neural tissue in the frog.^{41, 44} We microprobed the left D11 cell in $n = 5$ different 16-cell embryos and its vegetally located daughter cell (left D112) in $n = 7$ different 32-cell embryos, which we were able to accurately identify by fate mapping criteria (see earlier and Methods in the SI document). To test for cell heterogeneity over time, we also microanalyzed daughter cells of D112 in $n = 6$ different 64-cell embryos and grand-daughter cells of D112 in $n = 7$ different 128-cell embryos. Because the cleavages become irregular at these stages, the sampled cells were randomly chosen from the position of D112. Each analyzed cell was assigned an identifier (#1 to #25, Fig. 4B) in the order of sampling, although this information and cell identity were blinded during initial data analysis and revealed only to help results interpretation. About 1–5 ng of protein digest, or 0.006–0.25% of the total protein content, was analyzed in triplicate from each cell in the clone.

About 791 protein groups were identified between the single cells, providing rich information on the proteomic state of the cells upon formation of the clone. Identification numbers are summarized for each cell type in Table S5. Identified proteins are tabulated in

Table S6A. Based on gene ontology annotation (PantherDB), 26 of these proteins are related to developmental process, such as the homeobox protein C13 (HoxC13), stathmin 1 (STMN1), and adenylate cyclase associated proteins 1 (Cap1). Moreover, our data contained 14 transcription factors, such as high-mobility group protein b2 (Hmgb2), homeobox cut-like 1 (Cux1), and HoxC13. These results highlight notable sensitivity improvements for cell and developmental biology studies using single-cell MS, which so far required ~100 ng protein digest to identify ~450 protein groups from human oocytes²³ and several micrograms of proteins to identify ~4,000 proteins between multiple whole cells dissected from *X. laevis* embryos²².

We used label-free quantification (LFQ) to compare the proteomic state of the cells. About 460 protein groups had quantifiable reads using LFQ among at least half of the biological replicates at every developmental stage from the 16- to the 128-cell. The number of quantified proteins is compared in Table S5. Proteins that were identified and quantified are tabulated in Table S6A and Table S6B, respectively. For each identified protein, we calculated the LFQ intensity (MaxLFQ), which serves as a quantitative proxy for protein concentration.²⁰ These quantitative data were median-normalized and log-transformed for hierarchical cluster analysis (HCA). Figure 4B presents the HCA-heat map calculated for the 75 statistically most significant proteins. Intriguingly, the dendrogram clusters the samples (Cells #1 to #25) based on cell type and developmental stage (see horizontal axis); early cleavage-stage embryos (16- and 32-cells), in which all cells are on the surface, were distinguished from late blastula stages (64- and 128-cells), in which cell division generate internal cells. Likewise, the proteins are categorized based on expression profiles across the cell types (see vertical axis). For example, compared to the mean, expression levels were higher for proteins in group A in the 16- and 32-cell embryo, in group B in the 32- and 128-cell embryo, and in group C in the 64- and 128-cell embryo, whereas proteins in group D exhibited complex enrichment profiles across the clone.

In addition to revealing protein profiles for the cells, these data also informed us of detectable differences between cells of the D11 clone at the same stage of embryonic development. For example, closer inspection of the HCA plot revealed D11 cells #1, #11, and #16 of the 16-cell embryo to be different than the other cells for select proteins (see asterisks). The frequency of cell-to-cell variance over the clone appears to increase with development, as exemplified for cells #4, #5, #20, #22, #24 over their siblings in the 128-cell embryo (see asterisks). Future experiments may address a potential link between intra cell type heterogeneity (same developmental stage here) and divergent tissue specification from cells of the same clone (e.g., the D11 clone gives rise to the retina, brain, spinal cord, cement gland, olfactory placode⁴¹).

Proteomic profiles were extracted for each cell-stage. High Pearson correlations ($\rho = 0.81$ – 0.93) between the cells in the 16-cell embryo (Fig. S3) and 128-cell embryo (Fig. S4) confirmed sufficient quantitative repeatability across the biological replicates to explore protein patterns. Figure 4C groups proteins with similar expression trends into four clusters based on fuzzy c cluster analysis of the quantitative metadata. Exact cluster assignments are listed for each protein in Table S6C. Of the 456 different proteins quantified, 319 proteins presented with unique trends across the stages: Some proteins had strictly monotonal

increasing (cluster 1) or decreasing (see cluster 3) abundance in the measured cells of the clone, whereas others presented with transient enrichment only at the 32-cell (cluster 2) or 64-cell (cluster 4) stage. These results provide previously unavailable data on proteomic reorganization as cells begin their eventual restriction to neural-fated tissue.

The proteomic changes were evaluated for statistical significance. As reference, we selected the D11 cell, the precursor cell of the clone, and calculated fold changes for the 456 proteins that were reproducibly quantified from the 128-cell clone. Calculated ratios and corresponding statistical significance (P value, Student's t -test) are plotted in Figure 4D. Statistical significance and fold change values are reported for each quantified protein in Table S6D. While protein production was statistically nonvariant for 365 protein groups ($P > 0.05$), expression differences were significant for 91 proteins ($P \leq 0.05$). For example, the D11 cell rather than its 128-cell descendant cell was enriched in fascin actin-bundling protein 1 (Fscn1) and stathmin 1 (Stmn1), both of which promote disassembly of microtubules, and the transcription factors high mobility group box 2 (Hmgb2) and parkinson protein 7 (Park7).⁴³ In contrast, compared to the D11, the level of expression increased in the analyzed cell of the clone for the microtubule associated protein family member 1 (Mapre1, promotes cytoplasmic microtubule nucleation and elongation), the proliferating cell nuclear antigen (Pcna, involved in DNA replication), and the macrophage migration inhibitory factor (Mif) and prohibitin1 (phb1), which both are required for normal neural tissue development.⁴³ Combined, these results established reproducible and significant proteomic differences between cells as the clone developed. Detection of proteomic changes over development of the clone is surprising, because the D11 clone arises from the same founder cell, thus inheriting similar maternal cytoplasmic components, and there is little transcription known at such an early stage of development. This information was obtainable only using the quantitative microanalytical technology that we developed here, which in principle is scalable to any other clone in the embryo.

Scalable Single-cell Proteomics to other models: Live Zebrafish Embryos

Last, we demonstrated that microprobe CE-ESI-MS is scalable to smaller single cells and a different model system, the zebrafish (*Danio rerio*) embryo (Fig. 5A). Unlike in *X. laevis*, the cleavage-stage zebrafish embryo lacks visual indicators of the primary body axes; hence, it is not possible to confidently identify the same cell type or track the clone of the same precursor cell type in different embryos. Without cell identification, it is still unknown whether cells of the zebrafish embryo execute different gene regulatory programs at the cleavage stage. Detection of protein expression in cells of zebrafish embryos would provide a direct window to assess potential cell-to-cell differences.

To tackle this question, we scaled microprobe CE-ESI-MS to measure protein expression in single cells of the 2-cell stage embryo. Challenging the targeted analysis of cells, the zebrafish embryo undergoes free rotation inside the vitelline membrane. As a solution, we immobilized the embryo by inserting the capillary into the cell of interest through the vitelline membrane and then the yolk sac, essentially pinning the embryo in place. We used the microprobe platform to aspirate ~5 nL, or 5 ng protein content, from one of the cells using a capillary pulled to ~5 μ m tip diameter (vs. 10 ng collected using 20 μ m capillary in

X. laevis). Although this provided sufficient protein amount for our microanalytical CE-ESI-MS instrument, we deliberately analyzed only 33 pg of the resulting protein digest (or 0.13% of the total cell proteome) to demonstrate scalability to even smaller cells.

Even with this trace amount of material analyzed, ~330 proteins were identified (Table S7A) and ~320 proteins were quantified by LFQ (Table S7B). Following annotation for GO term enrichment, we found that most proteins identified were involved in translation, transcription, and cell division, as expected during early embryonic development closely following fertilization.^{45, 46} Roughly 25% of the proteins identified were structural components of the ribosome (n = 69), while another significant group regulated protein translation and folding (n = 22), such as translation elongation factors 1 α , 1 β , 1 δ , 1 γ , and 2b (*eef1a*, *eef1b2*, *eef1da*, *eef1g*, *eef2b*) and chaperone heat shock protein 90 (*hsp90ab1*, *hsp90b1*). Critical enzymes involved in energy production in both glycolysis and citric acid cycle were also enriched (n = 28 proteins with oxidoreductase activity), including pyruvate carboxylase (*pox*), phosphoglycerate kinase (*pgk1*), NADH dehydrogenase 1 alpha (*ndufa13*), and aconitase2 (*aco2*). Finally, multiple proteins were involved in cytoskeletal regulation during cell division [e.g., actin regulators non-muscle cell myosin (*myh9a*), cofilin 1 (*cfl1*) and profilin 2 (*pfn2*)] and DNA replication [e.g. helicase (*mcm9*)]. Identification of proteins with developmental relevance showcases the potentials of microprobe CE-ESI-MS for the zebrafish embryo.

The quantitative metadata were used to compare protein expression between the cells. As proof-of concept, we analyzed n = 3 cells, each from a different embryo. With fast and scalable sample collection afforded by microprobe sampling, these measurements can be readily scaled-up to more cells for biologically motivated studies. Based on mean-normalized log-transformed LFQ intensities, protein abundances varied by ~47% RSD, which far exceeds the technical quantitative repeatability of our technology (~7% RSD, Fig. 2B) and the biological quantitative variability observed between the same cell type in *X. laevis* (~23% RSD, Fig. 2B). Therefore, these data suggest quantitative proteomic differences between the cells that were analyzed in the embryo.

Protein expression was compared between the cells using Pearson cross-correlation analysis. As shown in Figure 5B, the protein composition of cell #1 was highly correlated with that of #2 (Pearson correlation coefficient, $\rho = 0.91$). Correlation was lower with cell #3 ($\rho = 0.84$), and notably lower between cells #2 and #3 ($\rho = 0.72$), revealing marked proteomic differences between the cells. Proteins underlying the observed differences are readily identified outside the 95% confidence prediction band in the plots. For example, among the poorly correlated proteins were three proteins of the ovostatin homolog-1-related signaling molecules (Zgc:171445, Zgc:165518, and Zgc:171426) that are involved in cell signaling (PatherDB). While the biological origin of these differences is yet unknown to us, future studies may use our data to design targeted experiments to test the developmental significance of stably or variably expressed proteins.

CONCLUSIONS

In conclusion, we here developed capillary microsampling single-cell CE-ESI-MS to enable, for the first time, label-free proteomic characterization of (identified) single cells directly in live embryos. With scalability to diverse cell sizes and adaptability to different animal models, the technology is able to identify and quantify hundreds of proteins in cells, capturing previously unknown reorganization of the proteome as cells give rise to clones in developing embryos. In principle, the analyzed protein amounts estimate to those afforded by larger mammalian cells⁴⁷, raising exciting technical capabilities to study the proteomic aspects of cell and developmental processes in other cell types (e.g., germ cells, neurons, glial cells). Finally, it does not escape our attention that the microsampling approach should be adaptable to other MS technologies, including liquid chromatography and matrix-assisted laser desorption ionization, to measure the proteomic state of cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Aparna B. Baxi and Reem Al-Shabeeb for help with data processing. We also thank the peer reviewers for valuable comments and suggestions on this manuscript.

Funding. This work was partially funded by the National Institutes of Health Award 1R35GM124755 (to P.N.), the Arnold and Mabel Beckman Foundation Young Investigator Grant (to P.N.), the American Society for Mass Spectrometry Research Award (to P.N.), and the DuPont Company Young Professor Award (to P.N.). The content of the reported work is the sole responsibility of the authors and does not necessarily represent the official views of the funding sources.

Abbreviations

CE	capillary electrophoresis
HRMS	high-resolution mass spectrometry
ESI	electrospray ionization
LC	liquid chromatography
LFQ	label free quantification

References

1. Llufrio EM; Wang LJ; Naser FJ; Patti GJ, Redox Biology 2018, 16, 381–387. [PubMed: 29627745]
2. Onjiko RM; Portero EP; Moody SA; Nemes P, Anal. Chem 2017, 89, 7069–7076. [PubMed: 28434226]
3. Rubakhin SS; Romanova EV; Nemes P; Sweedler JV, Nat. Methods 2011, 8, S20–S29. [PubMed: 21451513]
4. Lanni EJ; Rubakhin SS; Sweedler JV, J. Proteom 2012, 75, 5036–5051.
5. Zenobi R, Science 2013, 342, 1243259. [PubMed: 24311695]
6. Rubakhin SS; Lanni EJ; Sweedler JV, Curr. Opin. Biotechnol 2013, 24, 95–104. [PubMed: 23246232]

7. Passarelli MK; Ewing AG, *Curr. Opin. Chem. Biol* 2013, 17, 854–859. [PubMed: 23948695]
8. Chen X; Love JC; Navin NE; Pachter L; Stubbington MJT; Svensson V; Sweedler JV; Teichmann SA, *Nat. Biotechnol* 2016, 34, 1111–1118. [PubMed: 27824834]
9. Acunha T; Simo C; Ibanez C; Gallardo A; Cifuentes A, *J. Chromatogr., A* 2016, 1428, 326–335. [PubMed: 26296988]
10. Yang YY; Huang YY; Wu JH; Liu N; Deng JW; Luan TG, *Trends Anal. Chem* 2017, 90, 14–26.
11. Comi TJ; Do TD; Rubakhin SS; Sweedler JV, *J. Am. Chem. Soc* 2017, 139, 3920–3929. [PubMed: 28135079]
12. Yin L; Zhang Z; Liu Y; Gao Y; Gu J, *Analyst* 2018, In print, DOI: 10.1039/C8AN01190G.
13. Zhang LW; Vertes A, *Angew. Chem. Int. Ed* 2018, 57, 4466–4477.
14. Specht H; Slavov N, *J. Proteome Res* 2018, 17, 2565–2571. [PubMed: 29945450]
15. Zhou F; Lu Y; Ficarro SB; Adelmant G; Jiang WY; Luckey CJ; Marto JA, *Nat. Commun* 2013, 4, 2171. [PubMed: 23863870]
16. Valaskovic GA; Kelleher NL; McLafferty FW, *Science* 1996, 273, 1199–1202. [PubMed: 8703047]
17. Hofstadler SA; Swanek FD; Gale DC; Ewing AG; Smith RD, *Anal. Chem* 1995, 67, 1477–1480. [PubMed: 7741217]
18. Mellors JS; Jorabchi K; Smith LM; Ramsey JM, *Anal. Chem* 2010, 82, 967–973. [PubMed: 20058879]
19. Bendall SC; Simonds EF; Qiu P; Amir EAD; Krutzik PO; Finck R; Bruggner RV; Melamed R; Trejo A; Ornatsky OI; Balderas RS; Plevritis SK; Sachs K; Pe'er D; Tanner SD; Nolan GP, *Science* 2011, 332, 687–696. [PubMed: 21551058]
20. Lombard-Banek C; Moody SA; Nemes P, *Angew. Chem. Int. Ed* 2016, 55, 2454–2458.
21. Lombard-Banek C; Reddy S; Moody SA; Nemes P, *Mol. Cell. Proteomics* 2016, 15, 2756–2768. [PubMed: 27317400]
22. Sun LL; Dubiak KM; Peuchen EH; Zhang ZB; Zhu GJ; Huber PW; Dovichi NJ, *Anal. Chem.* 2016, 88, 6653–6657. [PubMed: 27314579]
23. Virant-Klun I; Leicht S; Hughes C; Krijgsveld J, *Mol. Cell. Proteomics* 2016, 15, 2616–2627. [PubMed: 27215607]
24. Zhu Y; Piehowski PD; Zhao R; Chen J; Shen YF; Moore RJ; Shukla AK; Petyuk VA; Campbell-Thompson M; Mathews CE; Smith RD; Qian WJ; Kelly RT, *Nat. Commun* 2018, 9, 10. [PubMed: 29295980]
25. Zhu Y; Clair G; Chrisler WB; Shen YF; Zhao R; Shukla AK; Moore RJ; Misra RS; Pryhuber GS; Smith RD; Ansong C; Kelly RT, *Angew. Chem.-Int. Edit* 2018, 57, 12370–12374.
26. Zhu Y; Podolak J; Zhao R; Shukla AK; Moore RJ; Thomas GV; Kelly RT, *Anal. Chem* 2018, 90, 11756–11759. [PubMed: 30269481]
27. Vogel C; Marcotte EM, *Nat. Rev. Genet* 2012, 13, 227–232. [PubMed: 22411467]
28. Lombard-Banek C; Baxi AB; Moody SA; Nemes P In *Microsampling for proteomic characterization of increasingly smaller cells in the vertebrate (frog) embryo*, The American Society for Mass Spectrometry, Indianapolis, IN, Indianapolis, IN, 2017; p Abstract ID 287805.
29. Lombard-Banek C; Onjiko RM; Plotnick DO; Shabeeb RQA; Moody SA; Nemes P In *Single-cell proteomics and metabolomics using microprobe CE-ESI-HRMS: towards single-cell systems biology*, The American Society for Mass Spectrometry, San Diego, CA, San Diego, CA, 2018; p Abstract ID 291375.
30. Sive HL; Grainger RM; Harland RM, *Early development of Xenopus laevis: A laboratory manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2000.
31. Dahm R; Nüsslein-Volhard C, *Zebrafish: A practical approach*. Oxford University Press: New York, 2002.
32. Sun LL; Zhu GJ; Zhao YM; Yan XJ; Mou S; Dovichi NJ, *Angew. Chem. Int. Ed* 2013, 52, 13661–13664.
33. Nemes P; Marginean I; Vertes A, *Anal. Chem* 2007, 79, 3105–3116. [PubMed: 17378541]
34. Cox J; Neuhauser N; Michalski A; Scheltema RA; Olsen JV; Mann M, *Proteome Res J* 2011, 10, 1794–1805.

35. Cox J; Mann M, Nat. Biotechnol 2008, 26, 1367–1372. [PubMed: 19029910]
36. Xia JG; Sinelnikov IV; Han B; Wishart DS, Nucleic Acids Res. 2015, 43, W251–W257. [PubMed: 25897128]
37. Rigbolt KTG; Vanselow JT; Blagoev B, Mol. Cell. Proteomics 2011, 10, 10.
38. Mi HY; Lazareva-Ulitsky B; Loo R; Kejariwal A; Vandergriff J; Rabkin S; Guo N; Muruganujan A; Doremioux O; Campbell MJ; Kitano H; Thomas PD, Nucleic Acids Res. 2005, 33, D284–D288. [PubMed: 15608197]
39. Mi HY; Muruganujan A; Casagrande JT; Thomas PD, Nat. Protoc 2013, 8, 1551–1566. [PubMed: 23868073]
40. Vizcaino JA; Csordas A; del-Toro N; Dienes JA; Griss J; Lavidas I; Mayer G; Perez-Riverol Y; Reisinger F; Ternent T; Xu QW; Wang R; Hermjakob H, Nucleic Acids Res. 2016, 44, D447–D456. [PubMed: 26527722]
41. Moody SA, Dev. Biol 1987, 119, 560–578. [PubMed: 3803718]
42. Cox J; Hein MY; Luber CA; Paron I; Nagaraj N; Mann M, Mol. Cell. Proteomics 2014, 13, 2513–2526. [PubMed: 24942700]
43. Karimi K; Fortriede J; Lotay V; Burns K; Wang D; Fisher M; Pells T; James-Zorn C; Wang Y; Ponferrada V; Chu S; Chaturvedi P; Zorn A; Vize P, Nucleic Acids Res. 2018, 46, D861–D868. [PubMed: 29059324]
44. Jacobson M, J. Neurosci 1983, 3, 1019–1038. [PubMed: 6842280]
45. Aanes H; Winata CL; Lin CH; Chen JQP; Srinivasan KG; Lee SGP; Lim AYM; Hajan HS; Collas P; Bourque G; Gong ZY; Korzh V; Alestrom P; Mathavan S, Genome Res. 2011, 21, 1328–1338. [PubMed: 21555364]
46. Vesterlund L; Jiao H; Unneberg P; Hovatta O; Kere J, BMC Developmental Biology 2011, 11, 18. [PubMed: 21401944]
47. Cohen D; Dickerson JA; Whitmore CD; Turner EH; Palcic MM; Hindsgaul O; Dovichi NJ, Chemical cytometry: Fluorescence-based single-cell analysis In Annu. Rev. Anal. Chem, Annual Reviews: Palo Alto, 2008; Vol. 1, pp 165–190.

Figure 1.

Label-free single-cell proteomics of an identified single cell in a live *Xenopus laevis* embryo. The example shows optical identification of the midline animal-dorsal cell (D11, white arrow) in the 16-cell embryo, whence ~20 nL of cell content were aspirated using capillary microsampling (microprobe). Proteins in the collected aspirate were extracted and digested. Peptides were separated in a microanalytical capillary electrophoresis (CE) platform and ionized in a custom-built electrokinetically pumped nano-electrospray ionization (EK-nanoESI) source that was monitored using a long-working distance camera for stable operation in the cone-jet regime (see Taylor cone, Tc) for efficient ionization. The resulting peptide ions were analyzed using a high-resolution Orbitrap mass spectrometer. Scale bars = 200 μm (black), 1 mm (white).

Figure 2. Benchmarking and optimization of technology performance against whole-cell dissection. **(A)** Comparison of the number of identified protein groups between microprobe (μ P) sampling in 50 mM ammonium bicarbonate (AmBic) containing 10% (v/v) acetonitrile (ACN), 0.05% (v/v) RapiGest, and no additive (AmBic) vs. whole-cell dissection in (Diss). **(B)** Comparison of quantification based on the number of quantified proteins (**left panel**) and repeatability (relative standard deviation, STD, **right panel**) between the approaches. Protein numbers are summarized in Table S1 with proteins listed in Tables S2A–D for identification and Tables S2E–H for quantification. The reproducibility of the CE-MS platform is shown for reference (“CE-MS”). Box-whisker plots comparing expression of representative proteins. Key: * P 0.05, ** P 0.01, *** P 0.001.

Figure 3.

Technology validation using the *X. laevis* zygote. **(A)** Microprobe CE-HRMS of random areas in the animal (AN, **top panel**) and vegetal (VG, **bottom panel**) pole. Scale bars = 200 μ m. **(B)** Unsupervised principal component (PC) analysis of the quantitative data. Scores plot (**top panel**) revealing systematic proteomic differences between the poles (each data point marks a different cell). The loadings plot (**bottom panel**) measuring the contribution of each quantified protein to the observed differences (each data point marks a different protein). Representative proteins (gene names) are highlighted. Gradients illustrate the observed abundance trends between the poles. Horizontal and vertical dashed lines mark PC1 = PC2 = 0 coordinates. **(C)** Statistical analysis revealing significant proteomic differences between the poles. Box-whisker plots highlight representative proteins (see text). Key: * P 0.05, ** P 0.01, *** P 0.001.

Figure 4.

Discovery single-cell proteomics of a neural-tissue fated clone in developing *X. laevis* embryos. **(A)** Optical imaging of microprobed cells in the left dorsal-animal cell (D11) lineage. Scale bar = 200 μ m. **(B)** Hierarchical cluster analysis-heat map of quantitative protein abundances considering the 75 most significant features (ANOVA). Embryonic developmental stages (top axis) and individual cell identifiers (bottom axis) are shown. Asterisks (*) mark examples of proteins with quantifiable cell heterogeneity in the 16-cell and 128-cell clone. **(C)** GProX cluster analysis of protein profiles showing 103 proteins in Cluster #1, 107 proteins in Cluster #2, and 109 proteins in Cluster #3 (see data in Table S6C). **(D)** Relative quantification of protein production between microprobed cell types revealing significant enrichment differences (see data in Table S6D). Dashed lines mark statistical ($P = 0.05$) and biological significance (fold change, FC = 1.5).

Figure 5.

Scalable single-cell proteomics for the zebrafish embryo at 2-cell stage. (A) Microsampling of 5 nL (~5 ng protein) from a cell of interest directly in the living embryo by piercing through the vitelline membrane (Vit.) and the yolk sac. Scale bar = 200 μ m. (B) Pearson correlation analysis (correlation coefficient, ρ) of detected protein abundances between $n = 3$ different single cells, each from a different embryo, indicating proteomic differences

between the aspirated samples. Each cell was marked with an identifier number (#) for reference.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript