Published in final edited form as:

J Proteomics. 2010 June 16; 73(8): 1641–1646. doi:10.1016/j.jprot.2010.04.004.

Towards proteome standards: The use of absolute quantitation in high-throughput biomarker discovery

Tzu-Chiao Chao^{a,b,*}, Nicole Hansmeier^{a,*}, and Rolf U. Halden^a

^aCenter for Environmental Biotechnology, Biodesign Institute, Arizona State University, PO BOX 875701 Tempe, AZ 85287-5701, USA

^bDepartment of Chemistry and Biochemistry, Arizona State University, PO Box 871604, Tempe, AZ 85287-1604, USA

Abstract

The use of proteomics to profile biological fluids and identify therein biomarkers for cancer and other diseases was initially received with considerable excitement. However, results have fallen short of the expectations. Traditionally, protein biomarkers have been identified by measurement of relative expression changes between case and control samples from which differentially expressed proteins are then considered to represent biomarker candidates.

We argue that current individual proteomics based biomarker discovery studies lack the statistical strength for the identification of high-confidence biomarkers. Instead, multi-group efforts are necessary to facilitate the generation of sufficient sample sizes. This is contingent on the ability to collate and cross-compare data from different studies, which will require the use of a common metric or standards.

Though profound, the technical challenges for absolute protein quantification can be overcome. The use of matrix specific, shared standards for absolute quantitation presents an opportunity to facilitate the much needed comparisons of different studies. In addition to community-wide approaches to standardize pre-analytical biomarker research studies, it is also important to establish means to integrate experimental data from different studies in order to assess the usefulness of proposed biomarkers with sufficient statistical certainty.

Keywords

proteomics; sample size; body fluids; standardization; peptide standard

Proteomics has matured into a valuable tool to analyze established model organisms on the systems level. Due to the high-throughput capacity of the methods involved, great expectations initially were placed on proteomic mass spectrometry (MS) for large scale identification of protein biomarkers for toxicant exposures, diseases, disease outcomes, and related physiological processes [1–4]. Much attention has been placed on the analysis of body fluids

© 2010 Elsevier B.V. All rights reserved.

Correspondence: Dr. Rolf U. Halden, Center for Environmental Biotechnology, Biodesign Institute, Arizona State University, PO BOX 875701 Tempe, AZ 85287-5701, USA, Tel: 480-727-0893, Fax: 480-727-0889, Rolf.Halden@asu.edu.

*both authors contributed equally

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

as sources for biomarkers due to the possibility of reducing risks and inconvenience to patients via non-invasive sampling and analysis [5]. However, the initial expectations have not been met with success. Despite an increasing number of reports on proteomic biomarker discoveries, only 15 protein biomarkers have been approved by the U.S. FDA between 1994 and 2008 [6, 7]. The benefits of accurate diagnostic or prognostic protein biomarkers are beyond doubt and have been reviewed extensively elsewhere [8]. In this article, we highlight some of the most pressing challenges in biomarker research, as well as present developments that may in the near future overcome these problems.

The path from biomarker discovery to clinical utilization can roughly be divided into three parts: discovery, verification and clinical validation (Fig. 1A). In the discovery phase, candidate biomarkers are identified usually based on a limited number of clinical samples. Most often the case and control samples are pooled and the relative protein expression of each group is assessed with MS-based techniques. Proteins found to be significantly modulated in their expression between case and control subjects or pools are then considered as biomarker candidates or putative biomarkers. However, this class comparison approach does not allow an evaluation of the actual diagnostic value of the identified proteins outside of the study cohort. In the qualification and verification stage the validity of the biomarker candidates are assessed in detail with a larger sample size. The goals here are to (i) ascertain that the proteins identified in the discovery phase are indeed differentially expressed in the diseased compared to the healthy state, and (ii) to further evaluate their potential for clinical utility. While a semi-quantitative method may still be applied to fulfill the first goal, a quantitative analysis is required to assess sensitivity and specificity of each respective biomarker for the detection of a given condition usually using ELISA, which is considered the gold standard in protein quantitation. In commercial settings this coincides with the development of a prototype assay with which the biomarkers are monitored. Finally, validated biomarkers that met clinical needs and standards are then monitored in a more extensive clinical trial. Each step is associated with a steep increase in cost, number of samples to be analyzed, and time invested [9,10]. Yet, experience shows that few, if any, of the candidates identified will meet the requirements of clinical biomarkers [11]. It has been acknowledged that the major bottleneck in this process is the transition from the discovery to the verification phase. Many potential biomarkers have been proposed but few were verified and validated in subsequent analyses. Of 1261 cancer biomarkers compiled by Polanski and Anderson [12] so far only 9 have been approved as tumor associated antigens by the FDA. Clearly, the success rate of proteomic biomarker discovery has to be increased. However, an alternative, more promising path can be envisioned using existing techniques and procedures (Fig. 1B).

One straightforward solution to increase the validation of biomarkers is simply to subject all proposed biomarkers to extensive verification and validation studies. However, the prohibitive time and financial investments associated with such an approach, especially if immunoassays first have to be developed, effectively prevent its execution. Recently, liquid chromatography-tandem mass spectrometry (LC-MS/MS) has emerged as an alternative to immunoassays. The main techniques employed for the validation of biomarkers are stable-isotope-dilution multiple reaction monitoring-mass spectrometry (SID-MRM-MS) [13] and stable isotope standards with capture by antipeptide antibodies (SISCAPA) [14,15]. While these methods are as of yet not among standard repertoire in clinical use, the reduced cost and the ability for highly multiplexed analyses promises a higher throughput of potential biomarker candidates in the verification and validation steps [7,16]. In the future, this may be a viable route to increase the rate of biomarker verification and validation, and ultimately may lead to a higher rate of clinical implementation.

Nonetheless, even at a lower cost high-throughput protein quantitation by targeted MS is still unlikely to be economically viable for an attempted validation of every proposed biomarker.

More importantly, many discovered biomarkers were found not to be specific for the disease or state of interest and hence are not useful for diagnostic or prognostic purposes [11]. This is closely tied to the high false discovery rate associated with most omics-based biomarker discovery studies. Three sources of error have been routinely submitted (i) low reproducibility of proteomics data, (ii) poor or inadequate statistical analyses, and (iii) high biological variability of protein expression.

Sources of technical variability are staggering in analytical pipelines employed in proteomic research. The vast amount of different pre-analytical procedures, including sampling, sample storage and sample preparation, as well as the analytical steps themselves, all contribute potentially to variation in the final data reported. While efforts have been made to minimize sources of variation by standardization workflows and data reporting [17–19], wide-scale implementation is still often lacking. Whether justified or not, omics studies are therefore often perceived as poorly reproducible and a number of studies support this claim [20–24]. It should be noted, however, that much of the perceived lack of reproducibility is based on comparing finalized literature data, whereas raw data are often rendered inaccessible for those interested in conducting meta-analyses. A notable exception is the multi-laboratory study by Bell et al. [25]. A total of 27 laboratories were tasked with the identification of 20 purified proteins using whatever technique they deemed appropriate. Initially, only seven groups correctly identified all proteins in question. However, a centralized analysis of the collective datasets revealed that with the acquired raw data in fact all groups could have been successful in identifying the complete set of proteins. This study demonstrated that data analysis as well as experience had a stronger impact on the ability of correctly identifying proteins than the technical equipment. It also illustrates the usefulness to share raw data in standardized formats with the research community as opposed to only publish the finalized reports.

The importance of proper data analysis is of even greater concern for quantitative analyses. One of the most common errors is to neglect the effects of multiple hypothesis testing. In most single hypothesis testing scenarios (e.g., using a *t*-test) a type I error threshold of 0.05 is chosen as a measure of statistical significance. However, in quantitative proteome analyses routinely hundreds of proteins are tested for differential expression from a sample size that is much smaller. *Per definitionem*, 5% of all *p*-values in a study will then be equal or lower than 0.05 just by chance. Given the high number of proteins routinely analyzed this will result in an unacceptable amount of false-positive discoveries. It is therefore necessary to correct for multiple testing. A conservative approach is to control the family-wise type I error using the classical Bonferroni approach [26]. Here the *p*-value cutoff is simply divided by the number of proteins to be tested. The main disadvantage of this approach is the low associated power. Less stringent variations include the Bonferroni Step-down [27] and Westfall and Young [28] permutation. Another approach is to control the false discovery rate according to Benjamini and Hochberg [29], which is routinely implemented in microarray based approaches and also has found its way into proteomic analyses. Fortunately, numerous reports have created awareness about this issue [30–32] and it is expected that in the near future all omics-based analyses will employ multiple testing correction routinely.

Another statistical aspect that is of special importance for biomarker research is the proper validation of potential biomarker candidates. The analyses during the discovery phase are usually of exploratory nature and are designed to identify proteins that distinguish case and control groups. In contrast, during biomarker validation the predictive value of the previously identified proteins has to be assessed. However, the large number of potential predictors (i.e., proteins) that are to be prognostic for a limited number of outcomes puts the model at an elevated risk of overfitting. The result is that, based on the training set, a predictive model would be established that is perfectly suited to distinguish between case and control groups; however, beyond this particular training set it will possess virtually no classification power. In

fact, a number of early biomarkers detected with omics-based analyses probably suffered from overfitting as in many cases non-reproducible sensitivities and specificities near 100% were reported. The subsequent failures to reproduce the performance of these biomarkers in extended studies are likely responsible for at least some of the doubts surrounding the utility of proteomics to generate new biomarkers. To overcome the problem of overfitting, it is necessary to determine whether the predictive model is able to discriminate between case and control in an independent dataset. The most straightforward approach is to randomly split the original sample of subjects and use one dataset, the training set, to create the predictor and use the remaining for validation. The key principle of this split-sample method is that samples used in the training set are not re-used in the validation set. The disadvantage of this approach is that it requires a larger amount of samples. Re-sampling methods as, e.g., cross-validation on the other hand allows for a more economical use of data and perform superior in smaller datasets [33]. An example is k -fold cross-validation in which the data are divided into k subsets of equal size. The predictor is developed k times, leaving out one of the subsets for training in each iteration. The validation is then performed on the remaining dataset. Similar to the multiple-hypothesis testing problem, the biomarker research community has been made aware of the dangers of inappropriate validation methods [34] and the integration into standard workflow is to be expected. These improvements will eventually lead to a higher confidence in biomarker discoveries.

Biological variability heavily influences the protein composition of tissues and body fluids. Without a careful control and evaluation of the biological background of the case and control subjects, it is possible that detected protein expression differences are not diagnostic for the disease under study and instead reflect other conditions and behaviors. In addition, the large variance of protein expression can also result in the exclusion of potentially useful biomarkers during the discovery phase. We will explore this using empirical data available for the established prostate cancer biomarker *prostate specific antigen* or PSA. The PSA-levels utilized here are based on the measurement of 2930 healthy individuals and 171 cancer patients [35]. Let us assume that a proteomic approach is conducted to detect biomarkers for prostate cancer in the same test population. We shall further assume that the technical variance does not play any role in this analysis. With an *a priori* power analysis, we can now calculate the sample size required for any given α and power. If we accept an α error probability of 0.05 and want to achieve a power of 0.8 for example, a total sample size of $N=42$ is needed to detect PSA as significantly elevated in the cancer patients with a t -test (Fig. 2A). This neither takes technical variance nor multiple testing corrections into account. Also it should be noted that t -tests require normal distribution of the data, which is not likely to be the case for PSA-levels found in human populations [36]. All these factors require a further increase of the required sample size to maintain power. As most discovery phase projects are conducted with around a dozen samples, the probability is high that using this population pool, the PSA-levels would not be detected as significantly different. Hence, omics-based techniques may not be able to identify PSA as a potential biomarker. Indeed, PSA-dependent screening has been criticized recently, in part based on the fact that PSA-levels alone were found to be of low specificity as a biomarker, resulting in potential false diagnosis and unnecessary treatment [37]. Nonetheless, PSA remains one of the few available protein biomarkers in use. Other biomarkers may exist but potentially were missed due to limited sample size in the initial phase.

In response to the challenges outlined above, we propose a promising and practical future approach for biomarker discovery: absolute quantitation of proteins in the discovery phase. Again, from the discussion above we concluded that the poor performance of proteome analyses can be linked to invalid results from false biomarker candidate discoveries on the one hand, and a potential inability to detect true biomarkers on the other. Both problems could be addressed potentially by increasing sample size. As discussed, in most cases it is not possible or practical to increase the sample size of a single study to the degree required. In response to

this conundrum, we propose as a solution allowing for a combined analysis of quantitative proteomics data from different studies to increase overall sample size and hence, statistical power (Fig. 1B). The approach involves the addition to proteomics samples of a neutral standard of multiple labeled peptides, each suitable for quantitatively tracking one of the proteins detectable in a tissue or biofluid. This tissue-specific absolute quantitation method (TEAM) would employ matrix specific TEAM standards (e.g., blood serum, blood plasma, amniotic fluid, urine, etc.), since the protein composition and abundance varies widely among the different biological materials. However, for the matrix of interest the TEAM standard would create a neutral, non-arbitrary benchmark against which the expression levels of all identified proteins could be compared. This would enable one to compare and integrate analyses across different studies and thus would increase the confidence of biomarker discovery. At present, a similar approach using existing datasets cannot be taken with current semi-quantitative studies, as these rely on a study-specific non-reproducible pooled control sample as the reference standard or benchmark. Another advantage of the TEAM standard is that it will allow for an absolute quantification of all identified proteins, provided each protein detected is matched in the standard mixture with at least one labeled peptide of known concentration. Measuring absolute concentrations will make it possible to tentatively create empirical models for case and control prediction instead on solely relying on class comparison. For instance, for each identified protein the specificity and sensitivity for the detection of a case sample for a range of concentrations can be determined. A priority list of biomarker candidates can then be generated based on their performance to distinguish case and control samples. Even biomarkers that alone do not possess sufficient predictive powers (i.e., relatively low sensitivity, specificity, or both) may still be interesting and integrated for further studies, if they are used in conjunction with other proteins. Predictive biomarker panels can then be created and evaluated. This is especially interesting in cases, in which only relatively weak effects on protein expression are expected, e.g., in early phases of a disease or effects based on moderate environmental exposure. It is also interesting for established biomarkers like PSA, as the use of additional biomarkers may greatly increase the specificity of blood tests and reduce the rate of unnecessary biopsies. An example is given in Fig. 2B based on data of the occurrence of prostate cancer in patients with PSA level equal or below 4 ng/ml [38]. The sensitivity of a PSA test to detect cancer with a threshold of, e.g., 2 ng/ml is for instance 0.75, whereas the specificity is 0.47. For clinical uses, these values are not acceptable. However, combined with another biomarker with a higher specificity a panel of biomarkers may emerge that is successful in classifying cancer patients. Finally, the quantitative data to be obtained with the TEAM standard will also enable cross-validation analyses, which will be mandatory for multi-dimensional protein panels. A robust biomarker panel should ideally be able to distinguish between case and control, regardless of minor differences in, e.g., sample preparation. While this approach also requires a large sample size, it ultimately provides metrics that will allow for a prioritization of candidate biomarkers and biomarker panels to be explored further in follow-up validation studies.

Unfortunately, at this point in time no such standard exist. However, the basic infrastructure for identification of diagnostic peptides and sharing of experimental data does already exist (e.g., PRIDE, PeptideAtlas) [39]. Moreover, an initiative termed the Human Proteome Detection and Quantitation Project (hPDQ) has been started with the explicit goal to create a complete suite of assays to quantify all protein products encoded in the human genome [40]. The realization of the ambitious hPDQ project would empower the whole proteomics community, while biomarker research will profit from it in particular. While it is intended as a platform for targeted proteomic approaches, it is likely to be adaptable for biomarker discovery approaches as outlined above.

However, should biomarker research be stalled until the hPDQ platform becomes widely available? This is surely not an option. The hPDQ project is envisioned to serve as a universal

platform, allowing for the quantification of all predicted human proteins. However, biofluids of interest usually are somewhat less complex, even while acknowledging that only a fraction of the proteins present will be detectable.

It may therefore be possible to create matrix specific standards based on available or newly generated proteomics data. Still, this is a daunting task considering that, e.g., in human plasma between 697 and 889 proteins were detected [32,41]. The costs of synthetic peptides have decreased over time and cost-effective but laborious approaches to express artificial proteins that reside on concatemers and express tryptic peptides which can then be used as defined standards have been developed (QconCAT) [42]. Yet, without extensive funding a single laboratory is likely to struggle to create and test a suitable standard for proteomic approaches.

However, due to the widespread interest to increase biomarker confidence, it would make sense for the biomarker community to start using defined standards, even if they do not cover the whole proteome to be assayed. While only few proteins would be quantified that way, it will still contribute to the knowledge regarding the biological concentration distribution of selected proteins, even if they should not turn out to be of interest for the given biomarker study. More importantly, if the standard is shared with other groups, the quantitative data for the covered proteins will provide important information regarding quality, reproducibility and potential biases in the different studies. Moreover, with sufficient participants eventually and successively a more complete standard can be created. An important prerequisite for these community based efforts is to adhere to standards for data representation to facilitate effective exchange of information, e.g. as defined by the HUPO Proteomics Standard Initiative [43].

Although proteomic approaches have not yet lived up to their initial expectations, progress has been made to define and formulate strategies to address the most pressing challenges. Technical improvements, guidelines pertaining to the standardization of data analysis and the here proposed collaborative TEAM approach of using matrix specific peptide standards have the potential to revitalize the field of biomarker discovery. Defining and adhering to the use of standards (both, in terms of protocols as well as physical standards), however, will require an effort from the research community. If the work is successfully undertaken, the reward may well be the long-anticipated but currently still elusive high-throughput, high-confidence biomarkers of exposure, effect and susceptibility.

Abbreviations

FDA	Food and Drug Administration
ELISA	immunosorbent assays
hPDQ	human proteome detection and quantitation project
MS	mass spectrometry
TEAM	T issue-specific A bsolute Q uantitation M ethod using shared peptide standards
PRIDE	Proteomics Identifications database
PSA	Prostate specific antigen
SID-MRM-MS	Stable-isotope-dilution, multiple reaction monitoring-mass spectrometry
SISCAPA	Stable isotope standards with capture by antipeptide antibodies

Acknowledgments

This research was supported in part by the National Institute of Environmental Health Sciences grant 1R01ES015445.

References

1. Wulfschuhle JD, McLean KC, Paweletz CP, Sgroi, Trock BJ, Steeg PS, Petricoin EF III. New approaches to proteomic analysis of breast cancer. *Proteomics* 2001;1:1205–1215. [PubMed: 11721633]
2. Adam BL, Vlahou A, Semmes OJ, Wright GL. Proteomic approaches to biomarker discovery in prostate and bladder cancers. *Proteomics* 2001;1:1264–1270. [PubMed: 11721637]
3. Bichsel VE, Liotta LA, Petricoin EF. Cancer proteomics: From biomarker discovery to signal pathway profiling. *Cancer Journal* 2001;7:69–78.
4. Lopez MF. Proteome analysis. I. Gene products are where the biological action is. *J. Chromatogr. B Biomed. Sci. Appl* 1999;722:191–202. [PubMed: 10068141]
5. Kennedy S. Proteomic profiling from human samples: the body fluid alternative. *Toxicology Letters* 2001;120:379–384. [PubMed: 11323197]
6. Gutman S, Kessler LG. The US Food and Drug Administration perspective on cancer biomarker development. *Nat Rev Cancer* 2006;6:565–571. [PubMed: 16794639]
7. Carr SA, Leigh A. Protein Quantitation through Targeted Mass Spectrometry: The Way Out of Biomarker Purgatory? *Clin. Chem* 2008;54:1749–1752. [PubMed: 18957555]
8. He QY, Chiu JF. Proteomics in biomarker discovery and drug development. *J. Cell. Biochem* 2003;89:868–886. [PubMed: 12874822]
9. Zolg JW, Langen H. How Industry Is Approaching the Search for New Diagnostic Markers and Biomarkers. *Mol. Cell. Proteomics* 2004;3:345–354. [PubMed: 14749446]
10. Paulovich AG, Whiteaker JR, Hoofnagle AN, Wang P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl* 2008;2:1386–1402.
11. Lescuyer P, Hochstrasser D, Rabilloud T. How Shall We Use the Proteomics Toolbox for Biomarker Discovery? *J. Proteome Res* 2007;6:3371–3376. [PubMed: 17655344]
12. Polanski M, Anderson LA. List of Candidate Cancer Biomarkers for Targeted Proteomics. *Biomarker Insights* 2006;1:1–48. [PubMed: 19690635]
13. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, Multiplexed Assays for Low Abundance Proteins in Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution. *Mol. Cell. Proteomics* 2007;6:2212–2229. [PubMed: 17939991]
14. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res* 2004;3:235–244. [PubMed: 15113099]
15. Whiteaker JR, Zhao L, Zhang HY, Feng LC, Piening BD, Anderson L. Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal. Biochem* 2007;362:44–54. [PubMed: 17241609]
16. Pan S, Aebersold R, Chen R, Rush J, Goodlett DR, McIntosh MW, Zhang J, Brentnall TA. Mass Spectrometry Based Targeted Protein Quantification: Methods and Applications. *J. Proteome Res* 2009;8:787–797. [PubMed: 19105742]
17. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Schuchard MD, Mehig RJ, Cockrill SL, Scott GB, Tammen H, Schulz-Knappe P, Speicher DW, Vitzthum F, Haab BB, Siest G, Chan DW. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* 2005;5:3262–3277. [PubMed: 16052621]
18. Taylor CF, Binz PA, Aebersold R, Affolter M, Barkovich R, Deutsch EW, Horn DM, Hühmer A, Kussmann M, Lilley K, Macht M, Mann M, Müller D, Neubert TA, Nickson J, Patterson SD, Raso R, Resing K, Seymour SL, Tsugita A, Xenarios I, Zeng R, Julian RK Jr. Guidelines for reporting the use of mass spectrometry in proteomics. *Nat Biotech* 2008;26:860–861.
19. Mischak H, Apweiler R, Banks RE, Conaway M, Coon J, Dominiczak A, Ehrlich JHH, Fliser D, Girolami M, Hermjakob H, Hochstrasser D, Jankowski J, Julian BA, Kolch W, Massay ZA, Neusuess C, Novak J, Peter K, Rossing K, Schanstra J, Semmes OJ, Theodorescu D, Thongboonkerd V, Weissinger EM, Van Eyk JE, Yamamoto T. Clinical proteomics: A need to define the field and to begin to set adequate standards. *Proteomics Clin. Appl* 2007;1:148–156.
20. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 2005;21:171–178. [PubMed: 15308542]

21. Lumbreras B, Miquel P, Soledad M, Marina P, Parker LA, Hernandez-Aguado I. Sources of error and its control in studies on the diagnostic accuracy of "-omics" technologies. *Proteomics Clin. Appl* 2009;3:173–184.
22. Dupuy A, Simon RM. Critical Review of Published Microarray Studies for Cancer Outcome and Guidelines on Statistical Analysis and Reporting. *J. Natl. Cancer Inst* 2007;99:147–157. [PubMed: 17227998]
23. Sorace JM, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 2003;4
24. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 2004;20:777–785. [PubMed: 14751995]
25. Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, Nilsson T, Bergeron JJM. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat Meth* 2009;6:423–430.
26. Dunnett CW. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *JASA* 1955;50:1096–1121.
27. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *SJS* 1979;6:65–70.
28. Westfall, PH.; Young, SS. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley-Interscience; 1993.
29. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *JEBS* 2000;25:60–83.
30. Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. *Adv. Genet* 2008;60:293–308. [PubMed: 18358325]
31. Dudoit S, Shaffer JP, Boldrick JC. Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science* 2003;18:71–103.
32. States DJ, Omenn GS, Blackwell TW, Fermin D, et al. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotech* 2006;24:333–338.
33. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21:3301–3307. [PubMed: 15905277]
34. Ransohoff DF. Lessons from Controversy: Ovarian Cancer Screening and Serum Proteomics. *J. Natl. Cancer Inst* 2005;97:315–319. [PubMed: 15713968]
35. Mettlin C, Littrup PJ, Kane RA, Murphy GP, et al. Relative sensitivity and specificity of serum prostate specific antigen (PSA) level compared with age-referenced PSA, PSA density, and PSA change. *Cancer* 1994;74:1615–1620. [PubMed: 7520352]
36. Li SJ, Peng M, Li H, Liu BS, et al. Sys-BodyFluid: a systematical database for human body fluid proteome research. *Nucl. Acids Res* 2009;37:D907–D912. [PubMed: 18978022]
37. Pienta KJ. Critical appraisal of prostate-specific antigen in prostate cancer screening: 20 years later. *Urology* 2009;73:S11–S20. [PubMed: 19375622]
38. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, et al. Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤ 4.0 ng per Milliliter. *N Engl J Med* 2004;350:2239–2246. [PubMed: 15163773]
39. Mead JA, Bianco L, Bessant C. Recent developments in public proteomic MS repositories and pipelines. *Proteomics* 2009;9:861–881. [PubMed: 19212957]
40. Anderson NL, Anderson NG, Pearson TW, Borchers CH, et al. A human proteome detection and quantitation project: hPDQ. *Mol. Cell. Proteomics*. 2009 R800015-RMCP200.
41. Schenk S, Schoenhals G, de Souza G, Mann MA. high confidence, manually validated human blood plasma protein reference set. *BMC Medical Genomics* 2008;1:41. [PubMed: 18793429]
42. Rivers J, Simpson DM, Robertson DHL, Gaskell SJ, Beynon RJ. Absolute Multiplexed Quantitative Analysis of Protein Expression during Muscle Development Using QconCAT. *Mol. Cell. Proteomics* 2007;6:1416–1427. [PubMed: 17510050]

43. Martens L, Orchard S, Apweiler R, Hermjakob H. Human Proteome Organization Proteomics Standards Initiative: Data Standardization, a View on Developments and Policy. *Mol. Cell. Proteomics* 2007;6:1666–1667. [PubMed: 17766694]
44. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotech* 2006;24:971–983.

A **B** **Fig. 1.**

The current process (A) for the development of protein biomarker candidates (modified from [10,44]) suffers from limitations in sample size, which may be overcome by the more collaborative TEAM workflow shown in panel B.

Fig. 2. Calculations based on empirical PSA levels **A.** Power analysis based on published PSA-levels by Mettlin et al. 1994. The required sample size is plotted against the achieved power based on either $\alpha=0.05$ or $\alpha=0.01$. **B.** Specificity and sensitivity of PSA-levels for the detection of cancer samples. The distributions are re-created from empirical data [38] of prostate cancer patients and healthy subjects with PSA-level equal or below 4 ng/ml.