Dynamic Organellar Maps for Spatial Proteomics

Daniel N. Itzhak,[1] Julia P. Schessner,[1] and Georg H. H. Borner[1,2]

[1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
[2]Corresponding author: *borner@biochem.mpg.de*

Eukaryotic cells are highly compartmentalized and protein subcellular localization critically influences protein function. Identification of the subcellular localizations of proteins and their translocation events upon perturbation has mostly been confined to targeted studies or laborious microscopy-based methods. Here we describe a systematic mass spectrometry-based method for spatial proteomics. The approach uses simple fractionation profiling and has two applications: Firstly it can be used to infer subcellular protein localization on a proteome-wide scale, resulting in a protein map of the cell. Secondly, the method permits identification of changes in protein localization, by comparing maps made under different conditions, as a tool for unbiased systems cell biology. © 2018 by John Wiley & Sons, Inc.

Keywords: mass spectrometry • organelle proteomics • protein dynamics • spatial proteomics • subcellular localization • systems biology

How to cite this article:
Itzhak, D. N., Schessner, J. P., & Borner, G. H. H. (2019). Dynamic organellar maps for spatial proteomics. *Current Protocols in Cell Biology*, *83,* e81. doi: 10.1002/cpcb.81

INTRODUCTION

Many biological processes involve the movement of proteins between different subcellular compartments. Monitoring the localization of proteins on a global scale is therefore a desirable tool for systems biology. Here we describe a method that permits identification of proteins that change subcellular localization under different experimental conditions using mass spectrometry-based proteomics (Fig. 1, workflow overview). In brief, cells are lysed and organelles are partially separated by differential centrifugation, according to size and density. Subcellular fractions are then analyzed by quantitative mass spectrometry, yielding an abundance distribution profile for each identified protein. Proteins associated with the same organelle(s) have similar profiles. A large set of marker proteins with known localizations is then used as reference to assign proteins to organelles by machine learning. Furthermore, profiles from maps made under different conditions can be subtracted to reveal proteins that undergo subcellular localization changes. Here, we first describe a gentle yet reproducible cell lysis protocol (see Basic Protocol 1) followed by subcellular fractionation by differential centrifugation (see Basic Protocol 2). Next, we describe the sample processing and mass spectrometry employed to acquire shot-gun proteomic data from these subcellular fractions (see Basic Protocol 3 and Basic Protocol 4). Processing of mass spectrometry .RAW files using MaxQuant software is described (see Basic Protocol 5) followed by data processing and visualization by principal component analysis using Perseus (see Basic Protocol 6), identification of moving proteins (see Basic Protocol 7), assignment of non-marker proteins to organelles using

CURRENT PROTOCOLS
A Wiley Brand
Current Protocols in Cell Biology e81, Volume 83
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpcb.81

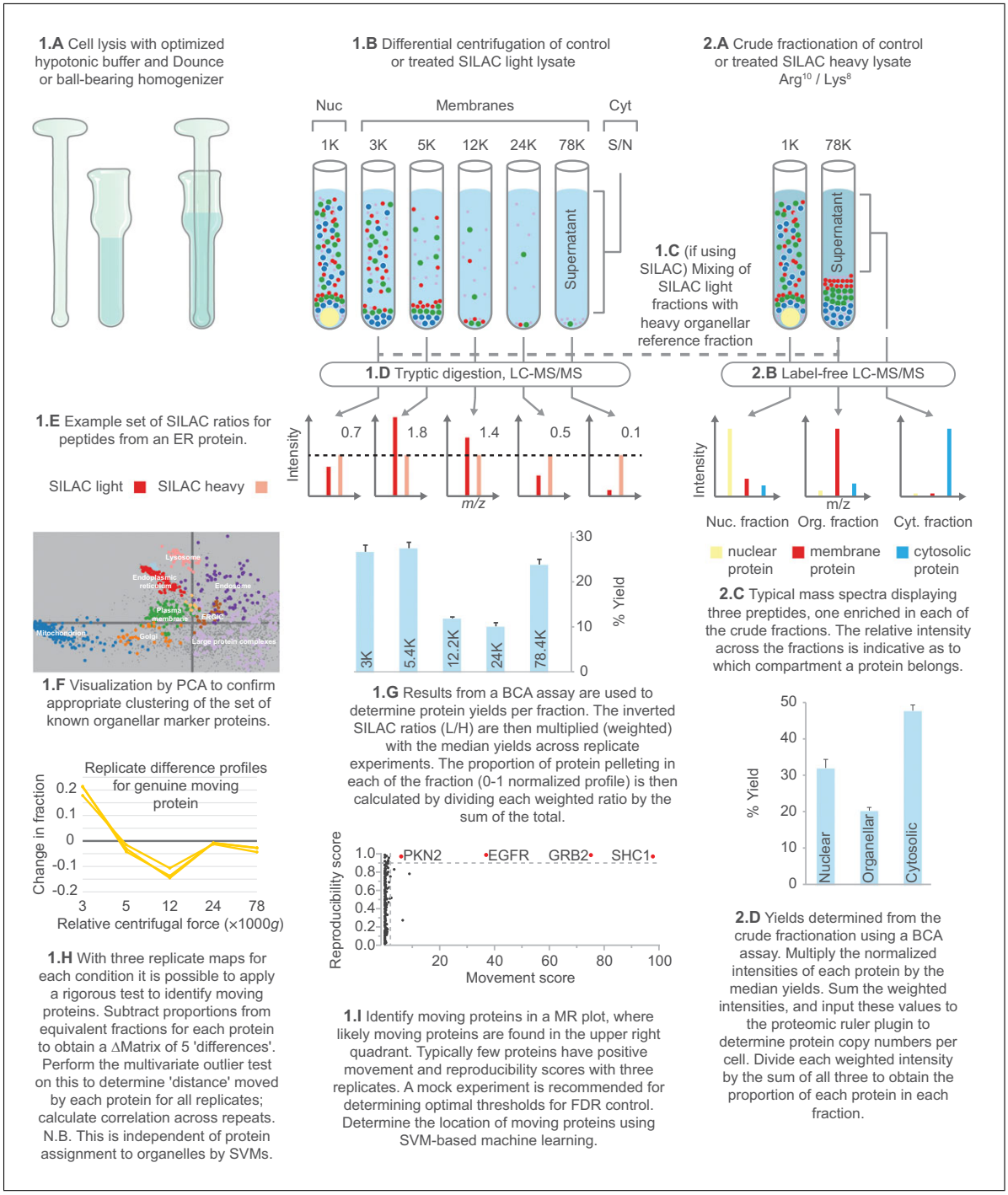

Figure 1 Overview workflow for the generation and analysis of organellar maps. See main text for details. Abbreviations: ER, endoplasmic reticulum; SVM, support vector machine; SILAC, stable isotope labeling by amino acids in cell culture; MR, movement reproducibility; PCA, principal component analysis; FDR, false discovery rate.

support vector machine-based classification (see Basic Protocol 8), and assignment of protein copy numbers (see Basic Protocol 9).

STRATEGIC PLANNING

Cells are fractionated by six differential centrifugation steps into six pellets and one supernatant, followed by identification and quantification of the proteins in each fraction. The method does not result in purified organelles but uses the distribution profiles across the differential centrifugation fractions to infer protein localization via comparison with

profiles from known organellar markers. To quantify relative protein abundance in each fraction by mass spectrometry, two methods are described: Label-free quantification (LFQ) implemented in MaxQuant software (Cox et al., 2014) or the spike-in stable isotope labeling by amino acids in cell culture (SILAC) approach (Geiger et al., 2011). In the latter method, which is recommended where possible, a heavy-isotope labeled reference fraction containing all membrane organelles is spiked into each of the five membrane containing sub-fractions during sample processing. The LFQ approach is more cost effective and can quantify more proteins but is not as accurate as SILAC and requires a highly robust liquid chromatography setup. It should therefore be used where it is not possible to metabolically label the cells using SILAC, e.g., primary cells.

GENERATION OF CELL HOMOGENATES

BASIC PROTOCOL 1

This protocol uses mild lysis conditions to release membrane-bound organelles without releasing their luminal content. Currently, this protocol is best suited to cultured mammalian cells or isolated cell types. In practice, the recommended amount of starting material would be no less than a single 10-cm dish of adherent cells or 1 mg protein mass of starting material but for comfort, 2 × 15-cm dishes of cells or >5 mg of protein mass is recommended.

The initial buffer for cell lysis is moderately hypotonic 50 mM sucrose (Graham & Rickwood, 1997; Itzhak, Tyanova, Cox, & Borner, 2016 & 2017), which permits facile lysis of cells using a Dounce homogenizer. The osmotic strength is subsequently restored to physiological strength to maintain organelle integrity until protein solubilization. The hypotonic strength of the buffer should be optimized to ensure sufficient swelling of the cells without inducing cell death. Two different homogenization methods are available: A traditional Dounce homogenizer with a tight pestle or a ball-bearing homogenizer, which tends to yield more consistent results.

Materials

DMEM without arginine, glutamine, lysine, and sodium pyruvate (Thermo Fisher Scientific, Gibco brand, #A14431-01)
Dialyzed FBS (PAA, #A11-107)
Sodium pyruvate (Sigma-Aldrich, #58636)
GlutaMAX (Thermo Fisher Scientific, Gibco brand, #35050-061)
$^{13}C_6$, $^{15}N_4$-L-arginine HCl (Silantes, #201604302)
$^{13}C_6$, $^{15}N_2$-L-lysine HCl (Silantes, #211604302)
Standard arginine monohydrochloride (Sigma-Aldrich, #A6969)
Standard lysine monohydrochloride (Sigma-Aldrich, #L8662)
SILAC medium: DMEM, 10% (v/v) FBS, 1 mM sodium pyruvate, 1× GlutaMAX, 42 mg/liter arginine monohydrochloride, 73 mg/liter lysine monohydrochloride
Your choice of cells (recommended 5 mg of protein yield)
Ice-cold PBS without $CaCl_2$ and $MgCl_2$ (Thermo Fisher Scientific, Gibco brand, #14190-094)
Tris·HCl, >99% (Sigma-Aldrich, #T6666)
Tris base, >99.9% (Sigma-Aldrich, #T6791)
Sucrose, >99.5% (Sigma-Aldrich, #84097)
Magnesium chloride ($MgCl_2$; Sigma-Aldrich, #M2670)
Ethylene glycol-bis(2-aminoethylether)-*N,N,N′,N′*-tetraacetic acid (EGTA; Sigma-Aldrich, #E4378)
Ice-cold hypotonic buffer: 25 mM Tris·HCl, pH 7.5, 50 mM sucrose, 0.5 mM $MgCl_2$, 0.2 mM EGTA
Trypan blue solution (Warner, Sakai, & Sandell, 2015)

Ice-cold hypertonic buffer (pre-aliquoted for step 8): 25 mM Tris base, pH 7.2, 2.5 M sucrose, 0.5 mM $MgCl_2$, 0.2 mM EGTA

10- or 15-cm cell culture dishes
Inverted bright-field microscope
Cell scraper (Sarstedt, #83.1831)
Dounce homogenizer with tight pestle (Sartorius, #8530700 and #8530807)

1. Culture dishes of cells in SILAC medium for at least seven doublings to ensure complete labeling. For heavy labeled cells, prepare SILAC medium with $^{13}C_6$, $^{15}N_4$-l-arginine and $^{13}C_6$, $^{15}N_2$-l-lysine; for light labeled cells, prepare SILAC medium with standard arginine and lysine.

   Non-dividing cells incorporate heavy amino acids only very slowly and some cell lines do not tolerate the dialyzed FBS used in SILAC medium. This can be overcome by using the LFQ method or potentially adding back certain growth factors and hormones. For details on how to cultivate mammalian cells and how to perform the below-mentioned trypan blue staining, refer to Warner et al., 2015.

2. Remove medium from adherent cultured cells at 70% to 90% confluence; non-adherent cells should be transferred to a benchtop centrifuge and gently pelleted followed by resuspension during each step.

   The minimum recommended starting material for cultured cells is a single 10-cm dish equating to around 1 mg of protein with HeLa cells. The lowest protein mass in any fraction is typically 1% of starting material, i.e., 1 mg would provide 10 µg, sufficient for mass spectrometry.

3. Store all buffers on ice and perform all work at 4°C. Rinse remaining medium on the plate using 25 ml ice-cold PBS buffer, immediately replace buffer with a further 25 ml PBS, and incubate on ice 5 min.

   This step is required to completely chill the cells before addition of hypotonic buffer because it is undesirable for cells to be exposed to hypotonic conditions at temperatures where proteins could remain substantially active. Pour the buffer onto the dish at the corner of the plate, so that cells are not displaced.

4. Rinse cells with 25 ml ice-cold hypotonic buffer and immediately replace buffer before incubation for 5 min, allowing cells to swell.

   Prior to making organellar maps or attempting this protocol, it is advisable to test the sensitivity of cells to different strengths of hypotonic buffer. This can be done by exposing the cells to the buffers for 5 min before restoring them to growth medium. Cells exposed to toxic strength hypotonic buffer will undergo rapid cell death that is readily evident by microscopy.

5. Remove buffer and stand plates vertically 1 min to allow excess buffer to drain into one corner of the plate; aspirate excess buffer. The final volume in the Dounce homogenizer should be 4 ml; a calculated volume of buffer should be added to the plates during scraping, taking into consideration that each drained 15-cm plate will contribute ~200 to 300 µl.

6. Homogenize cell suspension with fifteen strokes of the Dounce homogenizer using the tight pestle.

   A single motion to the base of the Dounce homogenizer and return to the meniscus constitutes a single stroke. Be careful not to introduce bubbles because this strongly reduces lysis efficiency. Do not pull too strongly because this can generate a vacuum and in the worst case implode the glass.

7. Check consistency of lysis by trypan blue staining of a small aliquot of the homogenate. If >10% unbroken cells are visible, it can help to increase the number of strokes.

   For small cells that do not lyse well with Dounce homogenization, it may be necessary to use a ball-bearing homogenizer (see Alternate Protocol 1).

8. Following homogenization in hypotonic buffer, the osmolarity is restored to 0.25 M sucrose with pre-aliquoted hypertonic buffer. When homogenizing in 4 ml, transfer homogenate to a 15-ml Falcon tube containing 356 µl hypertonic buffer.

   Consistent volumes are important for reproducible performance of differential centrifugation experiments. Because the sucrose is viscous, it is necessary to aliquot this at room temperature beforehand and add the homogenate to the chilled sucrose, pipetting vigorously.

CELL HOMOGENIZATION USING A BALL-BEARING HOMOGENIZER

ALTERNATE PROTOCOL 1

A ball-bearing homogenizer or 'cell-cracker' is an excellent alternative to the Dounce homogenizer for lysing cells and should lead to more consistent lysis. This is primarily because the passage between the ball and the case is engineered to a higher standard of precision than the between pestle and mortar of glass Dounce homogenizers. It is recommended to select a ball size/bore that results in proper lysis of cells, assessed by trypan blue staining and release but not lysis of organelles, assessed by mass spectrometry (see Commentary, Troubleshooting section), i.e., the widest bore that gives good lysis is best.

Additional Materials *(also see Basic Protocol 1)*

Scraped cells (generated as described in Basic Protocol 1, steps 1 to 5)
Ice-cold hypotonic buffer: 25 mM Tris·HCl, pH 7.5, 50 mM sucrose, 0.5 mM $MgCl_2$, 0.2 mM EGTA
Trypan blue solution (Warner et al., 2015)
Ice-cold hypertonic buffer (pre-aliquoted for step 5): 25 mM Tris base, pH 7.2, 2.5 M sucrose, 0.5 mM $MgCl_2$, 0.2 mM EGTA

Ball-bearing homogenizer (Isobiotec)
Luer lock syringes

1. Choose ball size and assemble homogenizer.

   For adherent HeLa cells: Recommended bore is 14 µm.

2. Flush homogenizer by attaching a syringe loaded with hypotonic buffer to one of the Luer fittings and push this buffer through to remove air from the device; this buffer will flow out of the open Luer fitting. Attach a second syringe loaded with hypotonic buffer to the open fitting and push this in the opposite direction to refill the first syringe, ensuring complete removal of air.

3. With buffer filling the device, detach filled syringe and attach a new syringe without plunger to the Luer fitting and fill this open syringe with the same volume of homogenate as for the Dounce homogenizer. Use an appropriately sized syringe.

   When attaching the new syringe, the homogenizer should be completely filled with buffer, but very little should be left in the second syringe. To ensure that no air is introduced to the device, a meniscus should be visible at the open Luer fitting.

4. With cells filled in the open syringe, introduce the plunger and force cell suspension past the bore five times in total. Take care not to introduce air into the device.

It may be necessary to pass the cells through the bore more than five times, checking lysis efficiency by trypan blue staining of a small aliquot after each passage.

5. Take out plunger of the syringe containing the lysate while it is still attached and collect cell lysate with a pipet. Immediately restore sucrose concentration to 250 mM, as described for Dounce homogenization above in Basic Protocol 1, step 8.

BASIC PROTOCOL 2

SUBCELLULAR FRACTIONATION BY DIFFERENTIAL CENTRIFUGATION

Subcellular fractionation should be done immediately after cell lysis to keep any proteolytic or metabolic activity to a minimum and to avoid leakage of proteins from organelles. The full fractionation of organelles is done in six centrifugation steps and yields one nuclear pellet, five organellar pellets, and the liquid cytosol fraction. When using SILAC labeling, apply this protocol to the SILAC light sample.

To generate the reference membrane fraction, apply Alternate Protocol 2 to the SILAC heavy-labeled sample. This describes the generation of the nuclear, organellar, and cytosolic fractions (nuclear-organellar-cytosol or 'NOC' split), from which the reference membrane fraction is derived. The two protocols can be done in parallel by keeping the post-nuclear supernatant of the heavy-labeled sample on ice until the final centrifugation step.

The recommended ultracentrifuge for this protocol is an Optima MAX Ultracentrifuge with the TLA-110 rotor. All standard centrifugal steps can be done with benchtop centrifuges. The precise speeds have been optimized for HeLa cells, but work well for other cell types tested, and rpm values are given for the recommended ultracentrifuge setup.

Materials

Crude cell lysates (see Basic Protocol 1 or Alternate Protocol 1)
Ice-cold isotonic buffer: 25 mM Tris·HCl, pH 7.2, 250 mM sucrose, 0.5 mM $MgCl_2$, 0.2 mM EGTA
>99% sodium dodecyl sulfate, (SDS; Roth, CN30.3)
SDS lysis buffer: 2.5% SDS, 50 mM Tris·HCl, pH 8.1
5× SDS lysis buffer: 12.5% SDS, 50 mM Tris·HCl, pH 8.1

15 ml-Falcon tubes
Bench-top centrifuge (for 15 ml-Falcon tubes; up to 3000 × *g*, e.g., Multifuge 1 liter, Heraeus)
Ultracentrifuge plus rotor (e.g., Optima MAX Ultracentrifuge with TLA 110 rotor, Beckman Coulter)
Ultracentrifuge thickwall polycarbonate tubes (e.g., Beckman Coulter, #362305)
Balance with ±1 mg accuracy
Sonicator Bioruptor Plus (Diagenode, #B01020001) including cooling unit

1. Immediately process crude cell lysates after restoring them to 250 mM sucrose (see Basic Protocol 1, step 8 or Alternate Protocol 2, step 5). Centrifuge in a Falcon tube 10 min at 1000 × *g* at 4°C to pellet nuclei and unbroken cells.

   The pellet from this step is designated the "nuclear fraction" since it contains almost all nuclei; there is only minimal nuclear contamination in the subsequent fractions. Thus, quick processing is required to limit the time that proteins have to leak out of the nucleus. Of note, the nuclear fraction itself is not pure; there is a substantial contamination from other organelles.

2. Remove supernatant from this first fraction, transfer it to another 15-ml Falcon tube, and centrifuge 10 min at 3000 × *g*. Do not perform this step for the NOC split: Place post-nuclear supernatant on ice until the final spin.

   The volume of supernatant at this step will depend on the volume of the nuclear pellet. Transferring an identical volume to the next step without touching the pellet is recommended. For example, transfer 3.8 ml using a microtiter pipet and remove the remaining supernatant carefully, taking note of this extra volume, ideally <200 µl.

3. Transfer 3.6 ml of the 3000 × *g* supernatant to an ultracentrifuge tube. Remove residual supernatant of the 3000 × *g* pellet and take note of this volume. Keep pellet on ice.

4. Prepare a counterbalance for the ultracentrifuge rotor by filling a second ultracentrifuge tube with an equal mass of isotonic buffer, using a weighing scale capable of milligram accuracy.

   When preparing a single map, adjust the mass of the second tube using a microtiter pipet and additional buffer where necessary. When preparing two maps from different conditions on the same day, use ice-cold isotonic buffer to top up one of the tubes if they have a different mass. With careful pipetting, this is <20 µl.

5. With a marker pen, mark one side on the top of the tubes and load into a pre-chilled rotor with the mark facing outwards. Place rotor into a 4°C ultracentrifuge, set time to 15 min, and set speed to 10,000 rpm (5400 × *g*). After evacuation of the rotor chamber, begin ultracentrifugation with the fastest acceleration and deceleration settings.

   The mark on the tube provides the user with a way to determine where the pellet is located; this is especially important for low sample starting amounts, where the pellet may be invisible to the naked eye.

6. Upon completion of the first ultracentrifugation step, transfer all supernatant to a second ultracentrifuge tube. By pipetting from the opposite side of where the pellet is located, the pellet will not become dislodged. Cover tube with Parafilm and keep on ice until SDS solubilization (see step 9).

7. Repeat steps 4 to 6, adjusting ultracentrifugation times and speeds for three additional centrifugation steps to: 20 min at 15,000 rpm (12,200 × *g*), 20 min at 21,000 rpm (24,000 × *g*), and 30 min at 38,000 rpm (78,400 × *g*). During the final spin, introduce the SILAC heavy tube(s) placed on ice in step 2.

8. The supernatant from the final centrifugation step is the cytosolic fraction. Transfer it to a 15-ml Falcon tube and place it on ice.

9. To each pellet (two in 15-ml Falcon tubes, four in ultracentrifuge tubes), add a volume of SDS lysis buffer to achieve a final protein concentration of 1 mg/ml. Heat to 72°C for 5 min. For the cytosolic fraction, add one-fifth the volume of 5× SDS lysis buffer.

   Based on the starting material and approximate yields per fraction, as determined in HeLa cells (Itzhak et al., 2016), it is possible to determine the approximate volume for the desired protein concentration. For 2 × 15-cm dishes of HeLa cells, 1 ml for the first (nuclear) fraction and 200 µl for all subsequent fractions, except the 24,000 × g fraction, which needs only 100 µl.

10. Following addition of SDS lysis buffer and heating, ensure complete solubilization as well as fragmentation of nucleic acids by sonication. This is achieved in a Bioruptor with fifteen 30 second on/off cycles at maximum intensity.

11. At this point, protein samples can be stored at −80°C or directly prepared for mass spectrometry (see Basic Protocol 3).

ALTERNATE PROTOCOL 2

CRUDE SEPARATION OF CELL HOMOGENATE TO GENERATE NUCLEAR, ORGANELLAR, AND CYTOSOLIC FRACTIONS

To generate the SILAC heavy reference organellar fraction or to make a crude pilot map, the cell homogenate is split into only three fractions by differential centrifugation. This is done in parallel to the more extensive differential centrifugation described above.

Additional Materials *(also see Basic Protocol 2)*

Crude cell lysates (see Basic Protocol 1 or Alternate Protocol 1)
Ice-cold isotonic buffer: 25 mM Tris·HCl, pH 7.2, 250 mM sucrose, 0.5 mM $MgCl_2$, 0.2 mM EGTA

15-ml Falcon tubes
Bench-top centrifuge (for 15-ml Falcon tubes; up to 3000 × *g*, e.g., Multifuge 1 liter, Heraeus)
Ultracentrifuge plus rotor (e.g., Optima MAX Ultracentrifuge with TLA 110 rotor, Beckman Coulter)
Ultracentrifuge thickwall polycarbonate tubes (e.g., Beckman Coulter, #362305)
Balance with ±1 mg accuracy

1. Immediately process crude cell lysates after restoring them to 250 mM sucrose. Centrifuge them in a Falcon tube 10 min at 1000 × *g*, to pellet nuclei and unbroken cells.
2. Transfer 3.6 ml supernatant to an ultracentrifuge tube. Remove residual supernatant of the 1000 × *g* pellet, taking note of this volume. Keep pellet and ultracentrifuge tube on ice.
3. Prepare a counterbalance and set up ultracentrifuge as described in Basic Protocol 2, steps 4 and 5.
4. Either centrifuge the supernatant together with an empty counter balance or together with the last centrifugation step of the SILAC light channel for 30 min at 38,000 rpm (78,400 × *g*).

   This step yields a pellet containing all organellar material that did not pellet together with the nucleus and a supernatant containing mostly cytosolic proteins.

5. Proceed with Basic Protocol 2, steps 8 to 11 to solubilize the two pellets and the cytosolic fraction.

BASIC PROTOCOL 3

PROTEIN EXTRACTION AND PREPARATION OF TRYPTIC PEPTIDES FOR LC-MS

Following solubilization, proteins must be prepared for digestion to peptides prior to mass spectrometry. Protein concentrations have to be determined for processing of equal amounts per fraction and for appropriate sample mixing in case SILAC is used. For protein determination, a BCA assay is recommended. Proteins are subjected to tryptic digestion, for which numerous protocols are possible although here in-solution digestion is described. Peptide cleanup via solid phase extraction with reversed-phase sulfonate (RPS) is recommended (Kulak, Pichler, Paron, Nagaraj, & Mann, 2014) and described here in detail, although other packing materials suitable for desalting or commercially available peptide cleanup solutions can be used. An advantage of RPS extraction is that the samples can be subjected to rapid, semi-orthogonal stage-tip fractionation without

the need for additional desalting, which yields increased numbers of identified proteins; this is described in Alternate Protocol 3.

Materials

SDS-solubilized protein samples in 2.5% SDS, 50 mM Tris·HCl, pH 8.1 (see Basic Protocol 2)
BCA protein assay kit (Thermo Fisher Scientific, #23225)
BSA standard for protein determination (Thermo Fisher Scientific, #23209)
100% acetone, pre-chilled (Thermo Fisher Scientific, Fisher Chemical brand, #A/0600/17)
Urea, >99% (Sigma-Aldrich, #U0631)
Dithiothreitol (DTT), >99.5% (Sigma-Aldrich, #43815)
Digestion buffer: 8 M urea, 50 mM Tris·HCl, pH 8.1, freshly added 1 mM DTT
Iodoacetamide, >99% (Sigma-Aldrich, A3221)
Lysyl endopeptidase (LysC; Wako, #129-02541)
50 mM Tris, pH 8.1
Trypsin (Sigma-Aldrich, #T6567)
Styrene-divinylbenzene reversed-phase sulfonate (SDB-RPS) solid-phase extraction material (VWR, #66886-U)
Trifluoroacetic acid (TFA), >99% (Merck, 808260)
Methanol, LC/MS grade (Thermo Fisher Scientific, #A456)
Ammonium hydroxide, 25% (Sigma-Aldrich, #44273)
Acetonitrile (Thermo Fisher Scientific, #A955)
10% and 0.2% trifluoroacetic acid (TFA) solutions
30% methanol, 1% trifluoroacetic acid (TFA) solution
Elution buffer: 5% ammonium hydroxide, 80% acetonitrile
Buffer A*: 0.1% trifluoroacetic acid (TFA), 2% acetonitrile

Stage-tipping device, or commercially available stage tips
96-well plates
96-well plate compatible spectrophotometer (e.g., Infinite M200, Tecan)
Table-top centrifuge with temperature control
Vacuum centrifuge
Tube adaptors for stage tips

1. Aliquot triplicates of 5 µl BSA standard solutions prepared according to the manufacturer's instructions and 2 µl protein samples into a 96-well plate. Use a repeating pipet to rapidly dispense 200 µl premixed assay reagents to each well.

2. The Tecan plate reader permits orbital shaking inside the device; perform 5 min at 300 rpm with temperature control at 37°C for a further 35 min followed by absorbance measurement at 562 nm. Determine sample concentrations by comparing to the linear regression of the BSA standard curve.

   Protein concentrations should be no less than 0.2 mg/ml to facilitate further processing in 1.5-ml tubes. If concentrations are above the linear range, dilute a small amount for concentration determination. If concentrations are too low to measure, increase sample volume per well to 10 µl; if still below the limit of detection, process the entire sample until the end of this protocol and use 1 µl of 0.5 mg/ml enzyme during the digestion steps.

SILAC fractions

3a. Transfer 25 µg of each SILAC light fraction to fresh 1.5-ml tubes and to each add 25 µg of the SILAC heavy organellar reference fraction. Increase volume to at least 100 µl with lysis buffer, add five volumes of ice-cold acetone, and vortex.

LFQ fractions

3b. For the LFQ protocol, aliquot 50 µg of each fraction to new tubes. Increase volume to at least 100 µl with lysis buffer, add five volumes of ice-cold acetone, and vortex.

4. Incubate samples at −20°C, 1 hr or overnight and pellet precipitated protein at 4°C, 10,000 × *g*, 5 min in a table-top centrifuge.

   All subsequent steps are performed at room temperature.

5. Remove supernatant (be careful not to disturb the pellet) and let pellets air dry 5 min.

6. Resuspend pellets in 50 µl digestion buffer containing 1mM DTT and incubate 15 min.

7. Alkylate cysteine residues by addition of 5 µl 55 mM iodoacetamide and incubate in the dark 20 min.

8. Add 1 µg of LysC (2 µl of a 0.5 mg/ml stock solution) and digest overnight.

   Do not incubate above 30°C because this will introduce chemical modifications on lysine residues and free N-termini (Stark et al., 1960).

9. Reduce concentration of urea to 2 M by dilution with 171 µl 50 mM Tris, pH 8.1.

10. Add 1 µg trypsin (2 µl of a 0.5 mg/ml stock solution) and digest a further 3 hr.

    The LysC digestion can be shortened to 3 hr, permitting digestion and subsequent peptide cleanup to be completed within one working day.

11. During tryptic digestion, prepare stage tips packed with three layers of SDB-RPS material as shown in the online video tutorial (*http://www.biochem.mpg.de/226863/Tutorials*). In brief, punch out discs of SDB-RPS with a stage-tipping device and force them to the bottom of a 200 µl pipet tip to create the stage tip.

12. Place stage tips in 1.5-ml tubes using an adaptor and then place in a benchtop centrifuge. Add 100 µl of 100% acetonitrile to each stage tip and centrifuge 30 sec at 500 × *g* to test their integrity.

    The acetonitrile should not pass completely through the stage tips in this time but some acetonitrile should pass. If all of the acetonitrile passes through or none of the acetonitrile passes through, then replace the stage tip.

13. Following the integrity check, continue equilibration of the stage tips by passing the remaining acetonitrile through the stage tip, followed by 100 µl of 30% methanol, 1% TFA solution, then 100 µl of 0.2 % TFA in HPLC-grade water.

    Add the wash volume on top of the column material and centrifuge through the column material in a stage-tipping block or a stage-tip centrifuge. If you are using C18, this should not run completely dry during the extraction process; leave a few µl of 0.2% TFA on the stage tip in the final step.

14. Acidify samples to a 1% TFA concentration (final pH <3) using 25 µl of 10% TFA stock solution and incubate on ice 5 min.

15. Pellet precipitate (mostly lipids) by centrifugation 5 min at 10,000 × *g*.

16. Load 10 µg peptide (equivalent to 50.6 µl supernatant, accounting for all the volumes added from steps 7 to 14) and centrifuge ∼5 min at 500 × *g*, or until the liquid has passed.

    It is also possible to load 20 µg on the stage tip and perform stage-tip fractionation (see Alternate Protocol 2).

17. Wash column twice with 100 µl 0.2% TFA.
18. Elute peptides with 60 µl freshly prepared elution buffer into collection tubes.

    Samples can be collected in 1.5-ml tubes using adaptors to make stage tips compatible with benchtop centrifugation (Yu, Smith, & Pieper, 2014) or directly into autosampler vials in a 96-well format rack comprising two layers, one holding stage tips and one beneath, holding autosampler vials and compatible with swing out centrifuges.
19. Dry samples under vacuum centrifugation ~45 min at 30°C.

    Drying samples to completion will introduce a small amount of oxidation, which can be minimized by halting vacuum centrifugation when sample volumes have reduced to <5 µl, i.e., after evaporation of acetonitrile and ammonia. Low levels of oxidation will have a minimal impact on peptide identification rates in MaxQuant.
20. Resuspend peptides in a final volume of 10 µl with buffer A* and freeze samples at −80°C for long-term storage or proceed directly to LC-MS.

SOLID-PHASE FRACTIONATION OF TRYPTIC PEPTIDES

ALTERNATE PROTOCOL 3

Where greater proteome depth is required, peptides can be eluted from stage tips in three fractions and measured in three independent runs. This reduces the complexity of the samples, allowing more proteins to be identified at the expense of increased measurement time (Kulak et al., 2014). Use this protocol in place of steps 16 to 20 of Basic Protocol 3.

Materials

Equilibrated styrene-divinylbenzene reversed-phase sulfonate (SDB-RPS) stage tips and acidified peptide samples (see Basic Protocol 3)
0.2% trifluoroacetic acid (TFA)
Ammonium formate (Sigma-Aldrich, #70221)
Formic acid (Merck, #100264)
SDB-RPS-1: 100 mM ammonium formate, 40% acetonitrile, 0.5% formic acid
SDB-RPS-2: 150 mM ammonium formate, 60% acetonitrile, 0.5% formic acid
SDB-RPS-3: 5% ammonia, 80% acetonitrile

Vacuum centrifuge
Table-top centrifuge
Tube adaptors for stage tips

1. Load 20 µl peptides, 101.2 µl accounting for all the volumes added from steps 7 to 14 and centrifuge ~5 min at 500 × *g* or until the liquid has passed.
2. Wash peptides on the column with two washes of 100 µl 0.2 % TFA.
3. Elute first fraction with 20 µl SDB-RPS-1.
4. Change collection tube and elute a second fraction with 20 µl SDB-RPS-2.
5. Change collection tube and elute a third fraction with 30 µl SDB-RPS-3 (equivalent to elution buffer in the Basic Protocol 3).
6. Dry samples under vacuum centrifugation ~30 min at 30°C to evaporate all elution buffer.

   When trying to minimize oxidation, not all buffers evaporate at the same rate and it is necessary to either remove samples from the centrifuge once the desired volumes have been reached or to top up their volumes with buffer A.*
7. Resuspend peptides in a final volume of 10 µl with buffer A* and freeze the samples at −80°C for long-term storage or proceed directly to LC-MS.

BASIC PROTOCOL 4

LC-MS ANALYSIS OF TRYPTIC PEPTIDES

A comprehensive description of maintaining and running a mass spectrometer is beyond the scope of this protocol. This section will provide a description of the key parameters and expected performance from running the samples on the instrument as described in our previous publications.

Materials

HPLC-grade water
Buffer A: 0.1% formic acid
Buffer B: 80% acetonitrile, 0.1% formic acid

In-house packed C18 columns or commercial alternative
EASY-nLC 1000 (Thermo Fisher Scientific)
Q Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer (Scheltema et al., 2014; Thermo Fisher Scientific) with nanoelectrospray source (Thermo Fisher Scientific)

1. Use a 50-cm HPLC column with 75-µm inner diameter and C18 beads as stationary phase. Either pack this column in-house with 1.8-µm C18 endcapped particles (Dr. Maisch, Germany) or get one of the commercial alternatives, such as EASY-spray.
2. Fit this column to a nano-LC system, such as the EASY-nLC 1000 system, with a mobile phase A of 0.1% formic acid and mobile phase B of 80% acetonitrile in 0.1% formic acid. Operate column at a constant flow rate of 250 to 350 nl/min and at constant temperature of 50° to 60°C.
3. Couple LC to a Q Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer via the nanoelectrospray source when using in-house packed columns or the EASY-spray source when using EASY-spray columns
4. Load 1 µg onto the column per injection. For the Q-Exactive HFX (Kelstrup et al., 2018), load 0.5 µg per injection.

   An assumption of a 50% efficiency in the entire procedure from converting proteins into peptides is made, thus from the presumed 10 µg of material loaded onto a stage tip and eluted in one step, only 5 µg peptide will be present in the autosampler vial at 0.5 µg/µl. For samples separated into three peptide fractions, it is evident that each vial should contain ∼3.5 µg. With these injection amounts, most of the peptide will remain in the vial. This peptide is reserved in case there are issues with the LC-MS or it can be stored at −80°C for repeated measurement if necessary.
5. Set a linear gradient from 5% to 30% buffer B on the LC to elute peptides from the column.

   The duration of this gradient step depends on access to MS time, but should be minimally 60 min and not exceed 4 hr. Sample loading takes ∼30 min, thus when the gradient step is 130 min, followed by washout (ramping up to 95% B in 5 min, constant at 95% B for 5 min, ramping down to 5% B in 5 min, constant at 5% B for 5 min), 3 hr total time per sample is expected or eight samples per day.
6. Acquire MS data in positive data-dependent mode using a top 15 method.

   Whilst many of the following parameters can be altered using the method editor in Xcalibur software, those in Table 1 have been shown to work effectively.

BASIC PROTOCOL 5

PROCESSING OF SPECTRA USING MAXQUANT SOFTWARE

The MaxQuant software package offers numerous features which are critical to the down-stream analysis. Therefore, the use of other software packages is not recommended.

Table 1 Recommended Mass Spectrometry (MS) Settings for a Thermo Fisher Scientific Orbitrap HF[a]

Full MS	
Resolution	120,000
AGC target	3e6
Maximum injection time	20 msec
Scan range	300 to 1650 *m/z*
Data-dependent MS²	
Resolution	15,000
AGC target	1e5
Maximum injection time	55 msec
Isolation window	1.4 *m/z*
Fixed first mass	100.0 *m/z*
Normalized collision energy	27
Data-dependent settings	
Minimum AGC target	2.9e3
Intensity threshold	1.0e5
Charge exclusion	Unassigned, 1, 6-8, >8
Peptide match	Preferred
Exclude isotopes	On
Dynamic exclusion	30.0

[a]AGC, automatic gain control.

Specifically, the 'MaxLFQ algorithm' for accurate label-free quantification, 're-quantification' for large SILAC ratios, and the match-between-runs feature are essential.

Materials

PC with minimally 8 logical cores
8 GB of RAM

1. For full explanations of each parameter in MaxQuant, refer to Tyanova, Temu, and Cox (2016).
2. Download the uncompressed .FASTA file for the proteome that matches the species used in your experiment from UniProt, only taking SwissProt entries.

   The decision to use canonical or canonical and isoform is up to the user.

3. Open the latest version of MaxQuant and go to the configuration tab → sequence databases → add.

   It may take up to 30 sec to load the entry form.

4. Select the location of the file. In 'Source' enter the date the file was downloaded and from where. Select the taxonomy of the .FASTA file. Change the parse rule by clicking select; UniProt is one of the selections available.

   It is possible to check if this was successful by using the test rules button to see how the file is parsed using the selected parse rule.

5. Click 'modify table' from the actions menu, then click 'save changes'.

6. Go to Raw files tab → load; route to the location where your .RAW files are stored; highlight all of them; and then click open.

   All .RAW files will now be in a table which needs to be completed by defining experiments, parameter groups, and fractions. An example is given in Table 2 .

7. The parameter group is used to indicate if certain files should be searched in a different way, for example if a different enzyme has been used for digestion. It can also be used if the MaxLFQ algorithm should be applied to independent subsets of the files. This is necessary when processing the files pertaining to the 'Nuclear-Organellar-Cytosolic' fractions. The nuclear, organellar, cytosolic, and high-resolution map fractions should each get their own parameter group.

8. The experiment name is the same for every MS run which belongs to the same starting protein digestion.

   For example, high-resolution map fraction 12K, if separated into three peptide fractions would all get identical Experiment names. It follows that without peptide fractionation, the Experiment name is the same as the .RAW file name (Table 2).

9. The fraction column is used for specifying which files will be matched to one another. Files will be matched to files that are one integer above or below. With fractions labeled 1, 2, and 3, those allocated fraction 1 will be matched to all other files allocated fraction 1 and fraction 2. Fraction 2 will be matched to all other files allocated fraction 1, 2, and 3. Files allocated fraction 3 will be matched to other files allocated fraction 3 and fraction 2. Because it is not warranted to match strongly different biological samples, as this may introduce false matches, choose different fraction numbering for those samples to ensure that they will not be matched (see Table 2).

   The reason for this matching strategy is that identical peptide fractions from similar starting samples have a highly similar peptide content and some overlap with adjacent fractions but very little overlap with fractions that are non-contiguous. This is true for in-gel digested fractions, stage-tip fractions, and concatenated basic reverse-phase fractions.

10. Go to the group-specific parameters tab. For the parameter group containing the high-resolution map fractions, in the 'Type' sub-section change the multiplicity to 2, and choose SILAC labels, Arg10 and Lys8. The only other parameter change required for this parameter group is in the miscellaneous tab; check re-quantify.

    The re-quantify feature is essential for obtaining SILAC ratios where no clear isotopic pattern was identified for either the heavy or the light labeled peptide. In these cases, the shapes of the identified isotope pattern will be translated to the place in the m/z retention time plane, where the missing SILAC partner is expected and intensities will be integrated over these regions.

11. For each of the other parameter groups when using the heavy-labeled cells for the NOC fractions, in the 'type' sub-section, change the labels to Arg10 and Lys8 but leave the multiplicity set to 1. In the label-free quantification sub-section, select LFQ, and change LFQ min. ratio count to 1. In miscellaneous sub-section, change match type to 'match from and to'.

12. Go to the global parameters tab, under the sequence sub-section, and locate the .FASTA file which was configured in steps 3 to 5.

13. In the advanced identification sub-section, check 'Match between runs'.

Table 2 Experiment Template for MaxQuant[a]

Name	Parameter group	Fraction	Experiment
MAP1_Nucleus_PEPTIDE FRACTION 1	Group 1	11	MAP1_Nucleus
MAP1_Nucleus_PEPTIDE FRACTION 2	Group 1	12	MAP1_Nucleus
MAP1_Nucleus_PEPTIDE FRACTION 3	Group 1	13	MAP1_Nucleus
MAP1_3K_PEPTIDE FRACTION 1	Group 2	21	MAP1_03K
MAP1_3K_PEPTIDE FRACTION 2	Group 2	22	MAP1_03K
MAP1_3K_PEPTIDE FRACTION 3	Group 2	23	MAP1_03K
MAP1_6K_PEPTIDE FRACTION 1	Group 2	31	MAP1_06K
MAP1_6K_PEPTIDE FRACTION 2	Group 2	32	MAP1_06K
MAP1_6K_PEPTIDE FRACTION 3	Group 2	33	MAP1_06K
MAP1_12K_PEPTIDE FRACTION 1	Group 2	41	MAP1_12K
MAP1_12K_PEPTIDE FRACTION 2	Group 2	42	MAP1_12K
MAP1_12K_PEPTIDE FRACTION 3	Group 2	43	MAP1_12K
MAP1_24K_PEPTIDE FRACTION 1	Group 2	51	MAP1_24K
MAP1_24K_PEPTIDE FRACTION 2	Group 2	52	MAP1_24K
MAP1_24K_PEPTIDE FRACTION 3	Group 2	53	MAP1_24K
MAP1_80K_PEPTIDE FRACTION 1	Group 2	61	MAP1_80K
MAP1_80K_PEPTIDE FRACTION 2	Group 2	62	MAP1_80K
MAP1_80K_PEPTIDE FRACTION 3	Group 2	63	MAP1_80K
MAP1_Cytosol_PEPTIDE FRACTION 1	Group 3	101	MAP1_Cytosol
MAP1_Cytosol_PEPTIDE FRACTION 2	Group 3	103	MAP1_Cytosol
MAP1_Cytosol_PEPTIDE FRACTION 3	Group 3	102	MAP1_Cytosol
MAP1_Organelles_PEPTIDE FRACTION 1	Group 4	1001	MAP1_Organelles
MAP1_Organelles_PEPTIDE FRACTION 2	Group 4	1002	MAP1_Organelles
MAP1_Organelles_PEPTIDE FRACTION 3	Group 4	1003	MAP1_Organelles
MAP2_Nucleus_PEPTIDE FRACTION 1	Group 1	11	MAP2_Nucleus
MAP2_Nucleus_PEPTIDE FRACTION 2	Group 1	12	MAP2_Nucleus
MAP2_Nucleus_PEPTIDE FRACTION 3	Group 1	13	MAP2_Nucleus
MAP2_3K_PEPTIDE FRACTION 1	Group 2	21	MAP2_03K
MAP2_3K_PEPTIDE FRACTION 2	Group 2	22	MAP2_03K
MAP2_3K_PEPTIDE FRACTION 3	Group 2	23	MAP2_03K
MAP2_6K_PEPTIDE FRACTION 1	Group 2	31	MAP2_06K
MAP2_6K_PEPTIDE FRACTION 2	Group 2	32	MAP2_06K
MAP2_6K_PEPTIDE FRACTION 3	Group 2	33	MAP2_06K
MAP2_12K_PEPTIDE FRACTION 1	Group 2	41	MAP2_12K
MAP2_12K_PEPTIDE FRACTION 2	Group 2	42	MAP2_12K
MAP2_12K_PEPTIDE FRACTION 3	Group 2	43	MAP2_12K
MAP2_24K_PEPTIDE FRACTION 1	Group 2	51	MAP2_24K
MAP2_24K_PEPTIDE FRACTION 2	Group 2	52	MAP2_24K
MAP2_24K_PEPTIDE FRACTION 3	Group 2	53	MAP2_24K

continued

Table 2 Experiment Template for MaxQuant[a], *continued*

Name	Parameter group	Fraction	Experiment
MAP2_80K_PEPTIDE FRACTION 1	Group 2	61	MAP2_80K
MAP2_80K_PEPTIDE FRACTION 2	Group 2	62	MAP2_80K
MAP2_80K_PEPTIDE FRACTION 3	Group 2	63	MAP2_80K
MAP2_Cytosol_PEPTIDE FRACTION 1	Group 3	101	MAP2_Cytosol
MAP2_Cytosol_PEPTIDE FRACTION 2	Group 3	103	MAP2_Cytosol
MAP2_Cytosol_PEPTIDE FRACTION 3	Group 3	102	MAP2_Cytosol
MAP2_Organelles_PEPTIDE FRACTION 1	Group 4	1001	MAP2_Organelles
MAP2_Organelles_PEPTIDE FRACTION 2	Group 4	1002	MAP2_Organelles
MAP2_Organelles_PEPTIDE FRACTION 3	Group 4	1003	MAP2_Organelles

[a]Fraction numbers ensure matching across experiments for each differential centrifugation step separately. Parameter groups ensure separate LFQ for the 'Nuclear-Organellar-Cytosol' split and the organellar SILAC pairs.

14. In the 'protein quantification' sub-section, change the label min. ratio count to 1 and make sure 'advanced ratio estimation' is checked.
15. In the Label free quantification sub-section, check the box to separate LFQ in parameter groups. All other parameters can be left as default.

    Separating LFQ in parameter groups in this case ensures that the MaxLFQ algorithm will be applied independently to the 'Nuclear,' 'Organellar,' and 'Cytosolic' fractions.
16. At the bottom of the interface, change number of processors to the number of processing threads to be used in parallel for the analysis.

    MaxQuant is very computationally intensive and ideally would not be run on the same machine that is used for day-to-day activities. Also, a few threads should be left unused to allow effective running of the operating system.
17. Run analysis by hitting start. This can take several days, depending on the data volume and the machine used. See Commentary, Time Considerations section for further details.

BASIC PROTOCOL 6

VISUALIZATION OF RESULTS WITH PRINCIPAL COMPONENT ANALYSIS (PCA)

As a first step in analyzing the profile data, it is useful to visualize them and look at the relative localization of marker proteins. In order to reduce the dimensionality of the data, principal component analysis (PCA) is applied. When looking at the result, keep in mind that protein groups overlapping in the first two components might still be well separated in one of the other dimensions. Thus, it is only suitable for crude visual assessment of organelle separation and reproducibility across replicates. For our publications (see Davies et al., 2018; Itzhak et al., 2016, 2017, Hirst, Itzhak, Antrobus, Borner, & Robinson, 2018), the SIMCA software package was used for PCA (Fig. 2); this can be freely downloaded as a trial version. To make things simpler for the user, PCA is described in Perseus and R (Alternate Protocol 4).

Materials

Proteingroups.txt from MaxQuant txt output
Latest version of Perseus (v1.6.2.2 was used here)

1. Open the latest version of Perseus and in the load tab click the green arrow which indicates 'generic matrix upload'. Go to the location of the MaxQuant proteingroups.txt file, which contains your data. Select SILAC ratios, not the normalized

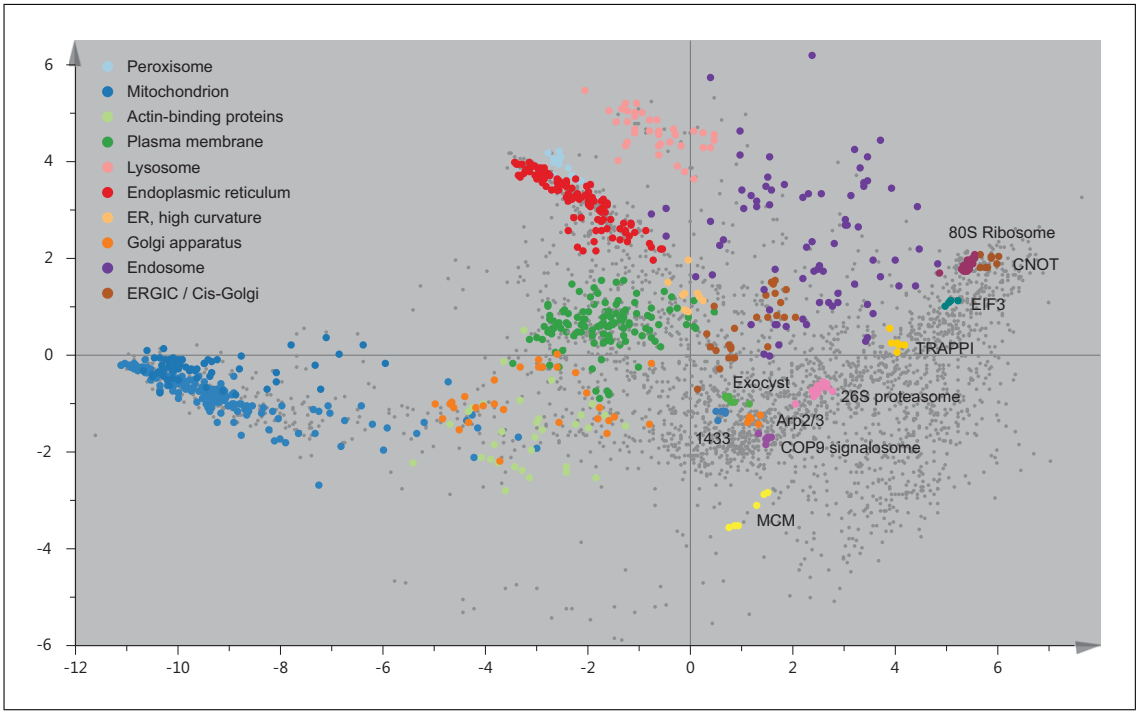

Figure 2 Typical organellar map obtained with the method described in this protocol, as visualized by principal component analysis (PCA, scores plot). Every scatter point represents a protein. Proximity of proteins suggests similar subcellular localizations. Established markers for various organelles are shown in color and indicate regions of the plot that correspond to organellar clusters. Small gray scatter points represent proteins that are not part of the organellar marker set. (Adapted from Itzhak et al., 2016.)

SILAC ratios, and click to put them into the 'main' box. Select 'Ratio H/L count' and 'Ratio H/L variability [%]' for each sample and click to put them into the numerical box. Transfer 'id' to the 'text' box and click OK.

2. Remove proteins that are only identified in a modified form, contaminants, and hits to the reverse database. Under 'processing', click 'filter rows' then 'filter rows based on categorical column', and select 'only identified by site', then click OK to remove these rows. Repeat this for 'Potential contaminant' and 'Reverse'.

3. Filter data for high quality measurements: In Perseus, under 'processing' click 'Quality', then select 'Create quality matrix'. Each sample should now be matched to its corresponding 'Ratio H/L count' using the drop down boxes. The quality matrix is now created. Click 'Quality', then select 'Filter Quality' to bring up a new dialog box, leave 'filter method' set to 'From quality matrix', set threshold to '2', leave 'values should be' as 'greater than', and click OK. Each SILAC ratio in the original matrix that had fewer than three ratio counts will now be converted to NaN.

   For more complex filtering steps, such as keeping all SILAC ratios whose ratio counts are equal to 2 and have a variability <30%, it is necessary to use another program, such as Excel, R, or Python. To export the matrix for this purpose, under the export tab click on the 'floppy disc' icon and save matrix as a .txt file.

4. Normalize data to account for different loading in the SILAC channels.

5. Even after very careful BCA assay and pipetting, there can still be up to a 20% difference in the amount of protein in the SILAC light and heavy channels. In Perseus, under 'Processing' click 'Normalization' and select 'Divide'; in the dialog box change 'Matrix access' to 'columns' and leave 'Divide by what' to 'Median'. Reformat data to generate proportion of protein per fraction.

Since all of the Ratios are H/L and the proportion of light relative to heavy is required in this instance, it is necessary to invert the ratios.

Transform the data as follows: 'basic', 'transform', then enter 1/(x), and make sure each of the SILAC ratios is on the right-hand side. These ratios now represent the fold enrichment of that protein over the reference spike-in.

6. In Perseus, under 'Analysis' click Clustering/PCA and select PCA. In the data part of the PCA transformed scatter plot, under 'Points' mark all lines from a cluster and select a color. Repeat this for all clusters to visualize cluster tightness and cluster overlap.

This procedure can be repeated following support vector machines (SVMs) classification in Basic Protocol 8 to view the predictions.

ALTERNATE PROTOCOL 4

PRINCIPAL COMPONENT ANALYSIS (PCA) IN R

As an alternative to principal component analysis (PCA; Fig. 2) in Perseus, this short protocol describes how to do it in R.

Materials

Proteingroups.txt from MaxQuant txt output
R version >3.0 including stats package (R Development Core Team, 2011)

1. Export matrix from Basic Protocol 6, step 5 from Perseus.
2. Open R session and load.txt file into a data.frame.

data <- read.table("filepath", sep="\t", header=TRUE)

3. Extract data columns of the map you want to visualize based on the indices of the column names.

names(data)

map1 <- data[,1:5]

map2 <- data[,6:10]

4. Run principal component analysis provided by the stats package of R and apply transformation to the data.

pca <- prcomp(map1, retx=TRUE, scale.=TRUE)

5. Plot variance per principal component.

plot(pca)

6. Plot PCA transformed data along principal components 1 and 3.

plot(pca$x[,1], pca$x[,3], pch=16)

7. Optional: Plot data including annotations or predictions as added by Basic Protocol 8. For this you need to change the color palette to comprise enough qualitative values. The code below is only a suggestion and the default palette can be restored by palette("default").

palette(c('#888888','#a6cee3','#1f78b4','#b2df8a','#33a02c','#fb9a99','#e31a1c', '#fdbf6f','#ff7f00','#cab2d6','#6a3d9a','#ffff99','#b15928'))

plot(pca$x[,1], pca$x[,3], pch=16, col=data$Compartment)

legend("bottomright", legend=levels(data$Compartment), col=palette(), pch=16)

Look at the graphics parameter reference to customize your plot: https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/par.html

IDENTIFICATION OF MOVING PROTEINS

BASIC PROTOCOL 7

NOTE: If you are using the mapping approach for generating 'static maps', i.e., for predicting subcellular localizations only but not for comparative purposes (± treatment), continue with Basic Protocol 8.

The movement analysis should be performed with three biological replicates, where the correlation between replicates is used as a filter for identifying consistent, and therefore likely, genuine translocations. However, it can also be performed with a single biological replicate, bearing in mind that the number of false positives will most likely be large. The purpose of performing the analysis with a single replicate is to check if positive controls (if available) are detected as 'outliers' at this stage and also to get an overview of new candidate outliers. (Further replicates will add stringency, and hence reduce the list of candidate outliers, but not add any new ones.)

The identification of proteins which have profiles that are statistically significantly different between two conditions may seem like it should be a later step in the analysis. However, the support vector machines for organellar predictions are trained on marker proteins, which are assumed to have invariant localizations; if one or more of the marker proteins shift localization between conditions, they may adversely affect organellar classification. Shifting markers should hence be identified first and excluded prior to training SVMs. Furthermore, an important feature of the outlier test described here is that it is independent of organellar assignments, as it works with the numerical abundance profiles. This has many advantages; among others, it allows the detection of partial organellar translocations and changes of localization within the same compartment (Itzhak et al., 2016).

The protocol should be applied following acquisition of three replicate measurements. This protocol will examine the steps required to convert the MaxQuant output into the format required to apply the statistical test.

Materials

Latest version of Perseus (v1.6.2.2 was used here)

1. Open Perseus, import the ratios from Basic Protocol 6, step 5 and correct them by the protein yield in those fractions. Use results of the BCA assay to obtain protein concentrations in the fractions and multiply concentrations by the volumes to obtain protein amounts. Convert protein amounts into yields by dividing the protein amount in individual fractions by the summed protein amount. Multiply inverted ratios by these yield values. These should be between 0.1 and 0.3. To multiply the inverted ratios by the yields in Perseus, click 'basic' then 'transform'; in the dialog box change the formula to (x)*yield, applying this transformation for the different fractions separately.

2. Normalize profiles to a sum of 1 so that they become comparable: Calculate sum of all fractions within a map. Using Perseus, click 'basic' then click 'summary statistics (rows)'; in the dialog box for 'expression column selection', select 'within groups' then choose the grouping performed above in Basic Protocol 6, step 5, and move everything except 'sum' to the left-hand side. Convert ratios to proportions by dividing each ratio by the sum of the ratios. In Perseus, click 'Basic' then 'Combine main columns'; in the dialog box change 'Operation' to 'x/y'. For 'x' move all of the ratios to the right-hand side and for 'y' move the respective sums for each ratio

to the right. Make sure that for every ratio there is a corresponding sum, i.e., five ratios need five complementary sums. Be sure to uncheck 'Keep original columns'.

The resulting matrix will now contain the proportion of each protein pelleted in each of the differential centrifugation fractions, from here on referred to as a 'profile'. Perseus will append the ratio names as follows: For Ratio H/L 03K 'Ratio H/L 03K_x/y_Sum_Control_map'.

3. Filter data for valid ratios in each sample of a map: In Perseus, under 'Processing' click 'annotate rows' then select 'categorical annotation rows'. Give all samples which belong to the same map the same identifier, e.g., "control map 1". Under 'Processing' click 'Filter rows' then select 'Filter rows based on valid values'. In the dialog box, change 'min. valids' to 5, set 'mode' to 'in each group'; 'values should be' to 'valid'; 'filter mode' to 'Reduce matrix', then click OK.

   This filtering will ensure that only proteins which have complete valid profiles in all replicates are retained in the dataset.

4. Generate difference profiles on which to perform the outlier test: Calculate difference between protein profiles in different conditions. When there are multiple replicates, subtract each profile from its cognate biological replicate. In Perseus, calculate the profile difference by clicking 'Basic', 'Combine main columns', then in the dialog box, leave 'Operation' as 'x-y'. Move ratios of one condition, e.g., the control map to the right-hand side of 'x' and ratios for the other condition, e.g., treated map to the right-hand side of 'y', uncheck 'Keep original columns', and click OK.

5. Use the multidimensional outlier test to identify proteins with significantly different profiles: In Perseus, click 'visualization' and then 'Histogram' to check that the difference profiles are approximately normally distributed. To perform the test, select 'Outliers' then select 'Multidimensional significance'. In the dialog box, move only one difference profile to the right-hand side for any one outlier test. Set 'Use for truncation' as *p* value, the threshold value to '0.01' (any cut-off can be used, it will not be used for distance score calculation). Uncheck the box marked 'Use standard covariance matrix'. Set the 'Quantile', i.e., the proportion of data to be used for determining the minimum covariance determinant (MCD is a robust measure for the scatter of the data; see Fauconnier & Haesbroeck, 2009).

   By choosing to use only 75% of the data to calculate the MCD, it ensures genuine outliers are less likely to be included in the estimate of data scatter. Use an odd number of iterations for this test, e.g., 101, this way the median test result will not be an average of two results. The output is a p value indicating the likelihood of observing a movement of this magnitude by chance.

6. Repeat step 8 for each of three biological replicates: A single replicate will already identify almost all outliers but there will also be considerable noise. Replicates help to remove most of the noise. If there is a positive control for a protein that is expected to have a profile difference, make sure it is identified by the outlier test in the first replicate before continuing MS measurement of the remaining replicates.

7. Calculate the movement score: The Movement score is the highest (i.e., least significant) *p* value from the three biological replicates, raised to the power of three, followed by the Benjamini-Hochberg correction, the –log10 transformation is then applied so that a lower q-value is expressed as an increased score. This is most easily calculated outside Perseus, e.g., in Excel, but can be achieved in Perseus as follows: Under 'processing', click 'rearrange' then select 'change column type'. In the dialog box, change 'source type' to 'Numerical', move the three 'Multidimensional significance *p* values' to the right-hand side, and change 'target type' to 'Main'.

Now it is possible to transform these p values into a movement score.

Go to 'Basic', 'summary statistics (rows)', then change the 'Expression column selection' to 'select columns'. A new section of the dialog box will appear. Move only the 'Multidimensional significance *p* values' to the right-hand side, and in the 'calculate' section, select 'Maximum'. Change column type of this new 'Maximum' column, as described above, from 'Numerical' to 'Main', then use 'Basic – Transform' and enter (x)^3, to cube the 'Maximum'.

This cubed p value needs to be false discovery rate (FDR) corrected and this matrix must be exported for this final step.

For the Benjamini-Hochberg correction, multiply each 'Maximum multidimensional significance p-value' by the number of rows in the table (the number of protein groups), and divide them by rank (ascending). Apply the –log10 transformation to obtain the Movement score.

The version of the outlier test described above aims at high sensitivity. Increased specificity, at the expense of decreased sensitivity, can be obtained by omitting the step where the highest p value of the three biological replicates is cubed. We recommend applying the sensitive version of this test and use three mock-treated maps to control and obtain an experimentally determined FDR at different threshold values of the Movement score (Itzhak et al., 2016).

8. Calculate the Reproducibility score.

   The Reproducibility score is the lowest Pearson correlation between difference profiles. This cannot currently be performed in Perseus.

   In Excel or other program of choice, calculate the Pearson correlation for the 5-datapoint difference profiles, for each of the three replicates, i.e., replicate 1 versus 2, replicate 1 versus 3, and replicate 2 versus 3. Determine the lowest of these three Pearson correlation values. This is the Reproducibility score.

   For each protein, plotting the Movement score on the abscissa and Reproducibility score on the ordinate of a scatter plot allows easy visualization of the data where genuine outliers tend toward the upper right corner of the plot.

9. Determine outliers based on significance thresholds or preferably a mock experiment (Itzhak et al., 2016).

   In our experience, using the same hard cut-offs for every experiment will not yield optimal results.

SUPPORT VECTOR MACHINES IN PERSEUS TO CLASSIFY THE SUBCELLULAR LOCALIZATION OF PROTEINS

BASIC PROTOCOL 8

Supervised classification with support vector machines (SVMs) works particularly well for classification of spatial proteomics data, where cluster boundaries are not linear, and cluster shapes do not fit standard distribution functions. Since supervised classification requires some proteins to be annotated with their subcellular localization, a marker set must be obtained. Great effort was undertaken to generate a marker set that is corroborated by the literature and effective in many cell lines tested in house, as published as an open access supplementary file (Itzhak et al., 2016 Supplementary File 1 'The HeLa spatial proteome', sheet 2 'Organellar Markers HeLa'). This protocol will go through the steps required to use these markers as a means to classify all other proteins using support vector machines implemented in Perseus software.

Materials

Perseus software

1. The classification should be applied on log transformed SILAC ratios that have been quality-filtered and corrected for loading, i.e., complete Basic Protocol 6, up to step 5. Under 'Processing', select 'Basic' and click 'Transform'. Leave log2(x) in 'Transformation' and ensure all SILAC ratios are on the right-hand side.
2. Remove any annotation for marker proteins, which were determined in Basic Protocol 7 to move under the different conditions.
3. Load marker .txt file into Perseus software. Select generic-matrix upload (green arrow) in the load section to open the dialog box, assign column names to the correct column type: Place 'compartment' into the 'categorical' type, and the remaining identifier column names, including 'Gene name', 'Protein name', 'Protein ID (canonical)', and 'Isoform ID used in this study' into 'Text' type.
4. Match markers into the matrix containing the SILAC ratio data. Under 'Multi-proc.' Select 'Basic', 'Matching rows by name', to open a dialog box.

   This brings up a matrix where it is necessary to choose the 'Base matrix', the one to which columns will be added and 'Other Matrix' from which information will be derived.

   Select 'Base matrix' as that containing the SILAC ratio and 'Other matrix' as that containing the marker protein annotation. Click OK.

   This brings up a second dialog box, where the identifiers that are the same in both matrices need to be matched and the columns to be added to 'Base matrix' are specified.

   Select column from each matrix which contains the 'Gene names', then place the 'Compartment' column onto the right-hand side of the categorical columns; this is the only column required. It is possible to match the two matrices and transfer the compartment information using UniProt ID, if so desired. Check that the additional column in the new matrix contains 'Compartment' information.
5. Optimize parameters for SVM classification. Under processing click 'Learning', 'Classification-parameter optimization'. In the dialog box, change 'items are in' to 'Rows' and in the 'Classes' drop-down selection make sure 'Compartment' is selected. Change 'Kernel' to 'RBF' (radial basis function).

   Ideally, perform a two-dimensional scan (grid-search) to find the optimum parameters of C and sigma.

   Change the starting value of sigma to 0.01 and the starting value of C to 1. Cycle through 10 values of C and sigma, in an additive manner, in steps of 10 for C and steps of 0.01 for sigma. This is computationally intensive; select as many threads as the computer has available.

   The parameter 'C' controls the tradeoff between a smooth decision boundary and classifying points correctly. The sigma parameter defines how far the influence of single training example reaches with respect to the boundary. For a description of the parameters that are optimized here, readers are directed to the online documentation of Perseus (http://www.coxdocs.org/doku.php?id=perseus:user:activities:matrixprocessing:learning:classificationparameteroptimization).

   CRITICAL: Use cross-validation to ensure models are not overfit. Use n-fold cross-validation to split the data into n stratified training and test sets.

This process of model building and prediction is repeated n number of times and a combined estimated misclassification rate is output. Ideally, n should be greater than the smallest cluster in the training set.

6. Use results of the grid-scan to identify parameters of C and sigma with low error rates.

7. Repeat optimization process, this time varying C and sigma in smaller steps, between values identified from the grid-search as having low error rates.

8. Apply optimized parameters to classify non-marker proteins. Click 'learning' and select 'Classification (cross-validation and prediction)', input values for C and sigma determined in step 6. Select 'leave one out' for cross-validation type and leave 'predict unassigned items' checked.

   Good predictions have values >1; values <0 are not useful for prediction, while values in between are indicative, i.e., prediction values closer to 1 are more likely to belong to that cluster. The SVM prediction can be done independently on each replicate or on the fifteen data points from three replicate maps simultaneously. The latter is easier to implement, however the former is superior, because poor data points have lower influence.

9. Predictions from independent maps can be combined into discrete classes as follows: Assign 1 to values between 0 and 0.5, the lowest confidence class; assign 2 to prediction values between 0.5 and 1; and assign 3 to prediction values >1, the highest confidence class. These classes can be summed for any number of maps.

   With three replicates, excellent predictions will have a summed prediction value of 9. Alternatively, positive prediction values can be summed over all replicates and converted to percentile scores based on the number of marker proteins that attain that value (Itzhak et al., 2016).

CALCULATION OF PROTEIN COPY NUMBERS USING THE PROTEOMIC RULER APPROACH

BASIC PROTOCOL 9

Mass spectrometry intensities from different peptides should not be compared directly owing to different physicochemical properties and thus different ionization efficiencies of distinct peptide sequences. However, averaging multiple peptides from the same protein overcomes this limitation to a large extent and it has been shown that such approaches can be as accurate as approaches using internal standards (Wiśniewski, Hein, Cox, & Mann, 2014). Summing intensities of all peptides belonging to the same proteins and dividing this by the summed intensity of all peptides in the sample is a means to convert intensity values into concentrations, where one knows the amount of peptide sample injected into the mass spectrometer. Alternatively, the summed intensity of the peptides can be scaled to any value, for example the mass of the cell. Using the proteomic ruler, the summed intensity is scaled to the mass of the cell via the summed intensity of all the histones in the sample (Wiśniewski et al., 2014). The histone mass per cell is roughly the same as the mass of DNA in the genome. Knowing the size of the genome, one can replace the summed intensity of the histones with the genome mass and extrapolate to all remaining intensities. Here, the technical details of normalizing the data from the nuclear-organellar-cytosol split are described followed by running the 'Proteomic Ruler' plug-in in Perseus.

Materials

Perseus version 1.5.5.0
Proteomic ruler v.0.1.6

1. Normalize intensities from the 'NOC' split samples by correcting for different summed intensity in each sample, i.e., different sample amount injected into the mass spectrometer.

   This is easier to implement in Excel, rather than Perseus.

   Calculate summed intensity for each of the intensity columns. Divide each of these summed intensities by the lowest summed intensity to generate a normalization factor for each column. Divide each value in the column by that column's normalization factor.

   Use the raw intensities not the LFQ intensities. Ideally, most correction factors will be around 1, and at the end of this process, the summed intensity of each column should be the same.

2. Use results of the BCA assay to determine protein mass in the 'NOC' fractions and how much protein mass each fraction contributes to the total. Accomplish by multiplying protein concentrations by the volumes of sample.

   For accuracy it is necessary to take into account the amount of sample lost at various steps and the volumes of the samples. Using the recommended volumes, the nuclear pellet is derived from 4.356 ml of cell lysate. The supernatant of this pellet varies depending on the cell type and cell number used: Typically 4.1 ml of supernatant remains. However, only 3.6 ml of this supernatant is used for generating the membrane and cytosolic fractions. Thus, the protein mass in the membrane pellet should be multiplied by 1.14 (4.1/3.6). Because the final volume of cytosol has been diluted with 5× lysis buffer, multiply the protein mass by this volume (4.5 ml) and account for the loss of volume (multiply by 1.14).

   In cell types tested, the cytosolic fraction contributes the most mass (∼50 %), with the nuclear and organellar fractions contributing between 20% to 30% each. Confirm that proportions for the different replicates are highly similar, within 2% of each other and calculate the mean proportion.

3. Adjust summed intensities so that each 'NOC' split is equivalent to a full proteome. Multiple each intensity in the column by the respective median contribution to the total cell mass.

   For example, each value in a cytosolic sample would be multiplied by 0.5.

4. Sum the intensity from fractions belonging to the same biological 'NOC' split to obtain a compound full proteome. Save this as a .txt file, retaining all annotation information, including gene names, protein names, and especially UniProt IDs.

5. Download Perseus version 1.5.5.0 and the associated proteomic ruler v.0.1.6. .dll file and copy it into the Perseus folder. Start the software.

   For full details, read the associated plugin documentation and accompanying publication on the proteomic ruler (Wiśniewski et al., 2014).

6. Import.txt file using the generic matrix upload function, placing the compound full proteome columns into 'main', and annotation columns into 'Text'.

7. Under 'Processing', click 'Proteomic ruler' and select 'annotate proteins'. Choose 'Majority Protein IDs' as the column which contains the Protein IDs and locate the .FASTA file that was used in the processing of the .RAW files for this project.

   It is possible to choose between annotating with information from all UniProt IDs in the Majority Protein ID or the Leading ID. Depending on the experimental question, one of these options may be preferable.

Calculate theoretical peptides, selecting the protease used for digestion; trypsin is selected by default. The minimum and maximum peptide lengths of 7 and 30, respectively, are empirical and based on what length peptides are most frequently identified in database searches. Click OK to complete the annotation.

This step will annotate the proteins based on the UniProt identifiers, using a stored list of histone UniProt IDs and the information from the fasta file including the molecular weights and sequence from which it will calculate the number of tryptic peptides.

8. Under 'Processing' click 'Proteomic ruler' → 'Estimate copy numbers and concentrations'. In the dialog box, for the 'Protein IDs' choose the column which contains 'Majority protein IDs'. If the data are logarithmized, check the box and choose the log base. For 'averaging mode', selection of 'all columns separately' is preferable for the purposes of this analysis because it enables determination of the precision of the method. Check the box for 'detectability correction'.

   The automatically selected correction factor is 'Number of theoretical peptides (trypsin/P, 7-30), calculated in step 7, which corrects the copy numbers of proteins based on the number of predicted peptides suitable for MS from tryptic digestion.

   For the scaling mode, select 'Histone proteomic ruler' and input the ploidy of the cell line. The default value for 'total cellular protein concentration [g/l] of 200' is true across a range of biological samples, but if more accurate determinations are available, substitute your own value. Click OK.

   Both copy numbers and concentrations will now appear in the matrix, in addition to the existing intensity values. Values for a second table, summarizing the protein amount per cell and the cell size, are calculated.

COMMENTARY

Background Information

The concept of organellar profiling goes back to Christian de Duve, who used ultracentrifugation to separate various cell components on the basis of their size and density (De Duve, Pressman, Gianetto, Wattiaux, & Appelmans, 1955). He observed that different enzymes had characteristic profiles across density gradients, suggesting their inclusion in different types of cytoplasmic granules, now known as organelles. For this remarkable discovery, among others, de Duve shared the 1974 Nobel Prize in Physiology or Medicine together with Albert Claude and George Palade, for discoveries "concerning the structural and functional organization of the cell" (NobelPrize.org, n. d.). It was not until almost 50 years later before enzymatic assays were replaced with mass spectrometry-based proteomics to identify proteins in organellar profiling experiments.

Localization of organelle proteins by isotope tagging (LOPIT) uses a combination of isopycnic separation and isotope-coded affinity tags for quantification (Dunkley, Watson, Griffin, Dupree, & Lilley, 2004, updated in Christoforou et al., 2016). A closely related approach published around the same time was protein correlation profiling (PCP, Foster et al., 2006, updated in Krahmer et al., 2018), which uses rate-zonal centrifugation for organelle separation and label-free quantification of proteins. However, the difficulty of reproducibly preparing and harvesting the gradients for organelle separation rendered these methods challenging to use in comparative experiments (Gatto et al., 2014). To overcome this problem, Borner et al. (2014) developed a simple yet robust organellar separation protocol that combined differential centrifugation with high accuracy SILAC proteomics, initially to classify clathrin-coated vesicle proteins from sub-proteomes. Buoyed by this success, the method was subsequently adapted to entire proteomes and for the first time allowed the global identification of protein translocations across MS-based organellar mapping experiments, exemplified by capturing the translocations during epidermal growth factor (EGF) signaling (Itzhak et al., 2016). In addition to the establishment of a reproducible separation workflow, the development of a complementary statistical framework for the sensitive and specific detection of protein movements was instrumental (Itzhak et al., 2016, 2017). The method has since been successfully applied as an unbiased discovery

tool in cell biology to reveal the cellular defects caused by the neurodegenerative AP-4 deficiency syndrome (Davies et al., 2018) and to uncover the function of the AP-5 protein sorting complex (Hirst et al., 2018). Furthermore, the method has been used to provide extensive databases of protein subcellular localization and copy number information, from HeLa cells (*www.MapOfTheCell.org*; Itzhak et al., 2016) and from mouse primary neurons (Itzhak et al., 2017).

Critical Parameters

In our experience, there have been very few instances where the protocol has not worked. However, the protocol has given rise to maps with poor resolution from cells undergoing acute cell death and it is recommended only to look at early signaling events during cell death, before membrane blebbing.

The most important factor for reproducibility and application of the comparative method is consistent cell lysis; Basic Protocol 1 and Alternate Protocol 1 describe how this can be achieved.

Key to reproducibility is to ensure that reasonable amounts of protein are pelleted in all fractions. Typically, none of the five membrane fractions contains $<10\%$ of total membrane protein. The lowest protein contribution is most frequently the 24,000 $\times$ g fraction.

Troubleshooting

If a large amount of organellar lysis is observed, which can be revealed by detecting substantial proportions of luminal contents in the cytosolic fraction, and visualized by PCA as a very obvious separation of membrane bound and luminal contents of the same organelle, then lysis conditions should be further optimized. Gentle lysis is achieved by swelling with hypotonic buffer followed by Dounce homogenization or a ball-bearing homogenizer. The strength of the hypotonic buffer can be varied and tested in simple experiments using Alternate Protocol 2 to monitor luminal leakage into the cytosol. With limited access to mass spectrometry, a panel of luminal marker proteins can be checked for cytosolic leakage by Western blotting, relative to the proportion in the other two nuclear and organellar fractions.

If no swelling is desired, it is necessary to use the ball-bearing homogenizer with small clearance (large ball bearing), subject to optimization. This will release organelles more consistently than the Dounce homogenizer.

Should insufficient protein be measured by BCA assay in any of the fractions, it may be necessary to optimize the centrifugation speeds. It is necessary to check the yields of the other fractions to determine where the protein is pelleting and use this information to adjust the speeds accordingly. For example, if no protein is obtained in fraction 2, yet fraction 1 is very rich in protein, then the speed of the first centrifugation could be decreased. In practice, there is variation between cell types using the speeds optimized for HeLa cells, with respect to the amount of pellets at each centrifugation step, but this has not resulted in completely empty fractions in our hands.

If the amount of starting material is very low and the protein concentration is lower than is possible to determine using BCA assay, do not abandon the preparation. Assuming the protocol was followed correctly, it is likely that the amount of material is still sufficient for mass spectrometry. In such cases, where using the label-free protocol, just process all material and inject a defined volume. The MaxLFQ algorithm will normalize for the different amount of peptide used for analysis. If using the SILAC protocol, divide the heavy reference equally among the five fractions. The unequal loading can be corrected as described in Basic Protocol 6, step 4.

Statistical Analyses

These have been described in Basic Protocols 6 through 8. The recommended data transformations are indicated as well as how to implement them using Perseus software, where the statistical tests are implemented at the click of a button.

Anticipated Results

The method separates the detection of outliers using multivariate statistics from the compartment prediction of the proteins using support vector machines. This can lead to situations where a protein is deemed to have moved but appears to remain in the same compartment. This can happen when only a portion of the protein has moved to a new compartment and because the method only reveals the average position of a protein, the organelle prediction does not change. In such circumstances it would be optimal to confirm the translocation by microscopy and get a visual understanding of the position of the protein under the different conditions. In addition, it is possible to infer some information

from the direction in which the protein shifts, either by looking at the proteins in a principal component analysis scores plot or by looking at the profiles. This can reveal if the protein now pellets earlier or later in the gradient, indicating transfer to a larger or smaller particle, respectively.

Compartment association prediction works best for proteins with a single predominant steady state localization. Low prediction scores or unexpected predictions can be a sign of multimodal organellar association, with mixed (and hence often unclear) profiles. These mixed profiles can be a result of inhomogeneous cell cultures or they can be due to genuine multilocalizations of proteins. Several types of multilocalization are commonly encountered. Some proteins have a cytosolic pool and a membrane associated pool. This case is easily dealt with by the approach described here because the cytosolic pool of each protein is determined independently from its organellar association. Even if only a small fraction of a protein is membrane associated, the organelle can still be predicted accurately. Likewise, a mixed localization between nucleus and cytosol is quantified directly. On the other hand, dual localization between different membrane organelles can pose a problem because the compartment assignment uses hard (single location) classifiers. If the protein is mostly associated with one organelle and has a minor second association, the method will assign the protein to the major localization only. If the protein is evenly split over two or multiple organelles, a mixed profile will be obtained, which usually has a low confidence prediction score or even a wrong classification as the profile does not match to any reference. The biological context and prior knowledge of the protein's localization, where known, together with an analysis of the closest neighbors of a protein, can help to understand these mixed profiles. It is important to note that dual localizations are only limiting for the absolute assignment of proteins to organelles but not for the identification of moving proteins because that part of the analysis depends only on the numeric values of the profiles. Thus, changes in the steady state mixed localization of a protein are still detectable, even if they may not be fully interpretable from the profiling data alone.

Perhaps the most confusing cluster to interpret is the 'large protein complex' cluster, since it does not reflect a membrane-bound organelle. The proteins belonging to this cluster pellet either because they are genuinely in large complexes (e.g., the proteasome) or because they are associated with a large structure (such as mRNA). The co-pelleting can be caused by direct or indirect associations; profiling does not make predictions about protein interactions per se but reveals sets of proteins with similar fractionation patterns, in turn suggesting physical proximity in cells. Inspecting a protein's profile 'neighborhood' can be very informative, using the distance between profiles across multiple experiments as a measure of similarity; this is explored further elsewhere (Borner et al., 2014; Itzhak et al., 2016). Importantly, proteins in the large protein complex cluster have relatively stable positions between maps; movements within this cluster can be readily detected using the outlier test (Basic Protocol 6).

Time Considerations

Starting at the point from which cells are in cultured dishes or have been extracted from some source, the protocol from cell lysis through fractionation and protein extraction takes 1 day in the lab (Basic Protocols 1 and 2). The subsequent day can be used for protein concentration determination and tryptic digestion (Basic Protocol 3). Because Basic Protocol 3 involves an overnight digestion step, the purification of peptides must be completed the following day (2 hr). Mass spectrometry measurements (Basic Protocol 4) take ~3 hr per sample when using 2.5-hr gradients, allowing 0.5 hr for sample injection. Thus, a single map, consisting of five membrane fractions, together with the three fractions from the 'Nuclear-Organellar-Cytosol' split, takes 24 hr in total.

We recommend processing a single pilot map using the cell type of interest to confirm the protocol gives the desired results. Following successful separation of organelles, confirmed by visual inspection using principal component analysis, and SVM classification accuracy, further maps can be measured. A complete binary comparison experiment (treated versus untreated) in biological triplicate, run in single-shot, takes 6 days of measurement time, whereas peptide fractionation increases this number linearly with the number of peptide fractions.

Processing of six maps with matching 'NOC' fractions processed in single peptide fractions would constitute 48 .RAW files. This should take between 3 to 4 days of processing time on a computer with eight logical cores. The time is substantially decreased when using solid-state hard drives and decreased

to 60% of the running time with the new Linux version of MaxQuant (Sinitcyn et al., 2018).

Basic Protocols 6 through 8 can be run in a single day, especially once the user achieves familiarity with the software.

Acknowledgement

We thank Matthias Mann for his continued support of this project. This work was funded by the German Research Foundation (DFG/Gottfried Wilhelm Leibniz Prize MA 1764/2-1), the Louis-Jeantet Foundation, the European Research Council Synergy grant "ToPAG—Toxic protein aggregation in neurodegeneration" (ERC-2012-SyG_318987-ToPAG), and the Max Planck Society for the Advancement of Science. We are grateful to Gabriele Sowa, Igor Paron, and Christian Deiml for outstanding technical support. We give special thanks to Jan Rudolph for Perseus support and all members of the Mann Department for valuable feedback.

Literature Cited

Borner, G., Hein, M., Hirst, J., Edgar, J., Mann, M., & Robinson, M. (2014). Fractionation profiling: A fast and versatile approach for mapping vesicle proteomes and protein-protein interactions. *Molecular Biology of the Cell*, *25*(20), 3178–3194. doi: 10.1091/mbc.e14-07-1198.

Christoforou, A., Mulvey, C., Breckels, L., Geladaki, A., Hurrell, T., Hayward, P., ... Lilley, K. (2016). A draft map of the mouse pluripotent stem cell spatial proteome. *Nature Communications*, *7*, Article number: 9992. doi: 10.1038/ncomms9992.

Cox, J., Hein, M., Luber, C., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics*, *13*(9), 2513–2526. doi: 10.1074/mcp.M113.031591.

Davies, A., Itzhak, D., Edgar, J., Archuleta, T., Hirst, J., Jackson, L., ... Borner, G. (2018). AP-4 vesicles contribute to spatial control of autophagy via RUSC-dependent peripheral delivery of ATG9A. *Nature Communications*, *9*, Article number: 3958. doi: 10.1038/s41467-018-06172-7.

De Duve, C., Pressman, B., Gianetto, R., Wattiaux, R., & Appelmans, F. (1955). Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *The Biochemical Journal*, *60*(4), 604–617. doi: 10.1042/bj0600604.

Dunkley, T., Watson, R., Griffin, J., Dupree, P., & Lilley, K. (2004). Localization of Organelle Proteins by Isotope Tagging (LOPIT). *Molecular & Cellular Proteomics*, *3*(11), 1128–1134. doi: 10.1074/mcp.T400009-MCP200.

Fauconnier, C., & Haesbroeck, G. (2009). Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology*, *6*(4), 363–379. doi: 10.1016/j.stamet.2008.12.005.

Foster, L., de Hoog, C., Zhang, Y., Zhang, Y., Xie, X., Mootha, V., & Mann, M. (2006). A mammalian organelle map by protein correlation profiling. *Cell*, *125*(1), 187–199. doi: 10.1016/j.cell.2006.03.022.

Gatto, L., Breckels, L., Burger, T., Nightingale, D., Groen, A., Campbell, C., ... Lilley, K. (2014). A foundation for reliable spatial proteomics data analysis. *Molecular & Cellular Proteomics*, *13*(8), 1937–1952. doi: 10.1074/mcp.M113.036350.

Geiger, T., Wisniewski, J., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., & Mann, M. (2011). Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature Protocols*, *6*(2), 147–157. doi: 10.1038/nprot.2010.192.

Graham, J., & Rickwood, D. (1997). *Subcellular fractionation: A practical approach*. Oxford: IRL Press at Oxford University Press.

Hirst, J., Itzhak, D., Antrobus, R., Borner, G., & Robinson, M. (2018). Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLoS Biology*, *16*(1), e2004411. doi: 10.1371/journal.pbio.2004411.

Itzhak, D., Davies, C., Tyanova, S., Mishra, A., Williamson, J., Antrobus, R., ... Borner, G. (2017). A mass spectrometry-based approach for mapping protein subcellular localization reveals the spatial proteome of mouse primary neurons. *Cell Reports*, *20*(11), 2706–2718. doi: 10.1016/j.celrep.2017.08.063.

Itzhak, D., Tyanova, S., Cox, J., & Borner, G. (2016). Global, quantitative and dynamic mapping of protein subcellular localization. *eLife*, *5*, e16950. doi: 10.7554/eLife.16950.

Kelstrup, C., Bekker-Jensen, D., Arrey, T., Hogrebe, A., Harder, A., & Olsen, J. (2018). Performance evaluation of the Q exactive HF-X for shotgun proteomics. *Journal of Proteome Research*, *17*(1), 727–738. doi: 10.1021/acs.jproteome.7b00602.

Krahmer, N., Florian Schueder, B. N., Quagliarini, F., Steger, M., Seitz, S., Kasper, R., ... Mann, M. (2018). Organellar proteomics and phospho-proteomics reveal subcellular reorganization in diet-induced hepatic steatosis. *Developmental Cell*, *47*, 205–221. doi: 10.1016/j.devcel.2018.09.017.

Kulak, N., Pichler, G., Paron, I., Nagaraj, N., & Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods*, *11*(3), 319–324. doi: 10.1038/nmeth.2834.

NobelPrize.org. (n. d.). The Nobel Prize in Physiology or Medicine 1974. Nobel Media AB 2018. Retrieved from *https://www.nobelprize.org/prizes/medicine/1974/summary/*

R Development Core Team. (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna: *http://www.R-project.org/*.

Scheltema, R., Hauschild, J.-P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., ... Mann, M. (2014). The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Molecular & Cellular Proteomics*, *13*(12), 3698–3708. doi: 10.1074/mcp.M114.043489.

Sinitcyn, P., Tiwary, S., Rudolph, J., Gutenbrunner, P., Wichmann, C., Yılmaz, Ş., ... Cox, J. (2018). MaxQuant goes Linux. *Nature Methods*, *15*(6), 401–401. doi: 10.1038/s41592-018-0018-y.

Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, *11*(12), 2301–2319. doi: 10.1038/nprot.2016.136.

Warner, D., Sakai, D., & Sandell, L. (2015). Mammalian cell culture. *Current Protocols Essential Laboratory Techniques*, *10*(1), 4.3.1–4.3.33. doi: 10.1002/9780470089941.et0403s10.

Wiśniewski, J., Hein, M., Cox, J., & Mann, M. (2014). A "Proteomic Ruler" for protein copy number and concentration estimation without spike-in standards. *Molecular & Cellular Proteomics*, *13*(12), 3497–3506. doi: 10.1074/mcp.M113.037309.

Yu, Y., Smith, M., & Pieper, R. (2014). A spinnable and automatable StageTip for high throughput peptide desalting and proteomics. *Protocol Exchange*. doi: 10.1038/protex.2014.033.

Key References

Davies et al., 2018. See above.

Application of proteomic methods to identify proteins whose subcellular localization are adaptor protein 4 (AP-4)-dependent.

Itzhak et al., 2017. See above.

Label-free quantification (LFQ) method to generate organellar maps and describe protein translocations.

Itzhak et al., 2016. See above.

Quantitative view of HeLa cell anatomy using Dynamic Organellar Maps method and application to study EGF stimulation.

Internet Resources

http://www.coxdocs.org/doku.php?id=perseus:sTart

Online documentation of Perseus.

http://www.mapofthecell.org/

Online resource to browse results of Itzhak et al., 2016.