Published in final edited form as:

J Am Soc Mass Spectrom. 2016 November ; 27(11): 1745–1751. doi:10.1007/s13361-016-1465-2.

Transitioning from Targeted to Comprehensive Mass Spectrometry using Genetic Algorithms

Jacob D. Jaffe^{1,*}, Caitlin M. Feeney^{1,†}, Jinal Patel¹, Xiaodong Lu¹, and D. R. Mani¹

¹The Broad Institute of Harvard and MIT, Cambridge, MA 02142 USA

Abstract

Targeted proteomic assays are becoming increasingly popular due to their robust quantitative applications enabled by internal standardization, and they can be routinely executed on high performance mass spectrometry instrumentation. However, these assays are typically limited to 100s of analytes per experiment. Considerable time and effort are often expended in obtaining and preparing samples prior to targeted analyses. It would be highly desirable to detect and quantify 1000s of analytes in such samples using comprehensive mass spectrometry techniques (e.g., SWATH and DIA) while retaining a high degree of quantitative rigor for analytes with matched internal standards. Experimentally, it is facile to port a targeted assay to a comprehensive data acquisition technique. However, data analysis challenges arise from this strategy concerning agreement of results from the targeted and comprehensive approaches. Here, we present the use of genetic algorithms to overcome these challenges in order to configure hybrid targeted/comprehensive MS assays. The genetic algorithms are used to select precursor-to-fragment transitions that maximize the agreement in quantification between the targeted and the comprehensive methods. We find that the algorithm we used provided across-the-board improvement in the quantitative agreement between the targeted assay data and the hybrid comprehensive/targeted assay that we developed, as measured by parameters of linear models fitted to the results. We also found that the algorithm could perform at least as well as an independently-trained mass spectrometrists in accomplishing this task. We hope that this approach will be a useful tool in the development of quantitative approaches for comprehensive proteomics techniques.

Graphical Abstract

*CORRESPONDENCE & REPRINT REQUESTS: Jacob D. Jaffe, The Broad Institute of Harvard and MIT, 415 Main St., Cambridge, MA 02142 USA, +1-617-714-7638, jjaffe@broadinstitute.org.

[†]Present address: Waters Corporation, Milford, MA 01757 USA

Introduction

Targeted proteomic assays are becoming increasingly popular due to their robust quantitative applications enabled by introduction of stable isotope-labeled (SIL) internal standards, reproducibility of analyte detection, and ‘designability’ to answer specific questions or monitor biological processes [1–4]. It is becoming common to execute these assays on high performance mass spectrometry instrumentation (high resolution, accurate mass) in addition to the traditional triple quadrupole instrumentation. When executed on high performance MS instruments, targeted approaches record the full MS/MS spectrum of all fragment ions generated from a given dissociation technique. The technique goes by various monikers such as PRM (parallel reaction monitoring), MRM-HR (multiple reaction monitoring, high resolution), HR-MRM, etc. Use of full scan MS/MS spectra in targeted assays allows for selection of many interference-free transitions for quantification and unambiguous identification of analytes, including localization of sites of post-translational modifications on peptides [5]. While perhaps not as sensitive as triple quadrupole instrumentation, high resolution instrumentation allows for more selectivity and less potential for interference.

Targeted methods have been used extensively for biomarker verification studies [6–10] and are emerging for acquisition of assay panels with specific analytical purposes. For example, the emergence of “sentinel assays” to study the changes in specific biological processes or activation of signaling pathways demonstrates an increasing role for targeted proteomics in biology [4, 11, 12]. In almost all cases, great care is placed in the selection of samples and their associated biochemical preparation prior to analysis. In some applications, specific biochemical enrichments are performed to isolate special populations of analytes, such as phosphopeptides. Yet targeted analyses are typically limited to detection of only ~100s of analytes in a single assay of reasonable laboratory time scale (~1 – 2 hours). The practical

limits are typically set by hardware duty cycle and scheduling requirements. Deeper interrogation of carefully prepared samples would be highly desirable.

Several methods exist to overcome these limits. The simplest method is to re-analyze a sample in multiple targeted assays, at the cost of increased time and subject to exhaustion of the sample. A new method has been proposed by Domon and colleagues that relies on real-time triggering of targeted assays based on detection of internal standard signals, but this method requires that all possible targets are determined prior to analysis and have readily available internal standard peptides and may be limited to certain hardware platforms [13].

Comprehensive mass spectrometry techniques such as DIA and SWATH [14, 15] hold great promise for deep interrogation of samples with reproducible detection of analytes across large sample sets. However, most quantitative applications of these comprehensive techniques have been limited to label-free modes with attendant difficulties in longitudinal quantitative sample comparison over months or years. We sought to take advantage of this mode of data acquisition but with robust quantification for a panel of analytes for which we had previously configured a PRM assay and had obtained SIL internal standard peptides [11]. This assay focuses on phosphosignaling and contains a phosphopeptide enrichment step prior to analysis. In a perfect world, we would run each sample twice. Once in PRM-mode (to get the best possible quantification on the selected panel) and once in DIA-mode (to identify and quantify additional analytes not found in the original panel design). However, this comes at a 2x cost in instrument time, and for high throughput labs that analyze large numbers of samples, this may not be feasible. So we searched for a compromise method where we could experience the benefits of SIL quantification for the selected panel of analytes plus the ability to mine for additional analytes after data analysis while keeping acquisition time the same. A hybrid targeted/comprehensive MS assay would allow us to quantify our assay panel while preserving the ability to reinterrogate for new analytes of high value to signaling studies after data acquisition.

As a first step, we wanted to ensure that the ratio-metric quantification of our analyte panel agreed between the targeted and comprehensive approaches. While in practice it was very easy to acquire data using the comprehensive “DIA” technique (with the added benefit of obviating any scheduling requirements), robust quantification of our analyte panel proved to be more difficult. More specifically, the quantitative ratios derived by comparing the signals from endogenous analytes to SIL standards in the DIA assay did not agree with the gold-standard fully targeted and scheduled PRM assay when attempting to use identical transitions for quantification. Several obvious factors contributed to this disagreement. First, in many cases both the endogenous and SIL standard are co-isolated in the same precursor window. Therefore, any fragments selected that did not carry a SIL amino acid in the internal standard were effectively unfit for quantification. Second, the wider precursor isolation windows attendant to the comprehensive technique introduce more possibilities for interference from co-eluting analytes.

We sought to develop an automated way to pick transitions for quantification from comprehensive MS data that would maximize quantitative agreement with the targeted implementation of our assay. We turned to genetic algorithms as a means of sampling a large

number of combinations of transitions without an exhaustive search. Genetic algorithms have been used since the 1970s in a wide variety of optimization problems. Here, we demonstrate their application to the problem of developing a quantitative framework for a hybrid targeted/comprehensive MS assay.

Methods

Cell Culture and Samples

HME-1 and MDA-MB-231 cells were obtained from ATCC (Manassas, VA). The compounds PD0325901, WYE125132, Torin, GDC0941, and MK2206 were the generous gift of Dr. Peter Sorger (Harvard Medical School). All other compounds were obtained from Sigma (St. Louis, MO). Cells were grown according to ATCC recommendations (MEBM medium + MEGM supplement kit (Lonza Biologics, Hopkinton, MA) for HME-1 and DMEM + 10% fetal bovine serum for MDA-MB-231) at 37 C/5% CO₂. Cells were grown to >90% confluence and then the media was supplemented with compounds. Cells were treated for 3 hours under each of the following conditions: PD0325901 @ 6, 30, and 150 nM, WYE125132 @ 20, 100, and 100 nM, Torin @ 10, 50 and 250 nM, GDC0941 @ 1, 5, and 25 μ M, MK2206 @ 2, 10, and 50 μ M, and a control condition consisting of 0.1% DMSO (equivalent to the final concentration of DMSO for the delivered compounds). Treatments were performed in biological triplicate. These samples were generated for another study, and not every sample was analyzed by DIA (see below).

Samples were processed exactly as in [11]. Briefly: cells were lysed, protein yield was measured, and 500 μ g of protein from each sample was subjected to proteolytic digestion with trypsin, peptide desalting, phosphopeptide enrichment, and a final desalting of the enriched phosphopeptides prior to introduction of SIL peptide standards in the final resuspension buffer (3% acetonitrile/5% formic acid) prior to MS analysis. In the end, 42 unique biological samples were analyzed by both the PRM and DIA acquisition methods for purposes of this study (a total of 84 LCMS runs, see below).

Parallel Reaction Monitoring (PRM) MS Data Acquisition

20% of each sample was subjected to targeted PRM LCMS analysis for the 96 core phosphopeptide analytes in our assay as described in [11] using a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The only key difference from the methods published in [11] is that Full MS resolution was set to 60,000 and MS/MS resolution was set to 15,000 due to the different operating resolutions of the Q-Exactive HF instrument as compared to the instrument used in [11].

Data Independent Acquisition (DIA) MS Data Acquisition

30% of each sample was subjected DIA LCMS analysis using the same Q-Exactive HF mass spectrometer as follows. LC conditions were identical to those stated in [11], and DIA analyses were performed “back-to-back” with their associated PRM analyses. Spray voltage was set to +2.0kV. The MS acquisition cycle was 1 full MS scan followed by 27 MS/MS scans using the “DIA” loop count method functionality (set to 27) of the XCalibur instrument control and acquisition software. Full scan parameters: range 300–1200 m/z ,

resolution 60,000, AGC target 3e6, maximum IT 20 ms, and data were acquired in profile mode. DIA MS/MS scan parameters: resolution 30,000, AGC target 1e6, maximum IT 40 ms, default charge state $z=4$, isolation width 22.0 m/z , normalized collision energy = 26, and data were acquired in centroid mode. The inclusion list consisted of 56 entries where the first entry was ~ 411.4 m/z , and the successive 27 entries added 22 m/z units to this value, then the 29th entry was 400.4, and the successive 27 entries (entries 30–56) added 22 m/z units to this value. Thus, the range from 400–1000 m/z is traversed twice per inclusion list cycle with a 50% offset in windows from the first instance to the second. Four peptides could not be detected using the DIA technique and thus subsequent analyses were limited to 92 analytes.

Analysis of PRM and DIA Data

PRM data were imported into a Skyline [16] template identical to that provided in [11] and manual peak boundary refinement was performed. The Light-over-Heavy ratios for the 92 remaining analytes for each of the 42 samples were exported from Skyline. For DIA data, we generated a Skyline document where all possible fragment ions (b, y, and neutral losses of phosphoric acid (–98 Da) therefrom where chemically possible) were allowed to populate the document (this resulted in a very large Skyline document). Raw data were then imported and peak boundaries were refined where necessary (again, the Skyline data file was very large). Spectral libraries (containing both observed transitions and retention times) were used in this study to assist with peak group assignment, but played no role in the process of selecting suitable transitions to create the best agreement between PRM-mode and DIA-mode peak area ratios. We subsequently exported the Light and Heavy intensities of each transition for each analyte in each sample. These data are available in the Supplemental Sourcecode and Data Examples.

Genetic Algorithm Implementation

The genetic algorithm for transition selection was implemented in R using the GA package [17]. The input to the algorithm was 1) a table of ratios for each of the 92 peptides in each of the 42 samples as determined by PRM from Skyline, and 2) a table of the fragment intensities for each of the 92 peptides in each of the 42 samples as determined by DIA from Skyline. Parameters for running the algorithm were as follows: maxiter=500, run=50, popSize=500, pcrossover=0.9, pmutation=0.5, elitism=0.1.

Each peptide in the data set was considered separately. One third of the data (14 samples) were held out (both PRM and DIA) at random as a test set. Trivial fragment ions (intensity of 0 in all samples) were discarded. After every round, ratios were computed from the selected subsets of DIA fragments in the same manner that Skyline computes ratios from light and heavy intensities (intensity-weighted average of ratios) and the sum-squared-difference between the corresponding ratios of PRM samples was computed. The objective of the algorithm was to minimize this quantity. A set of transitions is considered a “solution.” If, after convergence, more than one solution was found with equivalent optimal fitness, the solution with the least number of fragments was selected. If two or more solutions had equivalent optimal fitness and equivalent number of fragments, one solution was selected at random. More details are given in the Theory section below, and source code

and input files are available as supplementary materials. While third parties are free to run the algorithm, the exact nature of the final solutions for each peptide may differ due to the random selection of holdout data and the random nature of the genetic algorithm itself.

Theory

The formal task at hand is, for a given analyte, to identify a set of transitions in comprehensive MS (DIA) data that will reproduce the same ratios from the targeted (PRM) assay as calculated by our typical analytical software (Skyline). The set of transitions for the PRM assay has been previously determined and is published elsewhere [11, and online at <https://panoramaweb.org/labkey/project/LINCS/AbelinSupplemental/begin.view?>]. The entire set of theoretical transitions and their associated intensities (as extracted by Skyline from DIA data) are available for picking, from $b_1 \dots b_{n-1}$ and $y_1 \dots y_{n-1}$ where n is the length of the peptide. Transitions representing the neutral loss of phosphoric acid are also available where theoretically possible, and multiple charge states may be considered for each transition.

Consider a single peptide analyte. For a set of samples where measurements exist for both PRM and DIA acquisition methods, we can define an optimization problem where we seek to minimize the sum of squared differences between the log-transformed ratios of endogenous analyte to SIL internal standard from each method,

$$\min \sum_{i=1}^m (\log r_{Di} - \log r_{Pi})^2 \quad \text{Eq. 1}$$

where m is the total number of samples, r_{Di} is the ratio computed using a selected set of transitions from DIA data in sample i , and r_{Pi} is the ratio using the established set of transitions from PRM data in sample i .

Genetic algorithms simulate many rounds of breeding *in silico* and borrow concepts from real genetics, using probabilities of cross-over, mutation, and selection (survival) as parameters (Fig. 1A). For purposes of leveraging this technique, we define the following equivalencies: each transition is a gene, a set of transitions is a chromosome, a mutation is a random addition or subtraction of a gene (transition), and a cross-over adds or removes 1 or more genes (transitions) from a chromosome in analogy to a real meiotic cross-over event (Fig. 1B). Fitness is calculated as the reciprocal of the quantity calculated in Eq. 1, and the genetic algorithm maximizes fitness by successive rounds of breeding. At the first round, chromosomes are initialized with random genes (transitions). After each round, random mutations and cross-overs among chromosomes can occur. A percentage of the combinations are then selected for continuation to the next round (survival). The algorithm converges if no increase in fitness can be obtained after a set number of iterations.

Results and Discussion

We first acquired data from our targeted PRM assay and comprehensive DIA assay on the same samples (back-to-back injections to minimize systematic differences) to assess the scope of issues with porting from a PRM to a DIA data acquisition strategy. Importantly, there was a wide dynamic range of ratios (endogenous light-to-heavy internal standard) for each analyte across the sample set, which was useful for subsequent optimization problem and line fitting.

We present examples of two peptides from this study to illustrate the challenges in Fig. 2. Ideally, log-log plots of ratios between the PRM and DIA versions of the assay would have a slope of 1 with an intercept of 0, indicating perfect agreement. Fig. 2A shows that, for the peptide ALG(pS)PTKQLLPCEMACNEK, there is generally good correlation and linearity between the PRM and DIA ratios obtained from the assay ($r^2=0.78$ for a simple linear fit) without further transition optimization. However, there is a significant dampening of the response factor (slope = 0.39) and a substantial offset of the intercept (intercept = -1.43), with the DIA ratio being over-reported by >2.5-fold. Fig. 2B illustrates a similar pattern for the peptide SP(pS)PAHLPPDDPKVAEK, with a dampening of response by almost a factor of 10x, an offset of > 2-fold, and poor correlation between the ratios ($r^2=0.55$). The aggregate statistics across 92 peptides in the assay mirrored these trends, with the median slope ~0.3 and median intercept ~-1.0. We concluded that further optimization of the transitions selected for quantification by the comprehensive DIA assay.

To perform the optimization, we used a genetic algorithm as described above in the Theory section, with extracted ion intensities from Skyline as the underlying data. Using a total of 42 matched PRM and DIA data files, we considered each peptide independently, and trained on a random set of 2/3 of the data (28 samples) while holding out 1/3 of the data (14 samples) for assessment. The training and test sets were randomly selected (and different) for each peptide so as to avoid any systematic biases introduced by the samples themselves. The training algorithm was allowed to iterate (each iteration being a new generation) up to 500 times, but was considered converged if the fitness value did not increase for 50 consecutive generations. For a small number of peptides, we also investigated whether repeatedly running the algorithm for each peptide (bootstrapping up to 10 times to avoid “jackpot” effects and local minima) was necessary, but found that the overall fitness of the different repeats was generally the same and that the models converged to extremely similar solutions (data not shown). Due to the added time necessary for extensive bootstrapping, we discontinued this after our initial evaluations. In the cases of ties, where more than one set of transitions generated identical fitness values, solutions with fewer transitions were selected. For further ties, where identical fitness values and number of transitions were part of the solution, the first one from the default output order of the R package was selected.

In Fig. 2C–D, we illustrate the improvements after transition selection by the genetic algorithm for the same peptides shown in Fig 2A–B. For ALG(pS)PTKQLLPCEMACNEK, the new parameters of the linear fit were slope = 0.97, intercept = -0.06, with $r^2=0.98$, extremely close to the ideal parameters (Fig. 2C). This model contained 33 transitions and correctly excluded all transitions where the heavy standard did not bear an isotopically-

labeled amino acid. For this particular peptide, our heavy standard contains a labeled lysine in the 7th position. The first b-series ion selected in the model is $b[-98]_7^+$ (-98 indicating the loss of phosphoric acid). We also highlight the training data (black symbols) and the test data (red symbols) separately for the peptide, although the line fit parameters were computed using both sets together. In general, parameters of lines fit to training data alone were closer to ideal than those fit to test data or combined training and test data. Thus, the hold-out of test data gives us more realistic characterization of the performance of the selection algorithm and minimizes the effects of overtraining.

We can also see improvements for SP(pS)PAHLPPDDPKVAEK (Fig. 2D). For this peptide, the final line fit parameters were slope = 0.91, intercept = -0.23, with $r^2=0.95$, with the model containing 20 transitions. It should be noted that in this particular case, the hold-out set happened to contain 3 extremely low ratios (< 1/64-fold) where noise is likely to affect the quantification in any case (see red symbols in lower left of Fig. 2D). This case also illustrates that the selection algorithm is imperfect. While this peptide does again contain an internal isotopic label in the heavy standard at lysine 12, several b-series ions were selected prior to b_{12} . However, the total intensity contributed by these transitions was less than 0.1% of the total intensity of all transitions. Cases like these could be automatically flagged and removed from solutions if desired.

Beyond these individual peptide illustrations, we characterized the aggregate performance across the set of 92 peptides (Fig. 3). We show that the median values and overall distributions of slopes and intercepts improve towards the ideal values. The median slope improved from 0.27 to 0.92 (Fig. 3A), while the median intercept improved from -1.0 to -0.2 (Fig. 3B). The general correlation measure of r^2 improved from 0.65 to 0.94 (Fig. 3C). The number of transitions in the 92 models ranged from 4–57, with the median number of transitions selected at 15. We believe that these improvements are significant given that they were achieved automatically without further manual intervention and refinement. The set of all models and their pre- and post-modeling fitting parameters are available in the accompanying Supplemental Data 1.

We went on to compare these results to a set of transitions that had been selected independently by a targeted proteomics data analyst from our laboratory. To do this, we extracted ion intensities from transitions identified by the analyst as being acceptable for quantification of DIA data for these peptide analytes in other data sets. The aggregate performance characteristics between the quantification transitions selected by the genetic algorithm (GA) and those selected independently by an analyst (IA) for each peptide were essentially indistinguishable: median slope of 0.92 (GA) vs 0.94 (IA), median intercept of -0.2 (GA) vs. -0.3 (IA), and median r^2 of 0.94 (GA) vs. 0.92 (IA). In one case no ion intensity was present for the transitions selected by the analyst for a certain peptide but the genetic algorithm was still able to make successful model with a slope of 0.87, intercept of -0.61, and r^2 of 0.99. We found it encouraging that algorithm could perform at least as well as an extremely experienced mass spectrometrists in a fraction of the time.

While these early results are promising, the implementation of algorithmic transition selection could be further improved. As mentioned above, we could implement an explicit

filter to remove ‘disallowed’ transitions where no isotopic labeled fragment ions are present for heavy standards, or remove transitions from consideration that were not observed in spectral libraries generated from purified synthetic peptides. We have also noticed (in separate studies) that the biological background can contribute strongly to the optimal solutions due to the varying nature of interferences. Therefore, the optimal set of transitions from one cell line or tissue source may differ from those in another, and retraining of the model may need to be performed for specific biological matrices. We could also consider using an interference-detection tool (such as AuDIT) to help winnow the pool of transitions prior to execution of the genetic algorithm to reduce the search space and improve performance [18]. And finally, some peptides only show marginal improvement after transition selection with the algorithm, and will require manual reconfiguration. These cases tend to occur where the loss of sensitivity by going to DIA from PRM affects the ability to quantify low level analytes.

The ability to rapidly port a PRM assay to DIA modality places a premium on the biological sample collection and preparation that precedes analysis. The raw data can be mined for new analytes without sacrificing the original goals of the targeted assay that is presumably focused on analytes of high value. It also raises the possibility of using sets of analytes with robust ratio-metric quantification (those with internal standards from the PRM assay) for calibration of novel analytes that can be detected by DIA approaches. Use of these internal standards has precedent in other DIA analysis pipelines, mainly for use as retention time calibrants [19, 20]. However, we could envision using the internally standardized analytes for deeper quantification of novel analytes using a ratio-of-ratios approach. For example, if an internally standardized counterpart with similar intensity, m/z , and retention time properties can be identified for new analytes extracted from DIA data, perhaps this information can be leveraged for more robust longitudinal quantification across large data sets by comparing the intensity of the novel analyte to that of the established and standardized analyte.

Conclusions

The implementation of a genetic algorithm framework has allowed us to significantly streamline the process of porting a targeted PRM assay to a comprehensive DIA assay. In every case attempted we were able to improve the agreement between PRM and DIA results from the same samples, and we found that the automated selection of transitions performs equally well to a hand-curated set selected by an experienced mass spectrometrist. The major result of this methodology is to bring the ratio-metric quantification from data acquired by DIA modalities into line with the more established PRM approach. The automated selection of transitions for this purpose ensures backward compatibility of results from one version of the assay to another. Undoubtedly, future improvements to our implementation of the algorithm can streamline the process further. Still, matrix specific models and careful expert refinement will likely be required to decrease the chances of interference that are rampant in data collected with comprehensive MS approaches. We view the automated selection approach as a good first step in the process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This manuscript was prepared to celebrate Dr. Michael MacCoss's Biemann Medal award in 2015. In addition to leading the way in the nascent field of Computational Proteomics, I am privileged to be able to call Mike a collaborator and friend over many years. This work was very much inspired by his "munging" of computational and experimental techniques and I believe is apropos of Mike's overall philosophy. Congrats Mike! – *JDJ*. This work was supported by grants to JDJ: NIH U01CA164186 and U54HG008097.

References

1. Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, Grant RP, Hoofnagle AN, Huttenhain R, Koomen JM, Liebler DC, Liu T, MacLean B, Mani DR, Mansfield E, Neubert H, Paulovich AG, Reiter L, Vitek O, Aebersold R, Anderson L, Bethem R, Blonder J, Boja E, Botelho J, Boyne M, Bradshaw RA, Burlingame AL, Chan D, Keshishian H, Kuhn E, Kinsinger C, Lee JS, Lee SW, Moritz R, Oses-Prieto J, Rifai N, Ritchie J, Rodriguez H, Srinivas PR, Townsend RR, Van Eyk J, Whiteley G, Wiita A, Weintraub S. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol Cell Proteomics*. 2014; 13:907–917. [PubMed: 24443746]
2. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*. 2003; 100:6940–6945. [PubMed: 12771378]
3. Kuhn E, Addona T, Keshishian H, Burgess M, Mani DR, Lee RT, Sabatine MS, Gerszten RE, Carr SA. Developing multiplexed assays for troponin I and interleukin-33 in plasma by peptide immunoaffinity enrichment and targeted mass spectrometry. *Clin Chem*. 2009; 55:1108–1117. [PubMed: 19372185]
4. Soste M, Hrabakova R, Wanka S, Melnik A, Boersema P, Maiolica A, Wernas T, Tognetti M, von Mering C, Picotti P. A sentinel protein assay for simultaneously quantifying cellular processes. *Nat Methods*. 2014; 11:1045–1048. [PubMed: 25194849]
5. Creech AL, Taylor JE, Maier VK, Wu X, Feeney CM, Udeshi ND, Peach SE, Boehm JS, Lee JT, Carr SA, Jaffe JD. Building the Connectivity Map of epigenetics: chromatin profiling by quantitative targeted mass spectrometry. *Methods*. 2015; 72:57–64. [PubMed: 25448295]
6. Addona TA, Shi X, Keshishian H, Mani DR, Burgess M, Gillette MA, Clauser KR, Shen D, Lewis GD, Farrell LA, Fifer MA, Sabatine MS, Gerszten RE, Carr SA. A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat Biotechnol*. 2011; 29:635–643. [PubMed: 21685905]
7. Burgess MW, Keshishian H, Mani DR, Gillette MA, Carr SA. Simplified and efficient quantification of low-abundance proteins at very high multiplex via targeted mass spectrometry. *Mol Cell Proteomics*. 2014; 13:1137–1149. [PubMed: 24522978]
8. Gillette MA, Carr SA. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat Methods*. 2013; 10:28–34. [PubMed: 23269374]
9. Keshishian H, Burgess MW, Gillette MA, Mertins P, Clauser KR, Mani DR, Kuhn EW, Farrell LA, Gerszten RE, Carr SA. Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol Cell Proteomics*. 2015; 14:2375–2393. [PubMed: 25724909]
10. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*. 2006; 24:971–983. [PubMed: 16900146]
11. Abelin JG, Patel J, Lu X, Feeney CM, Fagbami L, Creech AL, Hu R, Lam D, Davison D, Pino L, Qiao JW, Kuhn E, Officer A, Li J, Abbatiello S, Subramanian A, Sidman R, Snyder EY, Carr SA, Jaffe JD. Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. *Mol Cell Proteomics*. 2016

12. Kennedy JJ, Yan P, Zhao L, Ivey RG, Voytovich UJ, Moore HD, Lin C, Pogossova-Agadjanyan EL, Stirewalt DL, Reding KW, Whiteaker JR, Paulovich AG. Immobilized Metal Affinity Chromatography Coupled to Multiple Reaction Monitoring Enables Reproducible Quantification of Phospho-signaling. *Mol Cell Proteomics*. 2016; 15:726–739. [PubMed: 26621847]
13. Gallien S, Kim SY, Domon B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Mol Cell Proteomics*. 2015; 14:1630–1644. [PubMed: 25755295]
14. Collins BC, Gillet LC, Rosenberger G, Rost HL, Vichalkovski A, Gstaiger M, Aebersold R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat Methods*. 2013; 10:1246–1253. [PubMed: 24162925]
15. Egertson JD, Kuehn A, Merrihew GE, Bateman NW, MacLean BX, Ting YS, Canterbury JD, Marsh DM, Kellmann M, Zabrouskov V, Wu CC, MacCoss MJ. Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods*. 2013; 10:744–746. [PubMed: 23793237]
16. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010; 26:966–968. [PubMed: 20147306]
17. Scrucca L. GA: A Package for Genetic Algorithms in R. *J Stat Softw*. 2013; 53:1–37.
18. Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin Chem*. 2010; 56:291–305. [PubMed: 20022980]
19. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics*. 2012; 12:1111–1121. [PubMed: 22577012]
20. Selevsek N, Chang CY, Gillet LC, Navarro P, Bernhardt OM, Reiter L, Cheng LY, Vitek O, Aebersold R. Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry. *Mol Cell Proteomics*. 2015; 14:739–749. [PubMed: 25561506]

Figure 1.

(a) Schematic illustration of the functionality of a genetic algorithm. The genetic operations are highlighted in the grey box. Adapted from [17]. (b) Illustration of a “crossover” event for transition selection. Each “chromosome” contains a set of “genes” (transitions). Here, the cyan chromosome “crosses over” in several places with the purple chromosome, recombining their “genes” to form 2 new chromosomes. The contributions of the original chromosomes can be seen from their backbone color in the new chromosomes. In this context, chromosomes are just proxies for sets of transitions to be evaluated for goodness of fit in comparing PRM to DIA data.

Figure 2.

Prior to selection of transitions with the genetic algorithm, linear fitting comparisons of PRM to DIA data are shown when all possible transitions are used to quantify the DIA data for the peptides (a) ALG(pS)PTKQLLPCEMACNEK and (b) SP(pS)PAHLPPDPKVAEK. (c) and (d) show the same linear fits after selection of transition sets with the genetic algorithm for the same peptides, respectively.

Figure 3.

Distributions of linear fitting statistics pre- (pink bars) and post-selection (green bars) of transition sets using the genetic algorithm. Line fitting parameters of PRM vs. DIA data (as in Fig. 2) were computed for (a) slope, (b) intercept, and (c) r^2 . Ideal values for these 3 parameters are $m=1$, $b=0$, and $r^2=1$, respectively.