

Proteomics

International Edition: DOI: 10.1002/anie.201510411

German Edition: DOI: 10.1002/ange.201510411

Single-Cell Mass Spectrometry for Discovery Proteomics: Quantifying Translational Cell Heterogeneity in the 16-Cell Frog (*Xenopus*) Embryo

Camille Lombard-Banek, Sally A. Moody, and Peter Nemes*

Abstract: We advance mass spectrometry from a cell population-averaging tool to one capable of quantifying the expression of diverse proteins in single embryonic cells. Our instrument combines capillary electrophoresis (CE), electrospray ionization, and a tribrid ultrahigh-resolution mass spectrometer (HRMS) to enable untargeted (discovery) proteomics with ca. 25 amol lower limit of detection. CE- μ ESI-HRMS enabled the identification of 500–800 nonredundant protein groups by measuring 20 ng, or <0.2% of the total protein content in single blastomeres that were isolated from the 16-cell frog (*Xenopus laevis*) embryo, amounting to a total of 1709 protein groups identified between $n=3$ biological replicates. By quantifying ≈ 150 nonredundant protein groups between all blastomeres and replicate measurements, we found significant translational cell heterogeneity along multiple axes of the embryo at this very early stage of development when the transcriptional program of the embryo has yet to begin.

Empowering high-resolution mass spectrometry (HRMS) to the characterization of the proteome in single cells is a long-standing challenge that promises to deepen our understanding of how functional differences between cells fundamentally contribute to health and disease.^[1] HRMS is label-free, quantitative, and allows for the unbiased measurement of the deep encoded proteome by averaging multiple, usually, millions of cells.^[2] However, to transform HRMS from a cell population-averaging tool into a “single-cell proteomic analyzer”, new approaches are needed with exceptional sensitivity and reproducibility. The prospect of proteomic single-cell analysis was pioneered in the 1990s when capillary electrophoresis (CE) and Fourier transform MS was combined to detect α - and β -globulins in 5–10 human erythrocytes^[3] and

carbonic anhydrase^[4] in their lysates diluted to single cells. HRMS-based detection of globulins was accelerated to 12 erythrocytes/min using microfluidic devices^[5] and, most recently, 34 targeted proteins contributing to hematopoiesis were detected at ≈ 1000 cell/s throughput using mass cytometry.^[6] Continuous developments in HRMS sensitivity enabled untargeted (discovery) proteomics on progressively smaller populations of cells^[1c,7] with representative successes including the identification of 2000 proteins from 2000–4000 cells from single Langerhans islands,^[8] 167 proteins from 500 breast cancer cells,^[1a] 109 proteins from 100 HeLa cells,^[9] and most recently, ≈ 12000 proteins in ≈ 50 fertilized^[10] and a single unfertilized egg of the South African clawed frog (*Xenopus laevis*).^[11] Using advanced microsampling and discovery HRMS,^[1c,12] neuropeptides were characterized in single molluscan^[13] and arthropod neurons^[12] as well as cells in the mammalian pituitary gland and Langerhans islets.^[13a] However, discovery measurements addressing the deep proteome have yet to be developed for single cells. To address this challenge, we here develop a bottom-up HRMS platform and protocol that enables identification–quantitation of a large number of proteins in single embryonic cells.

The new single-cell proteomic analyzer integrates a custom-designed single-cell CE platform and a CE-microflow electrospray ion source (μ ESI) to a high-resolution tandem mass spectrometer in a bottom-up proteomic workflow (Figure 1), in which proteins are identified/quantified based on unique peptide sequences. This design was inspired based on a CE platform that we recently developed for small

[*] C. Lombard-Banek, Prof. P. Nemes
Department of Chemistry, W. M. Keck Institute for Proteomics
Technology and Applications
The George Washington University
800 22nd Street, NW, Suite 4000, Washington, DC 20052 (USA)
E-mail: petern@gwu.edu
Prof. S. A. Moody
Department of Anatomy and Regenerative Biology
The George Washington University
Washington, DC 20052 (USA)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201510411>.

© 2016 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

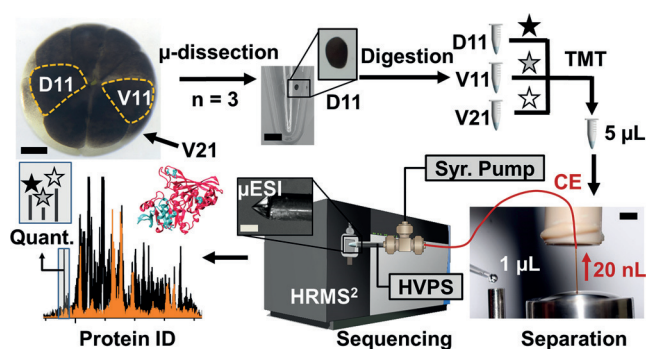


Figure 1. Microanalytical pipeline enabling multiplexing proteomic quantification of single embryonic cells in the 16-cell *Xenopus* embryo using microdissection, micro-scale bottom-up proteomics, and a custom-built single-cell CE- μ ESI platform for a high-resolution tandem mass spectrometer (HRMS²). Key: HVPS, high voltage power supply; Syr. Pump, syringe pump. Scale bars: 150 μ m (embryo and μ ESI, left-middle panels), 250 μ m (microcentrifuge vial), 1.5 mm (separation, right panel).

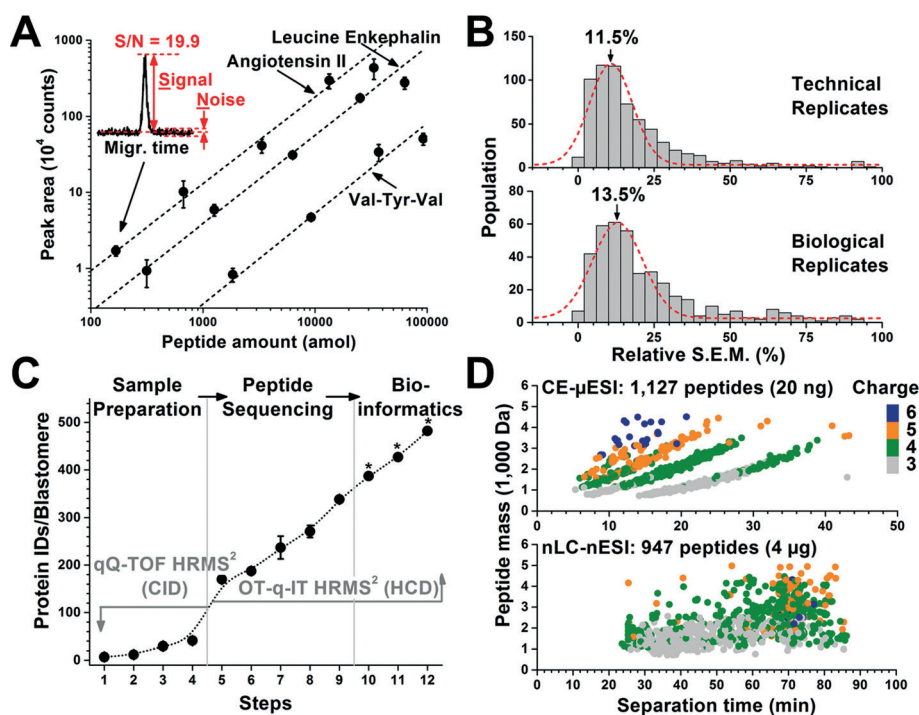


Figure 2. Advancing bottom-up discovery proteomics to single cells using CE- μ ESI-HRMS. A) Quantification curves for model peptides with 25-amol lower limit of detection and at least a 3-log-order linear dynamic range. B) Evaluation of technical and biological repeatability across a week of measurements. C) Proteomic coverage was enhanced using 20 ng digest by refining sample preparation-separation (Steps 1–4), peptide sequencing (Steps 5–9), and data analysis (Steps 10–12). Experimental conditions are in Table S2. D) Comparing peptide identifications by CE- μ ESI-HRMS with nanoLC-nanoESI-HRMS, the closest neighbor of bottom-up proteomic technology.

molecules (metabolites) in single neurons^[14] and *Xenopus* embryonic cells (blastomeres).^[15] The single-cell CE platform was able to electrophoretically separate 1–50 ng peptides into a CE- μ ESI interface, where peptides were efficiently ionized and sequenced on the order of milliseconds using HRMS (see Supporting Information). A 25-amol lower limit of detection and a 3–5 log-order quantitative range for standard peptides (Figure 2A) raised trace-level bottom-up proteomic capabilities.

We tested CE- μ ESI-HRMS on single embryonic cells of different developmental fates in the 16-cell embryo; the midline dorsal-animal cell (termed D11) reproducibly gives rise to the central nervous system; the midline ventral-animal cell (termed V11) gives rise to the neural crest and epidermis; and the midline ventral-vegetal cell (termed V21) gives rise principally to the hindgut^[16] (Figure 1). Although these blastomeres contain a considerable amount, $\approx 10 \mu\text{g}$ proteins (Figure S1 in the Supporting Information), $\approx 90\%$ (w/w) of the proteome is dominated by vitellogenin (yolk protein), essentially providing $\approx 1 \mu\text{g}$ yolk-free proteome/cell at this developmental stage.^[17] The volume of the blastomeres ($\approx 90 \text{ nL/cell}$) and their protein content are 1000- to 10000-times and 100- to 1000-times smaller than typically accessed in bottom-up HRMS, respectively. After downscaling a traditional bottom-up proteomic workflow by ca. 20-fold, we were able to extract and digest proteins from single D11, V11, and

V21 blastomeres and repeatedly analyze 20 ng, or 0.2% of the total protein content of these single cells with a quantitative technical reproducibility of 11.5% standard error of the mean (S.E.M.) (same extract measured multiple times across a week) and a biological reproducibility of 13.5% S.E.M. (different single blastomeres of the same type measured from different embryos) (Figure 2B). Technical details are provided in the Supporting Information. These results provide leverage to using this instrument for also smaller cells such as mammalian cells.

To access the single-cell encoded proteome deeper, we reconfigured the CE- μ ESI-HRMS instrument for higher-sensitivity detection (Figure 2C). Using digests of 16-cell embryos, we performed a 12-step refinement of system performance (see details in Table S1). As a result, we were able to identify 482 nonredundant protein groups (PGs) from 1630 proteins in 20 ng digest by acquiring 23 000 peptide spectral matches. In comparison with nano-liquid chromatography nano-ESI-HRMS, CE- μ ESI-HRMS identified a compar-

able number of proteins by sequencing a larger number of peptides while measuring 200-times less protein digest in half the time (Figure 2D). We attribute these performance improvements by CE to a synergy of factors: peptide loading directly into the separation capillary minimizes peptide losses; field-amplified sample stacking pre-concentrates peptides on-line; a higher electrophoretic separation efficiency distinguishes complex peptides in higher peak capacity; peptides separating as narrower bands (25-s vs. 75-s-wide peaks in nanoLC) better utilize the duty cycle of tandem HRMS, particularly benefiting parallelizable analyzer operation by the orbitrap-quadrupole-linear ion trap tribrid instrument design.

Next, we demonstrated CE- μ ESI-HRMS for querying protein expression in single blastomeres. Using new-generation nanoLC-nanoESI-HRMS^[18] capable of obtaining deep-to-near complete coverage of the encoded proteome,^[2a] recent studies on whole *Xenopus* embryos and eggs quantified the expression of 11 000 proteins with high precision,^[11] finding global translational changes among 4000 PGs during early embryogenesis^[19] and revealing even nucleocytoplasmic protein partitioning in the oocyte.^[20] Using CE- μ ESI-HRMS, we identified 1070 PGs in V11, 884 PGs in V21, and 853 PGs in D11 blastomeres, amounting to a total of 1709 different PGs (from 4846 proteins) identified among these cell types measured in $n = 3$ biological replicates with technical dupli-

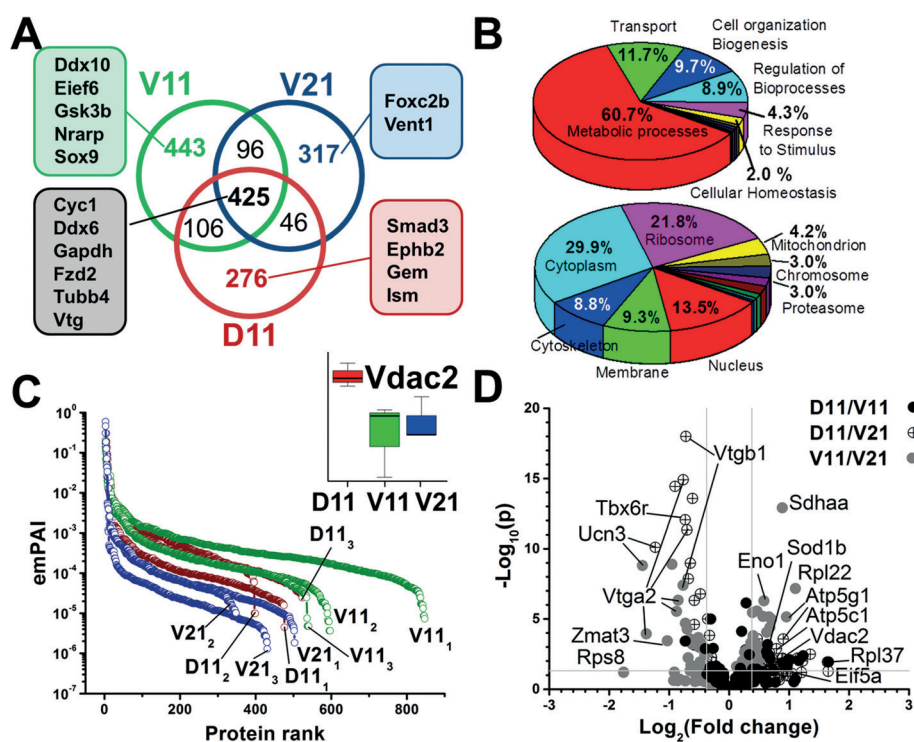


Figure 3. Single-cell measurements uncovering translational asymmetry in the 16-cell *Xenopus* embryo. A) Identification of 1709 different protein groups between D11, V11, and V21 cell types, suggesting proteomic cell differences (see peptide grouping in Figure S3 and proteins in Table S3). B) Gene ontology evaluation of biological processes (top) and sub-cellular location of identified proteins (bottom). C) Protein abundances covered a 5–6 log-order dynamic range. Differential expression shown for Vdac2 (Inset). D) Multiplexing quantification for 152 nonredundant protein groups between the D11/V11, D11/V21, and V11/V21 cell types. The volcano plot marks statistical and biological significance ($p < 0.05$, ≥ 1.3 -fold change) and labels select proteins. Significant protein differences are shown in Figure S4 and listed in Table S5.

brates (see also Table S2). A comprehensive list of identified proteins is provided in Table S3. About a quarter of these proteins were common to all the cell types, including Fzd2, Gapdh, and Vtg. Other proteins were identified in only one or two cell types (see Figure 3A), such as Gem and Ism in D11, Sox9 in V11, and Vent1 in V21. Results from independent studies corroborate these findings: Vtg is ubiquitous in the embryo and proteins such as Gapdh are involved in basic cell energy production; Gem and Ism participate in neuronal development,^[21] agreeing with the main fate of D11 blastomeres;^[16] Sox9 is involved in neural crest development, one of the fates of V11 blastomeres;^[16] and Vent1 is linked to the formation of the proctodeum,^[22] consistent with the reproducible hindgut fate of V21.^[16] Gene ontology annotation suggests that the identified proteins principally orchestrate metabolic processes, transport, and cell organization in various subcellular locations, including the cytoplasm, nucleus, and membrane (Figure 3B). In agreement, metabolism is particularly active in these blastomeres and different cell types harbor different metabolomes.^[15,23]

Exponentially modified protein abundance indexes (emPAI)^[24] calculated on the CE- μ ESI-HRMS data suggested that the proteins spanned a 5–6 log-order concentration range (Figure 3C), agreeing with abundances reported for whole

embryos using nanoLC-MS.^[11] Analysis of variance (ANOVA) on median-normalized emPAI scores suggested differential expression for several proteins between the cell types (see representative proteins in Table S4). For example, Vtg A2 and B1 were more abundant in the V21 and D11 blastomeres, and the D11 cells accumulated Vdac2 compared to the counterparts (see inset, Figure 3C). In agreement, Vtg is known to accumulate in blastomeres occupying the vegetal hemisphere, and Vdac2 is expressed in the brain and eyes of the embryo (see Xenbase Image ID 30480), which are derived from the D11 blastomere.^[16,25] With internal standards, it is possible to use these emPAI values to quantify single-cell protein expression with precision and accuracy in future studies. The proteins detected in this study correspond to ca. 30% of the encoded proteome of *Xenopus* embryos at this early stage of development,^[26] providing a deep coverage of the single-blastomere encoded proteome after cell cleavage begins.

Last, we validated translational cell-heterogeneity by adapting orthogonal multiplexing quantification to the single cells (Figure 1).

Digested D11, V11, and V21 blastomeres ($n = 3$) were labeled using distinct tandem mass tags, and the resulting peptides were pooled and measured together in technical quadruplets (see Supporting Information). Of the 1200+ PGs that were identified across all measurements, 138 PGs were quantified between D11/V11, 134 PGs between V11/V21, and 145 PGs between V11/V21 cell types, amounting to a total of 152 different PGs quantified between all cell types. Based on high-accuracy quantification of whole embryos by nanoLC-MS,^[11] these proteins are estimated to range from a few tens of nanomolar to tens of micromolar, demonstrating sufficient sensitivity by CE- μ ESI-MS also for single-cell analysis; low-abundance proteins included Zmat (≈ 13 nM), Tbx6r (≈ 20 nM), and Eno1 (≈ 90 nM) in our measurements. Figure 3D compares protein expression between the cell types, and differentially expressed proteins are compiled in Figure S4. Abundances were comparable between the blastomeres for the majority of the proteins (ratios centered at 1.0), whereas levels were significantly different for 16 PGs between D11-V11, 47 between D11-V21, and 45 between V11-V21 blastomeres (see proteins in Table S5). For example, expression was higher for Vdac2 and Eif5a in D11 as well as for Tbx6r and Vtga2 in V21. In agreement, we earlier found Vdac2 to be more highly abundant in D11 cells, Eif5a plays

a role in brain development and function,^[27] a stereotypical fate of the D11 cell type,^[16] and vegetal cells are known to contain more yolk platelets. Intriguingly, differential protein expression signifies that asymmetry along the animal-vegetal (V11 vs. V21) and dorsal-ventral (D11 vs. V11) axes of the 16-cell embryo are already established when transcriptional asymmetry is evidenced only along the animal-vegetal axis based on deep sequencing of the single-blastomere transcriptomes.^[28]

The qualitative and quantitative proteomic expression dataset that is provided here complements already available transcriptomic information on molecular processes during embryonic development and raises an opportunity to design targeted experiments to test the developmental roles of select proteins. Proteins that were identified or quantified during this work compare favorably to transcripts that were recently identified in the early developing embryo using next-generation sequencing technologies.^[28,29] Table S6 tabulates 998 transcripts for which we have manually confirmed the corresponding transcript in the published mRNA datasets. Additionally, this list expands to at least 1601 protein-transcript matches based on protein names that were derived from experimental mRNA datasets marking known expression of the gene.^[11,30] In addition to a large number of housekeeping gene products, this list includes at least 20 known signal transduction and transcription factors. Besides partial validation of performance for single-cell CE- μ ESI-HRMS, this agreement also suggests that several mRNAs inherited from the egg are translated prior to the onset of embryonic transcription, which occurs several hours later. We anticipate that the single-cell proteomic mass spectrometer that was developed here will expand the technological toolbox for cell and developmental biology to help derive a holistic understanding of basic cellular biochemical processes orchestrating normal development.

Acknowledgements

Research reported in this publication was supported by National Science Foundation Grants DBI-1455474 (to P.N. and S.A.M.), the George Washington University Department of Chemistry Start-Up Funds (to P.N.) and Columbian College Facilitating Funds (to P.N.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. We thank William F. Rutkowski (Department of Mechanical and Aerospace Engineering) for machining the sample-loading components of the CE platform.

Keywords: embryo · mass spectrometry · proteins · single-cell studies · *Xenopus laevis*

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 2454–2458
Angew. Chem. **2016**, *128*, 2500–2504

[1] a) N. Wang, M. G. Xu, P. Wang, L. Li, *Anal. Chem.* **2010**, *82*, 2262–2271; b) S. J. Altschuler, L. F. Wu, *Cell* **2010**, *141*, 559–

563; c) S. S. Rubakhin, E. V. Romanova, P. Nemes, J. V. Sweedler, *Nat. Methods* **2011**, *8*, S20–S29.

- [2] a) M. S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabudhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T. C. Huang, J. Zhong, X. Y. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda, A. Pandey, *Nature* **2014**, *509*, 575–581; b) P. Picotti, M. Clement-Ziza, H. Lam, D. S. Campbell, A. Schmidt, E. W. Deutsch, H. Rost, Z. Sun, O. Rinner, L. Reiter, Q. Shen, J. J. Michaelson, A. Frei, S. Alberti, U. Kusebauch, B. Wollscheid, R. L. Moritz, A. Beyer, R. Aebersold, *Nature* **2013**, *494*, 266–270; c) A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, J. J. Coon, *Mol. Cell. Proteomics* **2014**, *13*, 339–347.
- [3] S. A. Hofstadler, F. D. Swanek, D. C. Gale, A. G. Ewing, R. D. Smith, *Anal. Chem.* **1995**, *67*, 1477–1480.
- [4] G. A. Valaskovic, N. L. Kelleher, F. W. McLafferty, *Science* **1996**, *273*, 1199–1202.
- [5] J. S. Mellors, K. Jorabchi, L. M. Smith, J. M. Ramsey, *Anal. Chem.* **2010**, *82*, 967–973.
- [6] S. C. Bendall, E. F. Simonds, P. Qiu, E. A. D. Amir, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe'er, S. D. Tanner, G. P. Nolan, *Science* **2011**, *332*, 687–696.
- [7] M. K. Passarelli, A. G. Ewing, *Curr. Opin. Chem. Biol.* **2013**, *17*, 854–859.
- [8] L. F. Waanders, K. Chwalek, M. Monetti, C. Kumar, E. Lammert, M. Mann, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18902–18907.
- [9] Q. Chen, G. Q. Yan, M. X. Gao, X. M. Zhang, *Anal. Bioanal. Chem.* **2015**, *407*, 1027–1032.
- [10] L. L. Sun, G. J. Zhu, Z. B. Zhang, S. Mou, N. J. Dovichi, *J. Proteome Res.* **2015**, *14*, 2312–2321.
- [11] M. Wühr, R. M. Freeman, Jr., M. Presler, M. E. Horb, L. Peshkin, S. P. Gygi, M. W. Kirschner, *Curr. Biol.* **2014**, *24*, 1467–1475.
- [12] A. Buchberger, Q. Yu, L. J. Li, *Annu. Rev. Anal. Chem.* **2015**, *8*, 485–509.
- [13] a) T. H. Ong, E. G. Tillmaand, M. Makurath, S. S. Rubakhin, J. V. Sweedler, *Biochim. Biophys. Acta Proteins Proteomics* **2015**, *1854*, 732–740; b) E. J. Lanni, S. S. Rubakhin, J. V. Sweedler, *J. Proteomics* **2012**, *75*, 5036–5051.
- [14] P. Nemes, S. S. Rubakhin, J. T. Aerts, J. V. Sweedler, *Nat. Protoc.* **2013**, *8*, 783–799.
- [15] R. M. Onjiko, S. A. Moody, P. Nemes, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6545–6550.
- [16] S. A. Moody, *Dev. Biol.* **1987**, *119*, 560–578.
- [17] P. Jorgensen, J. A. Steen, H. Steen, M. W. Kirschner, *Development* **2009**, *136*, 1539–1548.
- [18] Y. Y. Zhang, B. R. Fonslow, B. Shan, M. C. Baek, J. R. Yates, *Chem. Rev.* **2013**, *113*, 2343–2394.
- [19] L. L. Sun, M. M. Bertke, M. M. Champion, G. J. Zhu, P. W. Huber, N. J. Dovichi, *Sci. Rep.* **2014**, *4*, 4365.
- [20] M. Wühr, T. Güttler, L. Peshkin, G. C. McAlister, M. Sonnett, K. Ishihara, A. C. Groen, M. Presler, B. K. Erickson, T. J. Mitch-

- ison, M. W. Kirschner, S. P. Gygi, *Curr. Biol.* **2015**, *25*, 2663–2671.
- [21] a) E. M. Pera, J. I. Kim, S. L. Martinez, M. Brechner, S. Y. Li, O. Wessely, E. M. De Robertis, *Mech. Dev.* **2002**, *116*, 169–172; b) S. Seo, A. Herr, J. W. Lim, G. A. Richardson, H. Richardson, K. L. Kroll, *Genes Dev.* **2005**, *19*, 1723–1734.
- [22] V. Gawantka, H. Delius, K. Hirschfeld, C. Blumenstock, C. Niehrs, *EMBO J.* **1995**, *14*, 6268–6279.
- [23] L. Vastag, P. Jorgensen, L. Peshkin, R. Wei, J. D. Rabinowitz, M. W. Kirschner, *PLoS One* **2011**, *6*, e16881.
- [24] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, *Mol. Cell. Proteomics* **2005**, *4*, 1265–1272.
- [25] S. A. Moody, *Dev. Biol.* **1987**, *122*, 300–319.
- [26] I. Yanai, L. Peshkin, P. Jorgensen, M. W. Kirschner, *Dev. Cell* **2011**, *20*, 483–496.
- [27] Y. F. Huang, D. S. Higginson, L. Hester, M. H. Park, S. H. Snyder, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4194–4199.
- [28] a) M. Flachsova, R. Sindelka, M. Kubista, *Sci. Rep.* **2013**, *3*, 2278; b) E. DeDomenico, N. D. L. Owens, I. M. Grant, RosaGomes-Faria, M. J. Gilchrist, *Dev. Biol.* **2015**, *408*, 252–268.
- [29] C. Collart, N. D. L. Owens, L. Bhaw-Rosun, B. Cooper, E. De Domenico, I. Patrushev, A. K. Sesay, J. N. Smith, J. C. Smith, M. J. Gilchrist, *Development* **2014**, *141*, 1927–1939.
- [30] A. H. Smits, R. G. Lindeboom, M. Perino, S. J. van Heeringen, G. J. Veenstra, M. Vermeulen, *Nucleic Acids Res.* **2014**, *42*, 9880–9891.

Received: November 20, 2015

Revised: December 21, 2015

Published online: January 12, 2016