Cell Host & Microbe, Volume *13*

Supplemental Information

The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of *Mycobacterium tuberculosis*

Olga T. Schubert, Jeppe Mouritsen, Christina Ludwig, Hannes L. Röst, George Rosenberger, Patrick K. Arthur, Manfred Claassen, David S. Campbell, Zhi Sun, Terry Farrah, Martin Gengenbacher, Alessio Maiolica, Stefan H.E. Kaufmann, Robert L. Moritz, and Ruedi Aebersold

Supplemental Information Inventory

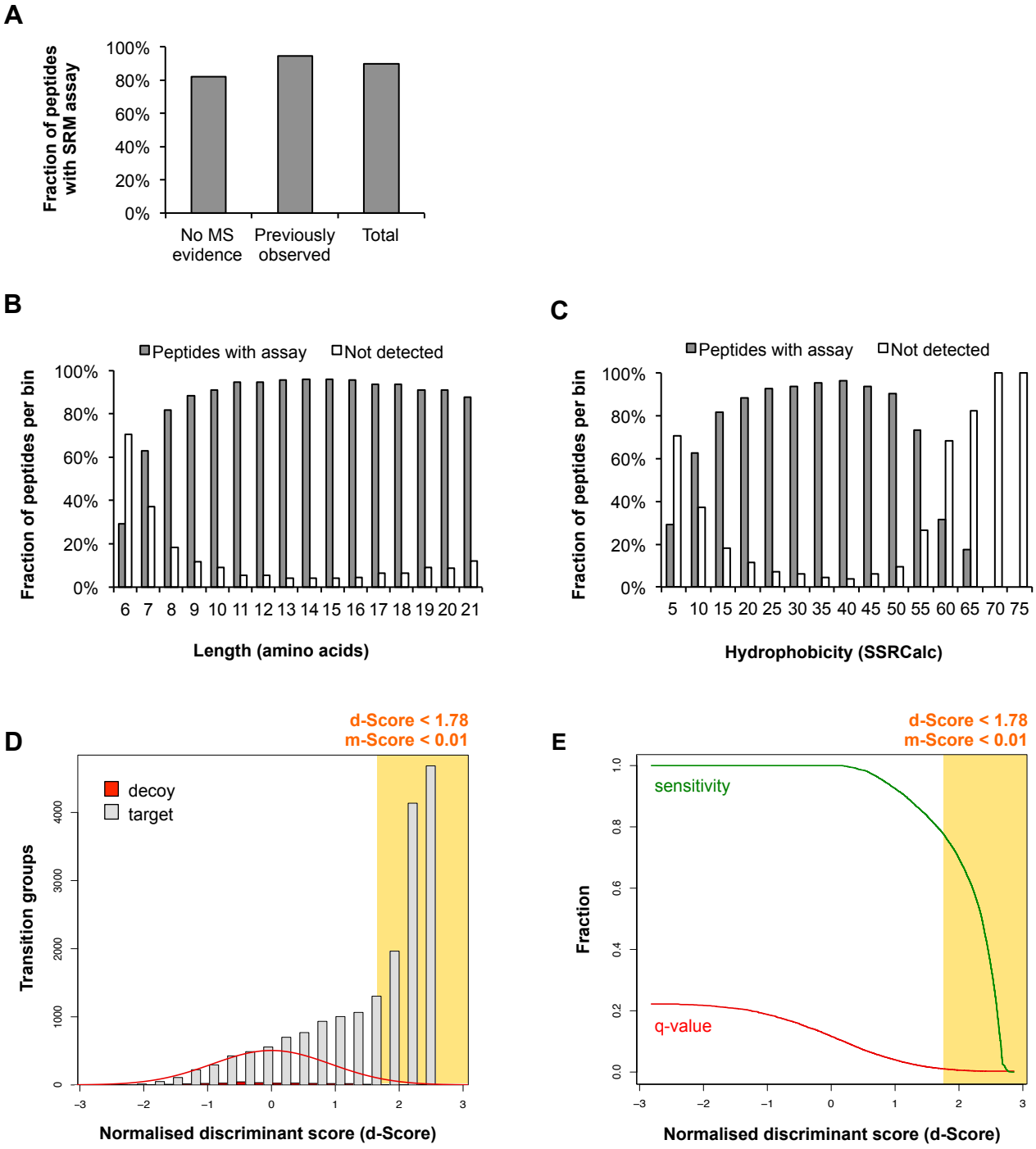
A
Fraction of peptides with SRM assay
100%
80%
60%
40%
20%
0%
No MS evidence
Previously observed
Total
B
Peptides with assay
Not detected
Fraction of peptides per bin
100%
80%
60%
40%
20%
0%
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Length (amino acids)
C
Peptides with assay
Not detected
Fraction of peptides per bin
100%
80%
60%
40%
20%
0%
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
Hydrophobicity (SSRCalc)
D
d-Score < 1.78
m-Score < 0.01
decoy
target
Transition groups
0
1000
2000
3000
4000
−3 −2 −1 0 1 2 3
Normalised discriminant score (d-Score)
E
d-Score < 1.78
m-Score < 0.01
sensitivity
q-value
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
−3 −2 −1 0 1 2 3
Normalised discriminant score (d-Score)

Figure S1. Evaluation of synthetic peptide properties and suitability for SRM assay generation and statistical analysis of large-scale SRM validation data by mProphet, related to Figure 3.

(A) Success rate of SRM assay generation from synthetic peptides with and without prior observations in the PeptideAtlas database. (B) Influence of the relative hydrophobicity (SSRCalc) of the synthetic peptides to yield a spectrum. Hydrophobicity affects the peptide synthesis, the behaviour of the peptides during sample preparation, and the reverse-phase chromatography. (C) Influence of the length of the synthetic peptides to yield a spectrum. Very short peptides generate less informative spectra which can often not be assigned to a peptide. (D) Results of the validation of SRM assays in whole cell lysates of Mtb as histogram plot. It shows the distribution of decoy and target transition groups according to their discriminant score as determined by the mProphet software, which was used to estimate the FDR of the underlying dataset. (E) Receiver operating characteristic (ROC) curve showing the sensitivity and the q-value as a function of the discriminant score.

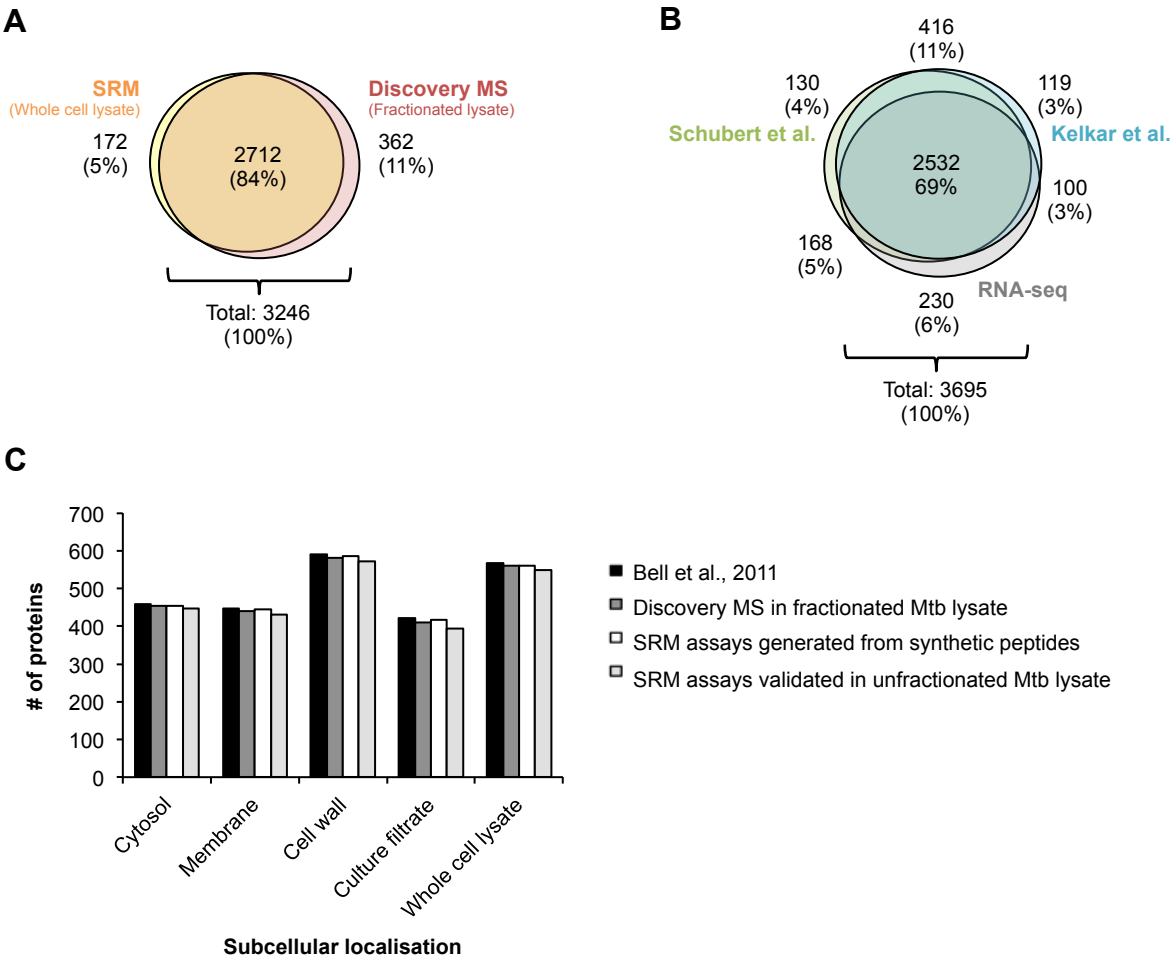
A
SRM
(Whole cell lysate)
Discovery MS
(Fractionated lysate)
172
(5%)
2712
(84%)
362
(11%)
Total: 3246
(100%)
B
416
(11%)
130
(4%)
119
(3%)
Schubert et al.
Kelkar et al.
2532
69%
100
(3%)
168
(5%)
RNA-seq
230
(6%)
Total: 3695
(100%)
C
# of proteins
700
600
500
400
300
200
100
0
Cytosol
Membrane
Cell wall
Culture filtrate
Whole cell lysate
Subcellular localisation
Bell et al., 2011
Discovery MS in fractionated Mtb lysate
SRM assays generated from synthetic peptides
SRM assays validated in unfractionated Mtb lysate

Figure S2. Mtb proteome coverage by different techniques, related to Figure 3.
(A) Comparison between proteome coverage achieved by targeted and discovery-driven MS. When comparing protein detectability by these two techniques it is important to note that the discovery MS measurements were done on extensively fractionated lysates, whereas for the SRM validation, the measurements were done on whole cell lysates. (B) Proteome coverage comparison of the present study (SRM and discovery MS combined) with a recent proteomic study on Mtb by Kelkar et al. and with the transcriptome of Mtb determined by RNA sequencing in exponential- and stationary-phase cultures of Mtb as determined by Arnvig et al. (Arnvig et al., 2011; Kelkar et al., 2011). (C) Distribution of the proteins identified by different MS techniques among various subcellular localisations as determined by Bell and colleagues (Bell et al., 2012).

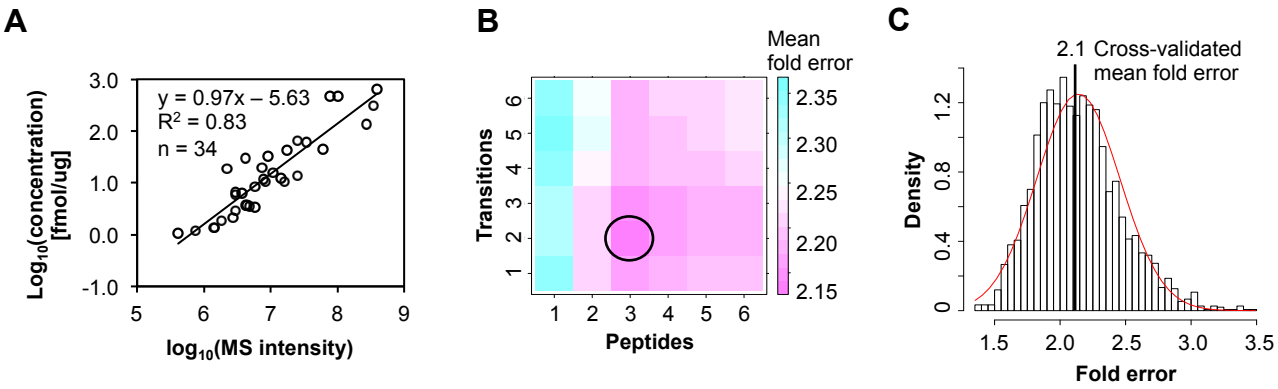
A
y = 0.97x – 5.63
R² = 0.83
n = 34
Log10(concentration) [fmol/ug]
log10(MS intensity)
B
Mean fold error
Transitions
Peptides
C
2.1 Cross-validated mean fold error
Density
Fold error

Figure S3. Determination of absolute proteome-wide protein concentrations, related to Figure 4.

(A) Plot showing for 34 anchor proteins the linear correlation between $\log_{10}$ of the SRM intensity and the protein concentration in fmol/µg protein extract. Protein concentrations were determined by spiking absolutely quantified heavy isotope-labelled reference peptides. The MS intensity is the sum of the two most abundant transitions of the three best-flyer peptides per protein. (B) Heat map showing the estimated mean fold errors, determined by Monte Carlo cross validation, of different models to summarise MS intensities of varying numbers of most intense transitions and peptides. The lowest mean fold error is achieved by summing the two most intense transitions of the three most intense peptides per protein. (C) Histogram of the cross-validated fold error distribution for the optimal model as shown in (B) (mean fold error = 2.1 ± 0.62, confidence interval with 0.95 significance level).

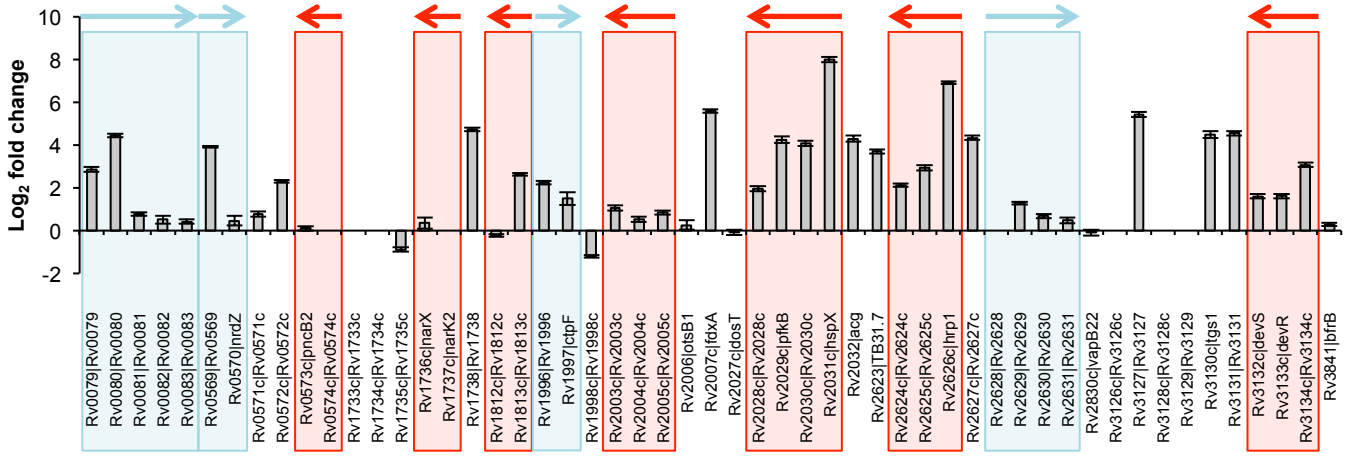
Log2 fold change
10
8
6
4
2
0
-2
Rv0079|Rv0079
Rv0080|Rv0080
Rv0081|Rv0081
Rv0082|Rv0082
Rv0083|Rv0083
Rv0569|Rv0569
Rv0570|nrdZ
Rv0571c|Rv0571c
Rv0572c|Rv0572c
Rv0573c|pncB2
Rv0574c|Rv0574c
Rv1733c|Rv1733c
Rv1734c|Rv1734c
Rv1735c|Rv1735c
Rv1736c|narX
Rv1737c|narK2
Rv1738|Rv1738
Rv1812c|Rv1812c
Rv1813c|Rv1813c
Rv1996|Rv1996
Rv1997|ctpF
Rv1998c|Rv1998c
Rv2003c|Rv2003c
Rv2004c|Rv2004c
Rv2005c|Rv2005c
Rv2006|otsB1
Rv2007c|fdxA
Rv2027c|dosT
Rv2028c|Rv2028c
Rv2029c|pfkB
Rv2030c|Rv2030c
Rv2031c|hspX
Rv2032|acg
Rv2623|TB31.7
Rv2624c|Rv2624c
Rv2625c|Rv2625c
Rv2626c|hrp1
Rv2627c|Rv2627c
Rv2628|Rv2628
Rv2629|Rv2629
Rv2630|Rv2630
Rv2631|Rv2631
Rv2830c|vapB22
Rv3126c|Rv3126c
Rv3127|Rv3127
Rv3128c|Rv3128c
Rv3129|Rv3129
Rv3130c|tgs1
Rv3131|Rv3131
Rv3132c|devS
Rv3133c|devR
Rv3134c|Rv3134c
Rv3841|bfrB

Figure S4. DosR regulon expression mapped to the genome, related to Figure 5.
Log_2 fold changes after four days (vs. day 0) of exposure to hypoxia in the standing culture model are shown for all members of the DosR regulon sorted according to their genome location. Blue boxes and arrows indicate operons on the forward strand, whereas red boxes and arrows indicate operons on the reverse strand. Operons were defined as described by Chauhan and colleagues (Chauhan et al., 2011). Error bars represent the standard error.

How to use the Mtb Proteome Library

List with proteins of interest

- Compile a list with proteins (Rv numbers) or peptide sequences of interest.

The upload from a file works only if line breaks are in Windows or Unix format.

Extraction of peptides and transitions from SRMAtlas

- Go to www.SRMAtlas.org.
- Click on “Search SRM Assays”.
- Select PABST build “**Mtb SRM Atlas 2012-10**”.
- Upload your protein or peptide list (.txt file)
- Select the appropriate parameters. For explanations click the respective “?”.
- Click “QUERY”.
- The output table contains transitions for each peptide. These transitions were either selected from a spectrum or predicted by the PABST algorithm (see “Source” column).
- Download this table in tsv format.

To examine the spectrum which was used to generate the assays, click on the little icon on the right.

Refinement of peptide list using experimental data in PASSEL

- Go to www.PeptideAtlas.org/passel.
- Click on “Query SRM results”.
- Enter each protein separately and inspect the peak groups for each of the peptides in your list.
- If possible, prioritise peptides that have an intense and high quality peak group (mScore ≤0.01). mScores are given also in Table S1.

Not for all peptides with an assay in SRMAtlas an SRM trace is available in PASSEL. The low-scoring hits of the first validation round were filtered out.

Depending on sample and instrument type SRM traces may look different. It might well be that under the conditions used for the validation, certain proteins were not expressed, but that they are detectable in your sample.

Preparation of an instrument method

- For scheduled SRM add the transitions of the retention time peptides to the instrument method and convert the iRT values of your target peptides into the corresponding retention time of your instrument platform. iRT values are also given in Table S1.
- If heavy isotope labelled reference peptides will be used for quantification, extract the heavy transitions together with the light transitions directly from SRMAtlas and add them to the instrument method.

The SRM transitions in PASSEL are not always identical with the transitions exported from SRMAtlas, because an earlier version of the Mtb SRMAtlas was used to select transitions for the validation.

iRTs are independent, unit-less retention times. To learn more about them, see Escher et al., Proteomics 2012.
If no iRT value is available you will have to run an unscheduled run or predict it with SSRCalc.

SRM measurement

Figure S5. How to use the Mtb Proteome Library, related to Figure 5.
Suggestions on how the Mtb Proteome Library could be used to build your SRM method starting from a list of proteins of interest.

Supplemental Experimental Procedures

Proteome definition

The complete Mtb H37Rv proteome was defined as annotated in the TubercuList database and contained 4012 proteins (TubercuList v2.3, April 2011) (Lew et al., 2011). The Mtb H37Rv genome encodes eight proteins mapping to more than one genome location and thus have different gene names (Rv numbers). In the Mtb Proteome Library these were counted as individual proteins to be able to refer back to the genome annotation (Table S1).

Organisms and culture conditions

Mtb H37Rv (ATCC #27294) and *M. bovis* BCG Danish SSI 1331 (BCG, ATCC #35733) were grown in Middlebrook 7H9 medium (Difco BD Biosciences) supplemented with 0.3% glycerol, 0.089% NaCl, and 0.05% of the detergent tyloxapol. Liquid cultures were incubated at 37°C on an orbital shaker with 100 rpm. Growth was monitored daily by measuring the optical density at 600 nm (OD_{600}). Protein samples were harvested during early exponential growth (OD_{600}=0.4), late exponential growth (OD_{600}=1.0–1.5), and stationary phase (one week after reaching maximal OD_{600}). For harvesting the bacteria, cultures were pelleted by centrifugation at 3000 g for 5 min, washed with ice-cold phosphate-buffered saline and stored at -80°C until use. To obtain hypoxic conditions, we employed a standing culture model, where agitation was stopped when cultures had reached exponential growth (Kendall et al., 2004). Thus, levels of dissolved oxygen in the medium decrease and bacteria settle to the bottom of the culture vessel. The screw caps were kept closed until harvesting to avoid fresh oxygen entering the vessel.

Sample preparation

Bacterial cell pellets were dissolved in lysis buffer containing 8 M urea and 0.1% RapiGest, (#186001861, Waters) in 0.1 M ammonium bicarbonate buffer. The cell suspension was thoroughly vortexed and incubated at room temperature for 10 min while shaking at 1000 rpm. Subsequently, Mtb cells were subjected to three 10-min cycles of sonication at 4°C (100% output, 50% intervals, Branson Sonifier 450, Emerson) while BCG cells were disrupted by three 10-min cycles of bead beating at 4°C using glass beads with a diameter of 0.5 mm (SIGMA #G8772). After each cycle, lysates were centrifuged for 10 min at 16,000 g and fresh lysis buffer was added. Protein concentration was determined using a BCA assay according to manufacturer's protocol (#23227, Thermo Fisher Scientific). Protein disulfide bonds were reduced by adding 5 mM tris(2-carboxyethyl)phosphine (TCEP) and incubating for 30 min at 37°C. Next, the free cysteine residues were alkylated by adding 10 mM iodoacetamide and incubating for 30 min in the dark at room temperature. Excessive iodoacteamide was captured by addition of 12.5 M N-acetyl cysteine and incubation for 10 min at room temperature. Extracted protein samples were diluted at a ratio of 1:5 with 0.05 M ammonium bicarbonate buffer to reach a urea concentration of <2 M. Sequencing-grade modified trypsin (#608-274-4330, Promega) was added at a ratio of 1:100 enzyme:substrate (weight/weight) and incubated for over night at 37°C with gentle shaking at 300 rpm. To stop the tryptic digest and to precipitate RapiGest the pH was lowered to 2 using 50% trifluoro acetic acid (TFA) followed by an incubation for 30 min at 37°C with shaking at 500 rpm. Water-immiscible degradation products of RapiGest were pelleted by centrifugation at 16,000 g for 10 min. The cleared peptide solution was desalted with C18 reversed-phase columns (Sep-Pak Vac C18, #WAT020805 and #WAT036820, Waters). Prior to use, the C18 columns were activated with 100% ACN, followed by equilibration with 2% ACN/0.1% TFA. After loading the sample, the columns were washed four times with 2% ACN/0.1% TFA. Finally, peptides were eluted with 50% ACN/0.1% TFA, dried under vacuum, and re-solubilised in 2% ACN/0.1% FA to a final concentration of 0.5–1.0 mg/ml.

Peptide fractionation by off-gel electrophoresis

Off-gel isoelectric focusing was used to separate peptides of the trypsin-digested mycobacterial lysates into 24 fractions according to their pI (Malmström et al., 2006; Picotti et al., 2009). 200 µg peptides from three growth phases of Mtb were pooled and solubilised in OGE buffer, which contained 5.6 M urea, 1.6 M thiourea, 5% v/v glycerol, 1% w/v dithiothreitol (DTT), and 1% v/v carrier ampholytes mixture (IPG buffer pH 3.0-10.0, #17-6000-87, GE Healthcare). The peptides were separated on a 3100 OFFGEL Fractionator (Agilent Technologies) using a immobilised pH gradient (IPG) strip of 24-cm length and pH 3–

10 (#17-6002-44, GE Healthcare) at a maximum of 8000 V, 50 uA, and 200 mW until 50 kVhrs were reached. After recovering the 24 fractions, they were desalted on C18 reversed-phase MicroSpin columns (#SEM-SS18V, The Nest Group Inc.) as described previously. Peptides were eluted with 50% ACN/0.1% TFA, dried under vacuum, and re-solubilised in 2% ACN/0.1% FA to a final concentration of 0.5 mg/ml.

Retention time calibration and conversion into iRTs

Eleven retention time calibration peptides (RT-kit WR, Biognosys) spanning a broad retention time range were added to all samples analysed by MS. These peptides allow the determination of system-independent retention times (iRT) for each peptide relative to these calibration peptides as recently described by Escher and colleagues (2012). For initial measurements on the synthetic peptides, a homemade mix of eight heavy-labelled synthetic peptides was used instead because the commercial peptides were not yet available.

Discovery MS data acquisition and analysis

One µg of each peptide sample was analysed on a nano-LC system (Eksigent Technologies) connected to an LTQ Orbitrap XL mass spectrometer equipped with a nanoelectrospray ion source (Thermo Fisher Scientific). Peptides were separated on a fused silica microcapillary column (10 cm x 75 µm, #PF360-75-10-N-5, New Objective) packed in-house with C18 resin (Magic C18 AQ 3 µm diameter, 200 Å pore size, Michrom BioResources) with a linear gradient from 95% solvent A (2% ACN/0.1% FA) and 5% solvent B (98% ACN/0.1% FA) to 35% solvent B over 60 min at a flow rate of 300 nl/min. The data acquisition mode was set to obtain one MS1 scan in the orbitrap at a resolution of 60,000 full width at half maximum (at 400 m/z) followed by collision induced dissociation of the five most abundant precursor ions with a dynamic exclusion for 30 s. MS2 spectra were acquired in the linear ion trap.
Thermo raw files were converted into mzXML format using ReAdW (version 4.0.2) (Pedrioli et al., 2004). The acquired MS2 spectra from the OGE samples were searched against an Mtb H37Rv protein database (TubercuList v2.3, April 2011) using SEQUEST (Sorcerer-SEQUEST, version 4.0.4) (Eng et al., 1994). Reversed sequences of all proteins were appended to the database to assess the number of false positive peptide identifications (Elias and Gygi, 2007). Search parameters were as follows: fully and semi-tryptic peptides (proteolytic cleavage after Lys and Arg unless followed by Pro) with zero or one missed cleavage were allowed, mass tolerance of the precursor ions was set to 20 ppm. Carbamidomethylation at cysteines was set as a fixed modification and oxidation at methionines as a variable modification. The output of the search engine was processed using PeptideProphet and ProteinProphet as part of the TPP (Deutsch et al., 2010; Keller et al., 2002; Nesvizhskii et al., 2003). Only peptides at a false-discovery rate of less than 1% were taken into consideration for further analysis.

Proteome coverage prediction

The PeptideProphet output from above, containing all peptide identifications of the 24 OGE samples, was processed with the software MAYU (v1.06) (Reiter et al., 2011). The command used was: Mayu.pl –A interact.pep.xml -C TubercuList_v2-3.fasta -E DECOY_ -P mFDR=0.01:td -H 100 –runR. Peptide-spectrum matches were selected at an FDR of 0.16% to obtain a protein FDR of 1%. Proteome coverage prediction was performed as described by Claassen et al. (2011). Briefly, this method models shotgun proteomics experiments by means of a recursive variant of the hierarchical Pitman-Yor process. The model extrapolates proteome coverage for further repetition of earlier experiments while capturing the redundancy across overlapping peptide sets in integrated datasets (Claassen et al., 2011).

Selection and preparation of synthetic peptides

For each protein several proteotypic peptides were selected using a recently developed algorithm (Campbell et al., unpublished data). Only fully tryptic peptides without missed cleavages and a length between 6 and 21 amino acids were allowed. Furthermore, peptides had to be unique to a particular protein and were not allowed to map to any human protein. Highest priority was given to peptides with the most previous observations in the PeptideAtlas database. If no or insufficient peptides have been previously observed, the most MS-suited peptides for a protein were predicted. Finally, 17,463 synthetic peptides were purchased in unpurified form in 96-well format (SpikeTides, JPT Peptide Technologies); all peptides are listed in Table S1. The lyophilised peptides were solubilized in 20% ACN/1% FA, vortexed for

10 min and subjected to water bath sonication for an additional 10 min. The peptides were pooled in mixes of maximally 96. After drying under vacuum and re-solubilising in 2% ACN/1% FA, the pools were desalted using reversed-phase C18 columns as described above. Finally, the peptides were re-suspended in 2% ACN/0.1% FA to reach a concentration of 1 pmol/peptide/µl.

Generation of the Mtb Proteome Library based on synthetic peptides

Synthetic peptide mixes were analyzed on three different mass spectrometer types: a Qtrap, an Orbitrap, and a TripleTOF. The Qtrap mass spectrometer was equipped with a nano-electrospray ion source (4000 QTRAP, ABSciex). Chromatographic separation of peptides was performed on a nanoLC ultra 1Dplus system (Eksigent) coupled to a 15-cm fused silica microcapillary column (75 µm inner diameter) packed in-house with C18 resin (Magic C18 AQ 5 µm diameter, 200 Å pore size, Michrom BioResources). Peptides were loaded onto the column from a cooled 10°C nanoLC-AS2 autosampler (Eksigent) and separated by a linear gradient from 98% solvent A (2% ACN/0.1% FA) and 2% solvent B (98% ACN/0.1% FA) to 35% solvent B over 35 min at a flow rate of 300 nl/min. The mass spectrometer was run in SRM mode in unit resolution of both Q1 and Q3 corresponding to a full width half maximum of 0.7. A full MS2 fragment ion spectrum was triggered when a threshold of 1000 ion counts per s was reached for the respective transition ion trace. The dwell time was 10 ms for each of the ~200 transitions per method resulting in a cycle time of ~2 s. The fragment ion of the y-ion series with an m/z value above the m/z value of the 2+ and 3+ precursor plus 15 Thomson, were used as triggering transitions in the first attempt to produce an MS2 spectrum of the peptide. MS2 spectra were acquired with positive polarity in enhanced product ion mode with collision-induced fragmentation in q2, low Q1 resolution, scan speed of 4000 Da/s, and an m/z scan range of 275 to 1450 Da. The collision energies were calculated according to the following equations: CE = 0.044 * (m/z) + 5.5 and CE = 0.051 * (m/z) + 0.5 for the 2+ and 3+ precursor charge ions, respectively. Raw data files (wiff) were converted into mzXML format using msConvert from ProteoWizard (version 1.6.1455) (Kessner et al., 2008).

For the measurements of the synthetic peptides on the Orbitrap, the peptides were pooled in mixes of ~500. The setup and parameters for the LTQ Orbitrap XL are described above. The data acquisition mode was set to obtain one MS1 scan in the orbitrap at a resolution of 60,000 full width at half maximum followed by collision induced dissociation of the five most abundant precursor ions with a dynamic exclusion for 30 s. MS2 spectra were acquired in the linear ion trap. Thermo raw files were converted into centroided mzXML format using ReAdW (version 4.0.2) (Pedrioli et al., 2004).

For the measurements of the synthetic peptides on the TripleTOF, the peptides were pooled in mixes of ~1000. The TripleTOF 5600 mass spectrometer (ABSciex) was coupled to a nanoLC 1Dplus system (Eksigent) and the chromatographic separation of the peptides was performed on a 20-cm emitter (75 µm inner diameter, #PF360-75-10-N-5, New Objective) packed in-house with C18 resin (Magic C18 AQ 3 µm diameter, 200 Å pore size, Michrom BioResources) as described above. A linear gradient from 2 – 35% solvent B (98% ACN/0.1% FA) was run over 90 min at a flow rate of 300 nl/min. The mass spectrometer was operated in IDA mode with a 500 ms survey scan from which up to 20 ions exceeding 250 counts per second were isolated with a quadrupole resolution of 0.7 Da, using an exclusion window of 20 s. Rolling collision energy was used for fragmentation and an MS2 spectrum was recorded after an accumulation time of 150 ms. Raw data files (wiff) were centroided and converted into mzML format using the ABSciex converter (beta version 2011) and subsequently converted into mzXML using openMS v1.8.

Qtrap MS2 spectra were analysed internally prior to submission to the PeptideAtlas database as follows: Qtrap MS2 spectra were assigned to peptide sequences using the SEQUEST algorithm and PeptideProphet as described above, but with an MS1 precursor mass tolerance of 1.2 Da and only fully tryptic peptides without missed cleavages allowed. All spectra identified with an FDR <1% were considered as true hits. For peptides that did not produce good MS2 spectra in the first attempt, additional experiments were performed where the y4 to y9 ions for both the 2+ and 3+ charged precursor ions were used as the triggering transitions. The software Skyline (MacLean et al., 2010) was used to generate a spectral library from the Qtrap mzXML files. To generate the SRM assays, for each peptide precursor the most intense y-ion fragment ions were extracted together with their relative intensities and charge states.

Retention times were extracted and converted into system-independent iRT values (see above) using an in-house written script. First, the lower median (sort numbers according to increasing values, for even n number of values, the value at position n/2 – 1 is taken, for odd n numbers the median value is taken) peptide retention time was calculated for each run separately and converted into iRT units using a run-specific linear correlation equation determined from the spiked-in iRT peptides. The iRT peptide LGGNEQVTR was excluded for the calibration. In the rare case an insufficient number of iRT peptides could be extracted for a specific run, the calibration of a neighbouring run was used instead. Because iRT peptides were not present in the first batch of synthetic peptide measurements on the Qtrap and the Orbitrap instruments, alternative peptides were selected for the iRT calibration. The peptide iRT values were subsequently averaged over the runs using the lower median. In a next step, the iRT values obtained for the three instrument types were compared and outliers removed. An iRT value was considered as an outlier if it differed from the other two iRT values by more than 10 units (~5% of the gradient), while the other two iRT values were less than 10 units apart. The average iRT value over the three instruments and its standard deviation are listed in Table S1. If no standard deviation is given, the peptide was identified by only one instrument platform and should be considered as less reliable. Also iRT values with a large standard deviation (>7) are less reliable and should be validated before use in scheduled SRM experiments. The peptide retention time, precursor mass and charge state, and fragment ion mass, charge state, and relative intensities together constitute an SRM assay.
For the version of the Mtb Proteome Library which is in the PeptideAtlas and SRMAtlas database, the data acquired from the synthetic peptides on the three different instrument platforms were analysed as follows: MS2 spectra generated from synthetic peptides were searched using X!Tandem (Craig and Beavis, 2003; Eng et al., 1994) with the k-score plug-in (MacLean et al., 2006) against a concatenated synthetic peptides plus decoy database. The parameters were parent monoisotopic mass error of 1.4 Da, fixed carbamidomethyl modification on cysteine, and variable methionine oxidation with refinement. The search results were processed using the PeptideProphet, iProphet, and ProteinProphet as part of the TPP (Keller et al., 2002; Nesvizhskii et al., 2003; Shteynberg et al., 2011). The Mtb Proteome Library was constructed using the iProphet results. iProphet probabilities 0.876, 0.836 and 0.817 were used for LTQ Orbitrap XL, 4000 QTRAP and TripleTOF respectively to reach <1% peptide FDR. The same probabilities were used to generate SpectrastST raw libraries for the three instruments. The raw libraries were converted into consensus libraries which were filtered against the synthetic peptide list to include only synthetic peptides in the libraries. The top 16 potential SRM assays then are generated for each peptide ion for each of the source instruments using SpectraST in SRM mode. The iRT values were determined as described above. Separately, all potential tryptic peptides from the target proteome are scored and annotated using the PABST algorithm (Campbell et al., unpublished data). The peptides and SRM assays are combined and loaded into a relational database, with theoretical fragments generated for any peptides lacking empirical data. These stored assays can be queried via a web page (www.SRMAtlas.org), wherein users can refine the suggested SRM assays using various peptide and fragment ion criteria. Results can be exported in a generic TSV format, or in a number of vendor-specific method file formats.

Validation of the Mtb Proteome Library by SRM

If available, the six highest fragment peaks belonging to the y-ion series and their corresponding intensities were extracted for 2+ and 3+ peptide precursors in the Mtb SRMAtlas. The SRM assay validation was based on y-ion transitions because y-ions are in general the most intense and consistently detected fragment ions (Holstein et al., 2011). Furthermore, it has been shown that the y-ion intensity distributions from different MS types correlate well and thus the SRM assays in the Mtb Proteome Library should be robust and transferable to comparable instrument platforms (Sherwood et al., 2009). The transition groups were measured in unfractionated Mtb H37Rv lysates (1:1 mixture of exponential and stationary phase cultures) in scheduled (± 2 min) SRM acquisition mode on a TSQ Vantage triple quadrupole mass spectrometer equipped with a nanoelectrospray ion source (Thermo Fisher Scientific). Peptides were separated online on a fused silica microcapillary column (10.5 cm x 75 µm) packed in-house with C18 resin (Magic C18 AQ 5 µm diameter, 200 Å pore size, Michrom BioResources) with a linear gradient from 98% solvent A (2% ACN/0.1% FA) and 2% solvent B (98% ACN/0.1% FA) to 35% solvent B over 35 min at a flow rate of 300 nl/min. The mass spectrometer was operated in positive mode using electrospray ionisation

with a voltage of 1,200 V. The capillary temperature was set to 280°C and the collision gas pressure to 1.5 mTorr. All transitions were monitored in scheduled mode with a retention time window of 240 s and a mass window of 0.7 half-maximum peak width (unit resolution) in Q1 and Q3. Per run ~450 transitions were acquired with a cycle time of 2 s and a dwell time of at least 20 ms. Collision energies were calculated as follows: CE = 0.034 * (m/z) – 0.848 and CE = 0.022 * (m/z) + 5.953 for 2+ and 3+ charged precursor ions, respectively. For each batch of measurements, 100 to 150 decoy transition groups were generated by subtracting or adding a random integer to precursor and fragment ion masses of the target list as described by Reiter and colleagues to allow statistical validation of the identified peak groups (Reiter et al., 2011). The SRM raw files were converted into mzXML format using ReAdW version 4.3.1 (Pedrioli et al., 2004) and peaks were detected and assigned a statistical confidence score using the mQuest/mProphet software (v2.0.2 Reiter et al., 2011). In a first screening phase all SRM assays generated from the synthetic peptides were acquired in 425 SRM runs and peak groups below an FDR of 20% were considered as potentially true. These candidates together with new decoy transition groups were re-acquired in 228 SRM runs. The resulting dataset was filtered at an FDR of 1% as determined by mProphet.

Prediction of transition specificity

Unique ion signatures were calculated as described by Röst and colleagues (Röst et al., 2012). *Homo sapiens* protein sequences were downloaded from http://www.ensembl.org, release 56_37a. Theoretical precursor ions were generated using trypsin for proteolysis, CAM as fixed modification, and charge states 2+ and 3+ for parent ions and considered up to three ^{13}C isotopic peaks (+0, +1, +2, +3 amu). For each of those precursor ions the complete set of y and b fragment ions was generated. The Mtb dataset gave rise to 74,921 tryptic peptides, 599,872 precursors, and 17,447,648 transitions. The human dataset gave rise to 677,150 tryptic peptides, 5,417,200 precursors, and 117,151,632 transitions. For each precursor in the Mtb Proteome Library the best n ions were selected, where n is in the range of 1 to 6. We then determined all transitions of the background proteome that were within a Q1 and Q3 tolerance of ± 0.35 m/z, as well as within a certain retention time tolerance for scheduled measurements (± 5 SSRCalc arbitrary units corresponding to approximately ± 2.5 min on a 30-min gradient) (Krokhin, 2006). The minimal number of transitions necessary to uniquely identify a precursor species was defined as the minimal n for which no other precursor existed in the background whose ions contained all n query ions.

Proteome-wide absolute label-free quantification

For absolute quantification 34 anchor proteins were selected covering a wide abundance range of the Mtb proteome. For each anchor protein two synthetic heavy isotope-labelled reference peptides in defined concentrations determined by amino acid analysis (AQUA QuantPro, Thermo Fisher Scientific) were added to the digested and purified lysate that was used for the SRM validation experiments. The concentrations of the synthetic peptides were roughly adjusted to the endogenous peptide abundance levels before. Data was analysed with the software Skyline (MacLean et al., 2010). Integrated peak areas of the reference and endogenous peptides were summed and from the obtained ratios the endogenous peptide concentration was determined in fmol/µg. MS intensities of the label-free proteome-wide SRM data were normalised using three positive control peptides that were acquired in each of the 228 runs. The optimal model to combine SRM intensities of best flying peptides and most intense transitions to a single MS signal was determined by Monte Carlo cross-validation (Ludwig et al., 2012). All anchor proteins were represented in the Mtb Proteome Library by at least two peptides. The resulting linear model with the highest quantification accuracy was used to estimate proteome-wide concentrations from the SRM signal intensities. Proteins quantified with less than two peptides were excluded from further analysis.

SRM analysis of the DosR regulon

For each protein of the DosR regulon SRM assays were selected from the Mtb Proteome Library. If possible, three peptides with five y-ion transitions were chosen. Heavy isotope-labelled, crude synthetic peptides were ordered in 96-well format (JPT Peptide Technologies). Peptide solubilisation and purification was done as described above. Concentrations of the peptides were adjusted to match the endogenous peptide concentration levels. For three proteins no heavy isotope-labelled reference peptide was available (Rv0573, Rv1735, Rv1998c). SRM analysis was done as described above. Non-detectable peptides and

interfered transitions were manually removed using Skyline and the resulting transition intensities were subjected to statistical analysis using linear mixed models with the SRMstats R package (Chang et al., 2012; MacLean et al., 2010). Hierarchical clustering (Euclidian distance, centroid linkage) and heat map visualisation were performed with the Cluster 3.0 (by Michael Eisen and Michiel de Hoon) and Java TreeView software (by Alok Saldanha).

Proteogenomic analysis

The genomic sequence of Mtb was translated in all six reading frames using the *Bacterial, Archaeal and Plant Plastid Code* provided by NCBI (http://www.ncbi.nlm.nih.gov). An in-house written python script was used to generate a database in FASTA format including all translated sequences from stop codon to stop codon and, in addition, for each potential start codon a peptide starting with the initiator methionine and ending at the next downstream tryptic site. Sequences of common contaminants were added and reversed sequences of all entries were appended to the database. The spectra from the above described 24 OGE fractions were searched against this six-frame translated genome database using SEQUEST and X!Tandem (Craig and Beavis, 2003; Eng et al., 1994). The search parameters were set as follows: fully tryptic cleavage, no missed cleavages, precursor mass tolerance 20 ppm and fragment mass tolerance 0.5 Da, carbamidomethylation of cysteines as a fixed modification and oxidation of methionines as variable modification. X!Tandem was run once in native mode, once with the k-score plugin (MacLean et al., 2006). In a second search, semi-tryptic peptides were allowed to find N-terminal peptides with cleaved initiator methionine. For statistical validation, PeptideProphet and iProphet were run as part of the TPP (Deutsch et al., 2010; Keller et al., 2002; Shteynberg et al., 2011) and only peptides with an FDR <1% were considered for further analysis. iProphet was used to combine search results of the three different search algorithms. After removal of peptides that map to annotated open reading frames, the Proteogenomic Mapping Tool (Sanders et al., 2011) was used to map the remaining peptides back to their genome location (NCBI RefSeq NC_000962.2 downloaded in November 2011). A new protein was only considered if there were at least two peptides mapping to the same region of the genome. The novel peptide sequences were blasted against all bacterial proteins using the UniProt BLAST tool to check if a protein containing them had already been predicted in another Mtb strain.

Transferability of the Mtb Proteome Library to other mycobacterial strains

Proteome fasta files containing NCBI RefSeq sequences were downloaded from the PATRIC database (March 2012, http://patricbrc.vbi.vt.edu/): Mtb H37Rv: NC_000962.2; Mtb CDC1551: NC_002755.2; *M. africanum*: NC_015758.1; *M. bovis*: NC_002945.3; *M. bovis BCG*: NC_008769.1; *M. leprae*: NC_002677.1; *M. marinum*: NC_010612.1; *M. smegmatis*: NC_018289.1; *M. ulcerans*: NC_008611.1.

Supplemental References

Bell, C., Smith, G.T., Sweredoski, M.J., and Hess, S. (2012). Characterization of the Mycobacterium tuberculosis proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. J Proteome Res *11*, 119–130.

Craig, R., and Beavis, R.C. (2003). A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid Commun Mass Spectrom *17*, 2310–2316.

Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., et al. (2010). A guided tour of the Trans-Proteomic Pipeline. Proteomics *10*, 1150–1159.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207–214.

Eng, J.K., McCormack, A.L., and Yates, J.R., III (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. *5*, 976–989.

Holstein, C.A., Gafken, P.R., and Martin, D.B. (2011). Collision energy optimization of b- and y-ions for multiple reaction monitoring mass spectrometry. J Proteome Res *10*, 231–240.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74*, 5383–5392.

Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics *24*, 2534–2536.

Krokhin, O.V. (2006). Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. Anal Chem *78*, 7785–7795.

MacLean, B., Eng, J.K., Beavis, R.C., and McIntosh, M. (2006). General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. Bioinformatics *22*, 2830–2832.

Malmström, J.A., Lee, H., Nesvizhskii, A.I., Shteynberg, D., Mohanty, S., Brunner, E., Ye, M., Weber, G., Eckerskorn, C., and Aebersold, R. (2006). Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. J Proteome Res *5*, 2241–2249.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Anal Chem *75*, 4646–4658.

Pedrioli, P.G.A., Eng, J.K., Hubley, R., Vogelzang, M., Deutsch, E.W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R.H., Apweiler, R., et al. (2004). A common open representation of mass spectrometry data and its application to proteomics research. Nature Biotechnology *22*, 1459–1466.

Sherwood, C.A., Eastham, A., Lee, L.W., Risler, J., Vitek, O., and Martin, D.B. (2009). Correlation between y-type ions observed in ion trap and triple quadrupole mass spectrometers. J Proteome Res *8*, 4243–4251.

Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics *10*, M111.007690.